



SAPIENZA
UNIVERSITÀ DI ROMA

Heterogeneity and noise in living systems: statistical physics perspectives

Scuola di dottorato Vito Volterra
Dottorato di Ricerca in Fisica – XXXII Ciclo

Candidate
Mattia Miotto
ID number 1454204

Thesis Advisor
Prof. Andrea De Martino

October 2019

Thesis defended on 21 January 2020
in front of a Board of Examiners composed by:
Prof. Andrea Crisanti (chairman)
Prof. Antonio Celani
Prof. Martin Weigt

Heterogeneity and noise in living systems: statistical physics perspectives
Ph.D. thesis. Sapienza – University of Rome

© 2019 Mattia Miotto. All rights reserved

This thesis has been typeset by L^AT_EX and the Sapthesis class.

Version: 21 January 2020

Author's email: mattia.miotto@roma1.infn.it

*Dedicated to
my family.*

Abstract

Even the most distracted observer could hardly miss noticing the extensive heterogeneity of traits and behaviors displayed by living systems. So great a variability is commonly ascribed to differences at the level of the genome, which originated from the evolution process to adapt the organisms to the different environments they live in. However, phenotypic heterogeneity is found even in genetically identical organisms, from monoclonal cellular populations to human twins. The multitude of microscopic causes that sum up to give such variability is commonly referred to as biological noise, coming both in the form of environmental fluctuations affecting the development of individual organisms (extrinsic noise) and as the unavoidable results of stochasticity at the level of molecular reaction (intrinsic noise). The latter persisting even when genetically identical organisms are kept under nearly identical conditions.

For quite a long time, such fluctuations were considered a nuisance that makes experiments just difficult to interpret, needing to enlarge the number of observations to have reliable outcomes, and from the point of view of cells, a disturbance cells need to deal with. In the last two decades, however, experimental progresses allowed to investigate the system at single-cell scale. The emerging view is that noise under some circumstances can have a beneficial role, like promoting survival to adverse environments or enhancing differentiation. Ultimately, evolution tunes the systems so they can take advantage of natural stochastic fluctuations. We will follow noise and fluctuation from the cellular level to the higher level of organization of the cellular population where heterogeneity in the molecular reactions translate in the variability of phenotypes. Biology is very broad though, and noise affects all biological processes. Time restraint and my limited knowledge of biological systems did not allow for an exhaustive discussion of all the aspects in which noise and the subsequent heterogeneity play a role. Instead, we will focus on the regulation of noise. More in details, the first part of the thesis introduces to the impact of noise on gene expression and the regulation mechanisms cells use to control it. The action of large regulatory networks is to coordinate a huge of number molecular interactions to obtain robust system-level outcomes. This capability can emerge even when individual interactions are weak and/or strongly heterogeneous. This is the case of post-transcriptional regulation driven by microRNAs (miRNAs). microRNAs are small non-coding RNA molecules able to regulate gene expression at the post-transcriptional level by repressing target RNA molecules. 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’, despite being typically weak, has been found to possess remarkable properties in the presence of extrinsic noise, where fluctuations affect all the components of the system. We will discuss crosstalk and illustrate how crosstalk patterns are enhanced by both transcriptional and kinetic heterogeneities and achieve high intensities even for RNAs that are not co-regulated.

Moreover, we will see that crosstalk patterns are significantly non-local, i.e. correlate weakly with miRNA-RNA interaction parameters. Since these features appear to be encoded in the network's topology this suggests that such crosstalk is tunable by natural selection.

Moving at the cellular level, we focus on the outcomes of gene expression, i.e. the observable phenotypes. Depending on the degree of regulation the cell manages to exert with respect to noise, the distribution of those phenotypes will display a certain extent of heterogeneity. Such cell-to-cell variability is found to have many implications especially for the growth of the whole population. In the second part of the thesis, we discuss some properties of those heterogeneous distributions. First, we focus on the dependence on the initial conditions for the different phases of growth, i.e. the adaptive phase and exponential growth phase. Since cellular populations grow in an exponential fashion, the size and composition of the inoculum shall matter. We discuss this following a novel extensive experimental investigation recently done on cancer cell lines in a controlled environment.

Finally, we focus on the effects that a heterogeneous phenotype has on the growth in hostile environments, i.e. environments fluctuating between states in which the growth is favored and others where growth is inhibited. In such a case, if cells can only replicate (by exploiting available resources) and modify their phenotype within a given landscape (thereby exploring novel configurations), an exploration-exploitation trade-off is established, whose specifics depend on the statistics of the environment. The phenotypic distribution corresponding to maximum population fitness requires a non-zero exploration rate when the magnitude of environmental fluctuations changes randomly over time, while a purely exploitative strategy turns out to be optimal in periodic two-state environments. Most notably, the key parameter overseeing the trade-off is linked to the amount of regulation cells can exert.

Acknowledgments

I would like to thank Andrea De Martino for his help when I most needed it and for having directed my research efforts with patience and essential advice.

Thanks to Giancarlo Ruocco.

Thanks to Edoardo Milanetti, Pier Paolo Olimpieri, Lorenzo Di Rienzo, and Francesco Ambrosetti.

Thanks to Chiara Enrico Bena, Marco Del Giudice, Carla Bosia and Thomas Guedre' for having made the time I spent in Torino fruitful and enjoyable.

Thanks to Lorenzo Monacelli for having shared many hours of study and work, making them lighter.

Finally, grateful thanks to my mother and my father for their support, advice, and trust in me. Thanks to Elena for always bringing me fun, happiness, and love.

Contents

1	Introduction and thesis outline	1
2	Regulation of gene expression	5
2.1	Noise in gene expression	6
2.2	The role of miRNA	7
2.2.1	The ceRNA hypothesis	7
2.3	Modelling post-transcriptional regulatory networks	10
2.3.1	RNA regimes	13
3	Molecular crosstalk	17
3.1	Measuring crosstalk	18
3.2	Crosstalk at system level	27
3.2.1	Probing CLASH crosstalk	30
3.2.2	Crosstalk asymmetry	36
3.3	Transcriptional noise processing	39
3.3.1	Tradeoff between crosstalk and noise buffering	41
3.4	Experimental validation of crosstalk	41
4	Phenotypic diversity	45
4.0.1	Maximum entropy distributions	46
4.0.2	A dynamical model of population growth	48
5	Population growth and survival	53
5.1	Phases of the cellular growth	54
5.2	The inoculum size influences the growth	57
5.2.1	Inoculum size effects on Lag phase	58
5.2.2	Inoculum effect of the population fitness	61
5.2.3	Memory and cooperation	63
5.3	Growth in a changing enviroment	65
5.3.1	Phenotypic distributions and fitness	69
5.3.2	Exploitation limit	70
5.3.3	Exploration limit	73
5.3.4	Fitness and regulation	74
5.3.5	Exploration-exploitation trade-off	76
5.3.6	Biological bet-hedging	78
6	Discussions	81

A Susceptibilities calculation	85
B Degree-preserving randomization	87
C Approximated phenotypic distribution	89
Acronyms and Glossary of Terms	91

List of Figures

- 1.1 **Representation of the genomic flow of information.** Living systems can be regarded as information-flow systems in which the genetic information, stored in the DNA, must be passed on to the phenotype, that is the ensemble of characteristics displayed by an organism within a certain environment [1]. The genetic information is transmitted through a cascade of bio-chemical reactions. From DNA is produced mRNA, mRNAs are converted into proteins. Proteins, constituting the molecular machinery of the cell, build up all the traits that constitute the phenotype, such as the shape, dimension, the motility of a cell. If the transmission of information would be totally efficient, all cells sharing the same DNA should have identical phenotypes, i.e. we should have a homogeneous population of cells. However, at almost every step of these cascades, both the environment and the limited number of molecules involved in the reactions introduce noise. Consequently, the output has some noise and populations manifest a certain level of heterogeneity. 3
- 2.1 **miRNA biogenesis and target binding.** Adapted from [2]. . . . 9
- 2.2 **Summary of the features of a miRNA-RNA network.** (a) Sketch of an interaction network formed by miRNAs and their targets (ceRNAs). The network is a weighted bipartite graph. Line thickness is proportional to the coupling strength (i.e. to the miRNA-ceRNA binding affinity). (b) Sketch of the individual processes lumped in each interaction represented in (a). (c) Sketch of the behavior of the level of free targets (ceRNA or miRNA) as a function of the level of free regulators (miRNA or ceRNA, respectively). Adapted from [3]. . . . 11
- 2.3 **Steady-state concentrations in a system with 2 ceRNAs, obtained by fixing all parameters but the transcription rate b_1 of ceRNA 1.** The dynamical range of the cross-talk interaction between the two ceRNAs corresponds to the window where the fraction of free and bound molecules (ϕ) are similar, i.e., to the \mathcal{S} -regime. Adapted from [4]. 14
- 3.1 **Sketch of the ceRNA mechanism: competition to bind a miRNA can induce an effective positive coupling between its targets.** Adapted from [3]. 18

- 3.2 **Susceptibilities χ_{ij} in a system of 4 ceRNAs, as a function of miRNA transcription rate β .** (All other parameters being fixed). In this example, ceRNAs are cast in two groups: group A, formed by ceRNAs m_1 and m_3 , and group B, formed by ceRNAs m_2 and m_4 . ceRNAs belonging to the same group share identical kinetic parameters. In particular, $\mu_1^0 = \mu_3^0 \ll \mu_2^0 = \mu_4^0$. For β smaller than ~ 500 , no cross-talk is observed; however, as β increases, a symmetric interaction between ceRNAs in group A (of magnitude comparable to the self-susceptibilities) appears. As β increases further, this interaction is switched off, and ceRNAs in group B begin to cross-talk instead. In this region, a change of transcription of a ceRNA in group A can affect the level of ceRNAs in group B, but not viceversa (asymmetric cross-talk). Finally, for sufficiently large β , no cross-talk takes place. Adapted from [4]. 21
- 3.3 **Real (CLASH) network features. (a)** Circular representation of a small part of the CLASH network. RNA species can crosstalk via miRNA-mediated effective interactions. Crosstalk can be established also between RNAs that do not share a miRNA regulator thanks to chains of miRNA-mediated couplings. **(b1-b4)** Classes of miRNA-RNA interactions proposed by [5]: (b1) perfect k-base-pairing in the seed region ('k-mer' mode); (b2) seed base-pairing with up to one mismatched or bulged nucleotide (non-canonical mode or 'seed-nc'); (b3) non-seed base pairing with bulged and/or mismatched nucleotides ('noseed-9nt' mode); (b4) non-seed binding within weak diffuse regions ('noseed' mode). **(c)** Frequency of each binding mode in the CLASH dataset (from [5]). 24
- 3.4 **Comparison between CLIP and CLASH experimental protocol.** 28
- 3.5 **Modeling heterogeneities. (a)** Distributions of RNA transcription rates used in this work: each rate is assumed to be drawn independently from a log normal distribution with given mean (same for each RNA species). Increased transcriptional heterogeneity (TH) corresponds to increased values of the relative fluctuations (CV_b). **(b)** Scenarios of miRNA-RNA binding heterogeneity (BH). From left to right: low BH, where each miRNA-RNA pair interacts with the same strength; medium BH, with k-mer interactions (stronger) distinguished from the rest (weaker); high BH, where the full structure is employed. **(c)** CLASH ceRNA and miRNA degree distributions in the CLASH network and in a randomized version having the same number of contacts and node degrees. Adapted from [6]. 29
- 3.6 **Relative overall molecular abundances in CLASH (a) and degree-preserving randomized networks (b).** Note that the susceptible regime in the latter is narrower compared to the original CLASH network. Adapted from [6]. 30

- 3.7 **Quantitative characteristics of RNA crosstalk in the CLASH network.** (a) Representative distributions of susceptibilities obtained for the CLASH interactome for five different realizations of parameters with different values of $\bar{\beta}$, $CV_{tr} = 0.4$ and maximal BH. (b) Mean susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity. (c) Maximal susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity. The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6]. 31
- 3.8 **Crosstalk in CLASH and random network.** (a) Mean susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$. (b) Maximal susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$. The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6]. 32
- 3.9 **Self and cross-susceptibilities.** (a) Mean self-susceptibility as a function of the mean miRNA transcription rate $\bar{\beta}$. (b) Mean maximal self-susceptibility as a function of the mean miRNA transcription rate $\bar{\beta}$. Results are shown for the 3 BH scenarios considered. The yellow shaded area qualitatively marks the region where the mean susceptibility is significantly different from zero, which coincides with the susceptible regime. In each case, the standard error of the mean is equal to or smaller than the size of the markers. The self-susceptibility is maximal when miRNA levels are low, in which case the availability of free RNA molecules increases roughly linearly with the transcription rate. As $\bar{\beta}$ increases, miRNA repression gets stronger and self-susceptibilities decrease until, at large enough $\bar{\beta}$, RNAs are fully repressed and therefore insensitive to small changes in their transcription rates. (c) Comparison between maximum self-susceptibility, mean self-susceptibility and $\bar{\chi}_{\max}$ for different degrees of TH in the high BH scenario. The intensity of crosstalk between different RNAs, measured by the latter quantity, is indeed of the same order of magnitude as self-susceptibilities. Adapted from [6]. 33

- 3.10 **Crosstalk selectivity in CLASH network.** Inverse of incoming (S_{in}^{-1} on left column) and outgoing (S_{out}^{-1} on right column) selectivities as a function of miRNA mean transcription rate, β . The inverses of S_{in} and S_{out} give a proxy of the number of mRNAs receiving or propagating the effects of a perturbation. CLASH network is very selective until the susceptible region is reached, i.e. both the number of nodes whose variation influences a given mRNA and the number of nodes that are influenced by variation of a certain node is very low. Upon reaching the susceptible threshold the network starts cross-talk and the number of nodes involved in perturbations drastically increases. The random network presents a lower selectivity with respect to real and reshuffled ones, with a peak of cross-talk in the susceptible region. Adapted from [6]. 35
- 3.11 **Measure of RNA asymmetry (a) and (b) CLASH and random asymmetries values, Δ , as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity.** The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6]. 37
- 3.12 **Degree of cross-talk localization (ρ) as a function of β .** a) CLASH network ρ for different values of transcriptional heterogeneity at fixed (high) BH. b) CLASH network ρ for different values of binding heterogeneity at fixed TH. c) Cross-talk localization for CLASH and random network. Adapted from [6]. 38
- 3.13 **Robustness of expression profiles from the CLASH interactome in the presence of crosstalk.** (a) Coefficient of Variation (CV) of RNA levels as a function of the overall mean miRNA transcription rate $\bar{\beta}$ for different degrees of TH. (b) Behaviour of the CV as a function of $\bar{\beta}$ in different BH scenarios in a given TH scenario (CV_{tr} = 0.4). (c) Comparison between the rescaled normalized maximal susceptibility χ_{max} (varying between 0 and 1) and the rescaled normalized Coefficient of Variation as a function of the overall mean miRNA transcription rate $\bar{\beta}$ in a given CV for different degrees of TH (CV_{tr} = 0.4) and BH (high) scenario. (d) χ_{max} vs CV transcriptional heterogeneity (CV_{tr}) and high BH. Results are obtained by averaging over 100 independent TH realizations. In each case the standard error of the mean is equal to or smaller than the size of the markers. Adapted from [6]. 40
- 4.1 **Growth rate distributions for a uniform sampling of E. coli's genome-scale metabolic network in a glucose-limited medium.** Inset: same on a log-log plot. Adapted from [7]. 47

- 4.2 **Empirical growth rate distributions (markers), together with the best fitting MaxEnt distributions (dashed lines in both panels) and the distributions derived from the dynamical model (straight lines in both panels).** Adapted from [7]. 48
- 4.3 **Results from the minimal population dynamical model. a)** Time-evolution of $p(\lambda)$. **b)** Stationary growth rate distributions obtained for different values of $\sigma = D/\lambda_{\max}^3$. **c)** Stationary mean growth rate $\langle \lambda \rangle$ as a function of σ . Adapted from [7]. 51
- 5.1 **Cartoons representative for the analysis of the logistic growth model. a)** The solution of the logistic model 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 emphasized by a dashed black line. 57
- 5.2 **Batch experiment protocol and results.** 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 new 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. 58
- 5.3 **Sketch of the lag time model. a)** Cartoon summarizing 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. 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. **b)** Schematic representation of the two possible scenarios for the population lag time. 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_i s. 60
- 5.4 **Experimental lag times. a)** Experimental lag times as a function of the initial seeding (N_0). Lighter color 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 61

- 5.5 **Experimental growth curves and growth rates.** **a)** Example of a growth curve as a function of time. **b)** Experimental growth rates as a function of the initial seeding (N_0). Lighter color 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 . In both plots the vertical dashed lines are located in correspondence of the carrying capacity value k 61
- 5.6 **Representations of the growth rate (dN/dt) of a population of size N as a 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 64
- 5.7 **Experimental data and model comparison.** Fit of the logistic (orange), weak Allee (blue) and modified logistic (red) models. 65
- 5.8 **Outcomes of the modified logistic model with uniform noise on a and b .** 66
- 5.9 **Putative validation of the modified logistic model.** **a)** Maximum growth rate as a function of N_0 for different values of the carrying capacity and $r(N_0) \sim a(N_0)^b$. **b)** Maximum growth rate as a function of N_0 for different values of the carrying capacity and $r(N_0) \sim a(\ln N_0)^b$ 66
- 5.10 **Representative behavior of the threshold x as a function of time in the different environments.** **(a)** a periodic two-state environment where x switches (in this case) between the values λ_{\max} and $x_{\min} = \lambda_{\max}/2$; **(b)** a periodically switching environment where x takes on random values drawn uniformly from $[x_{\min}, \lambda_{\max}]$; **(c)** a two-state environment where switches occur at exponentially distributed random times; **(d)** an environment where x behaves as in (b) but in which switches occur at exponentially distributed random times. In this example, the characteristic switching times ω_{ns} and ω_{s} are taken to be equal and fixed to 40 (a.u.). Adapted from [8]. 68
- 5.11 **Colormaps showing representative probability densities $p(\lambda, t)$.** Panels to the right of each map depict the density profile at different time points within the zoomed-in region, at time increasing from top to bottom. Results are shown for **a)** const- t and const- x environment, **b)** const- t and rand- x environment, **c)** rand- t and const- x environment, and **d)** rand- t and rand- x environment. Parameter values: $a = 20$, $x = 0.3\lambda_{\max}$, $D = 10^{-3}$. Adapted from [8]. 69

- 5.12 **Growth without phenotypic diffusion. (a and b)** Long-time phenotypic distributions (left) and time evolution of the population growth rate Λ (right) in the absence of diffusion in the different environments (represented by different colors and line widths) and for $x = 0.3\lambda_{\max}$ (panel (a)) and $x = 0.7\lambda_{\max}$ (panel (b)). In the former case ($x < \lambda_{\max}/2$), the distribution can achieve the highest possible CGR. In turn, the long term fitness Λ sets around $\lambda_{\max}/2$. In the latter case ($x > \lambda_{\max}/2$), the distribution peaks around the threshold CGR, while the population achieves a growth rate Λ larger than $\lambda_{\max}/2$. **(c)** Time-averaged \bar{f} as a function of the CGR for const- x (left) and rand- x (right) environments and for three different values of x_{\min} . One sees that the position of the maximum depends both on the chosen threshold and on the specific environment. Adapted from [8]. 71
- 5.13 **Phenotypic distributions and fitness in a symmetric environment. a,b)** Asymptotic, time-averaged phenotypic distributions obtained for a population evolving with a diffusive kernel in a uniform background phenotypic landscape ($a = 0$). **c)** Asymptotic population growth rate Λ (in units of λ_{\max}) as a function of D for the four types of environment. Vertical dotted lines mark the values of diffusion studied in panels (a) and (b). Horizontal lines at small and large D stand for the analytical estimates for the fitness obtained in the const- x regime (dotted blue line) and the rand- x regime, respectively. **d-f)** Same as a-c but with $q(\lambda)$ with $a = 20$ rather than uniform. Adapted from [8]. 75
- 5.14 **Phenotypic distributions and fitness in an asymmetric environment: bet-hedging. a)** Asymptotic, time-averaged phenotypic distributions (top panels) and asymptotic population growth rate Λ (in units of λ_{\max}) as a function of D (bottom panel) obtained as in Fig. 5.12d-f but in the presence of an asymmetric environment with characteristic switching times $\omega_s = 30$ and $\omega_{ns} = 50$ time units. Blue and green horizontal lines at small and large D show the analytical estimates of Λ in the exploration and exploitation limits, obtained in the const- x and rand- x regimes, respectively. Grey horizontal lines represent the same analytical estimates but in the presence of a symmetric environment with $\omega_{ns} = \omega_s = 40$ time units. **(b to d)** Same as (a) but with different choices of ω_{ns} and ω_s . Displayed curves are averaged over 100 independent realizations of the dynamics. Adapted from [8]. 77
- 5.15 **Cartoon of the multi-armed bandit problem. 78**

<p>B.1 Comparison between crosstalk patterns in the CLASH network and its edge-swapped randomized versions. a) Frequency of the shortest paths for CLASH (left) and edge-swapped (right) networks. b-c) Degree distributions of RNA (top) and miRNA (bottom) nodes in the CLASH and edge-swapped networks. d-g) Global crosstalk descriptors for the CLASH and edge-swapped networks as a function of the mean miRNA transcription rate $\bar{\beta}$: (d) mean susceptibility; (e) mean maximum susceptibility; (f) Pearson correlation coefficient between susceptibilities and local kinetic parameters; (g) Coefficient of variation of RNA levels. Adapted from [6].</p> <p>.....</p>	88
<p>C.1 Cartoon representation of the evolution of the distribution.</p>	90

List of Tables

2.1	Summary of parameter values.	15
-----	--------------------------------------	----

Chapter 1

Introduction and thesis outline

All the relevant information to build, sustain and replicate cells, the building blocks of life, is safely stored in the DNA molecule and constitutes the genotype of cells. For this information to be converted into observable traits, or phenotypes, it needs to be first transcribed into an intermediate molecule (RNA) and then translated into proteins. This flow of information, commonly referred to as the central dogma of the biology, involves many cascades of molecular reactions [1]. From the binding of transcription factors to particular loci of the DNA molecules, which initiate the transcription, to the action of ribosomes in translation and to the novel post-transcriptional layer of regulation constituted by micro-RNA molecules. The result is a complex network of interactions between molecules whose energies are often just a few times the thermal energy. In fact, along with strong interactions, such as covalent bonds, that guarantee the stability of molecules, weak molecular interactions (electrostatic, hydrophobic, van der Waals) stand, having typical energies falling in the $k_B T$ energy range. Being in the level of thermal noise fluctuations, they have a transitory nature which renders possible for the same molecule to take part in many reactions. The *contrappasso* for this is that an often not negligible amount of ‘noise’ is introduced in all biological processes [9, 10, 11]. Speaking of noise in biological systems requires particular care, in fact under the noise umbrella term, a lot of different processes and scenarios are considered. The basic idea is that we regard as sources of noise all the mechanisms and processes that introduce variability in the parameters of the system under investigation that we do not want/cannot consider. As much as we speak of thermal noise leading the Brownian motion of a power grain, instead of considering the multitude of interaction between the grain and the water molecules [12].

In practice, biological noise is usually divided into two contributions, an intrinsic component, due to the stochasticity of the biochemical processes in the system, and an extrinsic one, related to the coupling with the variability of the environment in which reactions take place. The functional roles of those noises in biological processes are very diverse. Along with expected entropy-increasing effects of limiting robustness, fidelity and channel capacity in signaling, under some circumstances, it plays more constructive roles, like accelerating the pace of evolution, increasing the fitness of the population in a dynamic environment (bet-hedging) and promoting heterogeneity [13]. Still, if not controlled, randomness may pose more damage

than benefit. Therefore biology had to evolve a number of strategies to cope with fluctuations. Since the relative magnitude of fluctuations in molecule numbers (N) is expected to scale as $\sim \frac{1}{\sqrt{N}}$, a direct solution to decrease noise is to increase the copy numbers. While this is possible for some processes, in general, spatial and energetic constraints go against this solution. An alternative, less trivial way of confronting noise is to use regulatory systems that can either suppress the noise or redirect perturbations. Overall, although living cells are a crowded spatially heterogeneous space which contains lots of different biomolecules whose concentrations and activities are subject to intrinsically random forces, from this randomness a vast array of precisely timed and intricately coordinated biological functions emerge. This seemingly paradoxical nature of life has drawn the interest of an increasing number of physicists, whose expertise contributed to disentangle the scenario of microscopic processes, experimental advances allow to probe in ever more details [14]. With all this in mind, along this thesis, we will try and see how physics, and in particular statistical physics tools can help to shed light on the effect of noise and the role of the subsequent heterogeneity in some key examples. More specifically:

In Chapter 2, we begin with the discussion of fluctuations at the cellular level, and in particular in gene expression, where stochasticity is mitigated by post-transcriptional regulation channels, that were optimized by evolution to stabilize the output protein levels. In particular, noise processing is thought to be the key function, miRNAs) exert in regulation. After an overview of the role of miRNAs as ‘micromanagers’ of gene expression, we will introduce the ceRNA effect and see how miRNA-RNA networks can be modeled.

In Chapter 3, we see how measuring the response of the system to specimen perturbation permits to grasp effective interactions between distant RNA molecules mediated by miRNAs. First, we explain the feature of such crosstalk in simplified cases of study and then we focus our attention on an experimentally validated large scale network and explore its features both with respect to its capability of processing the extrinsic noise on transcription and establish extended crosstalks, which are enhanced by heterogeneities and shaped by evolution.

In Chapter 4, we discuss the emergence of heterogeneous distributions of phenotype from considerations on the metabolic activity. In particular, we will see how both a maximum entropy scenario and a minimal dynamical model yield experimentally observed distributions that are ultimately dictated by the degree of regulation cells can attain over noise.

In Chapter 5, we definitively leave the molecular level moving forward to consider the growth of cellular populations. Here, heterogeneity in phenotypes like the lag time or the growth rate will be found to influence the fitness of the population in the presence of a controlled environment. In particular, we will describe an experimental setup that allowed to measure the growth rate of a cellular population which grows starting from different initial inoculum sizes. We will find that the fitness of those populations manifests non trivial features and minimal theoretical models manage to account for such characteristics. We will then devote the discussion to the problem of cell adaptability to hostile environments. In particular, in populations of bacteria or cancer cells facing variable and often unpredictable environmental changes, variability

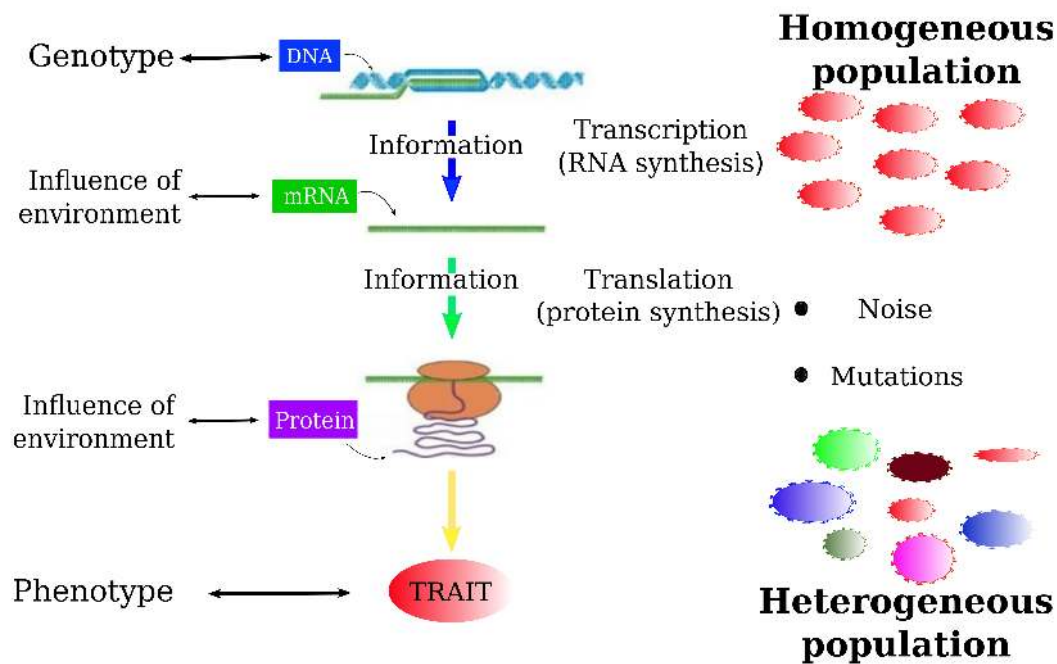


Figure 1.1. Representation of the genomic flow of information. Living systems can be regarded as information-flow systems in which the genetic information, stored in the DNA, must be passed on to the phenotype, that is the ensemble of characteristics displayed by an organism within a certain environment [1]. The genetic information is transmitted through a cascade of bio-chemical reactions. From DNA is produced mRNA, mRNAs are converted into proteins. Proteins, constituting the molecular machinery of the cell, build up all the traits that constitute the phenotype, such as the shape, dimension, the motility of a cell. If the transmission of information would be totally efficient, all cells sharing the same DNA should have identical phenotypes, i.e. we should have a homogeneous population of cells. However, at almost every step of these cascades, both the environment and the limited number of molecules involved in the reactions introduce noise. Consequently, the output has some noise and populations manifest a certain level of heterogeneity.

increases the probability that some individuals may survive the stress produced by a sudden change of the environment. In that respect, it has been speculated that those kinds of cell populations implement a risk distribution strategy: they may have evolved to regulate not only the averaged gene expression levels but also the extent of allowed deviations from such an average, setting it at the desired level in order to face (as population) each environmental condition [15].

Part of the work described in this thesis led to the publication of three papers:

1. (2019) **M. Miotto**, E. Marinari, and A. De Martino, "Competing endogenous RNA crosstalk at system level", *PLOS Computational Biology*, 15(11):e1007474
2. (2019) A. De Martino, T. Gueudre', and **M. Miotto**, "Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations", *Phys. Rev. E* 99, 012417, DOI: 10.1103/PhysRevE.99.012417
3. (2017) C. Enrico Bena, M. Del Giudice, T. Gueudre', **M. Miotto**, E. Turco, A. De Martino, C. Bosia, "Inoculum-density dependent growth reveals inherent cooperative effects and stochasticity in cancer cell cultures", *arXiv* preprint arXiv:1710.10978

An additional paper described in 3.4 is in preparation.

The research activity conducted during this Ph.D. involved also the collaboration with other projects not described in this thesis. These collaborations led to the following publications:

1. (2019) M. Monacelli and **M. Miotto**, "Gene heterogeneity drives the evolution of species", *arXiv* preprint arXiv:1912.01444 [16]
2. (2019) **M. Miotto**, L. Di Rienzo, P. Corsi, G. Ruocco, D. Raimondo and E. Milanetti, "Simulated Epidemics in 3D Protein Structures to Detect Functional Properties", *arXiv* preprint arXiv:1906.05390 [17]
3. (2018) **M. Miotto**, P. P. Olimpieri, L. Di Rienzo, F. Ambrosetti, P. Corsi, R. Lepore, G. G. Tartaglia, E. Milanetti, "Insights on protein thermal stability: a graph representation of molecular interactions", *Bioinformatics*, DOI: 10.1093/bioinformatics/bty1011 [18]
4. (2018) **M. Miotto**, L. Monacelli, "Entropy evaluation sheds light on ecosystem complexity", *Phys. Rev. E* 98, 042402, DOI: 10.1103/PhysRevE.98.042402 [19]

Chapter 2

Regulation of gene expression

The information to assemble all the proteins an organism needs in the various stages of life is encoded in its DNA. Anyway, neither all proteins are synthesized at the same time, nor all cells employ the same proteins. Each cell produces only the proteome it needs in relation to the environmental stimuli it receives. The control of gene expression, i.e. regulation of the conversion of genetic information into proteins, is a fundamental process to bring the genome to life and mis-regulation at any level is usually associated with disease [20]. Moreover, gene expression is regulated at different levels and there is increasing evidence that the diverse processes involved in this regulation are integrated with each other [21, 22, 23, 24].

The regulation mainly takes place at two levels, during transcription (the conversion of DNA into RNA) and before translation (before mRNA is turned into proteins). Furthermore, downstream of these two processes, expressed proteins can still be regulated by post-translational modifications (PTM) and protein degradation (post-translational control). Transcriptional control has received much attention, through both traditional single gene studies [25] as well as through genome-wide approaches such as expression profiling [26], transcription factor binding studies and identification of regulatory sequence elements [27], and chromatin remodeling and epigenetics [28]. Recently, there has been an increasing appreciation of the necessity and importance of post-transcriptional gene expression regulation. Post-transcriptional regulation mechanisms comprise various processes such as mRNA processing (ppolyadenylation, capping, and splicing), mRNA export and localization, mRNA decay, and mRNA translation. In particular, the regulation operated by miRNA is under much scrutiny. In fact, from 1993, year of their first observation, up to now, micro-RNAs were found to be involved in the regulation of a plethora of processes, from cellular development to proliferation, from survival to apoptosis, thus being regarded to play a central role in gene regulation both in health and disease [29]. The observation that, through a titration mechanism, miRNAs can act as mediators of effective interactions among their common targets (competing endogenous RNAs or ceRNAs) has brought forward the idea (known as ceRNA hypothesis) that RNAs can regulate each other in extended cross-talk networks. This ability is considered pivotal to shape a cell's protein repertoire. In fact, by being able to target different RNA species with different kinetics, they can act as mediators of an effective interaction between the RNAs, such that a change

in the transcription level of one RNA can result in an alteration of the levels of another RNA. Furthermore, the involvement of miRNAs in peculiar motifs of the post-transcriptional regulatory network suggests that they actively perform noise processing in gene expression [30]. Recent work has characterized the emergent properties of cross-talk and noise processing in small regulatory motifs in silico [4, 31]. Small motifs are however embedded in large, cell-scale networks of interactions and a strongly heterogeneous structure of targeting patterns. Perturbations could in principle propagate across this network and affect RNAs that are topologically distant through chains of miRNA-mediated couplings, creating a collective, long-range mode of regulation characterized by strong and highly selective cross-talks. In Chapter 3, we will see how to measure crosstalk and discuss its evolutionary signatures and biological consequences.

2.1 Noise in gene expression

An emerging view is that gene expression noise is not simply a necessary evil of the molecular computations of the central dogma, but is at times utilized by cells to achieve certain functions. In other words, evidence is mounting for specific physiological roles for gene expression noise [13]. For instance, stochasticity is the driving force that generates phenotypic heterogeneity in microbial populations. In fact, population heterogeneity in the expression of even one protein can have important consequences for a population of cells. For instance, [32] found that cell-to-cell variability in the expression of a single antibiotic resistant protein in *Saccharomyces cerevisiae* can lead to pronounced effects in the response of the whole population to an antibiotic challenge; at intermediate antibiotic concentrations, the strain with the largest phenotypic heterogeneity was best able to survive. In some instances, the effect of noise can be amplified by the presence of multi-stability in genetic networks. This can lead to multiple phenotypes coexisting in a cell population. Individual cells can make transitions between those phenotypes driven by fluctuations in the expression of certain key genes in the network [11].

An example is provided by the behavior of individual *Bacillus subtilis* cells when they are subject to stress; the cells must decide whether to enter a competent state or to start forming spores [33, 34]. The two phenotypes are mutually exclusive, and the decision between them is determined by transient values of the concentration of key proteins in the network, which can stochastically exceed a threshold value and thereby force the cells to assume the competence phenotype [35]. Not all gene expression noise, however, is beneficial. Some processes, such as the development of a multicellular organism, rely on precise spatial and temporal transmission of genetic information, and in these cases noise, in gene expression needs to be kept to a minimum [36]. Recent experiments have found that noise suppression mechanisms exist at the level of gene networks that control development [37]. Here, we will focus on one of the mechanisms cells evolved to process noise: the post-transcriptional regulation mediated by micro RNAs.

2.2 The role of miRNA

MiRNAs are endogenous 22 nucleotide-long noncoding RNA RNAs (ncRNAs), which play important regulatory roles in animals and plants by pairing to the messenger RNAs (mRNAs) of target genes and specifying mRNA cleavage or repression of protein synthesis [38]. Lin-4 [39, 40] and let-7 [41] were the first miRNAs to be discovered, identified studying the developmental stages of *Caenorhabditis elegans* (*C. elegans*). It rapidly became clear that miRNAs represented novel means of regulating developmental timing. Subsequently, hundreds of non protein-coding miRNAs started being identified in many other processes. Currently, thousands of miRNAs have been identified in humans and other species, and miRNA online sequences repositories, such as the miRbase database [42], are available as much as miRNA-mRNA binding prediction tools like TargetScan [43]. Researchers further gained interest in miRNAs when, in 2005, the first reports addressing the biological function of miRNAs in cancer were published, showing that miRNAs act as onco-genes or tumor suppressors and are involved in a huge variety of pathways deregulated in cancer [44]. To date, altered miRNA expression has been reported in almost all types of cancer.

2.2.1 The ceRNA hypothesis

With a lot of simplification, we can state that the phenotype of an organism is set by the amount of proteins the organism possesses. The proteome of the cell is not static but can change in response to both precise internal and environmental stimuli as much as due to fluctuations. The timing of responses can be of paramount importance for cell survival, especially in the presence of adverse fluctuating environments (as we will better discuss in Chapter 5). To reorganize the proteome, cells can tune the synthesis/degradation ratios of proteins and their precursors mRNAs. In particular, once mRNA molecules are transcribed, the cell can prevent their translation onto protein increasing their decay rate or acting at the level of ribosomes, those actions though inevitably affects all mRNA species. The employment of miRNA molecules allows a more fine layer of regulation. In fact, since miRNAs can be engineered to target specific RNA molecules in specific regions, cells can repress a certain RNA by increasing the number of miRNA species that bind to that RNA.

The interesting thing is that one miRNA can bind to several (up to many hundreds) different RNAs. In fact, the canonical RNA-miRNA binding takes place through the base-pairing of 6-9 nucleotides of the seed region of the miRNA. The sole fact that the pairing involves a small number of nucleotides explains why it is not so uncommon that the same miRNA can have more than one target. Whenever a molecule binds to other different kinds of molecules and the binding provokes sequestration of the complex from the system, a competition for the binding takes place between the targets. This equals to an indirect interaction among the molecules (see Figure 3.1). In fact, a variation in the concentration of one RNA induces a variation in the levels of the miRNA molecules that bind the RNA. This variation in the miRNA will in turn influence all the other RNAs the miRNA can bind to. This *crossstalk* introduces a further layer of regulation, in fact not only the cell can regulate the levels of different RNAs by varying the common

miRNA concentration. But it can act on one RNA and modify the level of other RNAs since the miRNAs propagate the perturbation. mRNAs become correlated. Therefore, looking at target concentrations (as experiments often do) it may be hard to disentangle the contribution to correlation due to competition from the more trivial one due to variations in the level of the common miRNA. Together with the great experimental effort spent in both identifying and characterizing new miRNA motifs, theoretical modeling has been pursued too. Mathematical models for ceRNA regulation have been developed by multiple groups, and include models for ceRNA co-regulation, for the effects of miRNA-target binding strength on ceRNA regulation, for the effects of the number of miRNA binding sites per ceRNA, and for the interplay between ceRNA regulation and other types of regulatory interactions [31, 4, 45]. In particular, kinetic modeling of a minimal network, in which one miRNA interacts with two competing RNAs, showed that ceRNA activity is determined by the relative abundance of ceRNAs and miRNAs as well by the type of their interaction (stoichiometric or catalytic) [46]. Further extension of the minimal model by considering interactions between multiple miRNAs and ceRNAs allowed for the characterization of mean and noise profiles of ceRNA network components and the response time of the network components required to resume their steady states upon perturbation [31]. Noorbakhsh et al. [47] studied noise characteristics within a ceRNA network composed by a miRNA, a protein-coding RNA and a non-coding RNA, showing that the noise is dramatically high when the combined transcription rate of the ceRNAs approximates the transcription rate of the miRNA (i.e. the cross-regulation of the two ceRNA happens). Later on, by integrating miRNA-mediated ceRNA crosstalk with transcription factor (TF) regulation, Martirosyan et al. [48] showed that miRNA regulation of a gene through the ceRNA network can outperform its regulation by a TF, suggesting the possible role of miRNAs as major regulators rather than fine-tuners of gene expression. While these models helped to improve our understanding of ceRNA regulation, they do not account for miRNAs that have hundreds of other targets or for ceRNA regulation by multiple shared miRNA species. Most importantly, they do not consider the effective coupling between ceRNAs that are linked by a long chain of miRNA-ceRNA interactions. Those aspects are the main topics of the following sections.

miRNA biogenesis in animals and interaction with target mRNAs

Genes coding for miRNAs can be found as autonomous units in intergenic regions of the genome [49, 50], in introns (i.e. non-coding regions inside a gene that are removed after splicing) of both protein coding and non-coding genes [51] and even in exons (i.e. regions coding for a portion of the mature RNA) of non-coding genes [52]. miRNA genes can also form clusters regulated by a common promoter, creating a polycistronic transcription unit [53]. The primary product of the transcription of a miRNA gene is a long transcript called pri-miRNA. Pri-miRNAs are processed in the nucleus by the enzyme Drosha, originating ~ 100 nucleotides long miRNA precursors with a hairpin structure, called pre-miRNAs. These precursors are transferred to the cytoplasm by the protein Exportin-5, where they undergo further processing by the Dicer enzyme. This enzyme cuts the hairpin loop of the pre-miRNA, leading to a double stranded RNA (dsRNA) duplex with an imperfect match between the sequences. The duplex is then incorporated into a protein complex called RNA-induced silencing complex (RISC). Within the RISC, only one of the two strands of the miRNA is loaded, the other being released and degraded.

Once loaded in the RISC, the miRNA is ready to exert its repressive action on its target RNA, whose sequence is recognized through Watson-Crick base pairing. The canonical binding mode between a miRNA and its target consists of a nearly perfect pairing with a small region, about 6 nucleotides long, contained in the 5' end of the miRNA, the so-called seed region [54]. Usually, the sequence of the target complementary to the seed region is located in the 3'-UnTranslated Region (3'UTR). In animals (plants behave quite differently), the formation of the complex promotes translational repression. Furthermore, mRNA bound to the RISC complex is subject to destabilization, both through direct cleavage and through deadenylation [29]. Both processes end with the degradation of the target and the release of the RISC complex. A further route is the trapping of the complex protein granular structures, called P-body [55]. In this case, both target and miRNA are removed from the system. The sequestration of target molecules, possibly combined to an enhanced degradation, is then at the basis of miRNA-mediated gene regulation. Computational models predict that more than 60% of human genes are targeted by miRNAs [54, 56].

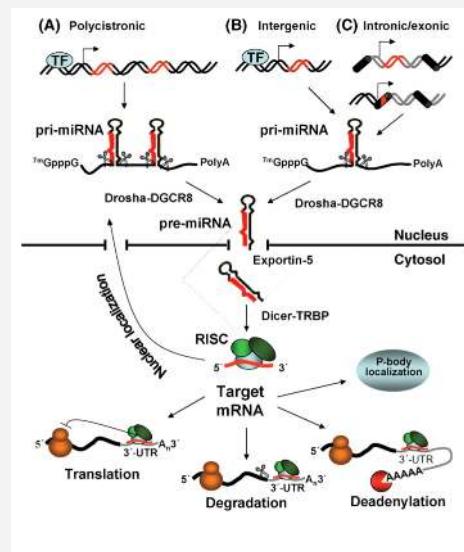


Figure 2.1. miRNA biogenesis and target binding. Adapted from [2].

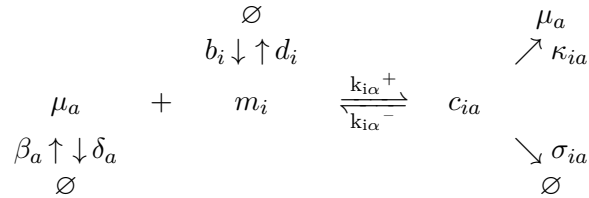
2.3 Modelling post-transcriptional regulatory networks

To build an efficient measure of crosstalk and see how this crosstalk is influenced by heterogeneity in important features of the network, the first thing we need is a model to describe the time evolution of the miRNA and RNA concentrations. When dealing with chemical reactions, the simplest model one can build is a deterministic mass-action scenario. In particular, the whole network can be modeled as a system of coupled ordinary differential equations (ODE) for the concentration levels of free molecules and complexes.

Let us consider the general case of a system composed of N RNA species $\{m_i\}_{i=1}^N$ (the ceRNAs) and M miRNA species $\{\mu_a\}_{a=1}^M$. Since we want to understand the basic features of crosstalk, we focus only on miRNA and RNA molecules. Consequently, we do not take into consideration all the processes that lead to the formation of miRNA and RNA molecules nor the pathways that lead to their degradation (as discussed in the Panel ‘miRNA biogenesis in animal and interaction with target mRNAs’). We only assume that free RNA i is synthesized and degraded with rates b_i and d_i respectively, while free miRNA a has β_a synthesis rate and δ_a degradation one.

RNA and miRNA can form complexes c_{ia} with rates k_{ia}^+ and complexes can dissociate with rate k_{ia}^- . (We neglect that in order to form a complex, miRNA molecules must bind to the RISC). While miRNA and RNA are bound, their complex is assumed to undergo two different degradation processes, either a catalytic process in which the miRNA returns to the cytosol or a stoichiometric route where both molecules are destroyed. The two different pathways the complex can take derives from biological information. In fact, it has been observed that the assembling of miRNAs in the RISC structure prevents their degradation, so that once the mRNA is processed the RISC (just the free miRNA for us) returns at disposal for binding another free mRNA. When miRNA and RNA are bound together, though, the whole complex can be sequestered by specific cellular structure-like P-bodies, resulting in effective removal of both miRNA and mRNA from the cytosol [55]. To model these two complex degradation pathways, we then call for two more rates. σ_{ia} quantifies the complexes full degradation route while κ_{ia} accounts for complex partial degradation: the release of miRNA and destruction of RNA.

The whole process can be schematize as



Since molecular interactions happen only inter species, the network of interactions has a bipartite structure (see also Figure 2.2a-b). In such a scenario, equations for

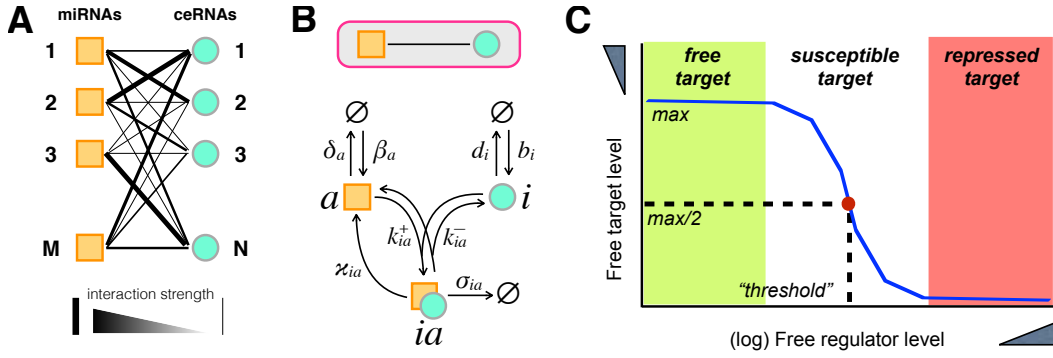


Figure 2.2. Summary of the features of a miRNA-RNA network. (a) Sketch of an interaction network formed by miRNAs and their targets (ceRNAs). The network is a weighted bipartite graph. Line thickness is proportional to the coupling strength (i.e. to the miRNA-ceRNA binding affinity). (b) Sketch of the individual processes lumped in each interaction represented in (a). (c) Sketch of the behavior of the level of free targets (ceRNA or miRNA) as a function of the level of free regulators (miRNA or ceRNA, respectively). Adapted from [3].

the time evolution of free molecules and complexes can be written as follows:

$$[\dot{m}_i] = b_i - d_i[m_i] - \sum_a k_{ia}^+[m_i][\mu_a] + \sum_a k_{ia}^-[c_{ia}], \quad (2.1a)$$

$$[\dot{\mu}_a] = \beta_a - \delta_a[\mu_a] - \sum_i k_{ia}^+[m_i][\mu_a] + \sum_i (k_{ia}^- + \kappa_{ia})[c_{ia}], \quad (2.1b)$$

$$[\dot{c}_{ia}] = k_{ia}^+[m_i][\mu_a] - (k_{ia}^- + \kappa_{ia} + \sigma_{ia})[c_{ia}]. \quad (2.1c)$$

where $[m_i]$ and $[\mu_a]$ stand for free RNA and miRNA concentrations, respectively. This set of equations can be numerically solved once parameters and initial conditions are specified. To understand how things work, we restrain to the steady state regime of the dynamics, so that we can not concern with initial values, and have a unique set of steady state free molecule concentrations. However, transient regimes are important and crosstalk can be characterized there as well. We briefly discuss this possibility in Panel ‘Crosstalk away from stationarity’.

Imposing nil time derivative in 3.41a, steady state expressions for the concentrations are given by:

$$[m_i] = \frac{b_i + \sum_a k_{ia}^-[c_{ia}]}{d_i + \sum_a k_{ia}^+[\mu_a]}, \quad (2.2a)$$

$$[\mu_a] = \frac{\beta_a + \sum_i (k_{ia}^- + \kappa_{ia})[c_{ia}]}{\delta_a + \sum_i k_{ia}^+[m_i]}, \quad (2.2b)$$

$$[c_{ia}] = \frac{k_{ia}^+[m_i][\mu_a]}{(k_{ia}^- + \kappa_{ia} + \sigma_{ia})}. \quad (2.2c)$$

Crosstalk away from stationarity

miRNA-ceRNA interaction strengths and silencing/sequestration mechanisms emerge, together with the relative abundance of regulators and targets, as key factors for the onset and character of ceRNA crosstalk, including its selectivity. Moreover, heterogeneities in kinetic parameters as well as in miRNA-ceRNA interaction topology are major drivers of ceRNA crosstalk in a broad range of parameter values. The picture obtained at stationarity can be extended to out-of-equilibrium regimes. In particular, one can characterize a 'dynamical' ceRNA effect, which can be stronger than the equilibrium one, as well as the typical timescales required to reach stationary crosstalk [57].

Eliminating the complexes $\{c_{ia}\}$, the $N \times M + N + M$ steady state equations eq. (2.2) reduce to a set of $N + M$ coupled equations for the free concentrations of species $\{\mu_a, m_i\}$:

$$[m_i] = \frac{m_i^*}{1 + \sum_a \frac{N_i [\mu_a]}{\mu_{ia}^0}} = m_i^* F_i, \quad (2.3a)$$

$$[\mu_a] = \frac{\mu_a^*}{1 + \sum_i \frac{M_a [m_i]}{m_{ia}^0}} = \mu_a^* F_a. \quad (2.3b)$$

in which $m_i^* = b_i/d_i$ and $\mu_a^* = \beta_a/\delta_a$ represent the concentrations of free compounds in the absence of inhibition; while M_a (N_i) is the number of RNA (miRNA) that bind miRNA a (RNA i). Finally, m_{ia}^0 and μ_{ia}^0 are given by:

$$m_{ia}^0 = \frac{\delta_a}{k_{ia}^+} \left(1 + \frac{k_{ia}^- + \kappa_{ia}}{\sigma_{ia}} \right), \quad (2.4a)$$

$$\mu_{ia}^0 = \frac{d_i}{k_{ia}^+} \left(1 + \frac{k_{ia}^-}{\sigma_{ia} + \kappa_{ia}} \right). \quad (2.4b)$$

m_{ia}^0 and μ_{ia}^0 are proportional to the inverse affinity of the complex (given by the ratio k^+/k^-), so they quantify the strength of the binding. Note that, put in this form, they can be estimated by novel miRNA transfection experiments [58].

Again, we stress that this modelization represents a coarse-graining of the real biological process, in fact, it does not consider all the catalyzing intermediate steps that involve the presence of proteins and other RNA molecules, as the formation of the RISC complex. Anyway, assuming that the only rate-limiting compounds are the miRNA molecules the underlining catalytic reactions may be ignored. Despite its essentialness, this approach allows to fully characterized analytically and numerically the steady state of the dynamics and consequently to precisely quantify the sensitivity of a ceRNA to alterations in the level of one of its competitors, sufficing to capture many of the central characteristics of miRNA-based regulation from basic assumptions about the underlying processes.

All reactions reported in 3.41a, i.e. transcription, degradation and titration events due to miRNA-ceRNA interactions, are intrinsically stochastic. This means in practice that molecular levels evolving in time are bound to be subject to random fluctuations, with the strength of the noise affecting each molecular species roughly proportional to the square root of its mean. After a transient, concentrations will stabilize and fluctuate around the steady state of the deterministic model. The deterministic model thereby yields a description of the miRNA-ceRNA network that is all the more accurate when the system is well mixed and concentrations are sufficiently large. In what follows, we will limit the discussion on results obtained only using the deterministic model in the steady state regime. This will allow to quantitatively measure the cross-talk between ceRNA in different parameter scenarios and to have strong analytical support. For a review on stochastic models see [3].

Parameter reduction

At the moment, the state of the system depends on $4MN + 2M + 2N$ parameters. While much biology is contained in each parameter and in specific situations all parameters should be accurately considered, here we want to focus on the main features of crosstalk and a reduction of the number of parameters to be explored is helpful. So, to shrink the parameter space we assume that $\sigma_{ia} = \sigma$, $\kappa_{ia} = \kappa$, and $d_i = d$ and $\delta_a = \delta$ for each i and a . Furthermore, we can also suppose that $k_{ia}^- \ll \sigma + \kappa$, which equals to say that complexes are degraded faster than the time thermal noise needs to separate the species. Under those assumptions, one gets $\mu_{ia}^0 = \lambda m_{ia}^0 \frac{d}{\delta}$ for all i and a , where $\lambda = \frac{\sigma}{(\sigma + \kappa)}$ is the ‘stoichiometric ratio’ measuring the relative degree of stoichiometricity of complex decay. The number of effective parameters in this case shrinks to $NM + N + M + 3$, namely the inverse affinities μ_{ia}^0 , the synthesis rates b_i and β_a , the stoichiometric ratio λ , and the rates, δ , d . Reasonable choices for all those parameters are reported in Table 2.1. Note that inverse affinities and synthesis rates will be the source of heterogeneity whose effect we aim to characterize. For this reason, in Table 2.1 we reported typical values that in practice can vary also by orders of magnitude. Finally, since miRNA molecules are used by cells as regulators, we use the miRNA transcription rate as a control parameter upon varying which crosstalk patterns will be analyzed.

2.3.1 RNA regimes

Once a reasonable choice of parameters is done, Eq. 2.3 can be numerically solved and the concentrations of free molecules can be studied varying miRNA transcription rate. Note that since equations for the concentration of miRNA and ceRNA have the same structure, qualitative similar results are found if one looks at the mirror system, i.e. if we vary RNA transcription rates [4, 57].

Varying β on a logarithmic scale, we can see that as β increases, the concentration of free miRNA in the system increases as well, while RNA concentrations go down in a sigmoidal fashion (see Figure 2.2c).

If the levels of all miRNA species interacting with ceRNA i are sufficiently low (specifically, much lower than the respective thresholds μ_{ia}^0 , so that $\sum_a \mu_a / \mu_{ia} < 1$),

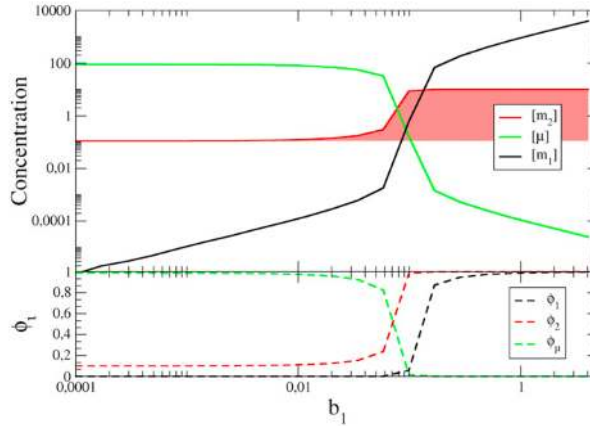


Figure 2.3. Steady-state concentrations in a system with 2 ceRNAs, obtained by fixing all parameters but the transcription rate b_1 of ceRNA 1. The dynamical range of the cross-talk interaction between the two ceRNAs corresponds to the window where the fraction of free and bound molecules (ϕ) are similar, i.e., to the \mathcal{S} -regime. Adapted from [4].

then the steady-state level of RNA i will be very close to the maximum possible, m_i^* . In such conditions, RNA species i will be roughly insensitive to changes in miRNA levels. We will call this the ‘unrepressed’ or ‘free’ regime for RNA i . As the quantity $\sum_a \mu_a / \mu_{ia}$ increases, e.g. following an increase in the level of one or more miRNA species, $[m_i]$ decreases. This occurs most notably when $\sum_a \mu_a / \mu_{ia} \sim 1$ (corresponding, for $M = 1$, to a miRNA level close to the threshold value μ_{ia}^0). Here RNA i is very sensitive to a change in miRNA levels. We shall therefore term this the ‘susceptible’ regime for RNA i . Finally, when miRNA levels become sufficiently large, ceRNA i will eventually become fully repressed. In order for this to occur, it suffices that $\sum_a \mu_a / \mu_{ia} \gg 1$ (which occurs e.g. when the level of at least one of the miRNA species targeting i significantly exceeds its corresponding threshold μ_{ia}^0). We then shall this the ‘repressed’ regime for ceRNA i (see Figure 2.2c for a sketch). Note that the sigmoidal shaped response curve is typical of titration mechanisms.

It is now very instructive to look at what happens if instead we modulate the transcription rate of one ceRNA. Let us consider a system with only two ceRNA molecules and one miRNA and look at concentrations upon varying the transcription rate of say ceRNA 1. As b_1 increases, m_1 grows as expected while the concentration of free miRNAs decreases as they increasingly engage targets. This in turn un-represses the other RNA species, whose level also increases as the transcription rate of ceRNA 1 is upregulated (see Fig. 2.3). The level of ceRNA 2 increases upon changing b_1 and this response happens in the susceptible region. This is the key signature of the miRNA-mediated crosstalk that can be established between competing RNAs. In what follows we will see how to quantitatively measure the crosstalk and why heterogeneity is a key ingredient.

Table 2.1. Summary of parameter values.

Parameter	Value	Description	Ref.
d	0.08 [h ⁻¹]	RNA degradation rate	[59]
δ	0.027 [h ⁻¹]	miRNA degradation rate	[59]
λ	0.2	stoichiometricity ratio	[59]
$\langle b \rangle$	8 [molecules/h]	mean RNA transcription rate	[59]
μ^0	4 [molecules]		

Chapter 3

Molecular crosstalk

Almost all processes inside cells involve bindings between molecules. If a molecule can bind to two possible target molecules, in such a way that once the molecule binds to one target it can not interact with the other until the complex dissociates, then a sort of competition for the binding of the common molecule is established between the targets. Such competition results in an effective, or mediated, interaction between the targets. In fact, if the concentration of one target increases, then part of the molecules targeting the second target will be diverted toward the first and the net result will be an increase of the second target free molecules. The effective interaction is termed crosstalk.

Here, we will focus on the crosstalk between RNAs originating from their competition in binding micro RNA [60]. However, this mechanism of ‘titrative’ interaction is found in processes as different as protein ubiquitination [61], growth factors signaling [62], and transcription factors sequestration [63]. Consequently, the picture we are going to discuss is expected to apply in general to all networks of molecular species competing for a common resource. As we saw in 2.1 and in 2.2, micro RNAs are important post-transcriptional regulators of gene expression. Since each micro RNA can bind to many targets, they originate a network of effective interactions among the whole population of RNAs.

The characteristic of the resulting crosstalk, as we show in 2.3, depends on three main ingredients: the concentrations of molecules, the kinetics of the binding between micro RNA and target and obviously the topology of the resulting network. The abundances of individual miRNA and target molecules in cells can span many orders of magnitude. Heterogeneities in transcriptional activities are the central mechanism behind this scenario. Furthermore, while cells can regulate the average level of transcription, variations from the average can happen both due to molecular noise and to perturbations.

miRNA binding kinetics depends on the degree of complementarity between the miRNA and its target RNA, with higher complementarities resulting in stronger repressive effects. Recent investigations highlighted a rich scenario of different binding modes and strength [5]. Finally, the wiring observed in real networks shows peculiar statistical features, like broad connectivity distributions.

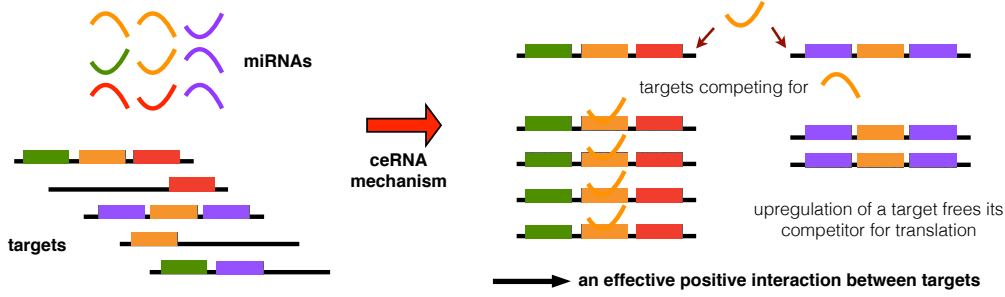


Figure 3.1. Sketch of the ceRNA mechanism: competition to bind a miRNA can induce an effective positive coupling between its targets. Adapted from [3].

3.1 Measuring crosstalk

Given the set of equations 2.3 for the free compound concentrations, we want to characterize the cross-talk established between RNAs by their interaction with miRNAs. Even if two RNAs can not directly interact, they can compete for binding a common miRNA, so that variations in the concentration of one RNA are transferred to the second RNA through the shared miRNA. We will say then that the two RNAs are connected by a path of length one miRNA, meaning that in the bipartite network we can move from the first RNA to the second going through the common miRNA. In principle, crosstalk can take place also between RNAs that are more than one miRNA apart, by a chain of miRNA mediated interactions. Generalizing, two RNAs can influence each other if they are connected by a n -miRNA-long path in the miRNA-RNA network.

The most intuitive measure of crosstalk is the computation of the Pearson correlation between the levels of the free RNA molecules, i.e.

$$\rho_{ij} = \frac{\langle m_i m_j \rangle - \langle m_i \rangle \langle m_j \rangle}{\sqrt{(\langle m_i^2 \rangle - \langle m_i \rangle^2)(\langle m_j^2 \rangle - \langle m_j \rangle^2)}} \quad (3.1)$$

where averages are taken over random fluctuations in the steady state of a stochastic dynamics or over different samples, as long as each sample can be considered to be stationary and the interaction network is conserved across samples. A large positive value of ρ_{ij} points to the existence of a positive (linear) correlation between m_i and m_j .

In such conditions, an increase in the level of ceRNA i , whichever its origin, will divert part of the miRNA population currently targeting ceRNA j to bind to i , thereby freeing up molecules of j for translation. In ρ is large, a perturbation affecting ceRNA i could be ‘broadcast’ to ceRNA j because of the miRNA-mediated statistical correlation existing between their respective levels. However, quantifying the crosstalk through ρ_{ij} may be misleading. To understand why we can imagine a system in the steady-state subject to some noise in the transcription rates of the species. In this situation, the level of RNA i is given by

$$m_i(\{m_j\}, \{\mu_a\}) \simeq \overline{\langle m_i \rangle} + \sum_j \frac{\partial m_i}{\partial b_j} \Big|_{m_i=\langle m_i \rangle} \delta b_j + \sum_a \frac{\partial m_i}{\partial \beta_a} \Big|_{m_i=\langle m_i \rangle} \delta \beta_a \quad (3.2)$$

where $\delta x = x - \bar{x}$ and the over-bar denotes average over transcription rates. Now if we assume that the transcription rates are mutually independent, the Pearson coefficient between any couple of RNAs assumes the form

$$\rho_{ij} \simeq \sum_k \tilde{\chi}_{ik} \tilde{\chi}_{jk} \overline{\delta b_k^2} + \sum_a \tilde{\chi}_{ia} \tilde{\chi}_{ja} \overline{\delta \beta_a^2} \quad (3.3)$$

where $\tilde{\chi}_{ik} = \frac{\partial m_i}{\partial b_k} \Big|_{m_i=\langle m_i \rangle}$, $\tilde{\chi}_{ia} = \frac{\partial m_i}{\partial \beta_a} \Big|_{m_i=\langle m_i \rangle}$ and the index k (resp. a) runs over all RNAs (resp. miRNAs).

From Eq. 3.3, we can see that ρ_{ij} is given by the contribution of both the fluctuations in RNAs and miRNAs. A problem of this measure relies on the fact that a finite value of ρ_{ij} may not be due to the response of ceRNA i to the variation in the level of ceRNA j . In fact, we can obtain a finite correlation also due to a variation in the level of one of the miRNA to which ceRNA j binds. So, ρ_{ij} quantifies a combination of different responses and its usage can be misleading. Anyway, equation 3.3 provides also the answer to the problem. In fact, to be sure that the crosstalk between the couple (i, j) is indeed solely due to the variation of the level of the involved RNA, we can directly compute the *susceptibility*, i.e.

$$\chi_{ij} = d_j \frac{\partial [m_i]}{\partial b_j} \quad (3.4)$$

where d_j serves the only purpose of rendering dimensionless the observable [7]. Using Eq. 2.3, it is possible to derive an exact analytic expression for the susceptibilities that depends from the model parameters, the free concentrations of the involved RNAs (e.g. $[m_i]$ and $[m_j]$) and most importantly from the concentrations of the miRNAs that are present along the path connecting the two RNAs. Note that with slightly different computation details (see Appendix A) we could obtain expression also for the miRNA-miRNA and miRNA-RNA susceptibilities:

$$\chi_{ab} = \delta_b \left(\frac{\partial [\mu_a]}{\partial \beta_b} \right), \quad \chi_{ia} = \delta_a \left(\frac{\partial [m_{ia}]}{\partial \beta_a} \right) \quad (3.5)$$

The scenario described by χ_{ab} is expected to be specular to the one depicted by χ_{ij} , while a proper characterization χ_{ia} is so far missing. Since the author has not properly worked on those quantities, they will not be discussed any further.

Computing the susceptibilities

Focusing on the RNA-RNA susceptibilities χ_{ij} we can explicit the calculations as:

$$\chi_{ij} = d_i \left(\frac{\partial [m_i]}{\partial b_j} \right) = \frac{\partial [m_i]}{\partial m_j^*} = F_i \delta_{ij} + m_i^* \sum_a \frac{\partial F_i}{\partial \mu_a} \sum_l \frac{\partial \mu_a}{\partial [m_l]} \frac{\partial [m_l]}{\partial m_j^*}, \quad (3.6a)$$

$$\chi_{ij} = \frac{[m_i]}{m_i^*} \delta_{ij} + \frac{[m_i]^2}{m_i^*} \sum_{a \in N_i} \frac{[\mu_a]^2}{\mu_{ia}^0 \mu_a^*} \sum_{l \in M_a} \frac{\chi_{lj}}{m_{la}^0} \quad (3.6b)$$

where F_i is defined in Eq. (2.3), N_i is the set of miRNA that target RNA i , M_a is the set of RNA that are bound by miRNA a and we used the identities:

$$\frac{\partial F_i}{\partial [\mu_a]} = -\frac{F_i^2}{\mu_{ia}^0}, \quad (3.7a)$$

$$\frac{\partial F_a}{\partial [m_i]} = -\frac{F_a^2}{m_{ia}^0}. \quad (3.7b)$$

$$(3.7c)$$

Eq. 3.6 can be re-casted in the compact form:

$$\sum_l (\delta_{il} - W_{il}) \chi_{lj} = \frac{m_i}{m_i^*} \delta_{ij}, \quad (3.8a)$$

$$(\widehat{1} - \widehat{W}) \widehat{\chi} = \text{diag} \left(\frac{m}{m^*} \right). \quad (3.8b)$$

where $\widehat{\chi}$ is the entire susceptibility matrix, $\text{diag}(\frac{m}{m^*})$ denotes the diagonal matrix with elements $\{m_i/m_i^*\}$ and

$$(\widehat{W})_{ij} = \frac{[m_i]^2}{m_i^*} \sum_{a \in (N_i \cap N_j)} \frac{1}{m_{ja}^0 \mu_{ia}^0} \frac{[\mu_a]^2}{\mu_a^*} \quad (3.9)$$

where $a \in (N_i \cap N_j)$ stands for all miRNA molecules that bind both RNA i and j . The $\widehat{\chi}$ matrix can be obtained from Eq. (3.8) through matrix inversion:

$$\widehat{\chi} = (\widehat{1} - \widehat{W})^{-1} \text{diag} \left(\frac{m}{m^*} \right) \quad (3.10)$$

Expressions (3.10) allows computing the susceptibility matrix exactly for any miRNA-RNA network. Anyway, expressed in such shape, it is quite hard to see which characteristic of the crosstalk the susceptibility can provide. To understand what we can learn about crosstalk using susceptibilities, let us focus on a minimal network where an approximated solution for the χ matrix can be evaluated.

Crosstalk is selective and directional

Eq. 3.10 provides a powerful recipe to compute the susceptibility, i.e. the competition-induced crosstalk, in an arbitrarily large network. In this form, though it is hard to understand the main features of such crosstalk. To gain some insight, it is more convenient to follow the original computation presented in [4], where we can obtain an approximated (but quite clarifying) expression for the χ_{ij} that depends explicitly on the states (\mathcal{F} , \mathcal{S} , or \mathcal{B}) the couple of RNAs are on.

We consider a system composed of one miRNA species, μ and N RNA species, $\{m_i\}$. Starting from the steady state concentrations of Eq. 2.3 we want to approximate:

$$\chi_{ij} = \frac{\partial [m_i]}{\partial m_j^*} = F_i([\mu]) \delta_{ij} + m_i^* \frac{\partial F_i}{\partial [\mu]} \frac{\partial [\mu]}{\partial m_j^*}, \quad (3.11)$$

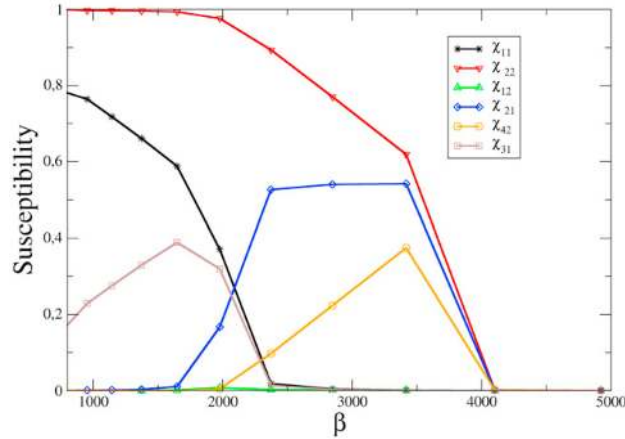


Figure 3.2. Susceptibilities χ_{ij} in a system of 4 ceRNAs, as a function of miRNA transcription rate β . (All other parameters being fixed). In this example, ceRNAs are cast in two groups: group A, formed by ceRNAs m_1 and m_3 , and group B, formed by ceRNAs m_2 and m_4 . ceRNAs belonging to the same group share identical kinetic parameters. In particular, $\mu_1^0 = \mu_3^0 \ll \mu_2^0 = \mu_4^0$. For β smaller than ~ 500 , no cross-talk is observed; however, as β increases, a symmetric interaction between ceRNAs in group A (of magnitude comparable to the self-susceptibilities) appears. As β increases further, this interaction is switched off, and ceRNAs in group B begin to cross-talk instead. In this region, a change of transcription of a ceRNA in group A can affect the level of ceRNAs in group B, but not viceversa (asymmetric cross-talk). Finally, for sufficiently large β , no cross-talk takes place. Adapted from [4].

where $[\mu]$ is the free concentration of the lone miRNA species and

$$F_i([\mu]) = \frac{\mu_i^0}{\mu_i^0 + [\mu]} . \quad (3.12)$$

Remembering that the three states in which one can find RNA i depend on the ratio $[\mu]/\mu_i^0$ being $\ll 1$ (\mathcal{F}), ~ 1 (\mathcal{S}) or $\gg 1$ (\mathcal{B}), we can express Eq. 3.12 in an approximated form for each of the three states, i.e.

$$F_i([\mu]) \simeq \begin{cases} 1 - \frac{[\mu]}{\mu_i^0} & \text{if } i \in \mathcal{F} \\ \frac{1}{2} - \frac{[\mu] - \mu_i^0}{4\mu_i^0} & \text{if } i \in \mathcal{S} \\ \frac{\mu_i^0}{[\mu]} & \text{if } i \in \mathcal{B} \end{cases} ; \quad (3.13)$$

similarly, also the ratio $\frac{\partial F_i}{\partial [\mu]}$ can be approximated as

$$\frac{\partial F_i}{\partial [\mu]} = -\frac{\mu_i^0}{(\mu_i^0 + [\mu])^2} \simeq \begin{cases} -\frac{1}{\mu_i^0} & \text{if } i \in \mathcal{F} \\ -\frac{1}{4\mu_i^0} & \text{if } i \in \mathcal{S} \\ -\frac{\mu_i^0}{[\mu]^2} & \text{if } i \in \mathcal{B} \end{cases} . \quad (3.14)$$

Finally, the term $\frac{\partial [\mu]}{\partial m_j}$ requires an explicit expression for $[\mu]$. To begin with, we note that the concentration of free miRNA in the steady state (Eq. 2.3b) can be

recast as

$$[\mu] \left(1 + \sum_{k \in \mathcal{F}} \frac{m_k^* F_k}{m_k^0} + \sum_{k \in \mathcal{S}} \frac{m_k^* F_k}{m_k^0} + \sum_{k \in \mathcal{B}} \frac{m_k^* F_k}{m_k^0} \right) = \mu^* \quad (3.15)$$

where we grouped together all the ceRNAs in the same state. The product $[\mu]F_k$ also assumes different forms depending on the state of the k -th ceRNA, i.e.

$$[\mu]F_k([\mu]) \simeq \begin{cases} [\mu] & \text{if } k \in \mathcal{F} \\ \frac{[\mu] + \mu_k^0}{4} & \text{if } k \in \mathcal{S} \\ \mu_k^0 & \text{if } k \in \mathcal{B} \end{cases} . \quad (3.16)$$

With these approximations, we can express $[\mu]$ as

$$[\mu] \simeq \frac{\mu^* - \sum_{k \in \mathcal{B}} m_k^* \frac{\mu_k^0}{m_k^0} + \frac{1}{4} \sum_{k \in \mathcal{S}} m_k^* \frac{\mu_k^0}{m_k^0}}{1 + \sum_{k \in \mathcal{F}} \frac{m_k^*}{m_k^0} + \frac{1}{4} \sum_{k \in \mathcal{S}} \frac{m_k^*}{m_k^0}} \quad (3.17)$$

and proceed to compute the last term of Eq. 3.11:

$$\chi_{\mu j} \equiv \frac{\partial[\mu]}{\partial m_j^*} \simeq -\frac{\chi_{\mu\mu}}{m_i^0} \times \begin{cases} [\mu] & \text{if } j \in \mathcal{F} \\ \frac{\mu_j^0 + [\mu]}{4} & \text{if } j \in \mathcal{S} \\ \mu_j^0 & \text{if } j \in \mathcal{B} \end{cases} \quad (3.18)$$

where

$$\chi_{\mu\mu} \equiv \frac{\partial[\mu]}{\partial \mu^*} = \left(1 + \sum_{i \in \mathcal{F}} \frac{m_i^*}{m_i^0} + \frac{1}{4} \sum_{k \in \mathcal{S}} \frac{m_k^*}{m_k^0} \right)^{-1} \quad (3.19)$$

We must now substitute in Eq. 3.11 the expressions found in Eq. 3.13 and 3.14 (which depend on RNA i) and Eq. 3.18 (depending on RNA j). After some calculations, we finally obtain

$$\chi_{ij} \simeq \left[F_i \delta_{ij} + \frac{m_i^* \chi_{\mu\mu}}{4[\mu]} M_{R(i)R(j)} \right] , \quad (3.20)$$

with

$$\hat{M} = \begin{pmatrix} 4 \frac{[\mu]^2}{\mu_i^0 \mu_j^0} & \frac{[\mu]}{\mu_i^0} \frac{[\mu] + \mu_j^0}{\mu_j^0} & 4 \frac{[\mu]}{\mu_i^0} \\ \frac{[\mu]^2}{\mu_i^0 \mu_j^0} & \frac{[\mu]}{\mu_i^0} \frac{[\mu] + \mu_j^0}{4\mu_j^0} & \frac{[\mu]}{\mu_i^0} \\ 4 \frac{\mu_i^0}{\mu_j^0} & \frac{\mu_i^0}{[\mu]} \frac{[\mu] + \mu_j^0}{\mu_j^0} & 4 \frac{\mu_i^0}{[\mu]} \end{pmatrix} = \begin{pmatrix} \mathcal{O}(\epsilon^2) & \mathcal{O}(\epsilon) & \mathcal{O}(\epsilon) \\ \mathcal{O}(\epsilon) & \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(\epsilon^2) & \mathcal{O}(\epsilon) & \mathcal{O}(\epsilon) \end{pmatrix} \quad (3.21)$$

where $R(k)$ represents the state (\mathcal{F} , \mathcal{S} , or \mathcal{B}) of RNA k .

The approximated solution required more calculation than the exact results. Anyway, it is worth the cost, in fact, put in this form, we can spot several features of the crosstalk. Firstly, the resulting cross-talk between RNAs is positively defined since all terms in Eq. 3.20 are positive.

This implies that cross-talk tends to positively correlate the levels of RNAs. Moreover, \hat{M} provides important information since it is the only quantity that

depends on the states of both RNAs. To begin with, the matrix is not symmetric. This means that given a certain level of miRNAs, if the RNAs have different thresholds (μ_0) the crosstalk between RNAs will be directional, with some RNAs influencing the others without being influenced themselves.

More specifically, if $i \neq j$, all the elements of \hat{M} are of order ϵ or smaller with the exception of $M_{\mathcal{S}\mathcal{S}}$ and $M_{\mathcal{S}\mathcal{B}}$ which are of order 1. This implies that, in this scenario, two types of effective interactions arise: the first one encodes the response of a ceRNA in the \mathcal{S} -regime to a perturbation of another RNA in the \mathcal{S} -regime, and it is symmetric; the second one encodes the response of an RNA in the \mathcal{S} -regime to a perturbation of an RNA in the \mathcal{B} -regime, and it is not symmetric (i.e., perturbing the susceptible RNA the bound one will not respond). RNAs in the \mathcal{B} -regime (higher binding affinity) can unidirectionally affect RNAs in the \mathcal{S} -regime, which in turn may influence other RNAs in the \mathcal{S} -regime. On the other hand, RNAs in the \mathcal{F} -regime (lower binding affinity) interact weakly with the rest of the system and fluctuations in their transcription rates do not propagate to other RNAs. All this means that the inverse affinities $\mu_i^0 \sim 1/k_i^+$, induce a hierarchy of interactions. If a couple of ceRNA bind to a miRNA with different affinities, then the response will be different. Furthermore, since cross-talk appears only when the miRNA level is in a specific range, this implies that the structure of the emergent interaction network is flexible and dynamical. The set of RNA species that interact may change upon varying $[\mu]$ or from the cellular point of view, upon changing its transcription/degradation rates.

Both selectivity and directionality as features of the cross-talk can be read looking at the susceptibilities computed for a small system of 4 RNAs as a function of the miRNA transcription rate, β (see Figure 3.2). One sees that different interactions are switched on in different ranges of values for the miRNA transcription rate, leading to a gradual modification of the structure of the interaction network as β changes. Clearly, heterogeneity in transcription rates of both RNAs and miRNA will influence the crosstalk. In fact, the regime an RNA assumes depends on the levels the other species can assume through the ratios μ_a/μ_{ia}^0 , m_i/m_{ia}^0 . So heterogeneity on the transcription rates b_i and β_a which determine the maximum levels of free compounds is expected to influence the crosstalk. At the same time, we can expect that heterogeneity in the quantities μ_i^0 will lead to interaction asymmetries. In Section 3.2, we will see the role both sources of heterogeneity exert on the crosstalk at system level.

The role of topology

Last but not least, topology, i.e. the architecture of the miRNA-RNA network strongly influences the establishment of crosstalk. To appraise its role, again it is better to consider an approximated case in which all ingredients, except the strictly topological ones, are as homogeneous as possible. Let us assume that there is no binding heterogeneity at all ($\mu_{ia}^0 = \mu^0$) and that miRNA levels are held the same for all miRNA species to a certain value $\mu_a = \mu$. In this scenario, the free concentrations of the RNA population are given by

$$[m_i] = \frac{m^*}{1 + n_i t} \quad (3.22)$$

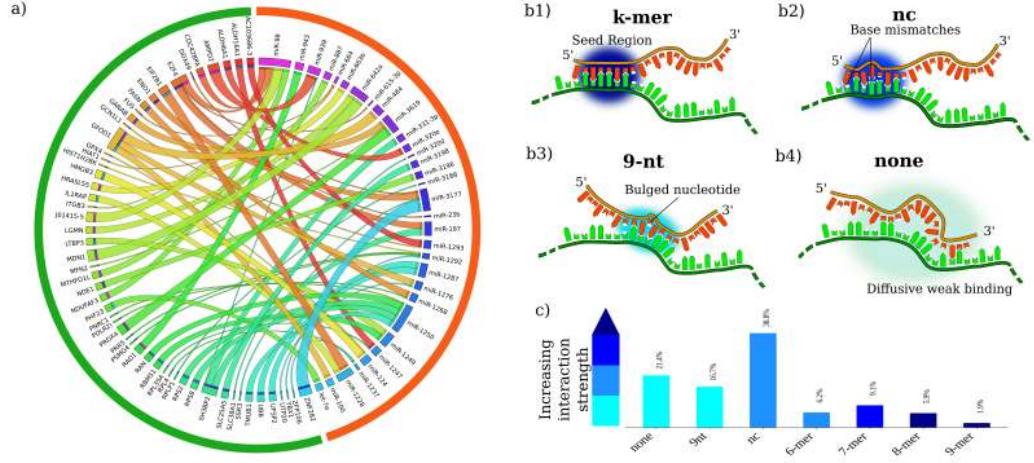


Figure 3.3. Real (CLASH) network features. (a) Circular representation of a small part of the CLASH network. RNA species can crosstalk via miRNA-mediated effective interactions. Crosstalk can be established also between RNAs that do not share a miRNA regulator thanks to chains of miRNA-mediated couplings. (b1-b4) Classes of miRNA-RNA interactions proposed by [5]: (b1) perfect k -base-pairing in the seed region (' k -mer' mode); (b2) seed base-pairing with up to one mismatched or bulged nucleotide (non-canonical mode or 'seed-nc'); (b3) non-seed base pairing with bulged and/or mismatched nucleotides ('noseed-9nt' mode); (b4) non-seed binding within weak diffuse regions ('noseed' mode). (c) Frequency of each binding mode in the CLASH dataset (from [5]).

where $m^* = b/d$, $t = \mu/\mu_0$, and $n_i = \sum_{\alpha} A_{i\alpha}$ is the number of miRNA that target i -th RNA (the degree of node i). We introduced the adjacency matrix $A_{i\alpha}$ as:

$$A_{i\alpha} = \begin{cases} 1 & \text{if RNA } m_i \text{ is targeted by miRNA } \mu_{\alpha} \\ 0 & \text{otherwise} \end{cases} \quad (3.23)$$

Under these approximations Eq. 3.9 becomes:

$$\left(\widehat{W}\right)_{ij} = \frac{t}{(1 + n_i t)^2} \sum_{\alpha \in (N_i \cap N_j)} \frac{1}{\mathcal{K}_{\alpha}} \quad (3.24)$$

with

$$\mathcal{K}_{\alpha} = \frac{m^0}{m^*} + \sum_k^{N^{\alpha}} \left(\frac{1}{1 + n_k t} \right) \quad (3.25)$$

Reminding the geometric series identity $(\widehat{1} - \widehat{Z})^{-1} = \sum_{n \geq 0} \widehat{Z}^n$, we can expand Eq. 3.10 as:

$$\chi_{ij} = \sum_{n \geq 0} \left(\widehat{W}\right)_{ij}^n \frac{[m_j]}{m_j^*} \equiv \sum_{n \geq 0} \chi_{ij}^{(n)} \quad (3.26)$$

Note that for this expansion to converge, we must have that $|\zeta_i| < 1$ for each eigenvalue ζ_i of Z .

Eq. (3.26) permits to obtain an analytical expression of the $\hat{\chi}$ matrix up to the chosen order. In particular, if we assume that the miRNA-ceRNA network is sufficiently sparse and there are no connectivity correlations we can stop to the first order of the series expansion, i.e.

$$\chi_{ij} \simeq \hat{\chi}_{ij}^{(1)} = \frac{t}{(1+n_it)^2} \left(\frac{1}{1+n_jt} \right) \sum_{\alpha \in (N_i \cap N_j)} \frac{1}{\mathcal{K}_\alpha} \quad (\text{if } i \neq j) \quad (3.27)$$

One easily sees that the susceptibility is symmetric only if $n_i = n_j$ for all i and j , i.e. only if RNA nodes have the same degree. Furthermore, as expected, the dilution increases upon increasing the number of ceRNAs interacting with a given miRNA species μ_a . In fact, each of RNA adds a positive term $(1+tn_k)^{-1}$ to \mathcal{K}_a thus making it large. The susceptibility diminishes also upon increasing n_i and n_j , since

$$\chi_{ij} \simeq \frac{1}{n_j n_i^2} . \quad (3.28)$$

To further get insights on the topology contribution, let us now focus on the particular case of a regular bipartite network with fixed RNA and miRNA connectivity so that $n_i = n$ for each i and $\nu_a = \sum_i A_{ia} = \nu$ for each a . We have that

$$\mathcal{K} = \frac{m^0}{m^*} + \frac{\nu}{1+nt} \quad (3.29)$$

From Eq. 3.29, it is clear the dilution effect of miRNA degree. Susceptibilities assume the form

$$\chi_{ij} \simeq \frac{n_{ij}}{\mathcal{K}} \frac{t}{(1+n_it)^3} = n_{ij} \chi_0 \quad (\text{if } i \neq j) \quad (3.30)$$

where $n_{ij} = \sum_\alpha A_{i\alpha} A_{\alpha j}$ is the number of miRNA shared by RNA i and j . We clearly see that the contribution of a single miRNA to the overall susceptibilities depend on the value of t . In particular, one finds that

$$\chi_0 = \begin{cases} \frac{t}{\mathcal{K}} \sim \mathcal{O}\left(\frac{\epsilon}{n}\right) & \text{for } t \ll 1/n \\ \frac{1}{\mathcal{K}n} \sim \mathcal{O}\left(\frac{1}{n}\right) & \text{for } t \simeq 1/n \\ \frac{t}{\mathcal{K}t^2 n^3} \sim \mathcal{O}\left(\frac{\epsilon}{n}\right) & \text{for } t \gg 1/n \end{cases} \quad (3.31)$$

Altogether, also for the topology we can identify three possible regimes. In fact, generalizing what we found varying miRNA transcription rate, one realizes that the case $t \ll 1/n$ (resp. $t \ll 1/n$ and $t \gg 1/n$) describes a ceRNA that is ‘globally free’ (resp. ‘globally susceptible’ and ‘globally bound’) with respect to the overall miRNA population.

Regarding topology, we can conclude that χ_{ij} (i) increases with the number n_{ij} of miRNA species shared by the ceRNAs m_i and m_j ; (ii) decreases if the shared miRNAs have many other targets; (iii) peaks when ceRNAs are ‘globally susceptible’ to the overall miRNA population, and it can be of the same order of magnitude

as the self-susceptibility, i.e. $\mathcal{O}(1/d)$, when $n_{ij} \simeq n$. Perhaps most remarkably, the crosstalk can be effective even among ceRNAs that are in the free regime with respect to individual miRNAs, provided they are commonly targeted by a large number of miRNA species thus becoming globally susceptible. However, in order to achieve efficient cross-talk strong correlations in the network connectivity are needed (large n_{ij}): highly clustered networks can allow for much stronger cross-talk than random graphs. Finally, the regular bipartite example can be used to appraise the contribution to the susceptibility given by higher orders of expansions of Eq. 3.26. In fact, if we look at the second order term

$$\begin{aligned} \widehat{\chi}_{ij}^{(2)} &= \left(\sum_l W_{il} W_{lj} \right) \frac{m_j}{m^*} = \\ &= \left(\frac{m^* \mu_0}{\mu^* m_0} \right)^2 \left(\frac{t}{1+nt} \right)^2 \left(\frac{1}{1+nt} \right) \sum_l n_{il} n_{lj} \left(\frac{t}{1+nt} \right)^2 \end{aligned} \quad (3.32)$$

we see that each order carries the contribution due to chains of miRNA-mediated interactions. Indeed, truncation at the first order is possible if correlations higher than degree-degree ones are negligible, i.e. $\langle n_{il} n_{lj} \rangle \sim 0$ as much as all higher correlations given by chains like $\langle n_{ij} n_{jk} n_{kl} \dots \rangle \sim 0$.

Summary

Summing up, when quantified through χ_{ij} , ceRNA crosstalk displays the following key features:

- **Plasticity:** The patterns of miRNA-mediated RNA crosstalk are modulated by kinetic parameters, and particularly by miRNA levels, i.e changes in miRNA availability modify the RNA crosstalk network;
- **Selectivity:** If a miRNA targets multiple ceRNA species, crosstalk may occur only among a subset of them and is enhanced by heterogeneities in the thresholds. This effect is related to the fact that different ceRNAs can have different thresholds for repression by the miRNA;
- **Directionality (asymmetry) :** In general, $\chi_{ij} \neq \chi_{ji}$, i.e. RNA i may respond to a perturbation affecting ceRNA j but not the reverse;
- **Dependency on stoichiometric processing:** If all miRNA-RNA complexes formed by RNA j are degraded in a purely catalytic way, then $\chi_{ij} = 0$ (i.e. stoichiometric processing is necessary for RNA crosstalk at stationarity).
- **Topology:** many weakly interacting miRNA species can collectively mediate efficient RNA crosstalk.

Now that we learned what features the crosstalk, as quantified by susceptibility, exhibits, one can wonder to which degree the above features persist in a real large-scale network, where transcriptional heterogeneity combines with binding heterogeneity and broad distributions of topological features like node degrees. In

the next sections, we will see how these three different kinds of heterogeneity affect the emergent cross-talk patterns and look for evolutionary signatures since both binding and network architecture are under the direct influence of evolution.

3.2 Crosstalk at system level

We now know how crosstalk looks like in particular situations of small motifs or in fictitious large regular graphs. To probe the modelization in more realistic scenarios we must both look for experimental miRNA-RNA networks and think about some proper measures to quantify features as the plasticity or the selectivity. In fact, we can not use the plots in Figure 3.2 for large networks. Fortunately, experimental techniques and bioinformatics tools provided the first ingredient. In fact, searching literature [5, 59] and databases like miRbase [42] or TargetScan [43], we can retrieve different interaction networks. For example, we consider the human miRNA-RNA interaction network proposed by [5], which was obtained using the novel cross-linking, ligation, and sequencing of hybrids (CLASH) experimental protocol. The network, to which we refer as CLASH network, is made of 17411 links shared between 6943 RNAs (N) and 383 miRNAs (M) (see Figure 3.3a for a sketch). Links are weighted according to the miRNA-RNA binding affinities, which strongly depend on the modes of binding. Adopting the same notation used in [5], we distinguish (see Figure 3.3b) four binding modes based on (i) the perfect pairing of k miRNA seed nucleotides (' k -mer', with $k = 6, 7, 8$ or 9), (ii) the presence of up to one imperfect pairing in the seed region (non-canonical, 'nc'), (iii) a no seed interaction, allowing bulged nucleotides in the target ('9nt'), or (iv) a no seed binding with distributed less stable pairings ('none'). Repression experiments, performed increasing the concentration of sets of miRNA with the same binding mode, ranked the repression strengths associated with each binding mode. In particular, [5] showed the seven kinds of interactions grouped in four sets according to their repression effect, namely 8 and 9mers, being the group with the strongest binding and repressive effect, followed by 7mers; 6mer and non-canonical interactions behave similarly in terms of repression capacity as well as 9nt and none binding modes (see Figure 3.3c).

Sources of heterogeneity

All the parameters introduced in the previous section present a certain degree of variability at physiological conditions. In what follows, we will try to understand how this heterogeneity influences the crosstalk scenario at system level. For simplicity of exposition, we can divide the source of heterogeneity in three distinct groups. The first source of heterogeneity arises in the transcription rates. In fact, the presence of extrinsic noise introduces a variability in the rates of synthesis and degradation of the molecules across cells. Following [6], we restraint to the sole variability in the transcription rates and assume that both RNA transcription rates b_i ($i = 1, \dots, N$) and miRNA transcription rates β_a ($a = 1, \dots, M$) can be modeled as random variables extracted from some distribution. In particular, the discussed results are obtained assuming log-normal i.i.d. random variables with means $\langle b \rangle$ and $\langle \beta \rangle$, and variances σ_b^2 and σ_β^2 , respectively.

CLASH - Crosslinking, Ligation, and Sequencing of Hybrids

CLASH maps RNA-RNA interactions [64, 5]. In this method, RNA-protein complexes are UV-crosslinked and affinity-purified. RNA-RNA hybrids are ligated, isolated, and reverse-transcribed into cDNA. Deep sequencing of the cDNA provides high-resolution chimeric reads of RNA-RNA interactions. The main advantages are that (i) it allows to map RNA-RNA interactions *in vivo* and (ii) provides binding site level resolution. On the other side, the CLASH protocol makes use of hybrid ligation which may be difficult to obtain between short RNA fragments. Furthermore, it has a relatively low efficiency [65]

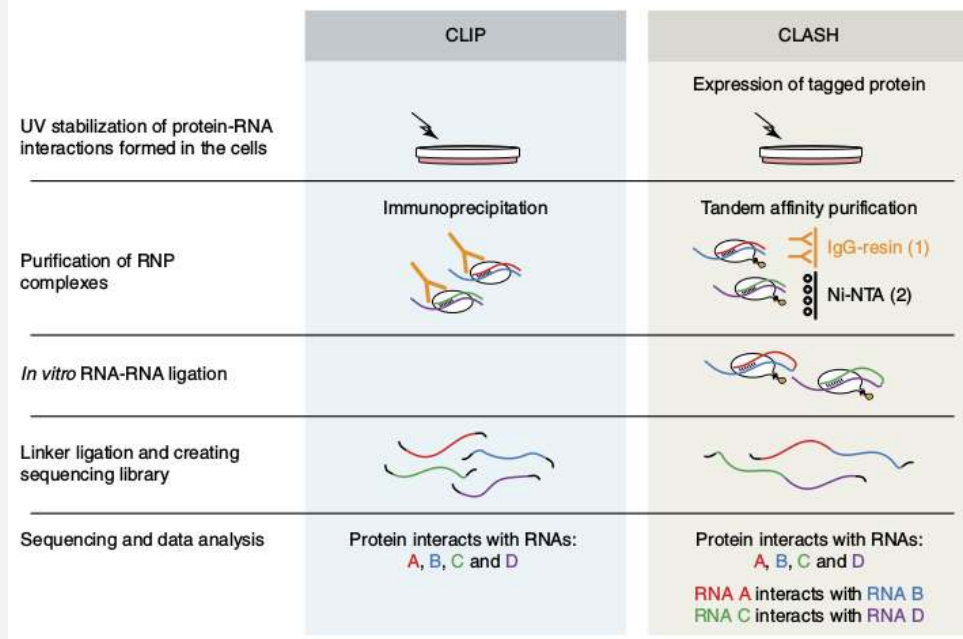


Figure 3.4. Comparison between CLIP and CLASH experimental protocol.

We will abbreviate the transcriptional heterogeneity as TH. A used measure for the magnitude of the fluctuation is the coefficient of variation of individual rates (standard deviation over mean), which we denote by CV_{tr} . So, assuming comparable magnitudes for fluctuations in the transcription rates of miRNAs and RNAs, we identify TH by the coefficient of variation.

Another source of heterogeneity in the system comes from the strengths of the bindings (binding heterogeneity, BH). The association (k^+) and dissociation rates (k^-), which quantify the binding strength between miRNA and RNA molecules, depend on the base-pairing between the molecules. Note that while cells can regulate transcription rates, the BH can not be regulated by the cells, but it is subject to evolutionary pressure on longer time scales.

Under the assumption made in the previous section, k^+ constitutes the main source of variability in the binding weights, μ_0 and m_0 . To appraise how heterogeneities in the miRNA-RNA interaction strengths affect the emergent crosstalk landscape, we can imagine three variants, with increasing diversity, of the structure of

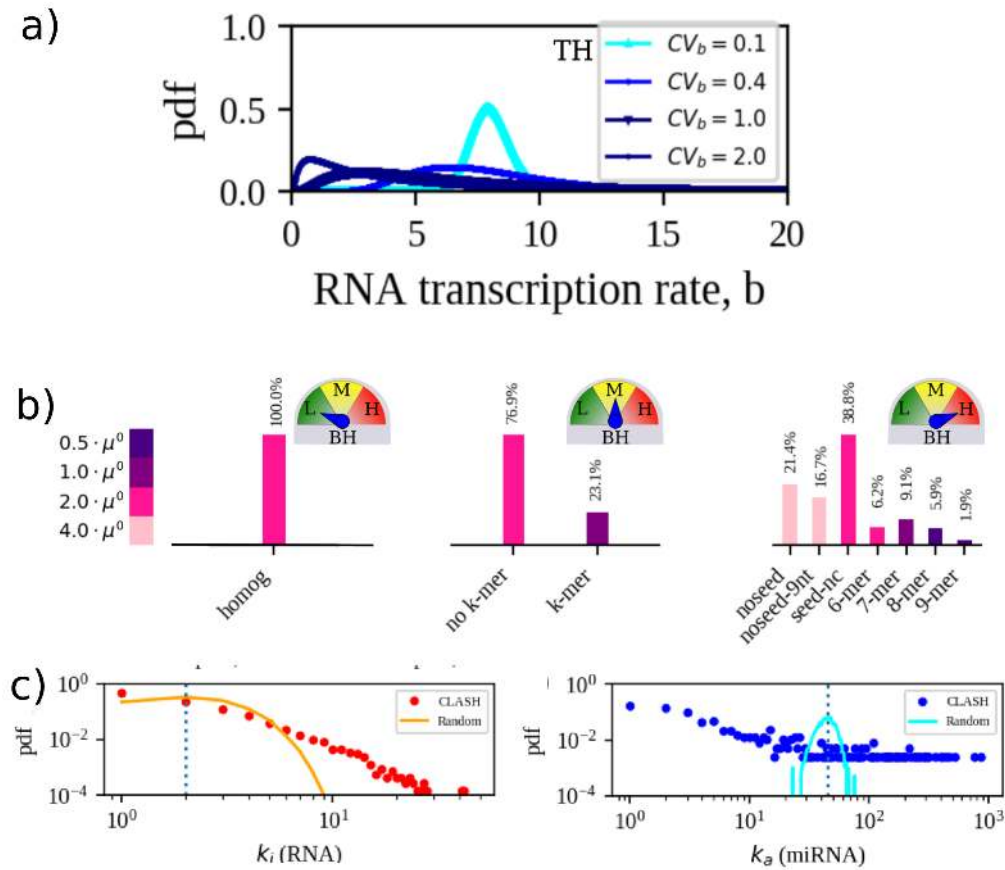


Figure 3.5. Modeling heterogeneities. (a) Distributions of RNA transcription rates used in this work: each rate is assumed to be drawn independently from a log normal distribution with given mean (same for each RNA species). Increased transcriptional heterogeneity (TH) corresponds to increased values of the relative fluctuations (CV_b). (b) Scenarios of miRNA-RNA binding heterogeneity (BH). From left to right: low BH, where each miRNA-RNA pair interacts with the same strength; medium BH, with k-mer interactions (stronger) distinguished from the rest (weaker); high BH, where the full structure is employed. (c) CLASH ceRNA and miRNA degree distributions in the CLASH network and in a randomized version having the same number of contacts and node degrees. Adapted from [6].

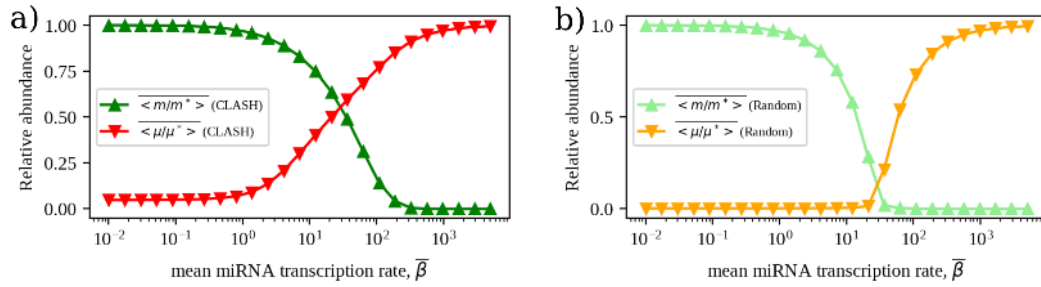


Figure 3.6. Relative overall molecular abundances in CLASH (a) and degree-preserving randomized networks (b). Note that the susceptible regime in the latter is narrower compared to the original CLASH network. Adapted from [6].

binding affinities encoded in the CLASH interactome (see Figure 3.5). At the lowest level, we assume a homogeneous network in which $\mu_{ia}^0 = \mu_0$ for each miRNA-RNA pair, with μ_0 a certain constant (see Table 2.1). To express an intermediate level of heterogeneity, we can imagine a bimodal distribution that discriminates (i, a) pairs having a k -mer interaction according to the CLASH data (i.e. stronger coupling rate) from the rest. At the highest level, we associate different binding strengths to each of the four types of miRNA-RNA pairs considered in CLASH, assuming a 2-fold change in μ_{ia}^0 between groups in agreement with estimates from [5].

Finally, the very same network topology can constitute another layer of heterogeneity. In fact, the network architecture can possess a broad node degree distribution which intuitively influences the interactions between compounds. Indeed, experimental inferred miRNA-RNA networks exhibit power-law distribution of degrees with exponents ranging from -3 to -1 both for miRNA and RNAs species (see Figure 3.5c).

To evaluate the role of topological heterogeneities, we can compare results for the real network with those found in ensembles of networks obtained (i) by re-assigning each (i, a) link to a miRNA-RNA pair drawn randomly among all possible pairs with equal probability; or (ii) by a more conservative procedure based on degree-preserving edge-swaps [66].

The first type of re-wiring disregards topological correlations of all orders, including node connectivities [67]. The second type instead preserves the real degree distributions, while it destroys all higher structures.

3.2.1 Probing CLASH crosstalk

Once reasonable choices for all the parameters have been made, we can use Eq. 2.3 and Eq. 3.4 to obtain the steady state concentrations of all miRNA and RNA molecules and the exact values of all the N^2 susceptibilities. Since we are dealing with $\sim 10^5$ RNA species, we can not hope to understand the behavior of the system by simply looking at plots of the susceptibilities as functions of the parameters like we did before. Instead, we must think at some clever system observables. To begin with, let us identify the region in which the network is globally susceptible. We can do so by looking at the behavior of the mean re-scaled populations:

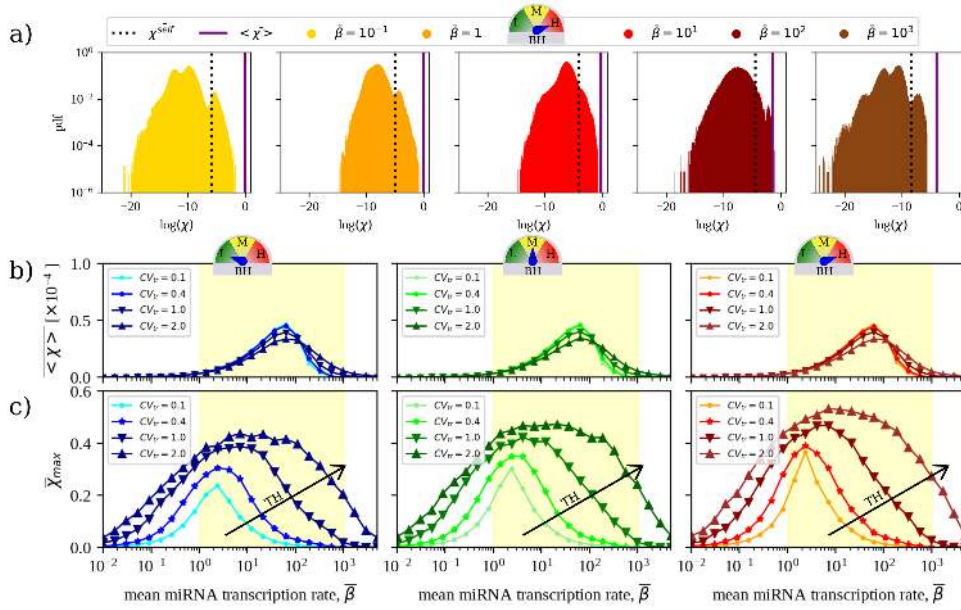


Figure 3.7. Quantitative characteristics of RNA crosstalk in the CLASH network. (a) Representative distributions of susceptibilities obtained for the CLASH interactome for five different realizations of parameters with different values of $\tilde{\beta}$, $CV_{tr} = 0.4$ and maximal BH. (b) Mean susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity. (c) Maximal susceptibility as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity. The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6].

$$\rho_m = \frac{1}{M} \sum_i \frac{[m_i]}{m_i^*} \quad \rho_\mu = \frac{1}{N} \sum_a \frac{[\mu_a]}{\mu_a^*} \quad (3.33)$$

Each specimen concentration is divided by the maximum value it can reach if all molecules were unbound. These two quantities range from 0, where all species are bound to 1 in which molecules are all in the free state. Figure 3.6 displays the mean re-scaled populations as a function of the mean miRNA transcription rate $\tilde{\beta}$. (The mean transcription rate of miRNA is often used as a control parameter in experiments since it is both simple to tune but also because it is indeed the parameter that cells can change.) As one would expect, the degree of RNA repression increases with β . In the region around miRNA-RNA equimolarity, the system becomes susceptible.

The position of the susceptible region along the $\tilde{\beta}$ axis is strongly influenced by the stoichiometricity ratio λ . An increase (resp. decrease) of λ shifts the susceptible region to higher (resp. lower) values of the mean miRNA transcription rate. This can be understood noting that changes in λ , in turn, effectively modify the ratio μ_{ia}^0/m_{ia}^0 of interaction thresholds. In particular, when λ increases (resp. decreases)

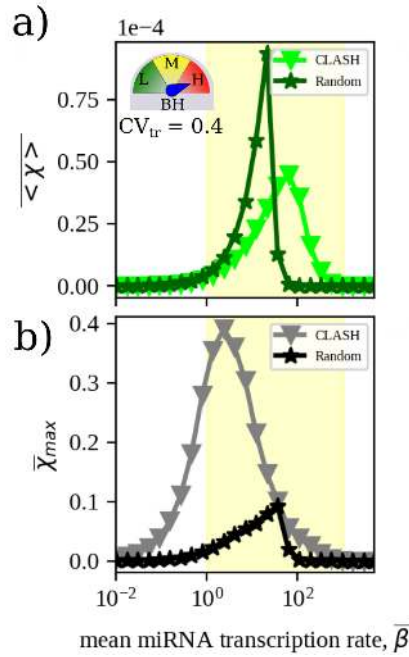


Figure 3.8. Crosstalk in CLASH and random network. (a) Mean susceptibility as a function of the overall miRNA mean transcription rate $\langle \beta \rangle$. (b) Maximal susceptibility as a function of the overall miRNA mean transcription rate $\langle \beta \rangle$. The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6].

ceRNAs become less prone to interact (resp. more prone to interact) with miRNAs, so that more (resp. less) miRNAs will be required to repress ceRNAs. The most intuitive measure to take is the mean susceptibility $\langle \chi \rangle$, where the brackets $\langle \dots \rangle$ denote an average over all pairs of RNA species such as $i \neq j$, while the over-bar stands for an average over different realizations of transcription rate profiles at fixed CV_{tr} .

$\langle \chi \rangle$ informs about the typical strength of RNA crosstalk in the network. In figure 3.7b, it is shown as a function of the mean transcription rate of miRNAs. We can note two things. Firstly, miRNA availability modulates $\langle \chi \rangle$ so that it peaks within the susceptible region and is vanishingly small outside of it, where molecular levels are practically unaffected by varying miRNA transcription rates. Secondly, this picture is substantially unchanged by modifying the degrees of TH and/or BH, save for a modest expansion of the susceptible region. Such a behavior therefore describes a ‘basal level’ of crosstalk that occurs in the network in any given condition. Indeed, because of miRNA titration, some degree of crosstalk is to be expected between randomly selected pairs of RNAs no matter what. In numerical terms, it is rather weak: a change in the transcription level of an RNA on average leads (via miRNAs) to a change in the level of another RNA by something of the order of 0.01% of the size of the perturbation. If we visualize susceptibilities, for example looking at the their distributions at different values of β and fixed values of all

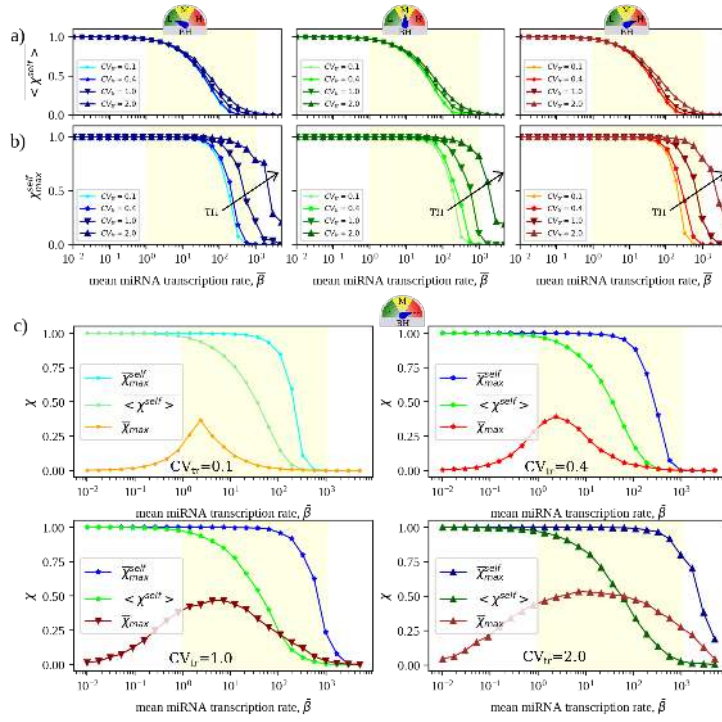


Figure 3.9. Self and cross-susceptibilities. (a) Mean self-susceptibility as a function of the mean miRNA transcription rate $\bar{\beta}$. (b) Mean maximal self-susceptibility as a function of the mean miRNA transcription rate $\bar{\beta}$. Results are shown for the 3 BH scenarios considered. The yellow shaded area qualitatively marks the region where the mean susceptibility is significantly different from zero, which coincides with the susceptible regime. In each case, the standard error of the mean is equal to or smaller than the size of the markers. The self-susceptibility is maximal when miRNA levels are low, in which case the availability of free RNA molecules increases roughly linearly with the transcription rate. As $\bar{\beta}$ increases, miRNA repression gets stronger and self-susceptibilities decrease until, at large enough $\bar{\beta}$, RNAs are fully repressed and therefore insensitive to small changes in their transcription rates. (c) Comparison between maximum self-susceptibility, mean self-susceptibility and $\bar{\chi}_{max}$ for different degrees of TH in the high BH scenario. The intensity of crosstalk between different RNAs, measured by the latter quantity, is indeed of the same order of magnitude as self-susceptibilities. Adapted from [6].

other parameters (see Figure 3.7a), the first thing one notes is that susceptibility values span several orders of magnitude. The mean is obviously not enough to properly describe the system. In this respect, it is interesting to see how extreme values behave. In particular Figure 3.7c displays the behavior of the mean maximum susceptibility

$$\bar{\chi}_{max} = \overline{\max_{(i \neq j)} \chi_{ij}} \quad , \quad (3.34)$$

where the maximum is taken over all pairs of different RNA species (i.e. with $i \neq j$) as a function of $\bar{\beta}$.

We can see that χ_{ij}^{max} increases going from $CV_{tr} = 0.1$ to $CV_{tr} = 2$ and going from low BH to high BH. This behavior suggests that heterogeneities both in transcription rates and in kinetic parameters enhance cross-talk intensity. In particular, the increase of

maximal susceptibility with the amplitude of RNA transcriptional fluctuations in contrast with the robustness of the mean susceptibility to parameter variability is a marker of the effect of heterogeneities in creating specific channels of RNA-RNA cross-talk: the mean susceptibility does not depend on noise amplitude, but some susceptibilities do increase, so only a limited number of RNA pairs reach sizeable coupling strength. $\bar{\chi}_{\max}$ quantifies the maximum achievable intensity of crosstalk interactions in each scenario, therefore providing a proxy for the strength of the most significant miRNA-mediated couplings arising between different RNA species in the network.

To appraise its significance, one can gauge $\bar{\chi}_{\max}$ against the self-susceptibility $\chi_{ii} = d_i \frac{\partial m}{\partial b_i}$, which quantifies the change in the level of free transcripts of species i induced by a small modification of its own transcription rate (note that by definition $\chi_{ii} \leq 1$). Self susceptibilities pose an upper bound against which we can evaluate the intensity of the relevant crosstalks for each choice of the parameters. In particular from Figure 3.9, one sees that $\overline{\langle \chi \rangle}$ is about four orders of magnitude smaller than the mean self-susceptibility. In this respect, basal crosstalk appears to be on average very weak. Also, the strongest crosstalk interactions are 4 orders of magnitude above basal and tend to occur between RNA pairs that are co-regulated (i.e. share some miRNA). Furthermore distant (not co-regulated) RNAs can still give rise to significant crosstalk, with susceptibilities that are 2 orders of magnitude above $\overline{\langle \chi \rangle}$.

The emerging picture is that the miRNA regulatory layer may be serving a two-fold role. On the one hand, it makes the system more homogeneous with respect to differences in the transcription rates, exerting a stabilizing crosstalk (mean susceptibility). On the other hand, it allows a certain number of mRNA pairs to be strongly interdependent (maximum susceptibility). Since placing both stabilization and enhanced interdependence of transcript levels under the term ‘crosstalk’ might be confusing, in the following, we will refer to crosstalk having in mind the extreme values of the susceptibility.

At last, to appraise the role of the specific wiring encoded by the CLASH data in determining the scenario described so far, we can compare our results against a null model obtained by randomly re-wiring the CLASH interactome. Specifically, if we re-assigned each link to a randomly chosen miRNA-RNA pair, thereby preserving only the overall numbers of links and nodes while altering all other topological features like node degrees, degree-degree correlations, etc we see (Figure 3.8) that randomized networks display a much larger (about two-fold) mean susceptibility for crosstalk than the CLASH interactome, possibly due to the fact that miRNA targets are generically closer in the randomized versions. However, the maximum achievable crosstalk strength $\bar{\chi}_{\max}$ is about 4 times smaller in the random networks compared to CLASH. Moreover, the susceptibility profile is more concentrated in the randomized network than it is for the CLASH network, reflecting a significantly narrower susceptible region. This picture is consistent with the fact that the susceptibility is proportional to the number of miRNA species co-targeting RNAs i and j [3]. Furthermore, since the wiring of the network is strongly determined by natural selection this suggests that natural selection fosters the emergence of stronger crosstalk links.

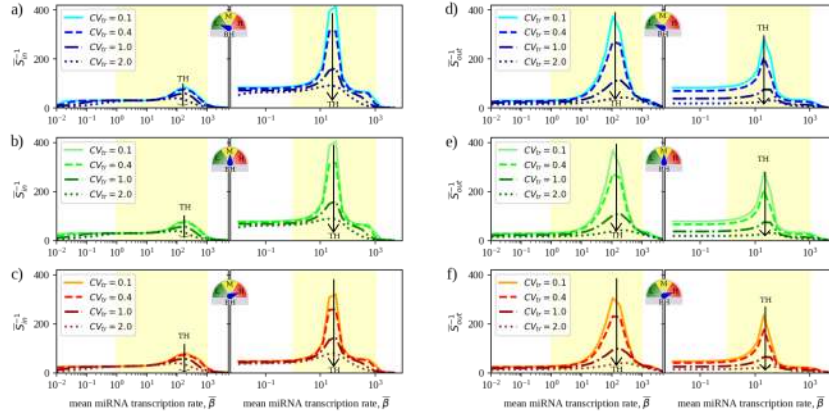


Figure 3.10. Crosstalk selectivity in CLASH network. Inverse of incoming (S_{in}^{-1} on left column) and outgoing (S_{out}^{-1} on right column) selectivities as a function of miRNA mean transcription rate, β . The inverses of S_{in} and S_{out} give a proxy of the number of mRNAs receiving or propagating the effects of a perturbation. CLASH network is very selective until the susceptible region is reached, i.e. both the number of nodes whose variation influences a given mRNA and the number of nodes that are influenced by variation of a certain node is very low. Upon reaching the susceptible threshold the network starts cross-talk and the number of nodes involved in perturbations drastically increases. The random network presents a lower selectivity with respect to real and reshuffled ones, with a peak of cross-talk in the susceptible region. Adapted from [6].

Quantifying the selectivity

Intuitively, if a cell wants to achieve a good degree of control over perturbations carried out on a certain mRNA, it is necessary that the number of mRNAs responding to the perturbation is not too large [68]. As we saw, analytical results obtained for small motif showed that competitive interactions can give rise to a rather selective RNA-RNA communication channel, as only RNAs in the susceptible regime are responsive to changes in the miRNA transcription rate and are coupled by cross-talk interactions at stationarity [4] (during transients, cross-talk is instead extended [57]). To quantify selectivity in a large network, we can introduce the following measures:

$$S_{in} = \frac{1}{M} \sum_i g_i \quad S_{out} = \frac{1}{M} \sum_j h_j \quad (3.35a)$$

$$g_i = \frac{\sum_{j \neq i} \chi_{ij}^2}{(\sum_{j \neq i} \chi_{ij})^2} \quad h_j = \frac{\sum_{i \neq j} \chi_{ij}^2}{(\sum_{i \neq j} \chi_{ij})^2} \quad (3.35b)$$

Note that both g_i and h_j vary between 0 and 1, as do S_{in} and S_{out} . The rationale for these quantities is the following. A value $g_i \approx 0$ indicates that a large number of RNA species can almost equally affect the steady state of RNA i , whereas a value of g_i close to 1 indicates that RNA i has a small number of RNA regulators. Likewise, when $h_j \approx 0$ a perturbation of the transcription rate of RNA j affects the steady state of a large number of other RNA species almost equally, while if $h_j \approx 1$ RNA j only has a small number of targets.

The quantities g_i and h_i are based on the standard definition of the inverse participation ratio (IPR), widely used in statistics to quantify the number of ‘states’ over which a certain variable is distributed. Given a probability distribution $\{p_i\}$ over N states, such that $\sum_{i=1}^N p_i = 1$, the IPR is given by

$$IPR = \left(\sum_{i=1}^N p_i^2 \right)^{-1} \quad (3.36)$$

One sees that if the variable is strongly concentrated, e.g. if $p_i = 1$ and $p_j = 0$ for $j \neq i$, then $IPR = 1$ (the variable occupies one state). If instead the variable is distributed equally over all states, i.e. if $p_i = 1/N$ for all i , then $IPR = N$ (the variable occupies N states). Therefore, it is the inverse of g_i (or of S_{in} and S_{out}) that provides the key information. In particular, S_{in}^{-1} (respectively S_{out}^{-1}) represents the average of the number of RNAs, a perturbation of which can considerably affect the level of a given RNA (resp. whose level can be affected by a perturbation of a given RNA). We will call S_{in} the incoming RNA-RNA selectivity and S_{out} the outgoing RNA-RNA selectivity.

In Figure 3.10, the inverse incoming (left) and outgoing (right) selectivities are showed as a function of the mean miRNA transcription rate $\bar{\beta}$. The network is rather selective outside the susceptible region, where both the number of RNAs that affect a certain node and the number of nodes being affected in return is comparable with the average degree of RNA nodes. On the other hand, where the system enters the susceptible region, it becomes distributed, i.e. perturbations on one RNA species propagates through the whole network. Notably, transcriptional heterogeneity has a strong impact on selectivity. In fact, the higher the degree of transcriptional heterogeneity the more crosstalk becomes selective at every level of miRNA concentrations. On the other hand, different degrees of binding heterogeneity (BH) appear to impact this scenario rather weakly.

3.2.2 Crosstalk asymmetry

Another feature of the crosstalk, we identified in the previous sections, is its directionality. In fact, we saw that in principle $\chi_{ij} \neq \chi_{ji}$, and this inequality originates from the states in which the couple of RNAs are. To measure crosstalk directionality, we can define the quantity

$$\Delta_{ij} = \left(\frac{\chi_{ij} - \chi_{ji}}{\chi_{ij} + \chi_{ji}} \right)^2, \quad (3.37)$$

such that $0 \leq \Delta_{ij} \leq 1$. In short, the closer Δ_{ij} is to zero (resp. one) the closer crosstalk between RNAs i and j is to being symmetric (resp. fully asymmetric). A global measure of asymmetry is conveniently obtained by computing the average asymmetry over all pairs of different RNAs in the network, i.e.

$$\Delta = \frac{2}{N(N-1)} \sum_{i,j>i}^N \Delta_{ij}. \quad (3.38)$$

Results for this quantity are reported in Fig 3.11 for both the CLASH network and its randomized variant. Crosstalk asymmetry is generically larger in the susceptible

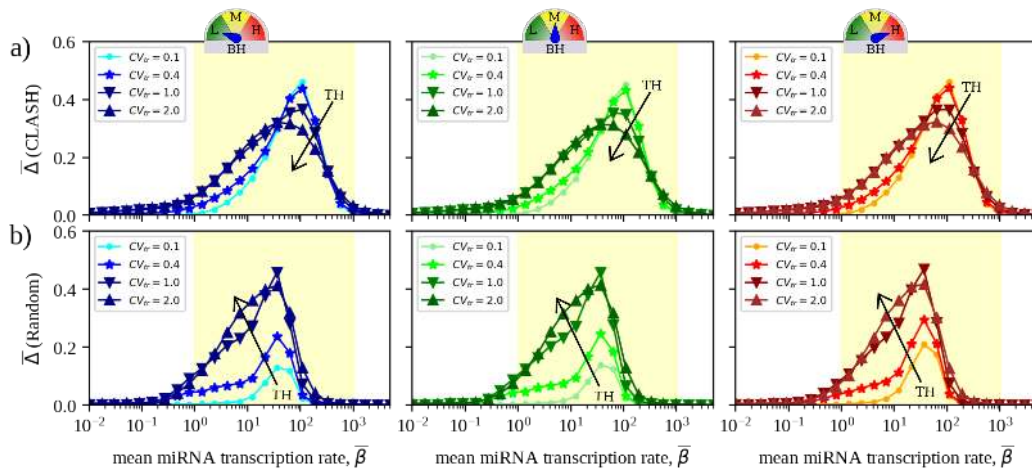


Figure 3.11. Measure of RNA asymmetry (a) and (b) CLASH and random asymmetries values, Δ , as a function of the overall miRNA mean transcription rate $\langle\beta\rangle$ in the 3 scenarios considered for binding heterogeneity. The yellow shaded area marks the region where the mean susceptibility is significantly different from zero (which coincides with the susceptible regime). Displayed curve points have a Standard Error of the Mean (SEM) comparable to the size of the markers. Adapted from [6].

regime, more pronouncedly so in the CLASH network than in its randomized version. Notably, the asymmetry profile is roughly independent of the degree of binding heterogeneity while it is only weakly modulated by transcriptional variability in the CLASH network. As seen for the mean crosstalk intensity, this state of things suggests that the way in which crosstalk asymmetry is tuned by the mean miRNA transcription rate $\bar{\beta}$ is an inherent property of miRNA-RNA networks, that is mainly encoded in their topology. The striking difference that can be seen between the behavior of Δ in real (CLASH) and random networks (see Fig 3.11b) supports this intuition.

Degree of cross-talk localization

The limited selectivity achievable in the susceptible regime suggests that the correlation length, i.e. the typical node-to-node distance (in terms of links of the miRNA-mRNA interaction networks) above which one node can be considered to be insensitive to perturbations carried out on another node, becomes comparable to the diameter of the network (that for the CLASH network amounts to 5 miRNAs). In this case, a perturbation in the transcription level of one mRNA could be broadcast (via a chain of miRNA-mediated effective interactions) across other mRNA nodes until it propagates over the entire network. Such a feature appears to be intrinsically due to the competition mechanism that generates effective couplings rather than to the topology of the underlying network. When network-scale effects dominate, however, higher-order interactions (passing via multiple miRNA-mediated steps) can become dominant. In such conditions, relying on local kinetic parameters to identify strongly coupled mRNAs may turn out to be a poor strategy. In order to dissect the

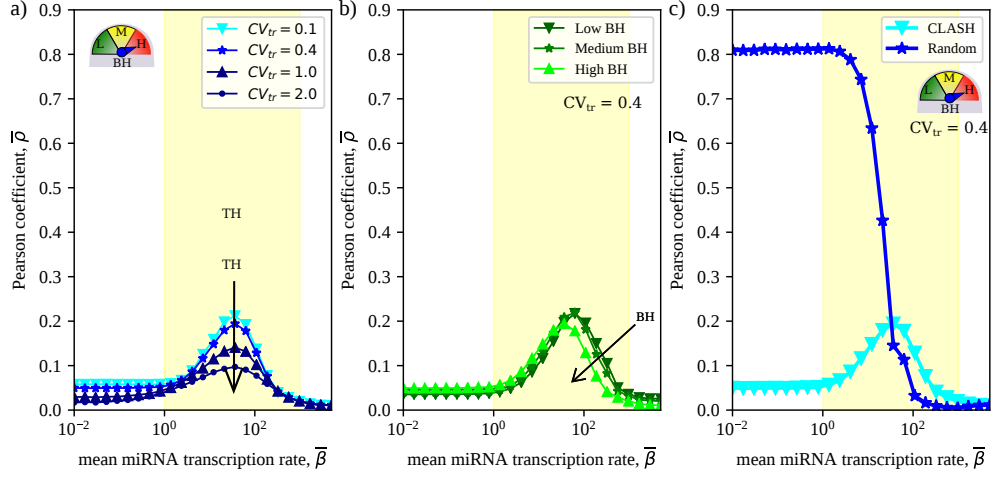


Figure 3.12. Degree of cross-talk localization (ρ) as a function of β . a) CLASH network ρ for different values of transcriptional heterogeneity at fixed (high) BH. b) CLASH network ρ for different values of binding heterogeneity at fixed TH. c) Cross-talk localization for CLASH and random network. Adapted from [6].

relationship between the local topology and the intensity of effective interactions we can estimate the correlation between the susceptibility of a couple of mRNAs and their local interaction, i.e. the strength of the parameters that link the mRNAs through the shared miRNAs.

A commonly used quantity, in estimating correlations, is the Pearson coefficient:

$$\rho = \frac{\langle \chi_{ij} K_{ij} \rangle - \langle \chi_{ij} \rangle \langle K_{ij} \rangle}{\sqrt{(\langle \chi_{ij}^2 \rangle - \langle \chi_{ij} \rangle^2) (\langle K_{ij}^2 \rangle - \langle K_{ij} \rangle^2)}} \quad (3.39)$$

where the brackets $\langle \rangle_{i \neq j}$ stand for the average over all ordered pairs of different mRNAs and K_{ij} quantifies the local interaction strength:

$$K_{ij} = \frac{1}{N} \sum_{\alpha} k_{i\alpha}^+ k_{j\alpha}^+ \quad (3.40)$$

Based on the locality of the sequence-based interaction mechanism, one would expect to observe a high correlation with local kinetics (corresponding to $\rho \approx 1$) when effective interactions are short-ranged and the ceRNA effect mostly couples nodes that are nearest neighbors on the network. On the other hand, a smaller correlation should characterize cases in which network effects become important. As shown in Figure 3.12, where the Pearson coefficient is plotted as a function of β , we see that while the correlation peaks in the susceptible region, crosstalk patterns generically appear to correlate poorly with local topology in the CLASH interactome, as $\rho \simeq 0.2$. Most notably, correlation decreases significantly as TH is strengthened. This implicates kinetic heterogeneities in the establishment of extended interaction paths that reduce the effective diameter of the interactome by

connecting distant RNAs via miRNA-mediated interactions. In this respect, miRNAs appear to operate on RNAs both as specific repressors of individual transcripts and as a diffuse regulatory layer affecting the transcriptome as a whole. This indicates that relying on local kinetic parameters to predict strongly coupled mRNAs or preferred miRNA targets is a rather poor strategy as network-scale effects are typically non negligible.

Things are different for a random network, in fact for the latter the correlation coefficient remains above 0.8 along all the selective region as shown in Figure 3.12, where the Pearson coefficient is plotted as a function of $\bar{\beta}$. It diminishes in the susceptible/distributed region and then goes to zeros in the bound region. miRNA-mediated crosstalk in random networks is therefore significantly more local, and thereby easily predictable by local interaction parameters, than it is in a network shaped by natural selection. Remarkably, by applying a more conservative protocol that reshuffles miRNA-RNA links while preserving node degrees, one retrieves a crosstalk scenario that is essentially identical to that found for the original CLASH interactome (see Appendix B). Overall, this picture indicates that degree sequences (i.e. the topology of miRNA-RNA interactions encoded by the different types of couplings), as opposed to e.g. degree-degree correlations or other higher-order topological features, are the key geometric controllers of RNA crosstalk patterns. Enhanced crosstalk and stronger non-locality therefore appear to be shaped by selection through the miRNA-RNA network interaction structure.

3.3 Transcriptional noise processing

Regulatory RNAs are known to be capable of processing the variability of transcript populations, leading, in several cases, to the fine tuning of expression levels [3]. In fact, the effective interactions that are established between RNA influence the fluctuations in the level of the free molecules. In particular, in our scenario, since all rates are independent random variables, fluctuations are expected to increase. In Figure 3.13, we show the coefficient of variation of RNA levels, averages being taken over many independent realizations of TH, as a function of the mean miRNA transcription rate $\bar{\beta}$ in different BH scenarios. Relative fluctuations exhibit a maximum at large values of $\bar{\beta}$ within the susceptible region and generically increase with the degree of TH. Variability in transcription rates therefore expectedly promotes variability in the resulting expression profiles. However, in an extended range of values of $\bar{\beta}$ within the susceptible region, the increase of fluctuations with respect to the unregulated case $\bar{\beta} \rightarrow 0$ is very modest. On the other hand, at fixed TH, different BH scenarios do not appear to modify the robustness of expression profiles. At the same time, expression profiles generated in the randomized network are more stable than those found in the CLASH interactome. This feature is however more marked at higher miRNA expression levels, where RNA crosstalk is generically weaker. The basic traits of the RNA crosstalk emerging in fully randomized versions of the CLASH data are hence substantially different from those characterizing the original interactome.

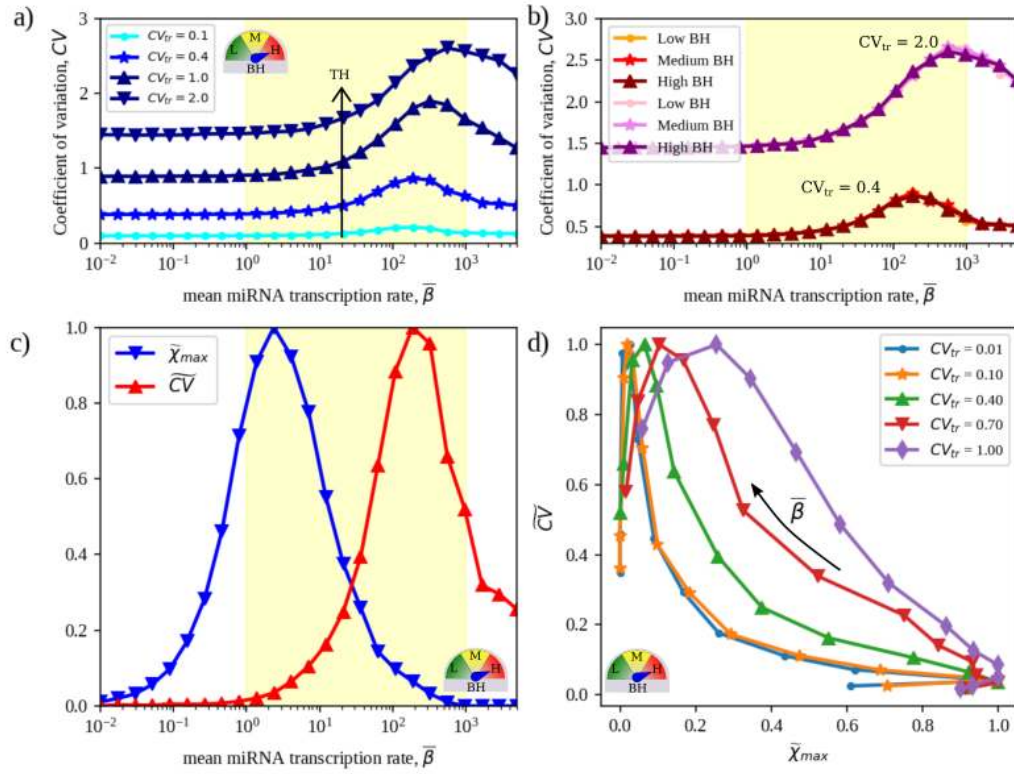


Figure 3.13. Robustness of expression profiles from the CLASH interactome in the presence of crosstalk. (a) Coefficient of Variation (CV) of RNA levels as a function of the overall mean miRNA transcription rate $\bar{\beta}$ for different degrees of TH. (b) Behaviour of the CV as a function of $\bar{\beta}$ in different BH scenarios in a given TH scenario ($CV_{tr} = 0.4$). (c) Comparison between the rescaled normalized maximal susceptibility $\bar{\chi}_{max}$ (varying between 0 and 1) and the rescaled normalized Coefficient of Variation as a function of the overall mean miRNA transcription rate $\bar{\beta}$ in a given CV for different degrees of TH ($CV_{tr} = 0.4$) and BH (high) scenario. (d) $\bar{\chi}_{max}$ vs \bar{CV} transcriptional heterogeneity (CV_{tr}) and high BH. Results are obtained by averaging over 100 independent TH realizations. In each case the standard error of the mean is equal to or smaller than the size of the markers. Adapted from [6].

3.3.1 Tradeoff between crosstalk and noise buffering

If we compare the behavior of the maximal susceptibility χ_{\max} (see Figure 3.7b) to that of the CV, we can notice that the strongest maximal crosstalk is attained within the susceptible region corresponding to more robust expression profiles and, vice-versa, stronger fluctuations in expression profiles occur when crosstalk gets weaker.

In other terms, uncorrelated transcriptional heterogeneities tend to be amplified when crosstalk is suppressed (higher miRNA expression levels), while they are more efficiently processed when the strongest crosstalk emerges (lower miRNA expression levels). This scenario is summarized in Figure 3.13a: for any given degree of TH, as miRNA availability increases, crosstalk intensity on one hand and fluctuations of the output levels on the other are subject to a tradeoff that becomes stronger and stronger as the transcription rates become more homogeneous. These results clearly implicate transcriptional heterogeneities as a key observation on small networks [3, 69]. It is however important to remark that this picture is obtained under the assumption of uncorrelated extrinsic fluctuations in RNA transcription rates.

3.4 Experimental validation of crosstalk

At last, we are left with the most compelling question: how do we validate the model? A first step in this direction was already done. In fact, results so far discussed were obtained using an experimentally validate network of interaction (i.e. using the CLASH protocol or the TargetScan database). Furthermore, the key parameters were taken from literature (Table 2.1). Anyway many assumptions were made and a more decisive validation is due. Clearly, the most direct route would be through experimental work. Although, there exists an intermediate route, i.e. make use of transcriptomic data and exploit the relationships between susceptibilities and correlation functions derived in [70] to locate strongly interacting ceRNA pairs from a statistical analysis of readouts. Indeed, such an analysis is currently in progress as this thesis is being written, so we are not able to present results. We want to discuss the basic idea.

Fluctuation dissipation relation revisited

As we already discuss, the physical meaning and therefore the crosstalk scenarios underlain by the Pearson correlation ρ_{ij} , and the susceptibility, χ_{ij} are rather different. The fact that χ_{ij} is asymmetric under exchange of its indexes (i.e. $\chi_{ij} \neq \chi_{ji}$ in general) whereas correlation is necessarily symmetric already pointed in this direction. If the two quantities are compared in greater detail other differences however emerge. In first place, susceptibility can be non zero (and possibly large) even for a completely deterministic system, as it simply measures how a target's steady state level is modulated by changes affecting the transcription rate of one of its competitors, independently of the presence of stochastic fluctuations around the steady state. In this sense, χ_{ij} focuses exclusively on the effects induced by competition. On the other hand, in absence of fluctuations ρ_{ij} is identically zero. Second, in general is true that if A is correlated with B and B is correlated to C, then A will result to be

correlated with C. In this case, one speaks of ‘indirect correlation’. It follows that a large value of ρ_{ij} can occur when both ceRNAs respond to fluctuations in miRNA levels. This however does not imply that m_i is responding to a direct variation of m_j . Operatively, in absence of a direct correlation between A and C, upon conditioning over the value of B one will observe that A and C are uncorrelated. The same holds in the presence of extrinsic noise, in which case averages are performed over different samples rather than over time in a single sample.

It remains that measuring correlation is the easier thing one can do from transcriptomic data readouts. Indeed, for physical system where a linear response theory can be applied, a relationship between the correlation function and the susceptibility can be derived. This relation is known as fluctuation-dissipation theorem [71]. Luckily, a similar relationship can be obtained also for the system under exam. To see how, let us follow the work done in [70]. We start from the set of equations 3.41a. Assuming that complex dynamics is much faster than those of the free molecules, we can suppose that complex achieved a steady state regime while free molecules still vary in time. If we also assume that stoichiometric degradation without recycling is the dominant channel of complex processing (i.e. $\kappa_{ia} + k_{ia}^- \ll \sigma_{ia}$), we can recast Eq. 3.41aa,b as

$$[\dot{m}_i] = -m_i \frac{\partial \mathcal{L}}{\partial m_i} \quad (3.41a)$$

$$[\dot{\mu}_a] = -\mu_a \frac{\partial \mathcal{L}}{\partial \mu_a} \quad (3.41b)$$

$$(3.41c)$$

where,

$$\mathcal{L} = - \sum_i (b_i \ln(m_i) - d_i m_i) - \sum_a (\beta_a \ln(\mu_a) - \delta_a \mu_a) + \sum_{i,a} (k_{ia}^+ m_i \mu_a) \quad (3.42)$$

Since \mathcal{L} decreases along the dynamics of molecules free concentrations (as one can easily see by total differentiation of \mathcal{L}), we can imagine that once the evolution reaches the steady state, i.e. \mathcal{L} is in its minimum, small perturbations of the concentrations (or equivalently the presence of some noise) make the system oscillate near the minimum. This behavior closely resembles that of a physical system at equilibrium inside a potential. In this framework, we can express ensemble averages of quantities like the mean level of free RNA as

$$\langle m_i \rangle = \frac{1}{Z} \sum_{m_j, \mu_a} m_i e^{-\mathcal{L}/T} \quad (3.43)$$

where T quantifies the strength of the noise (as temperature expresses the level of thermal noise).

Eq. 3.43 provides a direct way to link the (dimensional) susceptibilities to correlation functions. In fact, differentiating 3.43 by b_j we have

$$\chi_{ij} = \frac{\partial \langle m_i \rangle}{\partial b_j} = \frac{1}{T} \langle m_i \ln(m_j) \rangle - \langle m_i \rangle \langle \ln(m_j) \rangle = \frac{1}{T} \text{cov}(m_i, \ln(m_j)) \quad (3.44)$$

The system response to a small variation in the transcription rates is proportional (via the constant Γ) to the fluctuations induced by noise. Note that the presence of the logarithm in the covariance renders it asymmetric like the susceptibility. Surprisingly, if we differentiate according to the RNA degradation rate, d_j we find the real covariance between RNA levels

$$w_{ij} = \frac{\partial \langle m_i \rangle}{\partial d_j} = \frac{1}{T} \langle m_i m_j \rangle - \langle m_i \rangle \langle m_j \rangle = \frac{1}{T} \text{cov}(m_i, m_j) \quad . \quad (3.45)$$

In [70], the authors validated this model in the case of a small motif. Our aim is to extend the application to a system scale network like CLASH and look for crosstalk patterns between distant couples under physiological conditions, where the role of crosstalk is still under validation.

Chapter 4

Phenotypic diversity

The phenotypes of an organism result from the expression of the genetic information (its genotype) and the influence of environmental factors. To a great extent, the manifested phenotypes correlates with the set of proteins the cell manages to produce [72, 73]. For instance, an important phenotype like the cell growth rate is linked to the amount of nutrients the cell is able to import from the environment and its capability of turning it in biomass (mostly proteins). The biochemical processes leading to the synthesis of new proteins are noisy, as they typically involve a small number of diffusing molecules. This leads to fluctuations in the number of proteins in a single cell as a function of time and to cell-to-cell variability of protein abundances. These in turn bring cell population (even genetically identical one) to exhibit a heterogeneous distribution of their phenotypes. For many years the investigation of such heterogeneities was limited since experiments were able to probe just mean behaviors [74]. Advances at the experimental level mostly due to the possibility of labeling cells with fluorophores allowed to begin investigating cell-to-cell variability. Indeed, a lot of experimental work has been done to characterize the growth rate, which is considered one of the most important traits of cells since it gives a proxy for the fitness. The distribution of growth rates, in physiological conditions, turned out to have a unimodal shapes peaked at low values of growth rates but with tails toward much higher (up to two order of magnitude) rates [75, 76, 7]. In other words, they present a high degree of heterogeneity. The emerging view is that such phenotypic heterogeneity have important consequences for the development of multicellular organisms and the fitness of cell colonies, especially those subject to fluctuating adverse environments [77]. This enforced the question of how such heterogeneity arises and how/if it is regulated. In this Chapter, we will try to convey some ideas on those questions. In particular, we will discuss the work done in [7, 78], where the macroscopic features of phenotypic distributions are obtained starting from minimal assumptions on the network of metabolic molecular reactions (the microscopic states). The authors focused on the cell growth rate. Hence we will do the same, both at the level of single cell growth rate and as the growth rate of the whole cellular population.

4.0.1 Maximum entropy distributions

Intuitively (but with a lot of simplification), the growth rate depends on the amount of energy the cell manages to find, process and use to synthesize the required proteins. We know that all those metabolic processes are governed by a complex network of molecular reactions.

To model the cellular metabolic activity in a given medium, the standard *in silico* route relies on constraint-based models [79, 80]. To make a long story short, the basic assumption of those models consist in assuming that the intracellular reaction network that processes nutrients (e.g. glucose) to harvest free energy and synthesize macromolecular building blocks (e.g. amino acids) operates at a non-equilibrium steady state (NESS), such that instead of looking at the number of molecules at a certain time, it is more convenient to consider fluxes of molecules. Following [78], we can describe each viable metabolism by a vector $v = \{v_i\}$, representing the metabolic reaction fluxes ($i = 1, \dots, N$ with N the number of reactions). Note that N includes also the fluxes that determine exchanges of metabolites between the cell and the surrounding environment, thus defining the composition of the growth medium.

Clearly, the flux vectors can not take any possible values. Each flux has its own range of variability $[v_{\min}^i, v_{\max}^i]$, which accounts for thermodynamic irreversibility, kinetic bounds and other physiological or regulatory constraints [81]. Moreover, supposing that the cell is in a close system with respect to chemicals, fluxes must satisfy the mass-balance conditions

$$\mathbf{S} \cdot \mathbf{v} = 0 \quad (4.1)$$

where \mathbf{S} denotes the $M \times N$ stoichiometric matrix (M being the number of chemical species). In genome-scale models for specific organisms, like *E. coli*, \mathbf{S} is reconstructed from genomic data and represents a network with thousands of reactions and chemical species [82]. Likewise, upper and lower bounds for fluxes are normally available based on prior biochemical knowledge. The solutions of the mass-balance equation subject to the constraints on the maximum and minimum fluxes define the space \mathbb{F} of feasible flux vectors \mathbf{v} (the phase space). Therefore for a given organism and a given medium, each solution describes a feasible NESS of the network, or in other terms a ‘phenotype’. In fact, if we know the biomass output of all the reactions, we can found the corresponding growth rate $\lambda = \lambda(v)$ that a cell with those fluxes would have [83].

Now, if we consider a genetic identical population of cells embedded in the same environment they should in principle have the same flux vector and so the same growth rate. Though, we know that at least thermal noise affects all the reactions, so the resulting fluxes differ from cell to cell. Considering a cellular population and measuring the growth rate of each cells equals to perform a statistical sampling of \mathbb{F} , which in turn produces a distribution of growth rates relative to the specific environment defined in \mathbf{S} , to the constraints imposed by the bounds on fluxes, and to the chosen statistics.

The easier thing we can do is to regard all viable fluxes as equally probable and perform an uniform sampling of the phase space. With such a statistics, the growth-rate distributions retrieved in a rich and a poor growth medium are shown in Fig. 4.1. Both curves are fit by the formula

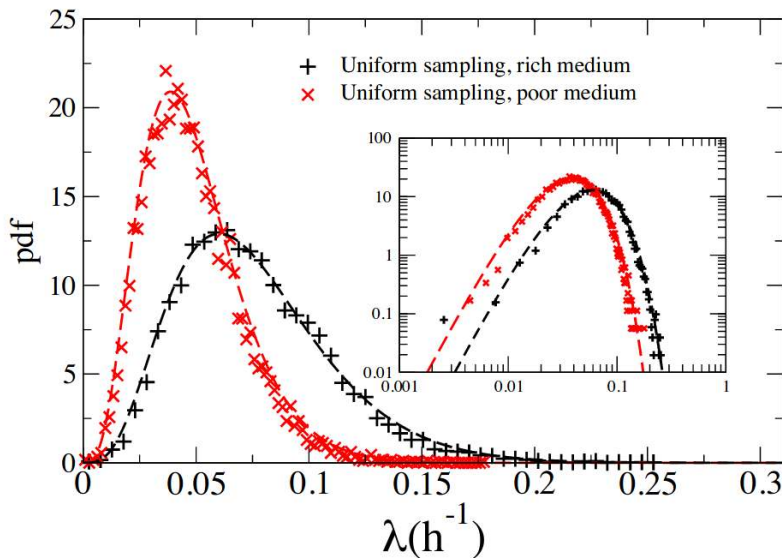


Figure 4.1. Growth rate distributions for a uniform sampling of *E. coli*'s genome-scale metabolic network in a glucose-limited medium. Inset: same on a log-log plot. Adapted from [7].

$$q(\lambda) \simeq \lambda^b \left(1 - \frac{\lambda}{\lambda_{\max}}\right)^a \quad (4.2)$$

where $a > 0$ and $b > 0$ are constants, while λ_{\max} is the maximum growth rate allowed by the constraints that define \mathbb{F} . A uniform sampling of the feasible space is the less biased choice we can make if we know nothing about the system (in fact for us all flux vectors should be good choices). It is remarkable that the resulting phenotype distribution turns out to be highly heterogeneous, meaning that if fluxes are random cells seldom can acquire high growth rates (and so high fitness). Another important feature is that in poor medium, the population is composed by cells with lower growth rates than in the presence of a favorable (nutrient rich) environment.

Since growth rates are easy enough to measure experimentally, we can compare predictions with experiments. Comparing the distributions obtained by the uniform sampling with experimental one in the same environmental conditions [75, 76], one find that although the shape of the distributions are quite similar, experimental curves settle on higher mean values (see Figure 4.2). A uniform sampling of the phase-space seen from the point of view of the cell means that cells reactions are completely dominated by noise and the only bounds are imposed by physical limits. In real cells, things works differently, in fact, regulatory networks, like the crosstalk we discussed before, have the precise task of controlling the noise and the rates at which reactions take place.

It seems quite reasonable then to assume that cells manage through regulation to constraint fluxes so to have a certain mean growth rate. Accordingly, we must require the sampling not to be uniform, but to ensure an average growth rate. In this case, the maximum-entropy distribution with prescribed mean assumes the form

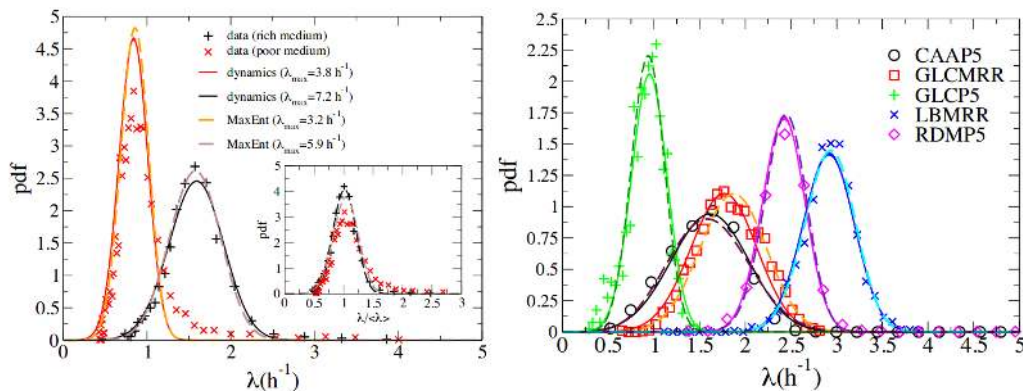


Figure 4.2. Empirical growth rate distributions (markers), together with the best fitting MaxEnt distributions (dashed lines in both panels) and the distributions derived from the dynamical model (straight lines in both panels). Adapted from [7].

of a Boltzmann distribution,

$$p(v) = \frac{e^{\beta\lambda(v)}}{Z(\beta)} \quad (4.3)$$

where β is the Lagrangian multiplier enforcing the constrain

$$\langle \lambda \rangle = \int_{\mathbb{F}} \lambda(v)p(v)dv \quad (4.4)$$

The case of a uniform sampling corresponds to having $\beta \ll 1$. On the other hand, if $\beta \gg 1$, one will tend to extract values of growth rates near to λ_{\max} .

From a biological point of view, flux vectors varies due to the presence of noise in the molecular processes. What regulation does is to select the optimal flux vector in response to different environmental stimuli. Here, we can see the parallelism with β . In fact, the case $\beta = 0$ equals to having no regulation. All fluxes are equally probable and noise differentiates the vectors. On the other hand, if the cell want to acquire higher growth rates ($\beta > 0$) it needs to invest energy in some degree of regulation. The higher the growth rate the more the regulation.

This gives also an hint on the effects of environmental changes: cells with high growth rate require high amount of nutrients to maintain the high level of regulation. If we suppose that the environment is represented by the amount of nutrients, a sudden negative fluctuation of the environment will affect more the cells with high growth rates than the one in low regulation regimes. We will see in 5.3 how a trade-off between low and high fitness cells is beneficial for the population living in a changing environment.

4.0.2 A dynamical model of population growth

Up to now, we discussed a rather static scenario where cells assume different growth rates depending on the degree of regulation (embodied by β), which fixes the mean growth rate of the population and on the noise that spreads the distribution. Apparently though, we are neglecting an important point. In fact, each cell of the

population grows and duplicates. Since $q(\lambda)$ admits cells with large growth rates, those cells should rapidly overgrow the slower ones. This apparent contradiction can be solved if one thinks at the probability to pass from one phenotype to the other. The maximum entropy framework that yielded the $q(\lambda)$ informs us that fast phenotypes are entropic unfavored states. In fact, since very few microscopic flux realizations correspond to high growth regimes, noise on reactions will continuously tend to divert those states toward slower one. Only a tight regulation can maintain those states. And such a regulation is expensive for the cell.

In terms of dynamics, all this can be seen as biased diffusion in the phenotypic space. To see how, let us for simplicity assume that each cell is fully characterized by a single phenotype, its GR λ , taking on values in $[0, \lambda_{\max}]$. And let us denoted by $n(\lambda, t)$ the density of cells having GR in $[\lambda, \lambda + d\lambda]$ at time t , such that the total number of cells in the population at time t is given by

$$N(t) = \int n(\lambda, t) d\lambda \quad . \quad (4.5)$$

Following e.g. [7], we require that n changes due to (a) replication events and (b) diffusion in the phenotypic space, whereby cells change their GR from λ to λ' . If the rate of the latter process is given by $W(\lambda \rightarrow \lambda')$, $n(\lambda, t)$ evolves according to

$$\frac{dn(\lambda, t)}{dt} = \lambda n(\lambda, t) + \int [W(\lambda' \rightarrow \lambda)n(\lambda', t) - W(\lambda \rightarrow \lambda')n(\lambda, t)] d\lambda' \quad . \quad (4.6)$$

Evolving the number of cells without imposing a limiting number soon produces huge populations, which are difficult to handle in simulations. We can solve the problem with a trick: introducing the population density

$$p(\lambda, t) = \frac{n(\lambda, t)}{N(t)} \quad , \quad (4.7)$$

we can directly evolve the distribution (which must be normalized at each time). Operatively, we re-cast Eq. (4.6) as

$$\frac{dp(\lambda, t)}{dt} = [\lambda - \langle \lambda \rangle] p(\lambda, t) + \int [W(\lambda' \rightarrow \lambda)p(\lambda', t) - W(\lambda \rightarrow \lambda')p(\lambda, t)] d\lambda' \quad , \quad (4.8)$$

with

$$\langle \lambda \rangle = \int_0^{\lambda_{\max}} \lambda p(\lambda, t) d\lambda \quad . \quad (4.9)$$

The time evolution of the phenotypic distribution is given by two contributions, a pure replication and a pure diffusive terms. In particular, first term on the r.h.s of Eq. (4.8) can be obtained noting that

$$\dot{n}(\lambda, t) = \dot{p}(\lambda, t)N(t) + p(\lambda, t)\dot{N}(t) \quad , \quad (4.10)$$

where, from Eq. (4.6), we have

$$\frac{\dot{N}(t)}{N(t)} = \int_0^{\lambda_{\max}} \lambda p(\lambda, t) d\lambda = \langle f \rangle . \quad (4.11)$$

A comparison between Eqs (4.6) and (4.10) immediately yields the replicator term.

To insert the diffusion bias toward lower values of λ we can require that transition rates satisfy a detailed-balance condition of the form

$$W(\lambda \rightarrow \lambda')q(\lambda) = W(\lambda' \rightarrow \lambda)q(\lambda') , \quad (4.12)$$

with $q(\lambda)$ the density of phenotypes, accounting for the fact that some phenotypes might be easier to attain than others. Although the detailed balance condition poses some constrain on the transitions, it is not sufficient to determine the whole W matrix. The most natural route to model the effects induced at phenotypic level by small random fluctuations in intracellular composition, is to assume that only transitions from phenotype λ to phenotypes $\lambda \pm \delta\lambda$ are allowed, with equal probability and small $\delta\lambda$. A diffusive transition kernel seems quite realistic since noise-driven transitions are unlikely to cause major gains or losses in terms of GR.

The last ingredient we need is a temporal scale, dictating the characteristic time of the diffusions. In this respect, we can introduce the mean waiting time τ characterizing transitions via

$$\int W(\lambda \rightarrow \lambda') d\lambda' = \frac{1}{\tau} . \quad (4.13)$$

If we assume the diffusive transition kernel and make use of the detailed balance condition (4.12), we can manipulate the second term on the r.h.s of Eq. (4.8) as

$$\begin{aligned} & \int \left[W(\lambda' \rightarrow \lambda) p(\lambda', t) - W(\lambda \rightarrow \lambda') p(\lambda, t) \right] d\lambda' \\ & \simeq \lim_{\delta\lambda \rightarrow 0} W(\lambda + \delta\lambda \rightarrow \lambda) p(\lambda + \delta\lambda, t) - W(\lambda \rightarrow \lambda + \delta\lambda) p(\lambda, t) + \\ & \quad + W(\lambda - \delta\lambda \rightarrow \lambda) p(\lambda - \delta\lambda, t) - W(\lambda \rightarrow \lambda - \delta\lambda) p(\lambda, t) = \\ & \quad = W(\lambda \rightarrow \lambda + \delta\lambda) q(\lambda) \left[\frac{p(\lambda + \delta\lambda, t)}{q(\lambda + \delta\lambda)} - \frac{p(\lambda, t)}{q(\lambda)} \right] \\ & \quad + W(\lambda - \delta\lambda \rightarrow \lambda) q(\lambda - \delta\lambda) \left[\frac{p(\lambda - \delta\lambda, t)}{q(\lambda - \delta\lambda)} - \frac{p(\lambda, t)}{q(\lambda)} \right] \\ & \quad \simeq \frac{(\delta\lambda)^2}{2\tau} \frac{\partial}{\partial \lambda} \left[q(\lambda) \frac{\partial}{\partial \lambda} \frac{p(\lambda, t)}{q(\lambda)} \right] , \quad (4.14) \end{aligned}$$

where the last step follows after a second order expansion in $\delta\lambda$ and we imposed that transitions from λ to $\lambda \pm \delta\lambda$ happen with the same probability (implying that $W(\lambda \rightarrow \lambda \pm \delta\lambda) = (2\tau)^{-1}$, see (4.13)). Defining $D = \frac{(\delta\lambda)^2}{2\tau}$, as the diffusion coefficient in the phenotypic space, one can recast Eq. 4.8 as

$$\frac{dp(\lambda, t)}{dt} = (\lambda - \langle \lambda \rangle) p(\lambda, t) + D \left[\frac{\partial^2 p(\lambda, t)}{\partial^2 \lambda} - \frac{\partial}{\partial \lambda} \left[p(\lambda, t) \frac{\partial}{\partial \lambda} (\ln q(\lambda)) \right] \right] . \quad (4.15)$$

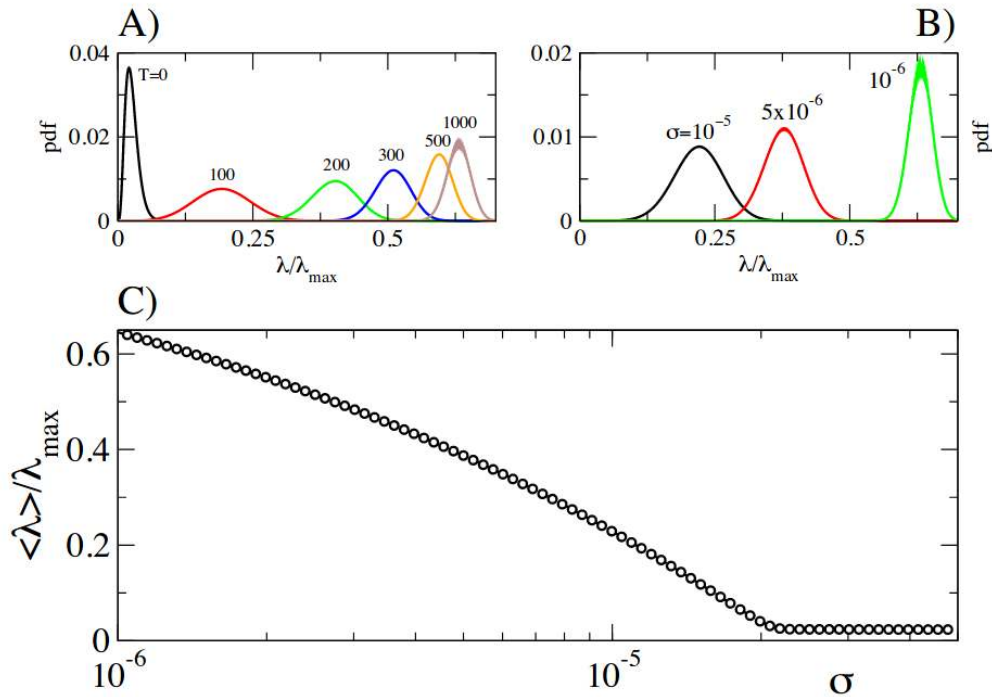


Figure 4.3. Results from the minimal population dynamical model. a) Time-evolution of $p(\lambda)$. b) Stationary growth rate distributions obtained for different values of $\sigma = D/\lambda_{\max}^3$. c) Stationary mean growth rate $\langle \lambda \rangle$ as a function of σ . Adapted from [7].

The time evolution of the phenotypic distribution in Eq. 4.15 follows a non-linear Fokker-Planck dynamics. In Figure 4.3a, we can see some snapshots of the evolution for a small value of the diffusion coefficient. The initial distribution, which peaks on small values of GR, rapidly shifts toward high growth rate values. This means that the replicator term dominates: cells with high GR replicates faster than the time required to change their phenotypes. As the diffusion increases, the distribution tends to the $q(\lambda)$ (see Figure 4.3b). If we look at the mean growth rate, we see that it monotonically decreases as a function of the diffusion coefficient (Figure 4.3c).

The diffusion coefficient behaves in opposition with the inverse temperature β we introduced in the maximum entropy framework. The uniform sampling limit corresponds, in fact, to having a vanishingly small β and a fast phenotypic diffusion and viceversa. Indeed, one could expect that, since in real systems the diffusion coefficient is proportional to the temperature of the system. In our model, the effective temperature or the diffusion coefficient represent the degree of regulation the cells manage to acquire. The population can achieve a high growth rate (i.e. a high fitness), if it can find enough nutrients to exert a tight regulation (and so reduce the diffusion) which allows the distribution to settle near λ_{\max} . If the environment is unfavorable, for example because the medium is poor of nutrients, the population can invest in regulation less energy and thus acquire lower fitness (see Figure 4.1). In the next Chapter, we will study the effects these heterogeneous distributions of phenotypes have on the growth of the population. In particular, in 5.3, we will see

that the model we have derived can be used to study the growth in a fluctuating environment, where favorable (nutrient rich) states alternate to unfavorable ones.

Chapter 5

Population growth and survival

Stochasticity at the molecular level produces diversity in cellular populations, as measured by the overall numbers of specific phenotypes different cells manifest. Such differences reflect in the fitness of the individuals to the environment it lives in. For instance, a cell that expresses a higher number of heat shock protein will withstand better temperature fluctuations and so it will possess a high fitness in a fast fluctuating environment; a cell with few mitochondria will produce less energy and thus will grow slowly, but since mitochondria are voluminous, less mitochondria means more space to accommodate other cellular components [84]. Conversely, the cell that invests much energy in producing heat shock proteins in an environment that remains stable at physiological conditions for long times will waste that energy fruitless most of the time. Nevertheless, that cell will be able to survive an abrupt change of temperature. Clearly, single cells can not be ready to face all possible adverse situations at the same time. Cells embedded in fluctuating environment evolved mechanisms like post-transcriptional regulatory networks that allow for a fast reorganization of the proteome in response to external stimuli [85, 86]. At the population level things are different. In fact, experiment and theoretical models provided evidence of different strategies cell population can adopt to face (as a population) adverse environments [87, 88, 77]. Essentially, instead of investing resources to render each cell capable of facing different conditions, population can be formed of sub-populations of cells with diverse phenotypes, each phenotype fitted for a particular environment. A sudden change of environment can kill part of the population, but there will be a resistant sub-population that will survive and reproduce. In this part of the thesis will see more in details how cell-to-cell variability influences the growth of the population and hence its fitness to the environment the population lives in. Basically, we will analyse two scenarios.

To begin with, we will look at the simple case where the environment is given by a fixed carrying capacity. In this situation, theoretical and experimental findings foretell highly heterogeneous distributions of phenotypes [7]. Such variability inevitably influences the growth especially in its initial phases. We discuss this first part mainly following the work done in [89], where the dependence of population phenotypes, like the growth rate and the lag time on the initial size of the population are studied. Then we will move to a more complex scenario where environment can fluctuate between two regimes, one which is favorable for the growth and another where fast

cells can not replicate. In this scenario, the fitness of the population strongly depends on the distribution of the population phenotype and on the capability of cells to differentiate. We will see that an exploration-exploitation trade off is established, whose understating is particularly important in cases where mechanisms controlling the growth are inhibited as in cancer cells. In fact, tumors evolve so rapidly that they manage to overcome suppressing mechanisms and develop resistance to anti-cancer drugs, in that respect exploiting the evolutionary response and the diversity of the cancer cells is believed to be one of the key-points to designing successful treatments. This second part of the discussion is mainly based on the work done in [8].

5.1 Phases of the cellular growth

In order to maintain their status cells must perform a wide range of functions, most of which require energy. Such energy can be provided either in the form of food molecules or by sunlight. Cellular nutrients come in many forms, the major part as sugars and fats. In order to provide a cell with energy, these molecules have to pass from the external medium across the cell membrane, which is semi-permeable. In much the same way that doors and windows allow necessities to enter a house, various proteins that span the cell membrane permit specific molecules into the cell, although they may require some energy input to accomplish this task. Depending on the environment composition and the amount of nutrients they manage to find, cells show different behaviors, e.i. they enter different phases of growth. In particular, if we seed a certain number of cells in a new environment with a limited amount of nutrients and leave them free to grow by consuming the growth medium three distinct phases representative of the growth can be identified [90].

The first growing phase cells undertake is termed *lag* phase. It enters this state whenever it is introduced into a medium richer of nutrients, such as glucose, than the one it comes from. Usually it takes some time for the cell to adjust with the new environment. During this time it accelerates its cellular metabolism, starting to absorb nutrients from the medium and to synthesize the necessary proteins, co-enzymes and vitamins needed for its growth. Hence cells begin to increase in size until they can duplicate themselves. The length of the lag phase depends directly on the previous growing condition of the cell. In fact if it comes from a similar medium it will already possess the right protein to continue its activity; on the contrary coming from an entirely different environment the cell will have to create all required proteins.

Once a cell has collected enough nutrients and stored sufficient energy, it enters the exponential growth phase. During the logarithmic or log phase, which follows the lag one, the microorganisms are in a rapidly growing and dividing state. The growth medium is exploited at the maximal possible rate. Metabolic activity increases and the cells begin the DNA replication at a constant rate, which reflects in a constraint growing rate with the number of specimen that exponentially increases.

If we look from the point of view of the population, the number of cells in the population, N , evolves in time simply as

$$\frac{dN}{dt} = \Lambda N(t) \quad (5.1)$$

where Λ is the population replication rate. Eq. 5.1 predicts that the population grows exponentially

$$N(t) = N_0 e^{\Lambda t} \quad (5.2)$$

Note that, usually what experiments measure is the generation time, which is the time taken by the population to double in number during a specified time period, i.e. $T = \frac{\ln 2}{\Lambda}$ defined by $N(T) = 2N_0$. For instance, the bacterium *E. coli* under optimal conditions (rich nutrient broth and a temperature of 37 °C), has a doubling time, $T \simeq 20$ minutes, which corresponds to a growth rate, $\Lambda \sim 2.1 \text{ h}^{-1}$.

Eq. 5.1 can be used to describe the system as long as we are dealing with a relatively large population ($\gg 10^3$ cells). For smaller populations, it may be important to consider that replication is not a continuous process but occurs as discrete events which may not be synchronous. In some cases, it may be more convenient to use as the dynamical variable the total biomass of the population rather than the number N of cells.

Coupling with nutrient dynamics

Saturating population growth can also be modeled in a more biologically consistent way by including the dynamics of the nutrient explicitly in the equations. The classic equation for the nutrient-concentration dependent growth of a bacterial population is:

$$\dot{N} = \frac{r_{max}s}{k_s + s} N \quad (5.3)$$

where s is the nutrient concentration, r_{max} is the maximal growth rate and k_s is the nutrient concentration at which the growth rate is half-maximal. In Eq. 5.3, the instantaneous growth rate is described by a 'Monod function' [90]:

$$g(s) = \frac{r_{max}s}{K_s + s} \quad (5.4)$$

which depends linearly on the nutrient concentration s for low nutrient concentrations, but becomes independent of the nutrient as $s \rightarrow \infty$. This captures the fact that for high nutrient concentration, growth is limited by the cell's capacity to import and use the nutrient, rather than by the availability of the nutrient in the environment. Eq. 5.3 must be coupled with a dynamical equation for the nutrient concentration:

$$\dot{s} = -\gamma \frac{r_{max}s}{K_s + s} N \quad (5.5)$$

where γ is a yield coefficient, describing the number of units of nutrient that are consumed to produce one cell (divided by the volume).

Once medium nutrients diminish and waste materials, toxic metabolites and inhibitory compounds such as antibiotics take their place in the medium, cells enter a stationary phase. This phase is characterized by a strong decrease of reproduction rate until the number of cells, that undergo division, becomes equal to the number of cells that die. This is due to the shift of conditions, which creates an unfavorable

environment for the population growth. The cell number is not increased and thus the growth rate is stabilized. At a certain point bacteria stop their division completely and the death phase begins. From now on, the cell completely loses its ability to reproduce and it begins to die due to the unfavorable conditions. The death is rapid and at uniform rate. Eq. 5.1 does not capture the transition to the stationary phase, where the population saturates, in fact the progressive lack of nutrients forces the population to slow its growth. A simple way to capture this saturation is to use instead a logistic growth equation [91]

$$\dot{N} = \Lambda N \left(1 - \frac{N}{k}\right) \quad (5.6)$$

where k is the maximal population size (or carrying capacity). The solution of the logistic equation is analytical:

$$N(t) = N_0 \frac{e^{\Lambda t}}{\left[1 + \frac{N_0}{k} (e^{\Lambda t} - 1)\right]} \quad (5.7)$$

This model is in quite good agreement with measured growth curves for experiments in simple nutrient media. At last, we note that the growth rate of the population does vary in time. The term $(1 - N/k)$ diminishes the effective growth rate when N becomes large, mimicking the effect of nutrient depletion or toxic waste product build up. In fact, in the long time limit it becomes vanishingly small, as the number of cells in the population approaches the carrying capacity. In particular, from Eq. 5.1, we can define the instantaneous growth rate as

$$\lambda(t) = \frac{\dot{N}}{N} = \dot{y} = \Lambda \left(1 - \frac{N}{k}\right) \quad (5.8)$$

where $y = \log(N/N_0)$. From Eq. 5.8, one sees that the population has a maximum growth rate,

$$\lambda_{\max} = \Lambda \left(1 - \frac{N_0}{k}\right) \quad (5.9)$$

for $N = N_0$ (see also Figure 5.1, for a recapitulation). We will see in the next section that Eq. 5.9 does not capture the right behavior for every inoculum size if memory and cooperation are present in the population.

Batch culture

In the laboratory, cells are often grown in liquid suspension under well-mixed conditions. One of the most used methods for growing cells in wet-laboratories is the so-called ‘batch-cell culture’. A sketch of a typical setup for a batch culture growth experiment is illustrated in Figure 5.2. A small number of cells are inoculated into a well-shaken container filled with liquid nutrient medium. Over a period of several days, the density of cells is measured and the results are plotted as a function of time. These ‘growth curves’ have a characteristic shape: an initial period in which no growth is detected (the lag phase), followed by a period of exponential growth (the log phase), followed by a slowing down and eventual cessation of net growth (the stationary phase).

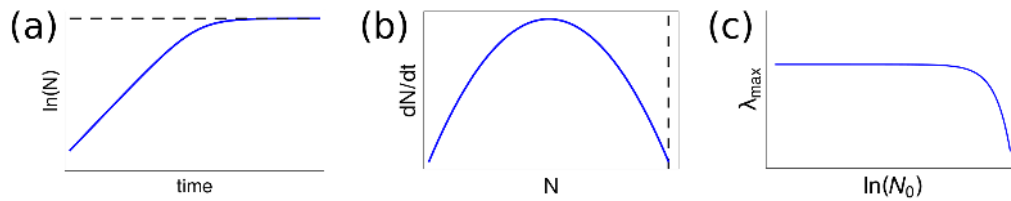


Figure 5.1. Cartoons representative for the analysis of the logistic growth model.

a) The solution of the logistic model 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 emphasized by a dashed black line.

It is generally stated that the lag phase happens because the cell needs to adjust to the liquid medium (having typically been stored under different conditions), while stationary phase happens when the population exhausts its nutrient supply, or builds up waste products. The lag time, the maximum growth rate and the carrying capacity of the medium are the most direct and used descriptors for the three phases of the growth, since they can easily enough be measure fitting the growth curves. This observables must be linked to the underlining phenotype distribution of the populations, in fact they are given by a often non-trivial function of the phenotypes of all the cells composing the population. Furthermore, the starting point of a batch experiment consists in transferring a certain number of cells from a certain culture to a fresh one, the one used in the experiment. This equals to sampling the old culture phenotype distribution.

Intuitively, the more heterogeneous the distribution the more the new population features will depend on the sampling size. If one takes a big sample such that the old distribution results well reproduced, then the new culture observables will be close to the old culture one and fluctuation will be small. If instead the sample is small, the system will be prone to large fluctuations, since the final outcome is the combination of many cellular replications, originating from a few growing trajectories.

5.2 The inoculum size influences the growth

In the previous Chapter, we discussed the role of noise in producing heterogeneous phenotypes. And we also saw that through regulation cells can interfere with noise. In the next sections, we will concern ourselves with the effects of such heterogeneity on the fitness of the population. First, we ask what happens if the growth starts from a bunch of cells randomly sampled from a certain $q(\lambda)$. This is quite a common scenario. In fact, from the experimental point of view, many experimental protocols start with an inoculum of cells taken from a previous culture (for example in the batch culture set-up). Similar scenarios may constitute also the start of an infection where some cells are transferred from one host to the next or in the more worrying case of cancer cell migrations.

Since the initial density N_0 of a population of cells (the inoculum density) can

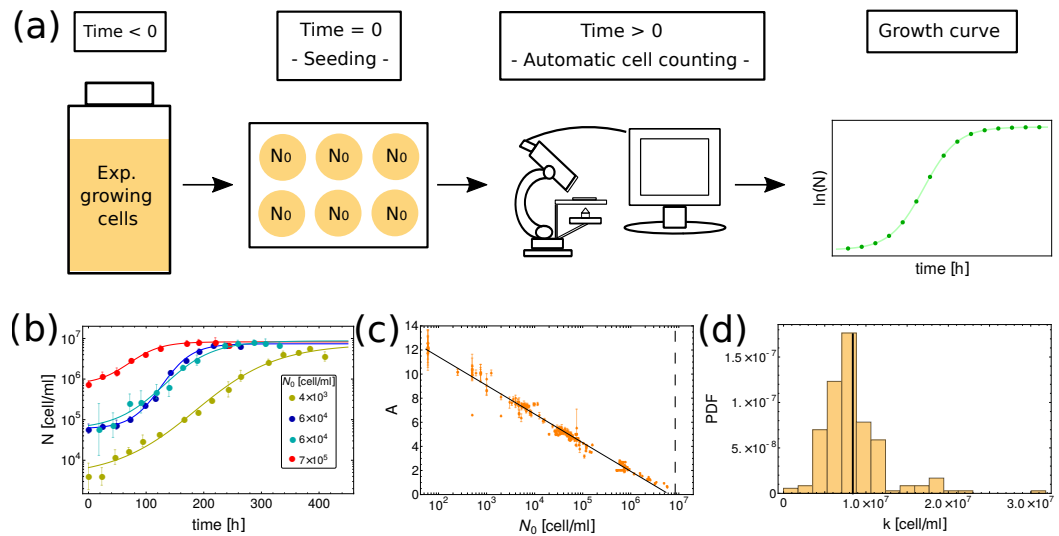


Figure 5.2. Batch experiment protocol and results. From left to right, from a flask with cells growing exponentially in their standard growth medium, at time $t = 0$ h, samples of N_0 cells have been taken and moved into new dishes supplied with the very same nutrient quality present in the flask. From 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.

influence several aspects of population growth in a variety of cell types, it has been subject of many investigations. Starting from the pioneering work of Rein and Rubin [92], inoculum-dependent traits have been observed, in populations as different as bacterial [93, 94], insect [95], plant [96], and cancer cells [97, 98], most notably affecting metabolism (specifically the ability to excrete specific compounds [97]), carrying capacity [97], and the duration of the lag phase [99]. For instance, N_0 -dependent lag times have been reported at very low inoculum densities [100] or under stress, when only small sub-populations can sustain growth [101]. In turn, the growth rate was found to increase, or decrease, with the initial density depending on the organism and the growth medium. Similar scenarios hold for its sample-to-sample variability [96]. Here, we will compare theoretical predictions against the data obtained in a novel experimental investigation by [89], clarifying the dependence of growth rate and lag time from the inoculum size (N_0) with quantitative accuracy across a broad range of initial densities for two widely used cancer cell lines (Jurkat and K562), growing in carbon-limited media with fixed carrying capacity. We will see that both lag time and the growth rate display a complex behavior as a function of N_0 , illustrated in Figures 5.4 and 5.5.

5.2.1 Inoculum size effects on Lag phase

Let us imagine to start a standard batch experiment. We take a bunch of N_0 cells from a system (in practice a flask) and we inserted them in a fresh medium at time $t = 0$. After the adaptation time, t_{lag} the population grows exponentially according to (by neglecting the carrying capacity)

$$N(t) = N_0 e^{\Lambda(t-t_{\text{lag}})} \quad (5.10)$$

where Λ is the 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 $i = 1, 2, \dots, N_0$. Indeed, each single cell needs a time τ_i to adapt to the new environment. In agreement with the existing models [102], we can 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 macroscopic population Λ and that τ_i are identical independent random variables. So, the size of each sub-colony (n_i) evolves in time according to:

$$n_i(t) = e^{\Lambda(t-\tau_i)} \quad (5.11)$$

At the same time, the size of the entire population of cells (N) is given by the sum of the n_i single colonies and grows as follows:

$$N(t) = \sum_{i=1}^{N_0} n_i(t) = \sum_{i=1}^{N_0} e^{\Lambda(t-\tau_i)} \quad (5.12)$$

Comparing Eq. 5.10 and 5.12, the population lag time can be written in terms of the single cell adaptation time τ_i as:

$$t_{\text{lag}} = -\frac{1}{\Lambda} \ln \left(\frac{1}{N_0} \sum_{i=1}^{N_0} e^{-\Lambda\tau_i} \right) . \quad (5.13)$$

Interestingly, t_{lag} carries 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 , two different regimes can be identified. If Λ is 'small' with respect to inverse adaptation times, i.e. if $\Lambda \ll \frac{1}{\tau_i}$, an asymptotic expansions of Eq. 5.13 yields:

$$t_{\text{lag}} \simeq \frac{1}{N_0} \sum_{i=1}^{N_0} \tau_i . \quad (5.14)$$

The population lag time is equal to the average of the single cell first division times and increasing N_0 , t_{lag} is driven by a Gaussian statistics. The value of t_{lag} is better defined, while its fluctuations decrease. If we now consider Λ assuming 'large' values ($\Lambda \gg \frac{1}{\tau_{\text{min}}}$), the sum is dominated by the minimum of τ_i for the saddle point method, leading to the following expression for t_{lag} :

$$t_{\text{lag}} \simeq \tau_{\text{min}} = \min\{\tau_1 \dots \tau_{N_0}\} \quad (5.15)$$

In this second case, the population lag time is described by the extreme value statistics. Thus, by increasing the sampling, namely increasing N_0 , there is a higher probability of sampling small τ . As τ_{min} decreases so do both t_{lag} and its fluctuations. The way t_{lag} decreases with N_0 depends on the distribution of the τ_i s. Figure 5.3 summaries the above-described model and the two possible scenarios for the population lag time derived from it.

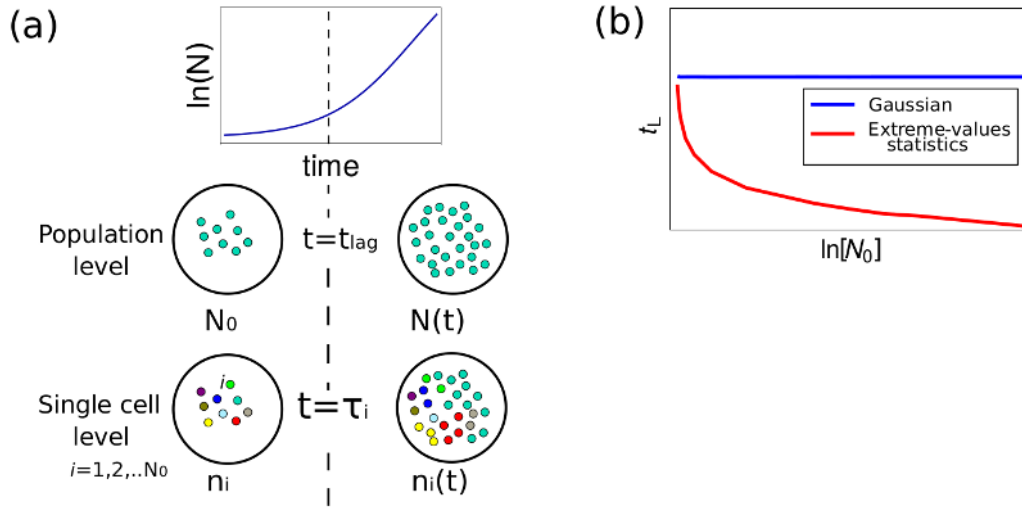


Figure 5.3. Sketch of the lag time model. **a)** Cartoon summarizing the model. At the population level, a population of N_0 cells (light blue dots in the top-left circle) needs a time t_{lag} before starting to grow exponentially over time. The same dynamics is valid at the single-cell level, where each single cell i starts after a time τ_i to give birth to sub-colonies whose sizes grow exponentially over time. **b)** Schematic representation of the two possible scenarios for the population lag time. 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_i s.

In conclusion, the behavior 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.

Experimental results

Figure 5.4 shows the lag time as a function of the initial density N_0 as measured in [89]. 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 . Fluctuations in the lag time behave similarly. The former is a non-trivial result as the composition of the growth medium used for growing cells before and after the seeding was the very same. 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. The simplest trend that describes our experimental data is a linear negative correlation between the population lag time

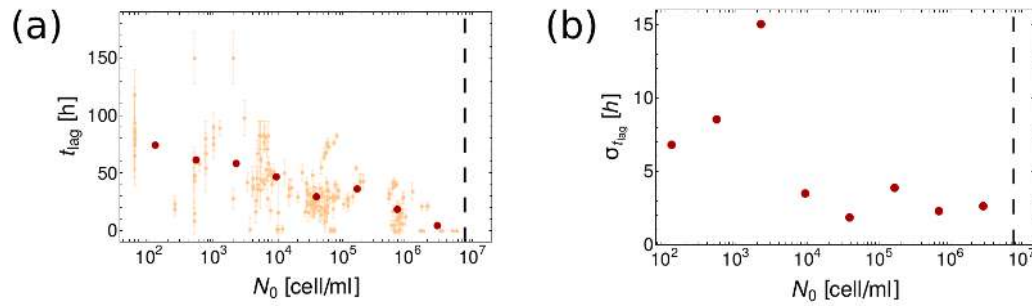


Figure 5.4. Experimental lag times. a) Experimental lag times as a function of the initial seeding (N_0). Lighter color dots are experimental data obtained through the fit. Their error bars are the error on the fit. The darkest red dots are the average of the smaller dots binned over N_0 . b) Standard 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 .

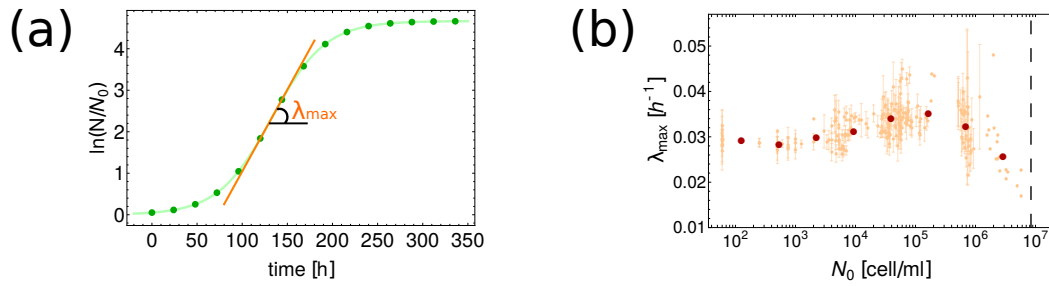


Figure 5.5. Experimental growth curves and growth rates. a) Example of a growth curve as a function of time. b) Experimental growth rates as a function of the initial seeding (N_0). Lighter color dots are experimental data obtained through the fit. Their error bars are the error on the fit. The darkest red dots are the average of the smaller dots binned over N_0 . In both plots the vertical dashed lines are located in correspondence of the carrying capacity value k .

and the initial cell density. Considering now, the two possible theoretical scenarios summarized in the previous section, we can reasonably say to be in the regime of large Λ 'large'. Thus, the extreme values statistics predicts a decreasing trend of the lag time when increasing N_0 , which is exactly what the experimental data show.

5.2.2 Inoculum effect of the population fitness

Following our experiment, once the population enters the exponential phase, the dynamics should be described by Eq. 5.6. Let us now see, what a simple model can tell about the dependence from the inoculum on the growth rate.

Consider a population of N_0 cells initially planted in a growth medium with finite carrying capacity k and assume that their intrinsic GRs λ_i are sampled independently from a distribution $q(\lambda)$ defined over the feasible space \mathbb{F} and describing the statistics of the population from which the initial inoculum was obtained (i.e. by the $q(\lambda)$). If the finite carrying capacity is the only growth-limiting factor, each of the initial seeds will expand in time according to its GR and the number n_i of cells with GR

λ_i will evolve in time according to

$$\frac{\dot{n}}{n} = \lambda_i \left(1 - \frac{N}{k}\right) \quad (5.16)$$

Overall, the total number of cells at each time will be given by

$$N(t) = \sum_{i=0}^{N_0} n_i(t) \quad . \quad (5.17)$$

This implies that we can obtain the ‘microscopic’ equivalent of Eq. 5.6, i.e.

$$\frac{\dot{N}}{N} = \langle \lambda \rangle_t \left(1 - \frac{N}{k}\right) \quad (5.18)$$

with

$$\langle \lambda \rangle_t = \frac{\sum_{i=1}^{N_0} \lambda_i n_i(t)}{\sum_{i=1}^{N_0} n_i(t)} \quad (5.19)$$

Eq. 5.16 can be formally solved by $n_i(t) = e^{\beta(t)\lambda_i}$ where

$$\beta(t) = t - \frac{1}{k} \int_0^t N(t') dt' \quad . \quad (5.20)$$

The formal solution of n_i closely resembles Eq. 4.3 even if it is not normalized. Going on with the analogy, if we consider a large enough inoculum, we can use an annealed approximation for $N(t)$, which yields

$$N(t) \simeq N_0 Z(\beta) \quad \text{with} \quad Z(\beta) = \int q(\lambda) e^{\beta\lambda} d\lambda \quad (5.21)$$

Using the expression for Z , we can find the stationary values β can assume. i.e

$$Z(\beta^*) = \int q(\lambda) e^{\beta^*\lambda} d\lambda = \frac{K}{N_0} \quad (5.22)$$

which fixes the asymptotic degree of optimization the regulation can provide given $q(\lambda)$, the carrying capacity k , and the size of the inoculum N_0 . Since

$$\langle \lambda \rangle = \frac{d \ln(Z)}{d\beta} \quad (5.23)$$

the growth rate of the population depends on those three quantities too. In particular, it decreases as a function of the inoculum size, in accordance with experimental results for large inocula (see Figure 5.5b for $N_0 > 10^5$). Anyway, if we look at Figure 5.5b, we see that the trend predicted by the logistic model (Eq. 5.9) does not match with data in an extended range of initial inoculum sizes. In particular, for small N_0 (more precisely, for $N_0 \ll k$), λ_{\max} is roughly constant. This agrees with the expectation that initial seeds far from the carrying capacity should generate populations growing at the (asymptotic) rate corresponding to medium with infinite carrying capacity. In this regime, sample to sample variability is small, i.e. different populations grow at very similar rates. As N_0 increases, however, one observes a significant increase in fluctuations as well as an overall

upward trend in the mean growth rate (averaged over different populations). In other words, larger seeds appear to yield a fitness advantage. Finally, when the initial seed gets closer to the carrying capacity, λ_{\max} expectedly decreases linearly with N_0 . Growth rate fluctuations in this regime are much smaller than those seen at intermediate values of N_0 . This behavior calls for the presence of memory and cooperation among cells. In fact, a decrease of growth rate with the initial cell density can occur whenever cells are subject to competition with density-dependent strength due e.g. to medium conditioning or finite resources. On the other hand, a positive correlation between the growth rate and the initial density points to the existence of cooperative, density-dependent effects. Interestingly, memory and history-dependence have recently been investigated as factors influencing exit from quiescence [103] and competence in bacteria [104]. In the latter case, quorum sensing was identified as a central coordinating mechanism. In the next section, we discuss some minimal modifications of the logistic model that allow to explain data along the whole range of N_0 .

5.2.3 Memory and cooperation

Many background causes can lead to memory in cell growth. For instance, the standard logistic growth model á la Baranyi [105] suggests that the growth medium can impose a well defined initial-density dependence on the growth rate through a finite carrying capacity k . In particular, the growth rate should be expected to decline as the initial seed approaches k , as quantified e.g. in [78]. The simplest explanation for the fact that, at intermediate N_0 , the growth rate increases with the inoculum density is provided by cooperative behavior mediated by cell-to-cell signaling that gets stronger as the population grows. In other words, the observations at a given time point would depend on what happened during the growth. More generally, initial density dependence can be understood in terms similar to those employed for generic density-dependent features described by the Allee effect [106]. In fact, to model cooperation, or more properly density-dependent growth, logistic growth is usually modified by an extra ‘Allee factor’

$$\frac{\dot{N}}{N} = r \left(1 - \frac{N}{k}\right) \left(\frac{N - q}{k}\right)^\alpha \quad (5.24)$$

Although no exact analytical solution is possible for Eq. 5.24, it can be easily solved numerically for any choice of the parameters, r, α, Q and k . As one can see in Figure 5.6, the instantaneous growth rate $\frac{dN}{dt}$, as a function of the number of cells, is mainly influenced by the exponent α quantifying the strength of the cooperation and by q , which instead dictates the effect (positive or negative) of cooperation. In fact, if $q > 0$ and $N_0 < q$, the population decreases and vanishes in time, so the effect of cooperation is detrimental. Even if no analytical expression for N is available, we can compute the maximum slope of y using a trick: the fact that $\dot{y} = \frac{\dot{N}}{N}$. So that the maximum slope of y equals the maximum of Eq. 5.24, i.e.

$$\lambda_{\max} = \begin{cases} r \left(\frac{1}{1+\alpha}\right) \left(\frac{\alpha}{1+\alpha}\right)^\alpha & \text{if } N_0 < \frac{\alpha}{1+\alpha}k \\ r \left(1 - \frac{N_0}{k}\right) \left(\frac{N_0}{k}\right)^\alpha & \text{otherwise} \end{cases} \quad (5.25)$$

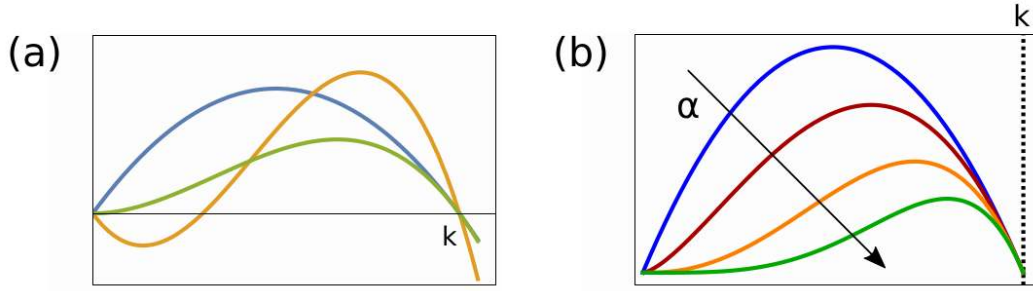


Figure 5.6. Representations of the growth rate (dN/dt) of a population of size N as a function of its size. a) Trends for logistic growth (blue solid line), weak (green) and strong (yellow) Allee effect. The value k corresponds to the carrying capacity. **b)** Examples of weak Allee effects with different 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 .

where $N_0 = \frac{\alpha}{1+\alpha}k$ corresponds to the flex point of y .

Anyway, fitting Eq. 5.25 over the data, we see from Figure 5.7 that the predicted behavior of λ_{\max} versus N_0 is qualitatively identical to that of the purely logistic model (see Eq. 5.25). Moreover, the instantaneous growth rate (dN/dt) in the exponential phase as a function of the population size N should show an Allee-like behavior, with an elbow for small N (see Figure 5.6) that deviates from a logistic map. However, when fitting the instantaneous growth rate in the exponential phase versus the population size at that time (Eq. 5.24) the exponent α that indicates the level of cooperativity is ~ 0 . This suggests that cooperativity is at least weak and that the dependence of the GR on N_0 should be interpreted more like a long-term memory effect.

The simplest modification of the logistic model allowing to recover empirical observations is obtained by assuming that the intrinsic growth rate r is a slowly increasing function of N_0 , $r = r(N_0)$. In turn, this implies

$$\lambda_{\max} = r(N_0) \left(1 - \frac{N_0}{k}\right). \quad (5.26)$$

In this way, the maximum growth rate is dominated by the intrinsic term (and thereby increases with N_0) for small enough inocula, while it gets more and more limited by the carrying capacity (and therefore decreases with increasing N_0) as N_0 gets closer to k . Upon assuming

$$r(N_0) = r + a(N_0)^b. \quad (5.27)$$

with r the asymptotic growth rate, this model replicates all the empirical observations by fitting parameters a and b to the observed profile of λ_{\max} versus N_0 .

We are left with the patterns of fluctuations to be explained. In fact, since experiments were repeated many times over approximately the same inoculum size, we have information also on the fluctuation away from the mean. Notably,

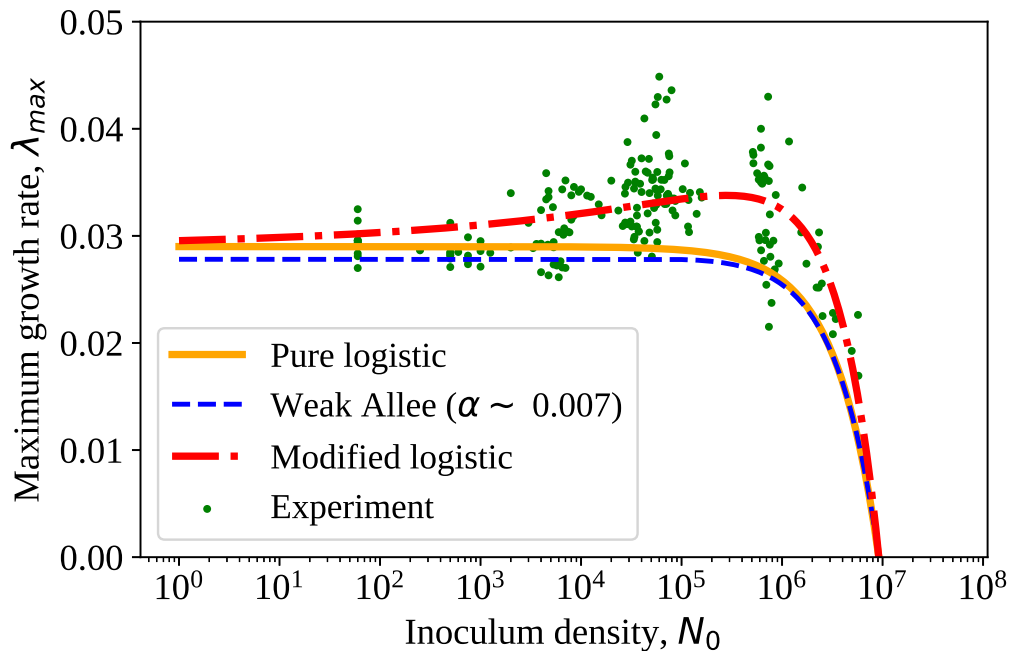


Figure 5.7. Experimental data and model comparison. Fit of the logistic (orange), weak Allee (blue) and modified logistic (red) models.

the patterns of fluctuations are recovered by assuming uniform fluctuation across experiments (see Figure 5.8) on the two parameters which govern $r(N_0)$, i.e. a and b (r is assumed fixed for simplicity). Finally, in the experimental setup discussed here, an N_0 -dependent r would imply that cell populations carry intrinsic memory of their initial condition into the exponential phase. Such a scenario could be validated by carrying out more experiments at different (e.g. larger) carrying capacities. Based on the above model, in fact, faster growth rates than those observed here should be achievable at larger values of N_0 , as one can see from Figure 5.9. Unfortunately, as this thesis is being written, those experiments are just being performed.

5.3 Growth in a changing environment

At last, we want to discuss the effects of a heterogeneous distribution of phenotypes on the growth in a changing environment. Up to now, we saw that the growth rate of a population (as so its fitness) is linked to (i) the distribution of phenotypes, $q(\lambda)$, which depends itself on the degree of regulation cells can impose, (ii) the initial number of cells and (iii) the carrying capacity of the system.

In the minimal model we presented before and described by Eq. 4.15, phenotypic diffusion is inversely proportional to the maximum fitness the population can acquire. In fact, the faster a cell can change its phenotype, the more the final phenotypic distribution tends to the underlining one (i.e. the $q(\lambda)$), which corresponds to a flat sampling of the flux vector space. Recapitulating in few words, once a cell manages to acquire a high growth rate, if it can maintain that rate, it will quickly dominate

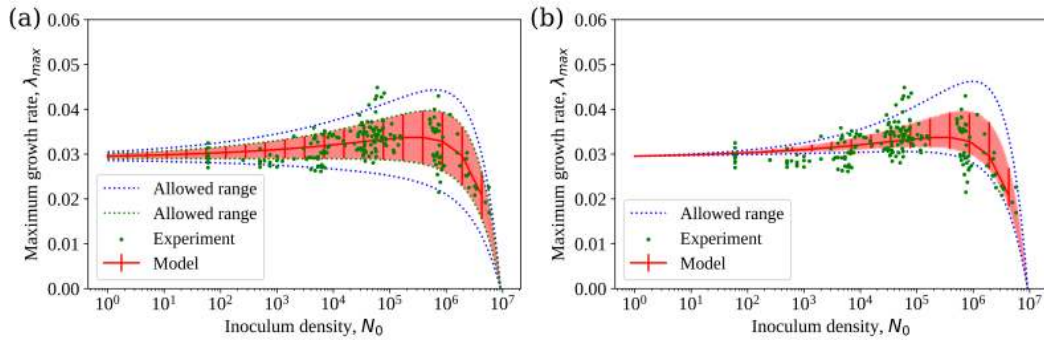


Figure 5.8. Outcomes of the modified logistic model with uniform noise on a and b .

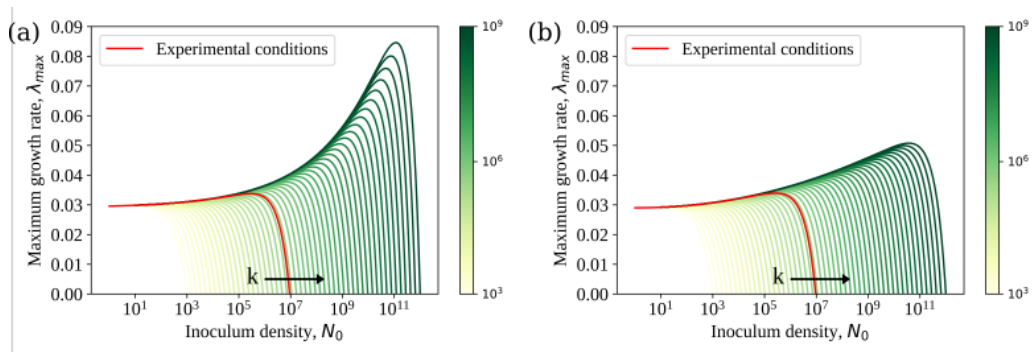


Figure 5.9. Putative validation of the modified logistic model. **a)** Maximum growth rate as a function of N_0 for different values of the carrying capacity and $r(N_0) \sim a(N_0)^b$. **b)** Maximum growth rate as a function of N_0 for different values of the carrying capacity and $r(N_0) \sim a(\ln N_0)^b$.

over slower growing one. This shifts the overall growth rate of the population toward higher values. High growth rates, though, require a lot of regulation to be maintained, and a tight regulation is expensive. Furthermore, it is experimentally known that fast phenotypes are more prone to suffer environmental changes. To give an intuition, a cell that replicates fast needs a lot of nutrients, if those nutrient become scarce in a sudden, then this cell will found itself with no energy to reorganize its proteome. Figure 4.3 clearly conveys the idea that the higher the fitness the lower must be the diffusion.

This scenario has been obtained in the presence of a ‘static’ environment. In the sense that while the population growth the environment remains either globally favorable (rich medium) or unfavorable (poor medium). The latter case producing a phenotypic distribution shifted toward lower values of growth rate. Let us see how the scenario described so far changes if we introduce a dynamical environment, i.e. an environment that changes in time from a state which is favorable for the growth of all the population to a state where the fast phenotypes are inhibited. The simplest route to couple the system to such an external environment is to assume that the effective growth rate (EGR), f , of a cell depends both on its constitutive growth rate, CGR, λ (which is the rate the cell would have in absence of external influences) and on the state of an exogenously varying medium which, for sakes of simplicity, we can describe by a single time-dependent variable x . In rough terms, we consider a fluctuating environment in which x describes the threshold fitness for replication under randomly occurring shocks. This corresponds to the choice

$$f(\lambda, t) = \begin{cases} \lambda & \text{if } \lambda \leq x(t) \\ 0 & \text{otherwise} \end{cases}, \quad (5.28)$$

according to which cells with CGR smaller than $x(t)$ can replicate at time t , while replication is inhibited for the others.

To see the impact of the environment on the population, we can look at two rather opposite scenarios: a predictable environment (switching periodically between two fixed states) and a most random one (switching after a random time and to a random value). More specifically following [8], we can assume that the threshold x fluctuate in time by switching between the value $x = \lambda_{\max}$, in which case all cells in the population can replicate, and a value $x = \lambda^* < \lambda_{\max}$, in which case replication can only take place for cells with $\lambda \leq \lambda^*$. We consider two choices for λ^* . In the periodic case, λ^* is a constant kept fixed throughout the dynamics, so that x takes the values λ^* and λ_{\max} alternately, leading to a two-state environment (labeled ‘const- x ’ in [8]). While in the random case, λ^* is sampled independently at every switch from a uniform distribution on the interval $[x_{\min}, \lambda_{\max}]$, leading to an environment with a continuum of states (‘rand- x ’ case in [8]). For simplicity, we set $\lambda^* = x_{\min}$ in the const- x environment. Switches from the non-selective environment where all cells replicate to the selective one where only some do ($ns \rightarrow s$) and *viceversa* ($s \rightarrow ns$) are assumed to occur with characteristic times ω_{ns} and ω_s respectively. Note that the statistic of the switching turned out to produce no appreciable differences to the fitness of the population (‘const- t ’ case or ‘rand- t ’ case according to the notation of [8]). With this modelization of the environment, we can still use Eq. 4.15 to describe the time evolution of the population growth rate distribution, with the only

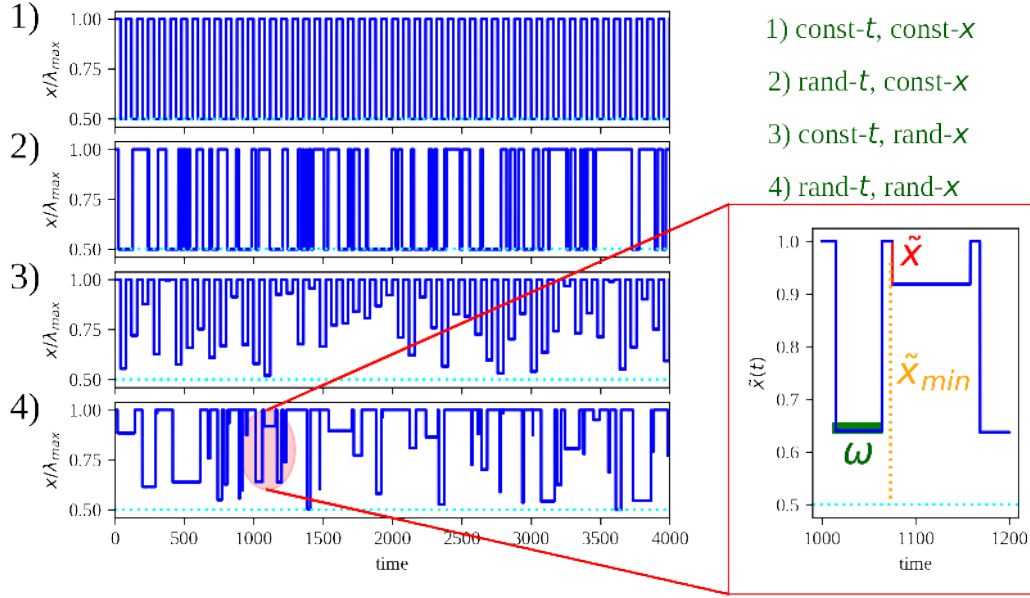


Figure 5.10. Representative behavior of the threshold x as a function of time in the different environments. (a) a periodic two-state environment where x switches (in this case) between the values λ_{\max} and $x_{\min} = \lambda_{\max}/2$; (b) a periodically switching environment where x takes on random values drawn uniformly from $[x_{\min}, \lambda_{\max}]$; (c) a two-state environment where switches occur at exponentially distributed random times; (d) an environment where x behaves as in (b) but in which switches occur at exponentially distributed random times. In this example, the characteristic switching times ω_{ns} and ω_{s} are taken to be equal and fixed to 40 (a.u.). Adapted from [8].

difference that f takes the place of λ in the replication term, i.e.

$$\frac{dp(\lambda, t)}{dt} = [f(\lambda, t) - \langle f \rangle] p(\lambda, t) + D \left[\frac{\partial^2 p(\lambda, t)}{\partial \lambda^2} - \frac{\partial}{\partial \lambda} \left[p(\lambda, t) \frac{\partial}{\partial \lambda} (\ln q(\lambda)) \right] \right], \quad (5.29)$$

where now the mean growth rate at time t obviously involves the effective growth rate, f , i.e.

$$\langle f \rangle = \int_0^{\lambda_{\max}} f(\lambda, t) p(\lambda, t) d\lambda. \quad (5.30)$$

We finally have to specify a form for the phenotypic landscape $q(\lambda)$. Instead of using the expression in Eq. 4.2, we define a slightly modified version, which is easier to treat analytically:

$$q(\lambda) = \frac{a+1}{\lambda_{\max}} \left(1 - \frac{\lambda}{\lambda_{\max}} \right)^a, \quad (5.31)$$

where the exponent $a \geq 0$ modulates the steepness of $q(\lambda)$. In short, the larger a , the more heterogeneous the landscape, with slow phenotypes being increasingly more frequent than fast ones as a increases. In [7, 78], the CGR landscape underlying

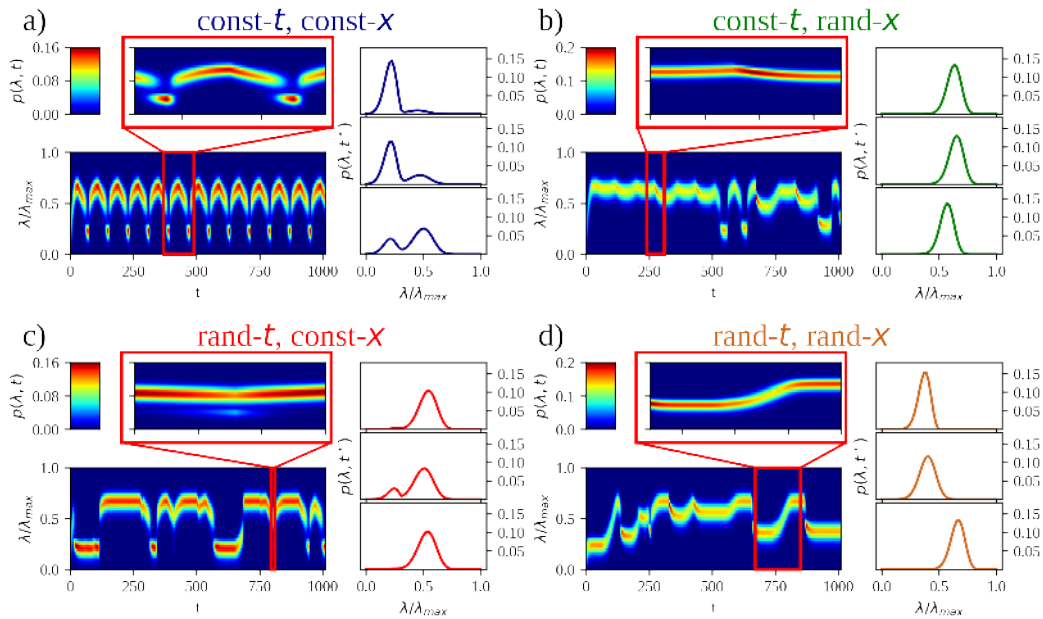


Figure 5.11. Colormaps showing representative probability densities $p(\lambda, t)$. Panels to the right of each map depict the density profile at different time points within the zoomed-in region, at time increasing from top to bottom. Results are shown for **a)** const- t and const- x environment, **b)** const- t and rand- x environment, **c)** rand- t and const- x environment, and **d)** rand- t and rand- x environment. Parameter values: $a = 20$, $x = 0.3\lambda_{\max}$, $D = 10^{-3}$. Adapted from [8].

bacterial metabolic networks has been found to be characterized by values of a lying between 200 and 300 depending on the specifics of the environment. Again to focus on two extreme cases, we can consider explicitly the cases $a = 0$ (uniform $q(\lambda)$) and $a = 20$ (strongly heterogeneous $q(\lambda)$).

5.3.1 Phenotypic distributions and fitness

The non-linear Fokker-Planck equation (5.29) can be solved numerically for any choice of the environment, of the diffusion coefficient and of the prior phenotypic density $q(\lambda)$. After a short transient, $p(\lambda, t)$ appears to settle in qualitative robust, environment-dependent patterns. Samples are shown in Fig. 5.11.

Different types of distributions emerge across the various environments, including bimodal distributions in which most of the population occupies the two peaks alternately (panel a) or in which one peak always dominates over the other (panel c), unimodal distributions with fluctuating positions (panel b) and unimodal distributions in which peaks drift in a specific direction (panel d). While all of these can occur in every type of environment, both their frequency of occurrence and the relative intensities of the peaks are strongly environment-dependent.

Such patterns provide hints about the way in which the population copes with environmental fluctuations. An important feature observed from data is that, independently of whether switches occur periodically or randomly, adaptation to two-state environments is achieved more efficiently by structuring the population in

a bimodal form, while complex environments favor unimodal distributions.

How can we assess which distribution, and/or which strategy provides the best chances against the environment? As we are mostly interested in understanding how the system behaves in the long-time limit, we surely must look at the long-term population structure, after it interacted with the environment sufficiently enough. Note that in this scenario, we can not hope for a steady state regime for the probability distribution as the one we had solving Eq. 4.15. In fact, here the environment continuously changes. However, since x takes on values in a finite range, we can expect that sampling after a certain transient time, we should start finding typical structures of the population. To assess whether we can start sampling typical configuration, we can look at the long-term growth rate, which is widely used to quantify the fitness of the population and it is expected to achieve a stationary value in the long term. The population growth rate is defined as a generalization of Eq. 5.1, i.e.

$$\Lambda \equiv \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{N(t)}{N(0)} . \quad (5.32)$$

Computing Λ for an exponentially growing population may result a difficult task. Luckily, from Eq. (4.6) and from the fact that $N(t) = \int n(\lambda, t) d\lambda$, in our case we can recast Eq. 5.32 as

$$\Lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \langle f(\lambda, t') \rangle dt' , \quad (5.33)$$

which can be computed directly from the phenotypic distribution at each time.

Figure 5.12a,b shows, for all environments, the long-term probability distributions $p(\lambda)$ obtained by averaging over time after Λ has reached its stationary value in the limiting case where replication is faster than diffusion. Comparing the distributions with the fitness, we can notice a quite peculiar behavior for the two thresholds: while as expected the more aggressive ($x_{\min} = 0.3$) environment produces a Λ which is lower than the one found in a softer environment ($x_{\min} = 0.7$); distributions behave the other way around. In the former case, the distribution of the phenotype sets near λ_{\max} , while for the softer case it peaks around the threshold. To understand why the system exhibits this feature, we can study more in details the so-called ‘exploitation limit’, i.e. the case where phenotypic diffusion happens on a scale bigger than the characteristic time of growth.

5.3.2 Exploitation limit

Let us now see how we can get some analytical insight in the case where cells grow without changing their phenotype ($D \ll 1$). The dynamics in this regime is given simply by first term in Eq. 5.34, i.e.

$$\frac{dp(\lambda, t)}{dt} = (f - \langle f \rangle) p(\lambda, t) . \quad (5.34)$$

In this pure replication dynamics, the distribution tends to concentrate on the highest achievable growth rate. In presence of a non-selective environment, distribution peaks on λ_{\max} . Instead, our environment switches from a selective state

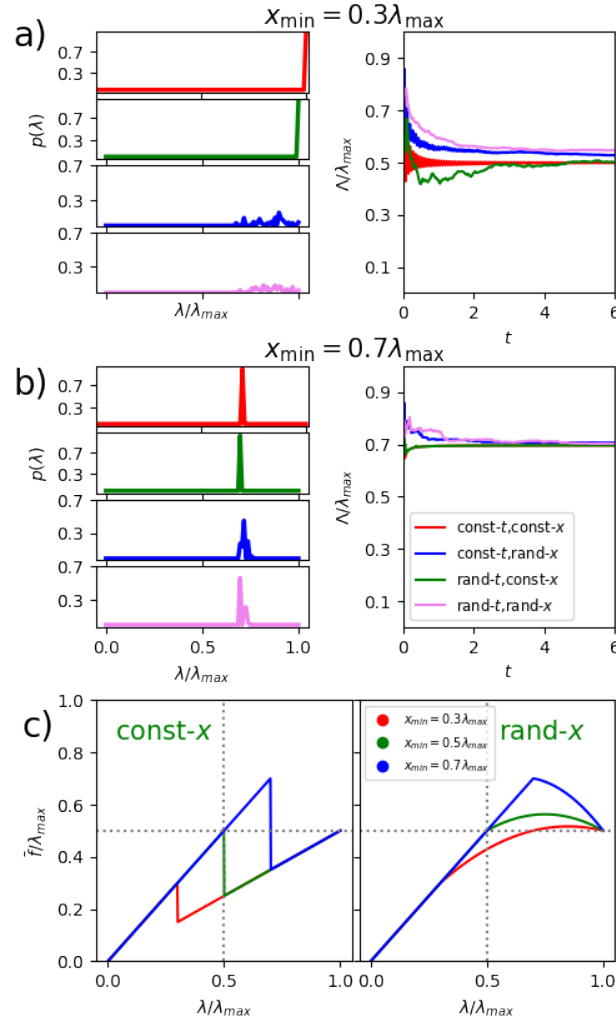


Figure 5.12. Growth without phenotypic diffusion. (a and b) Long-time phenotypic distributions (left) and time evolution of the population growth rate Λ (right) in the absence of diffusion in the different environments (represented by different colors and line widths) and for $x = 0.3\lambda_{\max}$ (panel (a)) and $x = 0.7\lambda_{\max}$ (panel (b)). In the former case ($x < \lambda_{\max}/2$), the distribution can achieve the highest possible CGR. In turn, the long term fitness Λ sets around $\lambda_{\max}/2$. In the latter case ($x > \lambda_{\max}/2$), the distribution peaks around the threshold CGR, while the population achieves a growth rate Λ larger than $\lambda_{\max}/2$. (c) Time-averaged \bar{f} as a function of the CGR for const- x (left) and rand- x (right) environments and for three different values of x_{\min} . One sees that the position of the maximum depends both on the chosen threshold and on the specific environment. Adapted from [8].

to a non-selective one with characteristic times ω_s and ω_{ns} respectively. So that observing the evolution for a long enough time, the population will spend a fraction $p_{ns} = \frac{\omega_{ns}}{\omega_{ns} + \omega_s}$ of time in the non-selective state and a fraction $p_s = 1 - p_{ns}$ of time in the selective one.

In the selective regime, cells with CGR higher than the threshold will be frozen, while cells with CGR near the x rapidly outnumber all growing cells.

Periodic environment

To gain an insight on the dynamics, let us see first what happens in the two-state environment with x oscillating between λ_{\max} and a constant value x_{\min} . In this case, f will equal λ at all times if $\lambda < x_{\min}$, while for $\lambda > x_{\min}$ it will be equal to λ for approximately a fraction p_{ns} of the total the time of the evolution and to zero for the other $p_s = 1 - p_{ns}$ of time. Therefore, the time average of f in this environment reads

$$\bar{f} \simeq \begin{cases} \lambda & \text{if } \lambda < x_{\min} \\ p_{ns}\lambda & \text{if } \lambda \geq x_{\min} \end{cases} . \quad (5.35)$$

In specific, for $x_{\min} > p_{ns}\lambda_{\max}$ (resp. $x_{\min} < p_{ns}\lambda_{\max}$), the mean IGR has a maximum for $\lambda = x_{\min}$ (resp. $\lambda = \lambda_{\max}$), where $\bar{f} = x_{\min}$ (resp. $\bar{f} = p_{ns}\lambda_{\max}$).

The mean IGR displays again a discontinuity at the threshold x_{\min} , and the value of λ for which it attains a maximum depends on both x_{\min} and p_{ns} (see Fig. 5.12c, left panel for a case in which $p_{ns} = p_s = 1/2$). In specific, for $x_{\min} > p_{ns}\lambda_{\max}$ (resp. $x_{\min} < p_{ns}\lambda_{\max}$), the mean instantaneous growth rate has a maximum for $\lambda = x_{\min}$ (resp. $\lambda = \lambda_{\max}$), where $\bar{f} = x_{\min}$ (resp. $\bar{f} = p_{ns}\lambda_{\max}$). Hence, at long times, we expect the population to grow at the fastest IGR achievable, with a phenotypic distribution $p(\lambda)$ peaked at $\lambda = x_{\min}$ (resp. $\lambda = \lambda_{\max}$) for $x_{\min} > p_{ns}\lambda_{\max}$ (resp. $x_{\min} < p_{ns}\lambda_{\max}$). This is in agreement with the numerical evidence shown in Fig. 5.12a,b for const- x environments. A more quantitative demonstration is provided in Appendix C.

This result can be used to obtain an analytical approximation for Λ . In fact, considering that the system spends roughly a fraction p_{ns} of time in the non-selective environment ($x = \lambda_{\max}$) and the other p_s in the selective one ($x = x_{\min}$), we have (see (5.33))

$$\Lambda \simeq p_{ns} \langle f \rangle + p_s \langle f \rangle_{x_{\min}} \quad (5.36)$$

$$\simeq \begin{cases} x_{\min} & \text{if } x_{\min} > p_{ns}\lambda_{\max} \\ p_{ns}\lambda_{\max} & \text{if } x_{\min} \leq p_{ns}\lambda_{\max} \end{cases} . \quad (5.37)$$

where

$$\langle f \rangle_z = \int_0^z f(\lambda, t) p(\lambda, t) d\lambda \quad (5.38)$$

From Fig. 5.13, we can see that the value of Λ estimated numerically agrees with the one just derived in the limit $D \rightarrow 0$ (horizontal blue line) for $x_{\min} = 0.3$. Note that Λ corresponds to the maximum of the time-averaged IGR, \bar{f} (see Fig. 5.12c),

confirming how, when diffusion is much slower than environmental fluctuations), fitness is ultimately limited by the environment alone.

Random environment

In the random case (x oscillating between λ_{\max} and a random value λ^* uniformly chosen from $[x_{\min}, \lambda_{\max}]$), f will again equal λ roughly a fraction p_{ns} of the time, while for the other time it will be randomly zero or λ depending on x_{\min} . In particular, one can calculate the probability that during the unfavorable period a cell with growth rate higher than x_{\min} is replicating, i.e. $\text{Prob}\{f = \lambda\} \equiv \text{Prob}\{x > \lambda\} = 1 - \phi$, with

$$\phi = \frac{\lambda - x_{\min}}{\lambda_{\max} - x_{\min}} . \quad (5.39)$$

Retracing the steps done for the periodic environment, we can evaluate the mean value of the IGR and look for the subpopulation of cells having the maximum IGR. One therefore finds

$$\bar{f} \simeq \begin{cases} \lambda & \text{if } \lambda < x_{\min} \\ \lambda(1 - p_s\phi) & \text{if } \lambda \geq x_{\min} \end{cases} , \quad (5.40)$$

from which we see that \bar{f} attains a maximum value \bar{f}_{\max} given by

$$\bar{f}_{\max} = \frac{1}{4p_s} \frac{(\lambda_{\max} - p_{\text{ns}} x_{\min})^2}{\lambda_{\max} - x_{\min}} , \quad (5.41)$$

at $\lambda = \frac{\lambda_{\max} - p_{\text{ns}} x_{\min}}{1 + p_s}$ if $x_{\min} < \frac{\lambda_{\max}}{1 + p_s}$, while $\bar{f}_{\max} = x_{\min}$ at $\lambda = x_{\min}$ if $x_{\min} > \frac{\lambda_{\max}}{1 + p_s}$. In complete analogy with the previous case, the population concentrates around phenotypes λ for which \bar{f} is maximum, while for the asymptotic growth rate of the population Λ one finds

$$\Lambda \simeq \bar{f}_{\max} . \quad (5.42)$$

The results displayed in Fig. 5.12c again confirm this conclusion. Let us now see what we can say on the other limiting case, i.e. the pure ‘exploration limit’ in which diffusion dominates over the replication.

5.3.3 Exploration limit

In the limit $D \rightarrow \infty$ and more generally whenever diffusion occurs on time scales much faster than those of environmental fluctuations, the growth term in Eq. 4.8 is negligible with respect to the diffusion one and population is rapidly redistributed according to the underlying phenotypic landscape described by $q(\lambda)$. As a consequence $p(\lambda) \rightarrow q(\lambda)$ asymptotically. Since also in this case we know the form the distribution assumes, it is again possible to derive an approximate expression for Λ from Eq. 5.33 following the lines traced in the previous section. In analogy with 5.36, we find

$$\Lambda \simeq p_{\text{ns}} \langle f \rangle + p_s \langle\langle f \rangle\rangle , \quad (5.43)$$

where

$$\langle\langle f \rangle\rangle = \int_{x_{\min}}^{\lambda_{\max}} [\langle f(\lambda, t) \rangle_x] \pi(x) dx \quad (5.44)$$

and $\pi(x)$ stands for the probability distribution of the threshold x . In particular, $\pi(x) = \delta(x - x_{\min})$ in the periodic case since the value of the threshold during the unfavorable regime is fixed. While $\pi(x) = (\lambda_{\max} - x_{\min})^{-1}$ for $x \in [x_{\min}, \lambda_{\max}]$ in the random case (by the definition we chose for the random environment). Because $p(\lambda) \simeq q(\lambda)$ and $f = \lambda$ (resp. $f = 0$) for $\lambda < x$ (resp. $\lambda > x$), the term inside the integral in Eq. 5.44 assumes the form

$$\begin{aligned} \langle f \rangle_x &\simeq \int_0^x \lambda q(\lambda) d\lambda = \\ &= \frac{\lambda_{\max}}{(a+2)} \left[1 - \left(1 - (a+1) \frac{x}{\lambda_{\max}} \right) \cdot \left(1 - \frac{x}{\lambda_{\max}} \right)^{a+1} \right]. \end{aligned} \quad (5.45)$$

Substituting this into Eq. 5.44 and then in Eq. 5.43 one obtains

$$\Lambda \simeq \begin{cases} \frac{\lambda_{\max}}{(a+2)} \left[1 - p_s(a+2) \frac{x_{\min}}{\lambda_{\max}} \left(1 - \frac{x_{\min}}{\lambda_{\max}} \right)^{a+1} - p_s \left(1 - \frac{x_{\min}}{\lambda_{\max}} \right)^{a+2} \right] & (\text{const-}x) \\ \frac{\lambda_{\max}}{(a+2)} \left[1 - p_s \frac{x_{\min}}{\lambda_{\max}} \left(1 - \frac{x_{\min}}{\lambda_{\max}} \right)^{a+1} - \frac{2p_s}{(a+3)} \left(1 - \frac{x_{\min}}{\lambda_{\max}} \right)^{a+2} \right] & (\text{rand-}x) \end{cases} \quad (5.46)$$

These formulas confirm the intuitive picture according to which the more the underlying distribution of phenotypes q concentrates on small values of CGR (i.e. the larger the value of a), the slower the population grows at fast phenotypic diffusion. Fig. 5.13c,f (horizontal lines at $D \gg 1$) show that the agreement between the long term population growth rate computed numerically and the theoretical estimate given above is excellent in both environments.

5.3.4 Fitness and regulation

The time evolution of $p(\lambda, t)$ is ruled by two different timescales, namely the mean environmental switching time, ω and the mean time to transition between different phenotypes, τ .

Generically, at sufficiently small values of D , phenotypes tend to concentrate close to λ_{\max} (see Fig. 5.13a,b and d,e). This situation reproduces the 'exploitation' limit $D \rightarrow 0$, where the time evolution reduces to the replicator dynamics. A population whose phenotypic diffusion occurs on exceedingly long time scales (compared to those characterizing environmental fluctuations) can only grow exploiting resources available from the environment and is therefore maximally sensitive to environment-derived shocks. In such a case, the population growth rate is significantly smaller than λ_{\max} , see Fig. 5.13c,f, due to the growth-curbing effect of environmental fluctuations. Upon increasing D (and therefore the relevance of diffusion in the phenotypic space), distributions start to acquire non-trivial traits, including bimodality (see Fig. 5.13a,d)

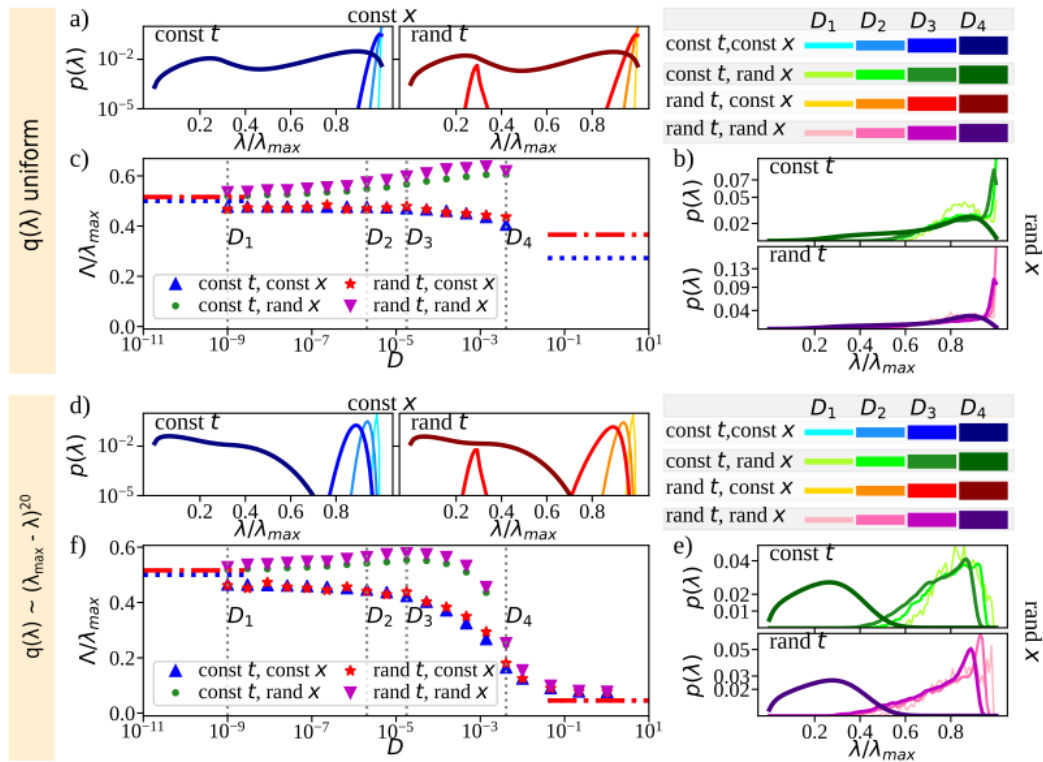


Figure 5.13. Phenotypic distributions and fitness in a symmetric environment.

a,b) Asymptotic, time-averaged phenotypic distributions obtained for a population evolving with a diffusive kernel in a uniform background phenotypic landscape ($a = 0$). **c)** Asymptotic population growth rate Λ (in units of λ_{\max}) as a function of D for the four types of environment. Vertical dotted lines mark the values of diffusion studied in panels (a) and (b). Horizontal lines at small and large D stand for the analytical estimates for the fitness obtained in the const- x regime (dotted blue line) and the rand- x regime, respectively. **d-f)** Same as a-c but with $q(\lambda)$ with $a = 20$ rather than uniform. Adapted from [8].

and extended tails (see Fig. 5.13b,e). The population growth rate Λ then increases with D with respect to the small-diffusion limit in complex environments, where the population structure develops tails. In such cases, Λ has a well-defined maximum at a specific value of D (which depends, as in [107], on the characteristic time of environmental switches), marking the existence of an optimal trade-off between diffusion (exploration) and growth (exploitation) in the given environment. On the other hand, the population growth rate decreases continuously with D , albeit slowly, in the simpler two-state environments, implying that any amount of exploration is detrimental to fitness in such contexts. When diffusion dominates the dynamics (larger values of D), Λ appears to drop rapidly in all environments. In such a case, which is close to the purely ‘exploration’ limit $D \rightarrow \infty$, cells explore the phenotypic space very efficiently, continuously redistributing their CGR among allowed states. The asymptotic behavior is hence dominated by the background provided by $q(\lambda)$. Indeed, the phenotypic distribution evolves towards its stationary limit $q(\lambda)$ due to the detailed balance constraint.

This suggests that in fluctuating environments the maximum of the fitness is not necessarily associated to the strongest regulation cells can exert. In fact, they can indeed tune the phenotypic diffusion through regulation to cope optimally with environmental fluctuations so as to ensure a significant gain in terms of fitness, provided the selective threshold of the environment changes randomly over time. Instead, in periodic environments, remains valid what we found in Section 4.0.2: the more the regulation the higher the fitness. In both cases, the fitness advantage that regulation can provide are enhanced by heterogeneity in the underlying distribution ($q(\lambda)$). In fact, the limiting case of an uniform distribution produces nearly no differences in fitness while phenotypic diffusion spans many orders of magnitude.

Finally, we can notice from Fig. 5.13a,b,d,e, that the population structures into an extended unimodal distribution of phenotypes in the case of a random environment. While, on the other hand, in an environment fluctuating between two well-defined states, bimodal phenotypic distributions occur. Notably, varying the fraction of time the population spends in the favorable/unfavorable regime, the relative intensity of the two peaks show a well defined trend. Ranging from a completely favorable environment to a fully selective one, the population distribution shifts from the right peak (high CGR) to the left one (low CGR) as we can see from Figure 5.14. This phenomenon is observed also in real system and it is known as ‘bet-hedging’. In the next two sections, we will spend a few more words on both the trade-off and the bet-hedging.

5.3.5 Exploration-exploitation trade-off

The exploration vs exploitation dilemma exists in many aspects of our life. Say, our favorite restaurant is right around the corner. If we keep going there every day, we can be quite confident of what we will get. Anyway, we miss the chances of discovering an even better option. On the other hand, if we try new places all the time, very likely we end having to eat unpleasant food from time to time.

The dilemma comes from the incomplete information: we need to gather enough information to make best overall decisions while keeping the risk under control. With exploitation, we take advantage of the best option we know. With exploration, we take some risk to collect information about unknown options. The best long-term strategy may involve short-term sacrifices. For example, one exploration trial could be a total failure, but it warns us of not taking that action too often in the future. Setting apart the delicate issue of restaurants, an especially significant effort to understand this trade-off is ongoing for biological systems, as seen e.g. in the recent interest about the ‘ecology of cancer growth’ [108, 109] (the strikingly diverse distributions of cell strains observed throughout different types of cancers) and its relationship to the timing of drug administration [109]. Microbial systems have also been a natural testing ground for the exploration-exploitation scenario for many years. It is empirically known that, in fluctuating environments, microbes tend to display a high degree of phenotypic heterogeneity driven by stochasticity in the regulation of gene expression and metabolism [110, 87, 111]. The ability to explore the space of allowed phenotypes ultimately provides an effective route to hedge against environmental noise [112, 113], favoring e.g. the persistence of a sub-population of resistant but slow-growing bacteria within a population subject to

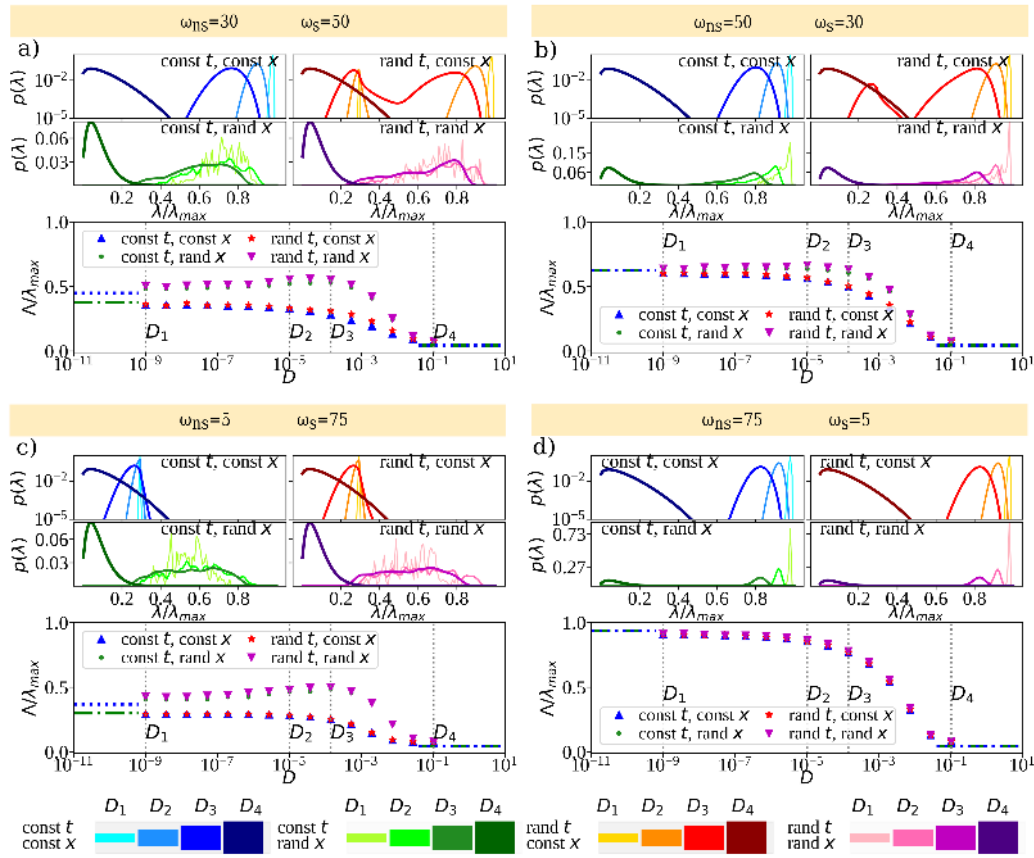


Figure 5.14. Phenotypic distributions and fitness in an asymmetric environment: bet-hedging. **a)** Asymptotic, time-averaged phenotypic distributions (top panels) and asymptotic population growth rate Λ (in units of λ_{\max}) as a function of D (bottom panel) obtained as in Fig. 5.12d-f but in the presence of an asymmetric environment with characteristic switching times $\omega_s = 30$ and $\omega_{ns} = 50$ time units. Blue and green horizontal lines at small and large D show the analytical estimates of Λ in the exploration and exploitation limits, obtained in the const- x and rand- x regimes, respectively. Grey horizontal lines represent the same analytical estimates but in the presence of a symmetric environment with $\omega_{ns} = \omega_s = 40$ time units. **(b to d)** Same as (a) but with different choices of ω_{ns} and ω_s . Displayed curves are averaged over 100 independent realizations of the dynamics. Adapted from [8].

high doses of antibiotics [114, 115]. Coming back to our case of study, the central question concerns which balance of replication (exploitation) and phenotypic change (exploration) will provide the population with the highest fitness (e.g. the fastest growth rate) in the long run. The optimal strategy is obviously interlocked with details like the statistics of resources and in more complex systems can be challenging to analyze at a quantitative level [107, 116, 117]. Still, fitness maximization is very often found to require a non-zero exploration rate.

The multi-armed bandit problem

The classic Exploration-Exploitation problem in mathematics is the multi-armed bandit [118], which is a slang term for a bank of slot machines. The problem affects players, who know that each machine has a variable payoff and they have a limit number of attempts before running out of money. The goal is to balance trying out new machines with unknown payoffs against exploiting the knowledge one already has from the earlier machines we have tried. When a player first starts on new bandits, he really does not know which will pay out and at what rates. So some exploration is necessary to know what the reward ratio in the territory will be. As the knowledge grows, the player gets to know which bandits are likely to pay, and which are not, and this later informs his choices as to where to place his dollars [119].

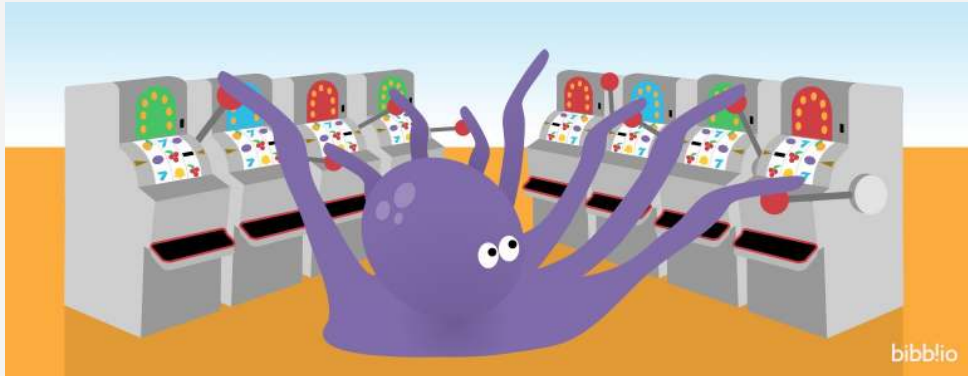


Figure 5.15. Cartoon of the multi-armed bandit problem.

5.3.6 Biological bet-hedging

Biological bet hedging was originally proposed to explain the observation of a seed bank, or a reservoir of ungerminated seeds in the soil [120]. For example, an annual plant's fitness is maximized for that year if all of its seeds germinate. However, if a drought occurs that kills germinated plants, but not ungerminated seeds, plants with seeds remaining in the seed bank will have a fitness advantage. Therefore, it can be advantageous for plants to 'hedge their bets' in case of a drought by producing some seeds that germinate immediately and other seeds that lie dormant. Examples of biological bet hedging are found in mechanism as different as female multiple mating [121] and bacterial persistence in the presence of antibiotics [77, 122]. The first formal study of the bet-hedging features dates 1956 [123]. In practice, we can speak of biological bet hedging when organisms suffer decreased fitness in

their typical conditions in exchange for increased fitness in stressful conditions [113]. Indeed, this is exactly what happens in our case. In fact, in a periodic environment (where periodicity must be solely relied to the position of the threshold and not on the times of the switches), the population splits in two well defined peaks. One remains near the threshold (where growth is always allowed) and another sets near λ_{\max} . As we can see from Figure 5.14, varying the fractions of time the system remains in the selective regime, interestingly, the weight of the slowest part of the distribution (smaller λ) follows the (mean) time spent in the selective state (i.e. it increases with ω_s). This means that the population does not just place some persister cells at low growth rates, but it ‘learns’ to invest the right fraction depending on the probability of having a selective environment.

Chapter 6

Discussions

Regulation has the role of conferring both robustness to gene expression and providing the flexibility cells need to rapidly adapt to sudden environmental fluctuations. Robustness and regulation capacity are both linked to how the network handles noise. In general, the intrinsic stochasticity in gene expression must be dampened in order to have stable output protein levels. However, cells often must respond to external fluctuations, with a specific reorganization of their proteome. So, the regulatory network has to process noise in a selective way, containing some kinds of noise while properly propagating some perturbations [124].

In order to do so, evolution selected many different molecular motifs, which intersect resulting in complex networks with lots of interactions. In those networks, competition to bind substrates, enzymes or gene expression machinery is ubiquitous and impacts regulatory processes in several ways [125]. For instance, the initiation and translation rates of different transcripts are effectively coupled by the competition for the ribosome pool, so that modifications of a given RNA species can alter the translational dynamics of other RNAs [126]. Quite generally, competition for limited and shared molecular resources induces effective interactions between the competing species, with signs (positive or negative) that depend on the specifics of the underlying processes [127]. Such interactions constitute an effective additional layer of indirect regulation, whose intensity is strongly context-dependent [128]. As we discussed, competition for miRNAs (or, more generally, small regulatory RNAs) among long transcripts is undergoing much scrutiny [129]. Following the work done in [4], we showed how miRNA-RNA interaction network can be characterized in terms of a set of coupled differential equations describing the concentrations of all the involved molecules. Such modelization permits to measure the crosstalk (through the evaluation of susceptibility functions, which quantify the response of each component of the system to the variation of any other) and analyze its features, like selectivity and directionality. Furthermore, whereas the typical crosstalk interaction generated by small changes in RNA availability is weak, non-local effects are found to be significant and crosstalk patterns can be modulated by cells, tuning the levels of free miRNA molecules. Collecting data from available databases [5, 43] and looking for a robust set of parameters [130], we probe the response to perturbations in system-level networks. In particular, we showed that all the identified features persist, are modulated by transcriptional and/or binding heterogeneities and appear

to be hard-wired in the topology of the underlying miRNA-RNA interactome. Most importantly, since both the strength of the bindings and the connections are under selective pressure, evolution seems to be playing a role in tuning the network in such a way to optimize noise processing. Probing the crosstalk scenario, we described, in experiments turned out to be challenging essentially due to weakness and non-locality of the interactions. To validate the picture, both in terms of individual interactions and of global features, we can however resort to transcriptomic data since a reformulation of fluctuation-dissipation [70] can relate, under certain conditions, correlations to the real susceptibilities. Evaluating such quantities on system-scale RNA readouts will provide a direct, data-driven test of the RNA crosstalk. Moving upward, the interplay between noise and regulation reflects on the phenotypes cells manifest. In particular, the differences in the microscopic states of each cell (intrinsic noise) give rise to variability at the level of the phenotype. Along the line of [7], we saw that the shape of experimental growth rate distributions can be recovered using both a maximum entropy approach and a minimal dynamical model. Control parameters (the MaxEnt inverse temperature, β and the phenotypic diffusion coefficient, D) can be introduced to shift the position of the mean growth rate. This informs that those two parameters are linked to the amount of regulation cells can exert, which depends on the environment the population lives in. Overall, those models told us that the growth of the population depends both on the characteristics of the cellular population and on the interaction with the environment. In fact, phenotypic heterogeneity influences the fitness in the case of a controlled environment where the only external effect is due to the presence of a carrying capacity [89]. In such conditions, the initial number of cells, N_0 , reflects on growth observables like the lag time and the population maximum growth rate as much as on their fluctuations. While initial-density dependencies are not totally surprising, since the growth rate of a population of non-interacting cells is naturally expected to decline as N_0 approaches the carrying capacity of the medium [78], increases of the growth rate with N_0 away from the carrying capacity are harder to understand. At the simplest level, they suggest the existence of a regime in which populations starting from larger inocula can achieve higher fitness e.g. by strengthening cooperation [106]. This however implies that populations can maintain a memory of their initial conditions well into the exponential phase. It is well known that randomness in cellular reproduction events can propagate to macroscopic parameters [131]. If such heterogeneities are putatively averaged out in large inocula, they become relevant when the initial cell density is small, so that even populations of identically prepared single cells can generate very different growth trajectories. The scenario complicates even further if we turn on fluctuations in the environment, in particular if those variations produce selective regimes where some cells can not replicate. In this situation, phenotypic variability is found to help the population to withstand external shocks. In fact, empirical data on phenotypic distributions, quantified from protein expression data, display a rich spectrum of behaviors of cellular populations, ranging from unimodal to bimodal distribution [132]. The question of when one type of distribution is favored naturally arises. The model we discussed suggests that the population structure is tightly linked to the specific features of the environment. In particular, generalizing the setup considered in [7, 133] and discussed in Chapter 4 to the case in which the replication rate depends on the coupling of cells to a fluctuating environment,

we saw that the structure of a population emerges from the balance between the term that rewards fast-growing states (which are however sensitive to environmental shocks) and the diffusion term favoring states with larger entropy in the phenotypic space (but slower replication rates). When the strength of the coupling between the environment and phenotypes takes on two distinct levels, bimodal distributions arise but exploration does not yield a fitness advantage to the population. On the other hand, under the more complex scenario in which the coupling strength varies randomly, the exploration-exploitation trade-off leads to a non-zero optimal search rate and unimodal phenotypic distributions are generically preferred. This picture is in complete agreement with the results obtained in [134], where the theoretical benefit of a bimodal distribution of stress response proteins was found to be highest in two-state environments, while more variable and structured environments allow for the selection of unimodal distributions. In addition, we saw that adding a small amount of diffusion to a purely exploitative strategy always leads to an increase of the fitness in random environments, while it is always detrimental in periodic environments. The fitness gain given by exploration also appears to be linked to the structure of the underlying phenotypic landscape $q(\lambda)$. In fact, the more $q(\lambda)$ is strongly heterogeneous, with rare fast phenotypes among a multitude of slow ones [7], the more significant the fitness advantage diffusion can provide. Since more heterogeneous populations are more likely to evolve in complex environments, this suggests that higher intra-tumoral heterogeneity may be the result of highly variable micro-environments. On the other hand, if we assume that the selective environment is caused by a therapeutic protocol, subjecting the population to a single repeated dose is effective in quenching its fitness irrespective of the timing of administration. Overall, such a framework sheds some light on the origin of phenotypically heterogeneous cell populations such as tumors and point to educated strategies to control their diversity.

Appendix A

Susceptibilities calculation

Dimensionless susceptibilities are defined as:

$$\begin{cases} \chi_{ij} = \frac{\partial[m_i]}{\partial m_j^*} & \chi_{ia} = \frac{\partial[m_i]}{\partial \mu_a^*} \\ \chi_{ab} = \frac{\partial[\mu_a]}{\partial \mu_b^*} & \chi_{ai} = \frac{\partial[\mu_a]}{\partial m_i^*} \end{cases} \quad (\text{A.1})$$

where i and j refer to RNA molecules while a and b to miRNA ones. We remind that m_i^* and β_a^* are the maximum level of free RNA and miRNA, respectively. Starting from the expression for the χ_{ij} of the main text (that we repeat for simplicity)

$$\chi_{ij} = \delta_{ij} \frac{[m_i]}{m_i^*} + \frac{[m_i]^2}{m_i^*} \sum_{\alpha \in (i \cap j)} \frac{[\mu_\alpha]^2}{\mu_{i\alpha} m_{j\alpha} \mu_\alpha^*} \frac{[m_j]}{m_j^*} \quad (\text{A.2})$$

we can compute the equivalent for the miRNA-miRNA susceptibility just carefully substituting indexes as

$$\chi_{ab} = \delta_{ab} \frac{[\mu_a]}{\mu_a^*} + \frac{[\mu_a]^2}{\mu_a^*} \sum_{i \in (a \cap b)} \frac{[m_i]^2}{m_{ia} \mu_{ib} m_i^*} \frac{[\mu_b]}{\mu_b^*} \quad (\text{A.3})$$

From here, the miRNA counterpart of Eq. 3.8 reads

$$\sum_c (\delta_{ac} - V_{ac}) \chi_{cb} = \frac{[\mu_a]}{\mu_a^*} \delta_{ab} \quad (\text{A.4})$$

where

$$V_{ac} = \frac{[\mu_a]^2}{\mu_a^*} \sum_{k \in (N_a \cap N_c)} \frac{1}{m_{ka}^0 \mu_{kb}^0} \frac{[m_k]^2}{m_k^*} \quad (\text{A.5})$$

The $\hat{\chi}$ matrix of the miRNA-miRNA susceptibilities can be obtained again through matrix inversion:

$$\hat{\chi} = (\hat{\mathbf{1}} - \hat{V})^{-1} \text{diag} \left(\frac{\mu}{\mu^*} \right) \quad (\text{A.6})$$

Once we know the susceptibilities among couples of the same species, we can easily compute also the cross-species susceptibilities since

$$\chi_{ai} = \mu_a^* \sum_k \frac{\partial F_a}{\partial m_k} \frac{\partial m_k}{m_i^*} = -\frac{\mu_a^2}{\mu_a^*} \sum_k \frac{\chi_{ki}}{m_{ak}^0} \quad (\text{A.7})$$

and

$$\chi_{ib} = m_i^* \sum_c \frac{\partial F_i}{\partial \mu_c} \frac{\partial \mu_c}{\mu_b^*} = -\frac{m_i^2}{m_i^*} \sum_c \frac{\chi_{cb}}{\mu_{ic}^0} . \quad (\text{A.8})$$

Appendix B

Degree-preserving randomization

To randomize the CLASH network while preserving the degree sequence we can employ a standard edge-swapping algorithm, i.e.

1. randomly select two links l_{ia} and l_{jb} from the miRNA-RNA network with uniform probability;
2. swap the links, obtaining new connections l_{ja} and l_{ib} while keeping the inverse binding affinities (μ_{ia}^0 and μ_{jb}^0) associated to RNAs i and j respectively;
3. discard the swap if it generates duplicate links or if the resulting network is not connected;
4. iterate steps 1-3 a number n of times much larger than the total number of links in the network.

The resulting edge-swapped network has the same number of links and the same one-point statistics (i.e. the node connectivities) of the original network, while higher-order (e.g. two-node) topological correlations are lost. As shown in Figure B.1, the structure of randomized networks differs only slightly from that of the original CLASH network in the distribution of shortest paths between RNA species, whereas degree distributions are expectedly unchanged. In such conditions, global crosstalk descriptors are nearly identical to those obtained in the original CLASH network (panels d through g). This confirms that node degrees are the key topological determinant of the crosstalk scenario derived from the CLASH data.

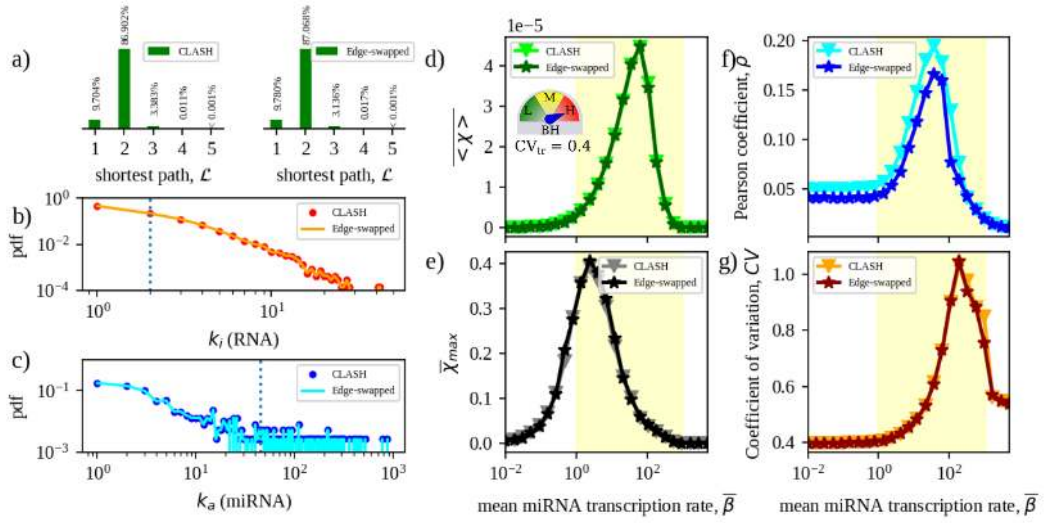


Figure B.1. Comparison between crosstalk patterns in the CLASH network and its edge-swapped randomized versions. **a)** Frequency of the shortest paths for CLASH (left) and edge-swapped (right) networks. **b-c)** Degree distributions of RNA (top) and miRNA (bottom) nodes in the CLASH and edge-swapped networks. **d-g)** Global crosstalk descriptors for the CLASH and edge-swapped networks as a function of the mean miRNA transcription rate $\bar{\beta}$: (d) mean susceptibility; (e) mean maximum susceptibility; (f) Pearson correlation coefficient between susceptibilities and local kinetic parameters; (g) Coefficient of variation of RNA levels. Adapted from [6].

Appendix C

Approximated phenotypic distribution

Based on the reasoning done in 5.3 and considering a symmetrical environment where on average the system spends half the time in a favorable regime and the other half in the selective one, we can approximate $p(\lambda, t)$ with the bimodal function

$$p(\lambda, t) \simeq \alpha(t)\delta(\lambda - \lambda^*) + (1 - \alpha(t))\delta(\lambda - \lambda_{\max}) \quad , \quad (\text{C.1})$$

with $0 \leq \alpha(t) \leq 1$ a time-dependent coefficient quantifying the fraction of cells with CGR equal to $\lambda^* < \lambda_{\max}$. (For sakes of simplicity, we shall henceforth omit to indicate explicitly the dependence of α on time.) We can then use (5.34), which in discrete time takes the form

$$p(\lambda, t + \delta t) \simeq \left\{ 1 + [f(\lambda, t) - \langle f \rangle] \delta t \right\} p(\lambda, t) \quad , \quad (\text{C.2})$$

to evolve the above ansatz for small time intervals δt during which the environment does not change.

In non-selective conditions ($x = \lambda_{\max}$), one can use the fact that

$$\langle f \rangle = \alpha\lambda^* + (1 - \alpha)\lambda_{\max} \quad (\text{C.3})$$

to arrive at

$$p(\lambda, t + \delta t) \simeq (\alpha - \delta\alpha_{\text{ns}}) \delta(\lambda - \lambda^*) + (1 - \alpha + \delta\alpha_{\text{ns}}) \delta(\lambda - \lambda_{\max}) \quad , \quad (\text{C.4})$$

where $\delta\alpha_{\text{ns}} = (\lambda_{\max} - \lambda^*)(1 - \alpha)\delta t$.

In a selective environment ($x = \lambda^*$), instead,

$$\langle f \rangle = \alpha\lambda^* \quad , \quad (\text{C.5})$$

and one finds

$$p(\lambda, t + \delta t) \simeq (\alpha + \delta\alpha_{\text{s}}) \delta(\lambda - \lambda^*) + (1 - \alpha - \delta\alpha_{\text{s}}) \delta(\lambda - \lambda_{\max}) \quad , \quad (\text{C.6})$$

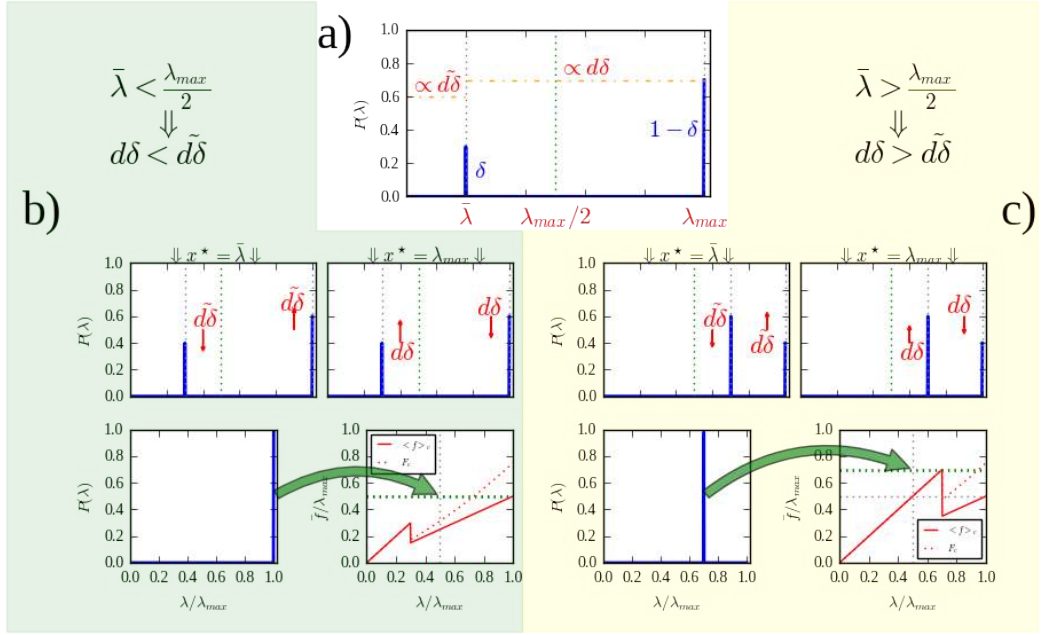


Figure C.1. Cartoon representation of the evolution of the distribution.

with $\delta\alpha_s = \lambda^*(1 - \alpha)\delta t$.

This shows that, at every switch, the population distribution will tend to shift from one threshold to the other, but the speed with which the two peaks grow or shrink are different. In particular, one has

$$\frac{\delta\alpha_{ns}}{\delta\alpha_s} = \frac{(\lambda_{max} - \lambda^*)}{\lambda^*}. \quad (C.7)$$

This implies that $\delta\alpha_{ns} < \delta\alpha_s$ for $\lambda^* > \lambda_{max}/2$. Hence the peak growing at speed $\delta\alpha_s$ is favored and the probability density will peak around λ^* in the long run. On the other hand, $\delta\alpha_{ns} > \delta\alpha_s$ when $\lambda^* < \lambda_{max}/2$, causing the population to concentrate around λ_{max} . In other terms,

$$p(\lambda) \simeq \begin{cases} \delta(\lambda - \lambda^*) & \text{if } \lambda^* > \lambda_{max}/2 \\ \delta(\lambda - \lambda_{max}) & \text{if } \lambda^* < \lambda_{max}/2 \end{cases}, \quad (C.8)$$

in agreement with the numerical picture for the two-state (const- x) environment shown in Fig. 5.12.

Acronyms and Glossary of Terms

apoptosis The death of cells which occurs as a normal and controlled part of an organism's growth or development. 5

capping Capping is a form of RNA processing in which the 5' end of the nascent pre-mRNA is capped with a 7-methyl guanosine nucleotide, 7-methylguanylate. Capping occurs shortly after initiation of transcription. The 5' cap is retained in mature mRNAs. Capping is required to protect the RNA transcript from degradation. 5

chromatin The material of which the chromosomes of organisms other than bacteria (i.e. eukaryotes) are composed, consisting of protein, RNA, and DNA. 5

cytosol The cytosol, also known as intracellular fluid (ICF) or cytoplasmic matrix, or groundplasm, is the liquid found inside cells. It is separated into compartments by membranes. 10

deadenylation The removal of an adenylate group from a protein. 9

mRNA export The process of transportation of the mRNA from the nucleus to the cytoplasm. 5

noncoding RNA A non-coding RNA (ncRNA) is an RNA molecule that is not translated into a protein. The DNA sequence from which a functional non-coding RNA is transcribed is often called an RNA gene. Abundant and functionally important types of non-coding RNAs include transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as small RNAs such as microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, exRNAs, scaRNAs and the long ncRNAs such as Xist and HOTAIR. 7

polyadenylation Polyadenylation is the addition of a poly(A) tail to a messenger RNA. The poly(A) tail consists of multiple adenosine monophosphates; in other words, it is a stretch of RNA that has only adenine bases. In eukaryotes, polyadenylation is part of the process that produces mature messenger RNA (mRNA) for translation. 5

proliferation Cell proliferation is the process that results in an increase of the number of cells, and is defined by the balance between cell divisions and cell loss through cell death or differentiation. Cell proliferation is increased in tumours. 5

PTM Post-translational modification (PTM) refers to the covalent and generally enzymatic modification of proteins following protein biosynthesis. Proteins are synthesized by ribosomes translating mRNA into polypeptide chains, which may then undergo PTM to form the mature protein product. PTMs are important components in cell signaling, as for example when prohormones are converted to hormones. 5

ribosome Ribosomes comprise a complex macromolecular machine, found within all living cells, that serves as the site of biological protein synthesis (translation). Ribosomes link amino acids together in the order specified by messenger RNA (mRNA) molecules. Ribosomes consist of two major components: the small ribosomal subunits, which read the mRNA, and the large subunits, which join amino acids to form a polypeptide chain. Each subunit consists of one or more ribosomal RNA (rRNA) molecules and a variety of ribosomal proteins. The ribosomes and associated molecules are also known as the translational apparatus. 7

splicing RNA splicing, in molecular biology, is a form of RNA processing in which a newly made precursor messenger RNA (pre-mRNA) transcript is transformed into a mature messenger RNA (mRNA). During splicing, introns (Non-coding regions) are removed and exons (Coding Regions) are joined together. 5

titration A titration is a technique where a solution of known concentration is used to determine the concentration of an unknown solution. Typically, the titrant (the know solution) is added from a buret to a known quantity of the analyte (the unknown solution) until the reaction is complete. 5

transcription factor In molecular biology, a transcription factor (TF) (or sequence-specific DNA-binding factor) is a protein that controls the rate of transcription of genetic information from DNA to messenger RNA, by binding to a specific DNA sequence. The function of TFs is to regulate—turn on and off—genes in order to make sure that they are expressed in the right cell at the right time and in the right amount throughout the life of the cell and the organism. 8

UTR In molecular genetics, an untranslated region (or UTR) refers to either of two sections, one on each side of a coding sequence on a strand of mRNA. If it is found on the 5' side, it is called the 5' UTR (or leader sequence), or if it is found on the 3' side, it is called the 3' UTR (or trailer sequence). 9

Watson-Crick base pairing In canonical Watson-Crick base pairing in DNA, adenine (A) forms a base pair with thymine (T) using two hydrogen bonds, and guanine (G) forms a base pair with cytosine (C) using three hydrogen

bonds. In canonical Watson-Crick base pairing in RNA, thymine is replaced by uracil (U). 9

Bibliography

- [1] W.S. Bialek. *Biophysics: Searching for Principles*. Princeton University Press, 2012.
- [2] Francesco Fazi and Clara Nervi. MicroRNA: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular Research*, 79(4):553–561, 06 2008.
- [3] 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*, pages 367–409. Springer New York, New York, NY, 2019.
- [4] 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, mar 2013.
- [5] Aleksandra Helwak, Grzegorz Kudla, Tatiana Dudnakova, and David Tollervey. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–665, apr 2013.
- [6] Mattia Miotto, Enzo Marinari, and Andrea De Martino. Competing endogenous RNA crosstalk at system level. *PLOS Computational Biology*, 15(11):e1007474, November 2019.
- [7] Daniele De Martino, Fabrizio Capuani, and Andrea De Martino. Growth against entropy in bacterial metabolism: the phenotypic trade-off behind empirical growth rate distributions in E.coli. *Physical Biology*, 13(3), may 2016.
- [8] Andrea De Martino, Thomas Gueudré, and Mattia Miotto. Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations. *Phys. Rev. E*, 99:012417, Jan 2019.
- [9] Mads Kærn, Timothy C. Elston, William J. Blake, and James J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, May 2005.
- [10] Dawn Fraser and Mads Kaern. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular Microbiology*, 71(6):1333–1340, March 2009.

- [11] Guilhem Chalancon, Charles N.J. Ravarani, S. Balaji, Alfonso Martinez-Arias, L. Aravind, Raja Jothi, and M. Madan Babu. Interplay between gene expression noise and regulatory network architecture. *Trends in Genetics*, 28(5):221–232, May 2012.
- [12] F. Cecconi, M. Cencini, M. Falcioni, and A. Vulpiani. Brownian motion and diffusion: From stochastic processes to chaos and beyond. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026102, June 2005.
- [13] Lev S Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, January 2014.
- [14] Nicole E. Radde and Marc-Thorsten Hütt. The physics behind systems biology. *EPJ Nonlinear Biomedical Physics*, 4(1), August 2016.
- [15] Juan I. Castrillo and Stephen G. Oliver, editors. *Yeast Systems Biology*. Humana Press, 2011.
- [16] Mattia Miotto and Lorenzo Monacelli. Genome heterogeneity drives the evolution of species. *arXiv:1912.01444*, 2019.
- [17] Mattia Miotto, Lorenzo Di Rienzo, Pietro Corsi, Giancarlo Ruocco, Domenico Raimondo, and Edoardo Milanetti. Simulated epidemics in 3d protein structures to detect functional properties. *arXiv:1906.05390*, 2019.
- [18] Mattia Miotto, Pier Paolo Olimpieri, Lorenzo Di Rienzo, Francesco Ambrosetti, Pietro Corsi, Rosalba Lepore, Gian Gaetano Tartaglia, and Edoardo Milanetti. Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics*, 35(15):2569–2577, December 2018.
- [19] Mattia Miotto and Lorenzo Monacelli. Entropy evaluation sheds light on ecosystem complexity. *Physical Review E*, 98(4), October 2018.
- [20] Tong Ihn Lee and Richard A. Young. Transcriptional regulation and its misregulation in disease. *Cell*, 152(6):1237–1251, March 2013.
- [21] Tom Maniatis and Robin Reed. An extensive network of coupling among gene expression machines. *Nature*, 416(6880):499–506, April 2002.
- [22] George Orphanides and Danny Reinberg. A unified theory of gene expression. *Cell*, 108(4):439–451, February 2002.
- [23] Nick J. Proudfoot, Andre Furger, and Michael J. Dye. Integrating mRNA processing with transcription. *Cell*, 108(4):501–512, February 2002.
- [24] Adrienne E McKee and Pamela A Silver. Systems perspectives on mRNA processing. *Cell Research*, 17(7):581–590, July 2007.
- [25] James T Kadonaga. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, 116(2):247–257, January 2004.
- [26] David J. Lockhart and Elizabeth A. Winzeler. Genomics, gene expression and DNA arrays. *Nature*, 405(6788):827–836, June 2000.

- [27] Sean E Hanlon and Jason D Lieb. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Current Opinion in Genetics & Development*, 14(6):697–705, December 2004.
- [28] Tony Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693–705, February 2007.
- [29] David P. Bartel. MicroRNAs: Target recognition and regulatory functions. *Cell*, 136(2):215–233, January 2009.
- [30] Matteo Osella, Carla Bosia, Davide Corá, and Michele Caselle. The role of incoherent MicroRNA-mediated feedforward loops in noise buffering. *PLoS Computational Biology*, 7(3):e1001101, March 2011.
- [31] Carla Bosia, Andrea Pagnani, and Riccardo Zecchina. Modelling competing endogenous RNA networks. *PLoS ONE*, 8(6):e66609, June 2013.
- [32] William J. Blake, Gábor Balázs, Michael A. Kohanski, Farren J. Isaacs, Kevin F. Murphy, Yina Kuang, Charles R. Cantor, David R. Walt, and James J. Collins. Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell*, 24(6):853–865, December 2006.
- [33] Gürol M. Süel, Jordi Garcia-Ojalvo, Louisa M. Liberman, and Michael B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550, March 2006.
- [34] G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–1719, March 2007.
- [35] H. Maamar, A. Raj, and D. Dubnau. Noise in gene expression determines cell fate in bacillus subtilis. *Science*, 317(5837):526–529, July 2007.
- [36] Alfonso Martinez Arias and Penelope Hayward. Filtering transcriptional noise during development: concepts and mechanisms. *Nature Reviews Genetics*, 7(1):34–44, January 2006.
- [37] Arjun Raj, Scott A. Rifkin, Erik Andersen, and Alexander van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918, February 2010.
- [38] Huili Guo, Nicholas T. Ingolia, Jonathan S. Weissman, and David P. Bartel. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840, August 2010.
- [39] 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, December 1993.
- [40] Bruce Wightman, Ilho Ha, and Gary Ruvkun. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *c. elegans*. *Cell*, 75(5):855–862, December 1993.

- [41] Brenda J. Reinhart, Frank J. Slack, Michael Basson, Amy E. Pasquinelli, Jill C. Bettinger, Ann E. Rougvie, H. Robert Horvitz, and Gary Ruvkun. The 21-nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. *Nature*, 403(6772):901–906, February 2000.
- [42] Ana Kozomara, Maria Birgaoanu, and Sam Griffiths-Jones. miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1):D155–D162, November 2018.
- [43] Vikram Agarwal, George W Bell, Jin-Wu Nam, and David P Bartel. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4, aug 2015.
- [44] Steven M. Johnson, Helge Grosshans, Jaclyn Shingara, Mike Byrom, Rich Jarvis, Angie Cheng, Emmanuel Labourier, Kristy L. Reinert, David Brown, and Frank J. Slack. RAS is regulated by the let-7 MicroRNA family. *Cell*, 120(5):635–647, March 2005.
- [45] Jun-Hao Li, Shun Liu, Hui Zhou, Liang-Hu Qu, and Jian-Hua Yang. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-seq data. *Nucleic Acids Research*, 42(D1):D92–D97, December 2013.
- [46] U. Ala, F. A. Karreth, C. Bosia, A. Pagnani, R. Taulli, V. Leopold, Y. Tay, P. Provero, R. Zecchina, and P. P. Pandolfi. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proceedings of the National Academy of Sciences*, 110(18):7154–7159, March 2013.
- [47] Javad Noorbakhsh, Alex H. Lang, and Pankaj Mehta. Intrinsic noise of microRNA-regulated genes and the ceRNA hypothesis. *PLoS ONE*, 8(8):e72676, August 2013.
- [48] Araks Martirosyan, Matteo Figliuzzi, Enzo Marinari, and Andrea De Martino. Probing the limits to MicroRNA-mediated control of gene expression. *PLOS Computational Biology*, 12(1):e1004715, January 2016.
- [49] P. Ramalingam, J. K. Palanichamy, A. Singh, P. Das, M. Bhagat, M. A. Kassab, S. Sinha, and P. Chattopadhyay. Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA*, 20(1):76–87, November 2013.
- [50] Chia-Hung Chien, Yi-Ming Sun, Wen-Chi Chang, Pei-Yun Chiang-Hsieh, Tzong-Yi Lee, Wei-Chih Tsai, Jorng-Tzong Horng, Ann-Ping Tsou, and Hsien-Da Huang. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Research*, 39(21):9345–9356, August 2011.
- [51] David P. Bartel. Metazoan MicroRNAs. *Cell*, 173(1):20–51, March 2018.
- [52] Izabella Slezak-Prochazka, Joost Kluiver, Debora de Jong, Gertrud Kortman, Nancy Halsema, Sibrand Poppema, Bart-Jan Kroesen, and Anke van den

- Berg. Cellular localization and processing of primary transcripts of exonic MicroRNAs. *PLoS ONE*, 8(9):e76647, September 2013.
- [53] V. Narry Kim. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology*, 6(5):376–385, May 2005.
- [54] R. C. Friedman, K. K.-H. Farh, C. B. Burge, and D. P. Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105, October 2008.
- [55] Anna Lavut and Dina Raveh. Sequestration of highly expressed mRNAs in cytoplasmic granules, p-bodies, and stress granules enhances cell viability. *PLoS Genetics*, 8(2):e1002527, feb 2012.
- [56] Yvonne Tay, Jinqiu Zhang, Andrew M. Thomson, Bing Lim, and Isidore Rigoutsos. MicroRNAs to nanog, oct4 and sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, 455(7216):1124–1128, September 2008.
- [57] Matteo Figliuzzi, Andrea De Martino, and Enzo Marinari. RNA-based regulation: Dynamics and response to perturbations of competing RNAs. *Biophysical Journal*, 107(4):1011–1022, aug 2014.
- [58] Rémy Denzler, Vikram Agarwal, Joanna Stefano, David P. Bartel, and Markus Stoffel. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Molecular Cell*, 54(5):766–776, jun 2014.
- [59] Hua-Sheng Chiu, María Rodríguez Martínez, Elena V . Komissarova, David Llobet-Navas, Mukesh Bansal, Evan O Paull, José Silva, Xuerui Yang, Pavel Sumazin, and Andrea Califano. The number of titrated microRNA species dictates ceRNA regulation. *Nucleic Acids Research*, apr 2018.
- [60] Changye Li, Zhenkai Wang, Jinjin Zhang, Xueying Zhao, Pan Xu, Xiangyong Liu, Minge Li, Changjun Lv, and Xiaodong Song. Crosstalk of mRNA, miRNA, lncRNA, and circRNA and their regulatory pattern in pulmonary fibrosis. *Molecular Therapy - Nucleic Acids*, 18:204–218, dec 2019.
- [61] Cecile M. Pickart and Michael J. Eddins. Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1695(1-3):55–72, nov 2004.
- [62] Daryl E. Klein, Valerie M. Nappi, Gregory T. Reeves, Stanislav Y. Shvartsman, and Mark A. Lemmon. Argos inhibits epidermal growth factor receptor signalling by ligand sequestration. *Nature*, 430(7003):1040–1044, aug 2004.
- [63] Nicolas E. Buchler and Matthieu Louis. Molecular titration and ultrasensitivity in regulatory networks. *Journal of Molecular Biology*, 384(5):1106–1119, dec 2008.
- [64] G. Kudla, S. Granneman, D. Hahn, J. D. Beggs, and D. Tollervey. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in

- yeast. *Proceedings of the National Academy of Sciences*, 108(24):10010–10015, May 2011.
- [65] J. Hausser and M. Zavolan. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat. Rev. Genet.*, 15(9):599–612, Sep 2014.
- [66] A. C. C. Coolen, A. De Martino, and A. Annibale. Constrained markovian dynamics of random graphs. *Journal of Statistical Physics*, 136(6):1035–1067, September 2009.
- [67] Ton Coolen. *Generating Random Networks and Graphs*. Oxford University Press, may 2017.
- [68] Beverly M Emerson. Specificity of gene regulation. *Cell*, 109(3):267–270, May 2002.
- [69] Davide Cora', Angela Re, Michele Caselle, and Federico Bussolino. MicroRNA-mediated regulatory circuits: outlook and perspectives. *Physical Biology*, 14(4):045001, June 2017.
- [70] Araks Martirosyan, Matteo Marsili, and Andrea De Martino. Translating ceRNA susceptibilities into correlation functions. *Biophysical Journal*, 113(1):206–213, July 2017.
- [71] U Marconi, A Puglisi, L Rondoni, and A Vulpiani. Fluctuation–dissipation: Response theory in statistical physics. *Physics Reports*, 461(4-6):111–195, June 2008.
- [72] François Bertaux, Samuel Marguerat, and Vahid Shahrezaei. Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits. *Royal Society Open Science*, 5(3):172234, March 2018.
- [73] Niclas Nordholt, Johan H. van Heerden, and Frank J. Bruggeman. Integrated biphasic growth rate, gene expression, and cell-size homeostasis behaviour of single *b. subtilis* cells. *BioRxiv*, January 2019.
- [74] Steven J. Altschuler and Lani F. Wu. Cellular heterogeneity: Do differences make a difference? *Cell*, 141(4):559–563, May 2010.
- [75] G. Ullman, M. Wallden, E. G. Marklund, A. Mahmutovic, Ivan Razinkov, and J. Elf. High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1611):20120025, February 2013.
- [76] Andrew S. Kennard, Matteo Osella, Avelino Javier, Jacopo Grilli, Philippe Nghe, Sander J. Tans, Pietro Cicuta, and Marco Cosentino Lagomarsino. Individuality and universality in the growth-division laws of single *e. coli* cells. *Phys. Rev. E*, 93:012408, Jan 2016.

- [77] E. Kussell. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743), sep 2005.
- [78] Daniele De Martino, Fabrizio Capuani, and Andrea De Martino. Quantifying the entropic cost of cellular growth control. *Physical Review E*, 96(1), July 2017.
- [79] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, February 2012.
- [80] Aarash Bordbar, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, 15(2):107–120, January 2014.
- [81] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, March 2010.
- [82] Adam M. Feist, Markus J. Herrgård, Ines Thiele, Jennie L. Reed, and Bernhard Ø. Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2):129–143, December 2008.
- [83] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349, June 2010.
- [84] Jennifer L. Rohn, Jigna V. Patel, Beate Neumann, Jutta Bulkescher, Nunu Mchedlishvili, Rachel C. McMullan, Omar A. Quintero, Jan Ellenberg, and Buzz Baum. Myo19 ensures symmetric partitioning of mitochondria and coupling of mitochondrial segregation to cell division. *Current Biology*, 24(21):2598–2605, November 2014.
- [85] A. Celani and M. Vergassola. Bacterial strategies for chemotaxis response. *Proceedings of the National Academy of Sciences*, 107(4):1391–1396, January 2010.
- [86] Christine Vogel and Edward M. Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4):227–232, March 2012.
- [87] M. Thattai. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, may 2004.
- [88] T. Lu, T. Shen, M. R. Bennett, P. G. Wolynes, and J. Hasty. Phenotypic variability of growing cellular populations. *Proceedings of the National Academy of Sciences*, 104(48), nov 2007.
- [89] Chiara Enrico Bena, Marco Del Giudice, Thomas GeudrÃl, Mattia Miotto, Emilia Turco, Andrea De Martino, and Carla Bosia. Inoculum-density dependent growth reveals inherent cooperative effects and stochasticity in cancer cell cultures. *arXiv:1710.10978*, 2017.

- [90] Jacques Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.
- [91] Nicolas Bacanr. Verhulst and the logistic equation (1838). In *A Short History of Mathematical Population Dynamics*, pages 35–39. Springer London, 2011.
- [92] A. Rein and H. Rubin. Effects of local cell concentrations upon the growth of chick embryo cells in tissue culture. *Experimental Cell Research*, 49(3):666–678, March 1968.
- [93] M Coleman. Influence of agitation, inoculum density, pH, and strain on the growth parameters of escherichia coli o157:h7—relevance to risk assessment. *International Journal of Food Microbiology*, 83(2):147–160, June 2003.
- [94] Konstantinos P. Koutsoumanis and Alexandra Lianou. Stochasticity in colonial growth dynamics of individual bacterial cells. *Applied and Environmental Microbiology*, 79(7):2294–2301, January 2013.
- [95] R.C.L. Marteiijn, 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. *Biotechnology Progress*, 16(5):795–799, October 2000.
- [96] Walter M. van Gulik, Anna Maria Nuutila, Ko L. Vinke, Hens J. G. ten Hoppen, and Joseph J. Heijnen. Effects of carbon dioxide, air flow rate, and inoculation density on the batch growth of catharanthus roseus cell suspensions in stirred fermentors. *Biotechnology Progress*, 10(3):335–339, May 1994.
- [97] Sadettin S. Ozturk and Bernhard Ø. Palsson. Effect of initial cell density on hybridoma growth, metabolism, and monoclonal antibody production. *Journal of Biotechnology*, 16(3-4):259–278, November 1990.
- [98] Ana C. Gregrio, Nuno A. Fonseca, Vera Moura, Manuela Lacerda, Paulo Figueiredo, Srgio Simes, Srgio Dias, and Joo 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):e0165817, November 2016.
- [99] C. Pin and J. Baranyi. Kinetics of single cells: Observation and modeling of a stochastic process. *Applied and Environmental Microbiology*, 72(3):2163–2169, March 2006.
- [100] J.-C. Augustin, A. Brouillaud-Delattre, L. Rosso, and V. Carlier. Significance of inoculum size in the lag time of listeria monocytogenes. *Applied and Environmental Microbiology*, 66(4):1706–1710, April 2000.
- [101] T Robinson. The effect of inoculum size on the lag phase of listeria monocytogenes. *International Journal of Food Microbiology*, 70(1-2):163–173, October 2001.
- [102] Jzsef Baranyi. Comparison of stochastic and deterministic concepts of bacterial lag. *Journal of Theoretical Biology*, 192(3):403–408, June 1998.

- [103] Xia Wang, Kotaro Fujimaki, Geoffrey C. Mitchell, Jungeun Sarah Kwon, Kimiko Della Croce, Chris Langsdorf, Hao Helen Zhang, and Guang Yao. Exit from quiescence displays a memory of cell growth and division. *Nature Communications*, 8(1), August 2017.
- [104] Stefany Moreno-Gámez, Robin A. Sorg, Arnau Domenech, Morten Kjos, Franz J. Weissing, G. Sander van Doorn, and Jan-Willem Veening. Quorum sensing integrates environmental cues, cell density and cell history to control bacterial competence. *Nature Communications*, 8(1), October 2017.
- [105] József Baranyi and Terry A. Roberts. A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, 23(3-4):277–294, November 1994.
- [106] P. A. Stephens, W. J. Sutherland, and R. P. Freckleton. What is the allee effect? *Oikos*, 87(1):185, October 1999.
- [107] Thomas Gueudré, Alexander Dobrinevski, and Jean-Philippe Bouchaud. Explore or exploit? a generic model and an exactly solvable case. *Physical Review Letters*, 112(5), February 2014.
- [108] Kirill S. Korolev, Joao B. Xavier, and Jeff Gore. Turning ecology and evolution against cancer. *Nature Reviews Cancer*, 14(5):371–380, April 2014.
- [109] Nicholas McGranahan and Charles Swanton. Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell*, 168(4):613–628, February 2017.
- [110] B. Hallet. Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr. Opin. Microbiol.*, 4(5):570–581, Oct 2001.
- [111] J. H. van Heerden, M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. M. Bollen, R. Planque, J. Hulshof, T. G. O'Toole, S. A. Wahl, and B. Teusink. Lost in transition: Start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114–1245114, January 2014.
- [112] Murat Acar, Jerome T Mettetal, and Alexander van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics*, 40(4):471–475, March 2008.
- [113] Ard Jan Grimbergen, Jeroen Siebring, Ana Solopova, and Oscar P Kuipers. Microbial bet-hedging: the power of being different. *Current Opinion in Microbiology*, 25:67–72, June 2015.
- [114] N. Q. Balaban. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, September 2004.
- [115] Iris Keren, Niilo Kaldalu, Amy Spoering, Yipeng Wang, and Kim Lewis. Persister cells and tolerance to antimicrobials. *FEMS Microbiology Letters*, 230(1):13–18, January 2004.

- [116] Thomas Gueudré and David G. Martin. Optimal growth entails risky localization in population dynamics. *EPL (Europhysics Letters)*, 121(6):68005, March 2018.
- [117] T. Dannemann, D. Boyer, and O. Miramontes. Lévy flight movements prevent extinctions and maximize population abundances in fragile Lotka-Volterra systems. *Proc. Natl. Acad. Sci. U.S.A.*, 115(15):3794–3799, 04 2018.
- [118] Michael N. Katehakis and Arthur F. Veinott. The multi-armed bandit problem: Decomposition and computation. *Mathematics of Operations Research*, 12(2):262–268, May 1987.
- [119] Donald A. Berry and Bert Fristedt. *Bandit problems*. Springer Netherlands, 1985.
- [120] Dan Cohen. Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12(1):119–129, September 1966.
- [121] Yukio Yasui. Female multiple mating as a genetic bet-hedging strategy when mate choice criteria are unreliable. *Ecological Research*, 16(4):605–616, December 2001.
- [122] J. S. Chuang, O. Rivoire, and S. Leibler. Simpson's paradox in a synthetic microbial system. *Science*, 323(5911):272–275, January 2009.
- [123] J. L. Kelly. A new interpretation of information rate. *The Bell System Technical Journal*, 35(4):917–926, July 1956.
- [124] Margaret S. Ebert and Phillip A. Sharp. Roles for MicroRNAs in conferring robustness to biological processes. *Cell*, 149(3):515–524, apr 2012.
- [125] I. L. Grigorova, N. J. Phleger, V. K. Mutalik, and C. A. Gross. Insights into transcriptional regulation and competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences*, 103(14):5332–5337, March 2006.
- [126] Alon Raveh, Michael Margaliot, Eduardo D. Sontag, and Tamir Tuller. A model for competition for ribosomes in the cell. *Journal of The Royal Society Interface*, 13(116):20151062, March 2016.
- [127] Lei Wei, Ye Yuan, Tao Hu, Shuailin Li, Tianrun Cheng, Jinzhi Lei, Zhen Xie, Michael Q. Zhang, and Xiaowo Wang. Regulation by competition: a hidden layer of gene regulatory network. *Quantitative Biology*, 7(2):110–121, June 2019.
- [128] William H. Mather, Jeff Hasty, Lev S. Tsimring, and Ruth J. Williams. Translational cross talk in gene networks. *Biophysical Journal*, 104(11):2564–2572, June 2013.
- [129] 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.

-
- [130] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38:D750–D753, oct 2009.
- [131] 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, March 2016.
- [132] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, jul 2010.
- [133] Daniele De Martino and Davide Masoero. Asymptotic analysis of noisy fitness maximization, applied to metabolism & growth. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(12):123502, dec 2016.
- [134] Javier Garcia-Bernardo and Mary J. Dunlop. Phenotypic diversity using bimodal and unimodal expression of stress response proteins. *Biophysical Journal*, 110(10):2278–2287, May 2016.