

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN BIOINGEGNERIA

Ciclo XXVIII

Settore Concorsuale di afferenza: 09/G2

Settore Scientifico disciplinare: ING-INF/06

# Effects of transcriptional and post-transcriptional control mechanisms on biological noise in synthetic gene circuits

Presentata da:

*Lucia Bandiera*

Coordinatore Dottorato

**Prof. Elisa Magosso**

Revisori

**Prof. Mauro Ursino**

**Prof. Jordi Garcia Ojalvo**

**Prof. Ian Stansfield**

Relatore

**Prof. Emanuele D. Giordano**

Correlatore

**Dott. Simone Furini**

**Esame finale anno 2016**



ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN BIOINGEGNERIA

Ciclo XXVIII

Settore Concorsuale di afferenza: 09/G2

Settore Scientifico disciplinare: ING-INF/06

# Effects of transcriptional and post-transcriptional control mechanisms on biological noise in synthetic gene circuits

Presented by:

*Lucia Bandiera*

PhD Coordinator

**Prof. Elisa Magosso**

Reviewers

**Prof. Mauro Ursino**

**Prof. Jordi Garcia Ojalvo**

**Prof. Ian Stansfield**

Tutor

**Prof. Emanuele D. Giordano**

Co-Tutor

**Dr. Simone Furini**

**Esame finale anno 2016**



# Acknowledgments

I would first like to thank my supervisor, Emanuele. As he promised me at the beginning of this path, he has been a mentor, motivating me and providing advices while preserving his good humour.

I am also obliged to Dr. Simone Furini, for endless help and support. I am afraid you will not get rid of me so easily.

I would like to give my thanks to all the guys and girls I met at the University of Bologna (too many to be listed in here), and particularly to Alice, Francesca and Marilisa: working with you was great and your presence highly improved the past three years.

I must highlight my gratitude to Prof. Peter Swain. Over the six months I spent in his Lab, he has provided me attention and has encouraged me in pursuing my interests.

Thanks to all the members of the Swain Lab, for welcoming me, for fruitful discussions and suggestions. A special thanks to Matt and Elco. Matt's enthusiasm was a driving force during my staying and I am sure he will be an astonishing PI. Elco... Thanks for sharing with me your project, training and supporting me with the microscope experiments. Thanks because your help overcame the geographical distance, for being a friend other than a colleague.

Infine, vorrei ringraziare di cuore la mia Famiglia. La vostra presenza dà senso a ciò che sono e a qualunque cosa faccia.

Grazie.



*Nothing in life is to be feared,  
it is only to be understood.*

*Now is the time to understand more,  
so that we may fear less.*

Maria Skłodowska.



# Abstract

Synthetic Biology is an interdisciplinary research field seeking to correct faulty cellular processes or implement predictable de-novo tasks by a meticulous engineering of biological systems. Over the past two decades, the evolution of this discipline witnessed a progressive shift from a qualitative to a quantitative approach in treating biological matter at the molecular scale. In the newly acquired perspective, the potential of developing biosynthetic devices of environmental, industrial and medical relevance is hindered by the requirement of accounting for, controlling and finally exploiting the randomness of biochemical events through which biological complexity is implemented.

In this thesis mathematical modelling and experimental acquisitions of basic synthetic circuits are adopted to address questions pertaining the selection of gene expression control mechanisms and network topologies in the design of synthetic devices able to reliably operate in the stochastic cellular context.

In the first chapter we will define biological noise and analyse the beneficial roles it exerts in naturally evolved systems, where its control has been achieved by means of a hierarchy of regulatory mechanisms. We will further describe the experimental methodologies and modelling techniques adopted so far to dissect and quantify stochasticity in gene expression.

In Chapter 2 we present the implementation of a noise tester circuits' catalogue which could provide a tool for quantitatively investigating the robustness of newly designed gene circuits and testing the reliability of available devices' performances. The topology of the synthetic circuits was derived from previous work carried out in the 'Laboratory of Molecular and Cellular Engineering' (ICM Lab) of the University of Bologna. Two synthetic gene circuits, implementing either a transcriptional or a post-transcriptional control in the expression of a green fluorescent reporter, were selected from the circuits' library for detailed characterization. Based on bulk measurements performed on populations of transformants growing in a microplate reader, we will present deterministic models defined to identify the kinetic rates of biochemical reactions governing the circuits' function. Stochastic models, based on these bulk measurements, are used in numerical computations of plasmid copy number effect on gene expression stochasticity.

In Chapter 3 flow cytometry analysis was used to experimentally quantify the steady-state dispersion in protein levels occurring among the individuals of an isogenic population of bacterial *E. coli* cells transformed with the circuits implementing either transcriptional or post-transcriptional control in the fluorescent reporter expression. To the best of our knowledge, the study undertaken in this chapter provides the first experimental, single cell comparison between synthetic circuits operating

through transcriptional and post-transcriptional control. An interesting feature of the stochastic models describing the observed variance in protein levels is the necessity of including extrinsic components, coming from the inclusion of cell division events. Numerical analysis, identifying in post-transcriptional control the best candidate for noise minimization, concludes the chapter.

Finally, we will report the results of research undertaken during a six months period staying at the “Centre for Synthetic and System Biology” (SynthSys) of the University of Edinburgh, under the supervision of Professor Peter Swain. In the project, the phenotypic consequences of a long-non coding RNA on the transcriptional activation of *GAL1-10* promoter in the well-studied eukaryotic model *Saccharomyces Cerevisiae* are investigated using fluorescence microscopy and microfluidics.

# Contents

Acknowledgments .....	i
Abstract .....	v
Contents.....	vii
List of Figures .....	ix
List of Abbreviations .....	xi
<b>1 Introduction.....</b>	<b>1</b>
1.1 Stochasticity in gene expression .....	4
1.1.1 Functional roles of biological noise.....	4
1.1.2 Sources of phenotypic variability.....	5
1.1.3 Quantifying phenotypic variability .....	7
1.2 Biological noise control: copy number and gene regulatory networks.....	10
1.2.1 Translational regulation of gene expression .....	12
1.3 Mathematical models in synthetic biology .....	14
<b>2 Development of a synthetic noise tester catalogue .....</b>	<b>19</b>
2.1 Methods .....	22
2.1.1 Plasmid construction and strain .....	22
2.1.2 Fluorescence assay .....	22
2.1.3 Data analysis .....	23
2.1.4 Numerical simulations.....	23
2.2 Results and discussion .....	26
2.2.1 Definition of the gene-circuits catalogue .....	26
2.2.2 Experimental characterization of synthetic circuits implementing transcriptional or post-transcriptional control of GFP expression .....	28
2.2.3 Macroscopic description of the TC and pTC gene circuits' function .....	32
2.2.4 Copy number effect on stochasticity in GFP expression under transcriptional and post-transcriptional control: a numerical investigation .....	36
2.3 Conclusions .....	39

<b>3</b>	<b>Experimental measurements and mathematical modelling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits</b> .....	<b>41</b>
3.1	Methods .....	44
3.1.1	Single cell fluorescence assays .....	44
3.1.2	Data analysis .....	45
3.1.3	Stochastic simulations .....	46
3.2	Results and discussion .....	48
3.2.1	TC gene circuit: flow cytometry and optical microscopy acquisitions .....	48
3.2.2	TC gene circuit .....	51
3.2.3	pTC gene circuit .....	55
3.2.4	Differences between TC and pTC gene circuits .....	57
3.3	Conclusions .....	59
<b>4</b>	<b>Investigation of <i>GAL10</i>-lncRNA effects on <i>GAL1</i> transcriptional activation using fluorescence microscopy and microfluidics</b> .....	<b>61</b>
4.1	The GAL Network in <i>Saccharomyces cerevisiae</i> .....	62
4.1.1	The <i>GAL</i> gene cluster .....	64
4.1.2	<i>GAL10</i> -lncRNA: evidences from scientific literature .....	66
4.2	<i>GAL10</i> -lncRNA effect on <i>GAL1</i> -GFP activation: single cell microscopy time-series data in microfluidic device .....	71
4.3	Induction of wild-type and <i>GAL10</i> -lncRNA $\Delta$ strains with SC 2% raffinose, 0.02% glucose and 0.04% galactose .....	74
4.4	Analysis of transcriptional dynamics using the UBI-M $\Delta$ k-GFP $\gamma$ reporter .....	80
4.5	Conclusions .....	84
<b>5</b>	<b>Final Remarks</b> .....	<b>85</b>
	<b>Bibliography</b> .....	<b>91</b>
<b>1</b>	<b>Appendix to experimental measurements and mathematical modelling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits</b> .....	<b>101</b>
1.1	Using the moment generating function to derive analytical expressions for protein mean and variance .....	101
	<b>Scientific writing</b> .....	<b>105</b>
	Publications in international journals .....	105
	Abstracts in proceedings of international conferences .....	105

# List of Figures

<b>Figure 1.1:</b> Flow chart for engineering gene circuits .....	3
<b>Figure 1.2:</b> Seminal papers for the experimental quantification of gene expression noise .....	7
<b>Figure 2.1:</b> Post-transcriptional controller characterized in [5] .....	20
<b>Figure 2.2:</b> Gene circuits catalogue .....	27
<b>Figure 2.3:</b> Genetic program for the gene circuits initially undergoing experimental characterization .....	29
<b>Figure 2.4:</b> Normalized steady-state fluorescence for TC and pTC gene circuits .....	30
<b>Figure 2.5:</b> Normalized fluorescence values dependence on the cloning vector's copy number for the TC and pTC gene circuits .....	31
<b>Figure 2.6:</b> Dose-response curves for the TC gene circuit cloned in pSB4A5 and pSB1A2 plasmids .....	33
<b>Figure 2.7:</b> Dose-response curves for the pTC gene circuit cloned in pSB4A5 and pSB1A2 plasmids .....	35
<b>Figure 2.8:</b> Results of the stochastic model for the TC gene circuit .....	37
<b>Figure 2.9:</b> Results of the stochastic model for the pTC gene circuit .....	38
<b>Figure 3.1:</b> Mean expression level and <i>squared coefficient of variation</i> measured with the microscope set-up .....	49
<b>Figure 3.2:</b> Results of the characterization of population size impact on the deviations of microscopy $CV^2$ from the value measured by flow cytometry .....	50
<b>Figure 3.3:</b> Genetic program and dose-response curves for TC and pTC gene circuits .....	51
<b>Figure 3.4:</b> <i>Fano factors</i> in gene circuits TC and pTC .....	53
<b>Figure 3.5:</b> <i>Squared coefficients of variation</i> in gene circuits TC and pTC .....	54
<b>Figure 3.6:</b> <i>Squared coefficients of variation</i> as a function of the average expression level .....	58
<b>Figure 4.1:</b> Functional scheme of the nested feedback loops characterizing the GAL network and the major interactions between the proteins encoded in it .....	63

<b>Figure 4.2:</b> Section of the GAL gene cluster affected by the GAL10-lncRNAs transcription .....	67
<b>Figure 4.3:</b> Results of wild-type and <i>GAL10-lncRNAΔ</i> cells, induced in accordance with Houseley et al. experimental conditions (SC with 2% raffinose, 0.01% galactose and 0.02 %) using the microfluidic device .....	72
<b>Figure 4.4:</b> Comparison of fluorescent histograms plot of wild-type and <i>GAL10-lncRNAΔ</i> cells exposed to SC containing 2% raffinose, 0.02% glucose and 0.04% galactose .....	73
<b>Figure 4.5:</b> Time-varying autofluorescence estimate .....	75
<b>Figure 4.6:</b> Results of wild-type and <i>GAL10-lncRNAΔ</i> induction with SC 2% raffinose, 0.02% glucose and 0.04% galactose .....	75
<b>Figure 4.7:</b> Comparison of wild-type and <i>GAL10-lncRNAΔ</i> responsive (ON) cells .....	76
<b>Figure 4.8:</b> Visual description of kinetics statistics adopted to explore the divergent induction of wild-type and <i>GAL10-lncRNAΔ</i> ON cells .....	77
<b>Figure 4.9:</b> Results of birth events detection performed using the automated daughter counts identification script .....	79
<b>Figure 4.10:</b> Birth events detection of wild-type* and <i>GAL10-lncRNAΔ</i> * strain .....	81
<b>Figure 4.11:</b> Results of the induction experiment (SC 2% raffinose, 0.02% glucose and 0.04% galactose) performed on wild-type* and <i>GAL10-lncRNAΔ</i> * strains .....	83

# List of Abbreviations

ALCATRAS	A Long-Term Culturing And TRApping System
aTc	Anhydrotetracycline
ATP	Adenosine TriPhosphate
BSA	Bovine Serum Albumin
ChIP	Chromatin ImmunoPrecipitation
CME	Chemical Master Equation
CV	Coefficient of Variation
Cy5	Cyanine 5
DNA	Deoxyribonucleic acid
<i>F</i>	Fano factor
GFP	Green Fluorescent Protein
eGFP	enhanced Green Fluorescent Protein
GMO	Genetically Modified Organism
HDACs	Histone DeACetylase
HXT	HeXose Transporter
IPTG	IsoPropyl $\beta$ -D-1-ThioGalactopyranoside
lncRNA	long non-coding RNA
LTA	Long Term Adaptation
miRNA	micro RNA
ODE	Ordinary Differential Equation

pTC	post Transcriptional Control
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
RNAP	RNA polymerase
SC (media)	Synthetic Complete
siRNA	small interfering RNA
sRNA	small RNA
TC	Transcriptional Control
UAS	Upstream Activating Sequence
UTR	UnTranslated Region

# 1 Introduction<sup>1</sup>

Synthetic biology is a relatively new research field seeking to implement de-novo cellular tasks or rewire faulty cellular processes by engineering complex biological architectures [8, 9]. In pursuing this goal, synthetic biology's multidisciplinary approach has the potential to both test and expand the present understanding of biology by means of a thorough application of the standardization, modularity and abstraction concepts. A plethora of molecular widgets [10-16] and potential biotechnology solutions [17-19] have so far been described in scientific literature. However, significant effort is still required to achieve a level of complexity commensurate with the natural biological landscape and thereby deliver biosynthetic devices of environmental, industrial and medical relevance.

The design and implementation of synthetic gene circuits with predictable functions often remains an error-prone and time-consuming process, relying on iterative cycles of design, implementation and revision. The application of an ideal pipeline for engineering gene circuits (Figure 1.1) is hampered by the frequent recurrence to a limited set of elementary components, which are assembled based on inadequate *a priori* mathematical modelling. The resulting networks barely perform as originally expected, due to both functional alterations introduced by the genetic context and a quantitative behaviour mismatching the requirements. These hurdles could be overcome through the expansion of the toolkit available to synthetic biologists, achievable coupling rational design and directed evolution approaches [20]. In addition, the design procedure of novel gene circuits would greatly benefit from the development of standard procedures for a meticulous and context dependent characterization of biological parts and modules. In fact, the improved reliability of parameters describing the properties of and the interactions among these parts would allow the development of more reliable computational models. As a result, the functionality of a device could be accurately predicted and only robust gene circuits would merit physical implementation.

It should also be considered that, so far, the attainment of optimized performances mainly relies on a re-engineering process which resorts to point-mutations or replacement of the originally selected biological parts. This prolongs the engineering procedure of synthetic tools and represents an additional obstacle to their actual use. A potential solution to this problem could be envisaged in the

---

<sup>1</sup> Part of the content of this chapter has been provisionally accepted for publication in *Frontiers in Microbiology- Microbiology, Ecotoxicology and Bioremediation* as Bandiera, L., Furini, S. and Giordano, E. , *Phenotypic variability in synthetic biology applications: dealing with noise in microbial gene expression.*, 2016.

design of molecular devices whose physical implementation allows for fine tuning of the performed function.

Another relevant hurdle to circuits' performance can be identified in the randomness of biochemical reactions through which the host machinery processes the encoded genetic program. This stochasticity, named biological noise, has proven to be an inherent feature of living systems, wherein it ensures fast phenotypic suitability to changing environments. At the same time, being the prime cause of phenotypic variability, e.g. the differential behaviour of single cell within an isogenic population, biological noise represents a challenge in the engineering of synthetic circuits with predictable functions. It thus becomes apparent that synthetic biology, whose aim is the design of gene circuits with well-defined functional properties, would greatly benefit from a quantitative understanding of cellular noise. While optimizing artificial gene circuitry for industrial applications, synthetic biology might also contribute to the understanding of the natural mechanisms underlying phenotypic variability through the engineering of networks for the analysis, control and exploitation of biological noise.

Some of the outlined challenges fostered the doctoral project presented in this thesis. In the present chapter, after a brief summary of the functional roles of noise in unicellular organisms, we will discuss its relevance in the design of synthetic networks. In particular, we will examine scientific efforts aimed at: identifying the sources of phenotypic variability, relating noise strength to regulatory mechanisms and network topologies and formulating a theoretical framework for the quantification of gene expression stochasticity.

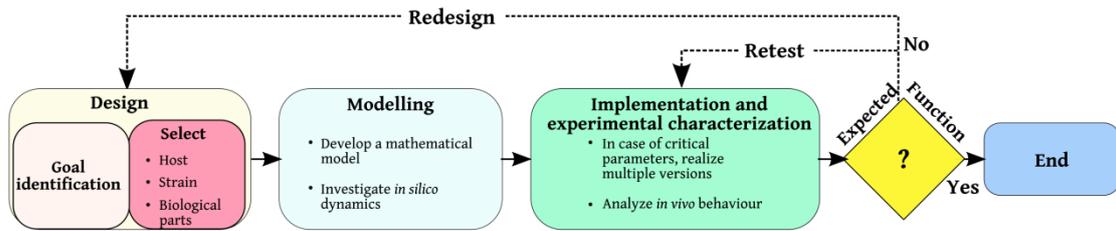


Figure 1.1: Flow chart for engineering gene circuits. The ideal engineering process (continuous connectors) starts with the design stage, during which the function that the artificial network is required to perform drives the selection of proper host, strain and biological parts. The choice of the circuit host (prokaryote/ eukaryote) and strain (genotype) usually relies on easy genetic manipulation, properties of the endogenous machinery that could enhance or impair the desired function and compatibility with operational conditions. The selection of biological components (promoters, operator sites, RBSs, genes, transcriptional terminators, fluorescent reporter proteins, etc.) should be performed based on their characterization or simplicity of synthesis as well as the risk of undesired interferences with the host. The inclusion of fluorescent reporters is advantageous as it supports the *in vivo* investigation of the function encoded in the genetic program. The design phase supports the implementation of mathematical models (deterministic/ stochastic) aimed at defining the optimal network topology and numerically investigating the circuit dynamics dependence on parameters value. Computational results subsequently help the physical implementation of the designed network, which can be performed either integrating the gene circuit into the host genome or recurring to plasmid expression vectors. In the latter case, the selection involves considerations inherent to the proper antibiotic resistance marker and origin of replication. In the case when preliminary sensitivity analysis revealed the existence of parameters critical for circuit performance, an expansion of the sample space, via implementation of multiple versions, is strongly encouraged. Other than maximizing the probability of obtaining a functional device, the experimental characterization of multiple variants provides a tool for assessing the real criticality of a genetic parameter. The coherence between the observed behaviour and the expected one proves a successful engineering procedure, which can be extended with numerical interpretations of the investigated biological phenomenon. Different, unforeseen functions require a retest of the circuit, performed in different genetic contexts or adopting modified experimental conditions, or a redesign of the implemented device (replacement of biological parts or network topologies). The actual strategies for reaching a functional gene circuit, indicated with dashed connectors in the diagram, make the engineering of network architectures a time consuming procedure. As indicated in the main text, the availability of a vast set of well characterized biological parts, supporting predictable mathematical models, would optimize the procedure.

## 1.1 Stochasticity in gene expression

In 1957 Novick and Weiner first observed the differential ability of individual cells within an isogenic population to respond to the environment when they revealed the variability in beta galactosidase synthesis in *E. coli* cells induced with lactose [21]. This noise, the stochasticity in cellular response due to the infrequent collisions among low copy number molecules subjected to Brownian motion within the cellular compartment, has been identified as an inherent feature of living systems [22].

### 1.1.1 Functional roles of biological noise

The overall single cell variability in gene expression within an isogenic population (i.e. biological noise) is generally considered to hamper the outcome of cellular processes relying on fine control of molecular fluxes [23]. However, a plethora of studies has attributed beneficial functions to noise-driven phenotypic variability. For example, the noise in gene expression introduces dynamic phenotypic heterogeneity within clonal populations, allowing species survival in time-varying environments by implementing cellular decision-making strategies. Indeed, fluctuations on the time-scale of environmental changes might divide a clonal population into phenotypic subpopulations, providing an evolutionary advantage without the burden of sensing and reacting [24]. A classic example of this logic is represented by the phage  $\lambda$  choice between lytic and lysogenic cycles [25]. The probabilistic fate commitment has been attributed to the overwhelming abundance of one of two key repressors (Cro/C1), interacting through nested positive and negative feedback loops constituting a genetic switch [26]. The final fraction of lysogens is determined by multiple factors, such as the nutritional state of the host and the multiplicity of infection [27], but the fate of any single cell results from a random, noisy process. The described genetic switch effects a bistable system in which the phenotype decision is memorized in each cell, preventing reversion of fate commitment [28].

Another case of noise-driven differentiation is observed in *B. subtilis*; a fraction of the population becomes competent after entering a stationary phase as a stress response induced by limited nutrient availability [29]. This dynamic transition is triggered when the expression level of the regulator *ComK* exceeds a threshold value, leading to the activation of downstream genes responsible for the uptake of extracellular DNA. The noisiness of the system, which maximizes the efficiency of nucleic acid import over a wide range of environmental conditions, impacts both the percentage of cells entering the competent state and its duration. The direct proportionality between the amplitude of *ComK* fluctuations and the fraction of *B. subtilis* undergoing the competent commitment has been reported as a consequence of an increase in transcription, concomitant with a decrease in

translation, of *ComK* [30, 31]. Natural variability in the duration of competence events has been related to the architecture of the molecular loop controlling the stress response. Indeed, a rewired network where the end of competence events occurs at high *ComS* concentrations, rather than at low ones as in the wild-type configuration, exhibits a reduced variability in their duration while preserving the behaviour in the deterministic limit. This evidence outlines that a low noise regime is evolutionary accessible but has not been selected, suggesting that cells have evolved mechanisms for tuning and exploiting biological noise within a defined spatial and temporal frame. In addition, it shows that noise control is often encoded in simple network topologies, where nested positive and/or negative feedback loops support the coexistence of alternative states and ensure the stochastic achievement of a functional optimum for at least a proportion of the cells, in physiological and pathological conditions [32].

### **1.1.2 Sources of phenotypic variability**

Owing to its pivotal role in biological processes, stochasticity in gene expression has been the focus of intense research. To date, experimental and theoretical studies have elucidated the prime causes of phenotypic variability and their impact on cell fitness [33-35].

Biological noise is usually described as the sum of two orthogonal components: intrinsic and extrinsic stochasticity. Intrinsic stochasticity arises from the random occurrence of biochemical events inherent to gene expression processes (e.g. the burst-like synthesis of mRNA and protein molecules) within the system of interest. Extrinsic fluctuations result from the system interacting with the intra- and extracellular environments (e.g. the concentration of available polymerases, ribosomes, metabolites and the micro-environmental conditions).

The empirical distinction of intrinsic and extrinsic stochasticity dates back to 2002, when Elowitz et al. [1] used fluorescence microscopy to analyse the expression of two distinct but identically regulated fluorescent reporter genes integrated into the *E. coli* chromosome (Figure 1.2). Intrinsic noise was quantified as the degree of uncorrelated fluctuations among the two fluorescent reporters over time. Beyond the technological advance represented by the definition of the dual reporter gene assay, this study evidenced that noise magnitude scales with increasing promoter strength and that the relative contribution of intrinsic and extrinsic components to the overall stochasticity varies with the expression regime. Indeed, when the fluorescent reporters' expression was transcriptionally downregulated, intrinsic noise monotonically decreased upon induction [36, 37], while extrinsic fluctuations reached a maximum at intermediate transcription rates. As a result, in the low expression regime intrinsic noise proved to be a prominent component of global stochasticity, otherwise dominated by extrinsic fluctuations. Finally, the study showed that

transcriptional regulation of fluorescent reporters by a plasmid-encoded protein repressor amplified stochasticity, compared to the genomic integrated equivalent, as a consequence of increased copy number variability [38].

Although extrinsic fluctuations, denoted by a timescale comparable to the cell cycle duration [39], appear to often be the dominant component of gene expression stochasticity [1, 39, 40], we lack a precise characterization of their significant contributors and scientific studies have mainly investigated intrinsic noise.

In 2002, Ozbudak et al. [6] used a genetically modified organism (GMO) to experimentally investigate the impact of genetic parameters, such as transcription and translation rates, on phenotypic variability. To facilitate extension of their results to native genes, present in one or two copies depending on the cell cycle stage, the authors integrated a gene encoding the green fluorescent protein (GFP) into the *B. subtilis* chromosome. The transcription rate of the fluorescent reporter was tuned by means of an inducible promoter or by mutating the promoter sequence, while translational regulation was achieved by inserting point mutations in either the ribosome binding site (RBS) or the first codon of GFP. Their experimental and numerical results highlighted that the fluorescence distributions were over-dispersed of a term, named burst size, representative of the average number of proteins translated from a single mRNA. Hence, the study empirically validated theoretical predictions which identify translational efficiency as a major determinant of prokaryotic gene expression noise (Figure 1.2). In identifying the dependency of expression noise on translation rate, the authors provided an explanation as to why essential genes (e.g. *malT*<sup>2</sup>, *tetR*<sup>3</sup> and *cya*<sup>4</sup>) [41-43], whose expression level requires tight control, are translated from low-efficiency RBSs. This noise-control strategy has been subsequently validated in yeast for both essential genes and gene encoding subunits of protein complexes [44]. It is worth considering that the weak positive correlation between noise strength and transcriptional efficiency shown in [6], originally perceived as a marked difference between prokaryotic and eukaryotic gene expression [45], has been ascribed to the index adopted for quantifying stochasticity. Indeed, theoretical studies identified in burst size greater than two, as it usually is in *E. coli*, a condition for transcription dominating intrinsic noise. Under this perspective, a reduced translational efficiency constrains biological noise via buffering mRNA fluctuations [36].

The seminal work by Ozbudak et al. outlined that the two-step process of gene expression endows cells with the ability to independently control mean expression level and stochasticity. Indeed, a given protein concentration can be attained coupling either low transcription rates with high

---

<sup>2</sup> Gene encoding the positive regulator of *E. coli* maltose regulon.

<sup>3</sup> Repressor of genes conferring tetracycline resistance.

<sup>4</sup> Gene involved in the synthesis of cAMP.

translation ones or transcribing high levels of poorly translated mRNA. The latter, consistent with a steady state reduction in protein fluctuations, incurs a higher metabolic cost: hydrolysing ATP molecules for the transcription of scarcely used mRNA. This consideration supports the idea of an evolutionary shaped trade-off between accurate and energetically advantageous control of gene expression [46].

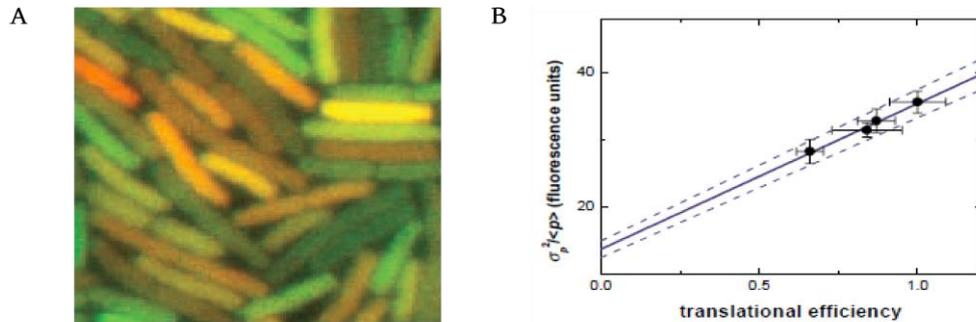


Figure 1.2: Seminal papers for the experimental quantification of gene expression noise. Panel A shows a fluorescence image from Elowitz et al. [1], obtained merging CFP (reported in green) and YFP (reported in red) channels. Cells expressing equal amounts of the two fluorescent reporters appear yellow, while the presence of green and red cells reveals high intrinsic noise. Panel B shows the strong positive correlation between noise strength and translational efficiency measured by Ozbudak et al. [6] with alternative translational mutants. Both images are reproduced with permission.

### 1.1.3 Quantifying phenotypic variability

The variability in the expression of a gene within an isogenic population can be assessed quantifying the amplitude of single cell protein fluctuations compared to the population mean concentration. Experimental investigations of gene expression stochasticity were hence fostered by both the engineering of fluorescent proteins variants and the progress in single cell measurement methodologies. Indeed, fluorescent proteins, denoted by various excitation and emission spectra, allow for the simultaneous *in vivo* monitoring of multiple genes expression levels. More specifically, fluorescent proteins act as readouts of protein levels when adopted in translational fusions or indirectly report on promoter kinetic if polycistronically transcribed with the gene of interest. So far, the acquisition of the emergent single cell fluorescent signal has been performed using flow cytometry or fluorescence microscopy. These observational methodologies provide experimental fluorescence distributions whose first and second order moments, respectively mean and variance of gene expression levels, support the definition of indexes quantifying stochasticity.

The probabilistic interpretation of gene expression process identifies the protein as a random variable ( $X$ ). The variance of  $X$ ,  $\sigma^2$ , is the dispersion index of its probability distribution and represents the average squared distance of  $X$  from its expected value  $E(X)$ . Being measured in the same units as  $X$ , the standard deviation  $\sigma$ , e.g. the variance square root, provides a more intuitive quantification of randomness. However, the standard deviation scaling with the data makes the interpretation of stochasticity dependent on the scale on which  $X$  is measured, thereby limiting the comparison of stochasticity measurements performed with alternative instruments. To circumvent this problem, the adoption of a dimensionless index of stochasticity would be beneficial in experimental acquisitions.

The *coefficient of variation* (CV) of a random variable  $X$  satisfies this requirement and is defined by

$$CV(X) = \frac{\sigma}{|\mu|}. \quad 1.1$$

It is adopted for non-negative random variables with positive expectation and it is scale-invariant. Albeit being denoted by the same properties, its inverse, named signal-to-noise ratio, is less frequently adopted in studies pertaining stochasticity because it counterintuitively associates small values to high noise levels.

The *Fano factor*<sup>5</sup> ( $F$ ), characterized by the same scaling properties affecting the standard deviation, is defined by

$$F(X) = \frac{\text{Var}(X)}{|E(X)|} = \frac{\sigma^2}{|\mu|}. \quad 1.2$$

This index, adopted for discrete random variables, provides a measure of stochasticity relative to a Poisson distribution with the same expectation. Indeed, the Poisson distribution is the reference for discrete variables and has equal mean and variance. In this case, variables with *Fano factor* higher (lower) than one are described as over- (under-) dispersed relative to the Poisson.

A review of scientific literature highlights that stochasticity in gene expression is generally quantified using the *coefficient of variation*, the *squared coefficient of variation* or the *Fano factor*. The first two indexes are mainly adopted in experimental studies while the *Fano factor*, used to represent noise strength, proves useful in theoretical ones to reveal trends which might otherwise be obscured by noise scaling due to finite number effect. It is worth noting that, when applied to the same dataset, an increase in *Fano factor* does not imply an increase in *coefficient of variation*. In fact, saying

---

<sup>5</sup> Also known as *dispersion index*, *coefficient of dispersion*, *index of dispersion* or *variance-to-mean ratio*.

that two genes have high and low noise strengths only implies that the first will be characterized by a higher variability, when they are expressed at similar levels [47].

## 1.2 Biological noise control: copy number and gene regulatory networks

The integration of gene(s) encoding fluorescent reporter(s) into the host genome was selected in the studies described in section 1.1.2. This experimental design simplifies results extension to native genes, which usually occur in few copies. However, synthetic gene circuits are often harboured on plasmids: circular DNA molecules first identified with the discovery of bacterial conjugation. Plasmids, retained by cells thanks to the selective pressure exerted administering antibiotic to which they encode resistance for, inhabit cellular compartment with a copy number ranging from units to hundreds, as determined by the properties of their origin of replication. The use of plasmids mirrors the applicative perspective of synthetic biology, where consistent protein yields achievable with high copy number plasmids might be desirable. In addition, cloning vectors are supposed to be orthogonal to the endogenous machinery and therefore facilitate the mathematical formalization of gene network behaviour.

Although plasmid-encoded fluorescent reporters hinder the deciphering of chromosomal gene expression randomness, as transcription and translation are averaged over multiple gene copies, an accurate control of gene copy number would allow a reduction of biological noise thereby providing a tool for network robustness optimization. Indeed, noise magnitude scaling with the inverse of the square root of gene copy number justifies both the spread of polyploidy [40] and evolutionary gene redundancy, a strategy through which the robustness of naturally occurring networks is enhanced [48, 49]. While high copy number plasmids proved suitable for noise reduction [50], the considerable metabolic burden imposed on transformants could impair cell growth and lead to aberrant network behaviour [38]. Moreover, variation in randomly fluctuating plasmid counts during cell growth and division is predicted to act as an additional source of extrinsic stochasticity [37, 51]. To the best of our knowledge, increasing copy number of plasmid-encoded synthetic circuits as a potential control knob for noise reduction, has been experimentally considered only in Guido et al. [50]. In that paper, however, the characterization of circuit behaviour, when cloned in a low copy number plasmid, was directed to test the mathematical model predictive capabilities. For this reason, the authors did not highlight the impact of plasmids counts on noise features.

Another control variable of network dynamics and robustness is its architecture. Regulatory mechanisms that cells evolved to tune gene expression in response to physiological and environmental variations [52, 53] are implemented through complex networks. Their properties have been theoretically and experimentally investigated, permitting the characterization of elementary synthetic circuits such as feedback loops and transcriptional cascades.

Positive feedback loops, in which a protein upregulates its own synthesis, have been associated with increased phenotypic variability. A hallmark of this positive autoregulation, generally implemented through protein-mediated transcriptional activation, is bistability. The bimodal distribution of protein levels reflects the coexistence of high- and low-expression states, between which single cells can stochastically switch. Analysing the positively regulated expression of a GFP reporter in *S. cerevisiae*, Becskei et al. [54] attributed the unobserved ON/OFF switch to a hysteretic component, which could reduce GFP fluctuations by 'remembering' past states. Later, an inverse proportionality between positive feedback strength and switching frequency was theoretically proven by numerical simulations [55].

Randomness analysis in the expression of genes regulated through negative feedback loops has attracted particular interest as approximately 40% of *E. coli* transcription factors undergo negative autoregulation [56]. In addition, relevant fluctuations in transcription factors concentration have been related to developmental disorders [57].

The general idea that negative feedback loops enhance system robustness while reducing gene expression noise [46, 58-60] has been empirically demonstrated by Becskei and Serrano [61]. Comparing the stochastic expression of the tetracycline repressor co-transcribed with enhanced GFP (TetR-eGFP) from a TetR repressible promoter with an equivalent unregulated system, the authors measured reduced fluctuations, retained within physiologically meaningful parameters range, for the negatively autoregulated repressor.

Specifically, the less noisy behaviour was observed at maximal feedback strength, while administering the chemical inducer anhydrotetracycline (aTc) resulted in weaker feedback and noisier expression. In contrast, Dublanche et al. [62] observed optimal noise suppression at intermediate feedback strengths. Their results agreed with theoretical analyses indicating that negative feedback has the ability to reshape the noise spectrum through a shift from low to high frequency components. The latter can easily be suppressed by downstream molecular cascades acting as low pass filters. In particular the extent of the shift, a function of the feedback strength, was maximal at intermediate strengths [63]. The most prevalent form of negative feedback in natural networks is protein-mediated transcriptional downregulation [64-67]. Alternative negative-feedback topologies can be implemented through transcriptionally-/translationally-regulated expression of a gene mediated by mRNA [68, 69]. In fact, mRNA-operated translational gene downregulation is indicated as the best noise suppression strategy by mathematically controlled comparison of efficiency in alternative regulatory mechanisms of noise minimization. It is worth noting that the disruption of this type of negative feedback, e.g. intron-encoded micro-RNA (miRNA) regulation of its coexpressed target gene, has been associated with pathological states and improper stress-related responses [70, 71]. Although mRNA-based feedback proved optimal for minimizing noise under the

constraint of fixed feedback strengths, it is important to consider that when the protein products translated from the target mRNA regulate the strength of the feedback via their multimerization, this introduces a cooperative regulation which might render transcription/translation ultrasensitive to protein levels.

The effect of the length of a transcriptional cascade on noise propagation has been investigated by Hooshangi et al. [72], who compared the magnitude of fluctuations in networks with up to three stages. The authors observed higher stochasticity at intermediate inducer concentrations, revealed by bimodal fluorescent distributions. Furthermore, the addition of a transcriptional layer approximately doubled gene-expression noise, resulting in the noisiest output at maximal cascade length. The increasing number of stages improved the hypersensitivity of the network at intermediate induction, leading to a more precise steady-state switch between low and high expression levels, but it also extended the time required for network activation. This caused decreased synchronization within the population, as transient intercellular variability in the activation times increased. Analogous results were obtained by Blake et al. [45] and Pedraza et al. [73]. Remarkably, while theoretical studies have shown that elongating a transcriptional cascade leads to low-pass filter activity, preventing network activation from short, noisy inputs [74], long transcriptional cascades rarely occur in short living organisms such as bacteria and lower eukaryotes [75].

### **1.2.1 Translational regulation of gene expression**

Owing to the awareness that protein mediated transcriptional repression constitutes the prevalent form of gene regulation in natural occurring systems, as exemplified by the extensively studied lactose operon in *E. coli*, scientific studies have primarily focused on transcriptional negative feedbacks as a strategy for noise suppression. However, protein mediated negative autoregulation, where the protein binding to operator site located within its promoter region prevents RNA polymerase (RNAP) binding or progressing through the gene sequence, indirectly controls fluctuations in mRNA counts, thought to be the major source of intrinsic noise. Reasoning that a direct control would better buffer mRNA fluctuations, Swain theoretically compared stochasticity arising from transcriptional and translational negative autoregulation [76]. His results identified in the latter, independently of the way through which the protein downregulates translation of its coding mRNA, an optimal strategy for stochasticity control. Furthermore, it provided plausible justification as to why the synthesis of bacterial global regulators, whose stochasticity would challenge cell fitness, is controlled by means of this mechanism [77-79]. This study anticipated a number of others, published in the following years, relating post-transcriptional regulation to gene

expression randomness. In particular, Levine et al. [80], while exploring the features of small non-coding RNAs (sRNAs) regulation of gene expression, predicted lower stochasticity for downregulation based on sRNAs compared to protein-mediated transcriptional repression. In another theoretical study, translational control by an upstream regulator caused higher stochasticity compared to transcriptional one [81]. Remarkably, the authors' results were based on the assumption of invariant transcription and translation rates, set via different mechanisms, rather than equal mean expression level of the target gene as hypothesized in previous works which drew contradictory conclusions. This difference highlighted the relevance of comparing the stochastic effects arisen by alternative regulatory mechanisms on the same expression regime.

The use of mRNA mediated translational regulation to control gene expression stochasticity meets synthetic biologists' interest in the expansions of the available toolkit through engineering of novel RNA-based biological parts [82]. Both the chemical nature of this nucleic acid and the functions performed by RNA molecules in naturally occurring contexts constitute reasons for interest. Indeed, the rational design of RNA-based biological parts is fostered by reliable predictions of their secondary structures coming from base-pair ruled intramolecular interactions. Furthermore, directed evolution strategies need to sample a space of reduced extension compared to the one required for a protein of similar length, as each position can be occupied by one out of four nucleotides (instead of one amino acid out of twenty) [20].

It is interesting to note that, due to these properties, the engineering of RNA molecules capable of enzymatic activity or metabolite-driven regulation of gene expression preceded the discovery of their natural counterpart. Riboswitches and riboregulators are classes of functional RNAs implicated in gene expression regulation [83]. The former are usually found within the 5' UTR of genes, whose transcription or translation is modulated due to conformational changes following ligand binding. The fast and flexible response of natural riboswitches prompted the engineering of a theophylline responsive variant by the Smolke group [84]. Riboregulators, envisioned as the prokaryotic analogue of miRNA and siRNA, are small regulatory RNAs which inhibit translation through ribosome docking. Their mode of action involves the activity of the Hfq cofactor. This protein mediates the interaction between riboregulator and target mRNA and protects the regulatory RNA from ribonucleases degradation. Following hybridization, both the riboregulator and target mRNA are degraded. Riboregulator's features inspired the engineering of a post-transcriptional controller of gene expression by Isaacs et al. [85, 86].

While the aforementioned studies evidenced the potential of using modular RNA-based biological parts to regulate the mean expression level of genes, none of these tools has been characterized at the single cell level. The results of such experiments would allow theoretical exploration of post-transcriptional regulation as a suitable strategy for engineering stochasticity control.

## 1.3 Mathematical models in synthetic biology

The engineering of biological processes pursued in synthetic biology largely founded on the development of mathematical models which should ideally support both the design procedure and the interpretation of experimental results. As previously mentioned, the lack of a large and quantitatively characterized toolkit available to synthetic biologists has so far hampered the extensive use of *in silico* predictions of networks behaviour, constraining the adoption of computational models to an *a posteriori* and *ad-hoc* description of the function implemented by gene circuits. Irrespective of their actual use and the abstraction level they are based on, theoretical models in synthetic biology, likewise in all other engineering fields, are expected to stem from a trade-off between simplicity and ability to capture the dynamic of the investigated phenomenon. Before dealing with the properties of deterministic and stochastic kinetics, theoretically outlined in this paragraph and applied to the investigated gene circuits in the following chapters of this thesis, we would like to emphasize two assumptions recurrently adopted in modelling biochemical networks. First, the mathematical formalization of biological processes taking place in a cell conceives it as a biochemical reactor. Second, the multitude of occurring chemical reactions is conceptually subdivided into independent modules, whose dynamics can be studied in isolation.

Deterministic models provide a macroscopic description of the system dynamics and are structured in a set of ordinary differential equations (ODEs). The ODEs formulation depicts the time-evolution of the concentration of chemical species as the sum of the chemical reactions contributions assessed by applying the law of mass action. As a result, the system dynamics, solution of the ODEs system, is completely specified by the initial concentration of each species and the reactions rate constants. Deterministic models, neglecting the stochasticity inherent to biological processes, usually provide a correct description of the gene circuits' population level performance. Extensively, and sometimes arbitrarily, adopted in the early infancy of modelling efforts in synthetic biology, the deterministic formulation constitutes the actual reference in the design procedure of synthetic networks thought to be unaffected by biological noise. As an example, ODEs based models are currently used for synthetic devices aiming at the spatial or temporal coordination of cell populations dynamics [87, 88]. A key feature of deterministic models is their simplicity, which nurtures sensitivity and bifurcation analysis aimed at evaluating how the gene circuit behaviour could be affected by changes in genetic parameters.

Considering that the law of mass action validity is restricted to chemical species counts of the order of the Avogadro's number, the deterministic approach fails to capture the stochasticity of biochemical reactions whose reagents are present in traces. In such cases the transition to a probabilistic description of the time-evolution of the system, subsumed in stochastic models

theoretically founded on the chemical master equation (CME), is required. Stochastic models, essential when studying gene expression noise, provide a tool to quantify the impact of fluctuations in chemical species counts on gene circuits' robustness, thereby enabling the identification of strategies for noise minimization and exploitation.

Stochastic chemical kinetics consider a well-stirred and thermally equilibrated system, in which  $N$  chemical species  $\{S_1, S_2, \dots, S_N\}$  react through  $M$  reaction channels  $\{R_1, R_2, \dots, R_M\}$ . The state of the reactor is described by the vector  $\mathbf{X}(t) = \mathbf{x}$ , whose entries are the molecules counts of each species at time  $t$ . The objective is to estimate the distribution  $P(\mathbf{x}, t)$ , given the system in state  $\mathbf{X}(t_0) = \mathbf{x}_0$  at the initial time. Each reaction channel  $R_j$  is mathematically described by the state-change vector, whose elements are the integer variations in molecules counts associated with the occurrence of reaction  $j$ , and its propensity function  $a_j(\mathbf{x})$ , defined as the probability per unit time of  $R_j$  occurring within the system. Given the propensity function dependence on the actual state, the evolution in time of the reactor state can be interpreted as a time continuous Markov process with a discrete sample space. The time-evolution of the probability of being in state  $\mathbf{x}$  at time  $t$  is provided by the CME:

$$\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^M [a_j(\mathbf{x} - \mathbf{v}_j)P(\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0) - a_j(\mathbf{x})P(\mathbf{x}, t | \mathbf{x}_0, t_0)] . \quad 1.3$$

As the CME can be solved only in special cases, numerous Monte Carlo approaches have been developed to simulate exact numerical realizations of  $\mathbf{X}(t)$ . To this aim it is necessary to consider the probability function  $p(\tau, j | \mathbf{x}, t)$ , representing the probability that, given the system in state  $\mathbf{x}$  at time  $t$ , the next reaction will occur in the time interval of duration  $\tau + d\tau$  and will be reaction  $j$ .  $p(\tau, j | \mathbf{x}, t)$  is the joint probability density function of the random variables  $\tau$  (time to next reaction) and  $j$  (index of the next reaction channel) and is given by:

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) \exp(-a_0(\mathbf{x}) \tau) = a_j(\mathbf{x}) \exp \left[ \left( - \sum_{j'=1}^M a_{j'}(\mathbf{x}) \right) \tau \right] . \quad 1.4$$

Hence  $\tau$  is an exponential variable with mean  $1/a_0(\mathbf{x})$  and  $j$  an independent random variable with probability  $a_j(\mathbf{x})/a_0(\mathbf{x})$ .

Though numerous Monte Carlo procedures have been developed for the extraction of the parameters  $\tau$  and  $j$  according to their relative distributions, we will consider only the Gillespie algorithm<sup>6</sup> [89]. Indeed, beyond being the basis on which all the subsequent methods were developed, this procedure was selected for its simplicity in the stochastic simulations run in our projects. According to the Gillespie's direct method the time to next reaction and the index of the next reaction channel are

---

<sup>6</sup> Also known as Gillespie's direct method.

$$\tau = \frac{1}{a_0(\mathbf{x})} \ln \frac{1}{r_1} ,$$

$$j \text{ the smallest index for which } \sum_{j'=1}^M a_{j'}(\mathbf{x}) > r_2 a_0(\mathbf{x}) .$$

1.5

In equations 1.5,  $r_1$  and  $r_2$  are random numbers drawn from the uniform distribution in the unit interval. The stochastic simulation algorithm for generating exact trajectories  $X(t)$  can be summarized as:

1. Initialize at the initial time  $t=t_0$  the rate constants and the system state.
2. Based on the current state  $\mathbf{x}$  of the system, compute the propensity function  $a_j(\mathbf{x})$  for each reaction channel and their combination  $a_0(\mathbf{x})$ .
3. Identify the time to next reaction,  $\tau$ , and the index of the next reaction channel,  $j$ , using equations 1.5.
4. Update time and the system's state according to  $t=t + \tau$  and  $\mathbf{x} = \mathbf{x} + \mathbf{v}_j$ .
5. Record  $\mathbf{x}$  and  $t$  and return to step 2 or end the simulation.

The Gillespie algorithm is advantageous in that it is both easy to implement and it generates correct trajectories of the stochastic process  $X(t)$ , even when the underlying CME proves analytically intractable. However, it becomes computationally too expensive and slow whenever the species populations are large due to the evaluation, at each step, of the inverse of the combined propensity function,  $1/a_0(\mathbf{x})$ . While these drawbacks justified the development of several approximated algorithms [90-92], the conditions under which they can be applied without incurring in relevant errors remain unclear. It is worth noting that the expectation of the system trajectories generated by means of this Monte Carlo procedure, conceivable as the time-evolution of single cell, converges to the solution of the deterministic ODEs formulation when all the propensity functions are linear in the chemical species.

## Summary

The literature review undertaken in this chapter evidences that stochasticity, pervading biochemical processes at the cell scale, constitutes the key to deepen a quantitative understanding of biology. In particular, both the counterintuitive, beneficial roles of gene expression stochasticity and its ability to shape the dynamics of biological systems have so far nurtured scientific efforts aimed at exploring the role of biological noise as a signal in engineered gene networks. In the emerging framework, the assessment of stochastic effects might constitute a specific design goal of next generation synthetic circuits, driving the selection of biological parts and network topologies enabling the implementation of robust and optimized devices.

A significant body of work has investigated how phenotypic variability relates to alternative network topologies. The theoretical picture that has arisen identifies in translational feedback loop the best candidate for biological noise suppression. Despite being supported by observations in naturally occurring systems, it seems surprising that stochasticity control via mRNA-based translational downregulation has attracted little attention from the synthetic biology community. Indeed, RNA-based biological parts are predicted to play a central role in the future toolkit available to synthetic biologists. Furthermore, while the potential of mRNA driven gene regulation in engineered circuits has been witnessed in several papers, the function implemented by these synthetic tools has been characterized only at the population level.

Inspired by these observations, in the remainder of this thesis we will compare, numerically and experimentally, the variability in the expression of a gene subject to alternative regulatory mechanisms encoded in simple and similar network architectures. In particular, we will focus on genetic cores in which the transcriptional or mRNA-based translational regulation in the expression of a fluorescent reporter can be tuned by means of an exogenous signal. Cloning these gene circuits in plasmids with various origin of replication made the comparison of the regulatory mechanisms over a wider range of mean expression levels and the evaluation of copy number effect amenable.



## 2 Development of a synthetic noise tester catalogue

The ubiquitous stochasticity of gene expression processes is traditionally envisioned as a hurdle for engineering synthetic devices with predictable functions. While the thorough characterization of biological parts and genetic programs noise features would encourage more precise, *a priori in silico* predictions of the behaviour of novel molecular widgets, the development of synthetic circuits for biological noise tuning would provide a tool for testing the robustness of available devices. The applicability of such noise generators/testers widely exceeds the synthetic biology field. Indeed, investigating the effect of stochastic fluctuations in natural gene networks' key regulators on biological processes would likely provide insights on evolutionary selected strategies for gene expression noise control. Such investigations, coupled with the ability to dissect beneficial or detrimental effects related to the noise level, would facilitate the optimization of procedures for correcting faulty cellular processes, thereby providing advanced medical solutions. As an example, a noise tester might be used to elucidate and exploit the impact of stochasticity on stem cell differentiation, leading to more efficient protocols for cellular reprogramming.

One of the basic requirements for a valuable noise tester is the possibility of independently tuning the mean expression level and noise profile of a target gene, over a wide regime, by means of external, hence controllable, signals. Considering that the dual-step nature of gene expression, structured in transcription and translation, endowed cells with the ability to achieve a direct control of both the average concentration and the dispersion of gene products, the simplest design of a noise generator could evoke the simultaneous transcriptional and translational regulation of a target gene. This possibility has been theoretically proven in [93], where the authors analysed the effects of transcriptional and translational regulation on gene expression stochasticity. While the use of regulated promoters ensures a simple and systematic transcriptional control via the administration of inducers or repressors, the external regulation of translation could be implemented resorting to sRNA, such as riboswitches and riboregulators. Several synthetic devices emulating natural riboregulators proved useful for implementing an external post-transcriptional control in the expression of a target gene. Among these, Isaacs et al. [86] proposed a post-transcriptional controller in which the direct interaction between the coding mRNA and the synthetic analogue of a *trans*-acting sRNA implemented an inducible OFF-to-ON switch in the expression of a reporter gene. Later, a conceptually equivalent post-transcriptional controller was developed in the ICM Lab for exploring

the feasibility of a pure bottom-up approach in the modular design of synthetic circuits [5]. In this system, the hybridization of a *cis*-acting non-coding sequence (CIS), constitutively transcribed with the target gene, with a *trans*-acting oligoribonucleotide (TRANS) mediated an ON-to-OFF switch in gene expression. As shown in Figure 2.1, the CIS fragment comprises a non-coding sequence and a RBS and is complementary to the TRANS element. Upon transcription, the molecular annealing of the two RNAs causes a partial occlusion of the RBS, preventing ribosomal access. The resulting translational inhibition, applicable to any gene cloned downstream of the CIS element, can be tuned by modulating the ratio of the regulatory RNA sequences. Specifically, the ratio was varied acting on the cytoplasmic abundances of TRANS sequence, whose transcription occurs from a lactose repressible promoter. The gene regulation function implemented by the synthetic device was assessed by exposing *E. coli* CSH126 cells, transformed with two plasmids harbouring the described genetic program and an additional module for the controlled synthesis of the LacI repressor respectively, to various concentrations of the gratuitous inducer Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG).

We envisioned in this post-transcriptional controller a toy model for the translational regulation required, in association with a transcriptional control, to implement a catalogue of gene circuits acting as a noise tester.

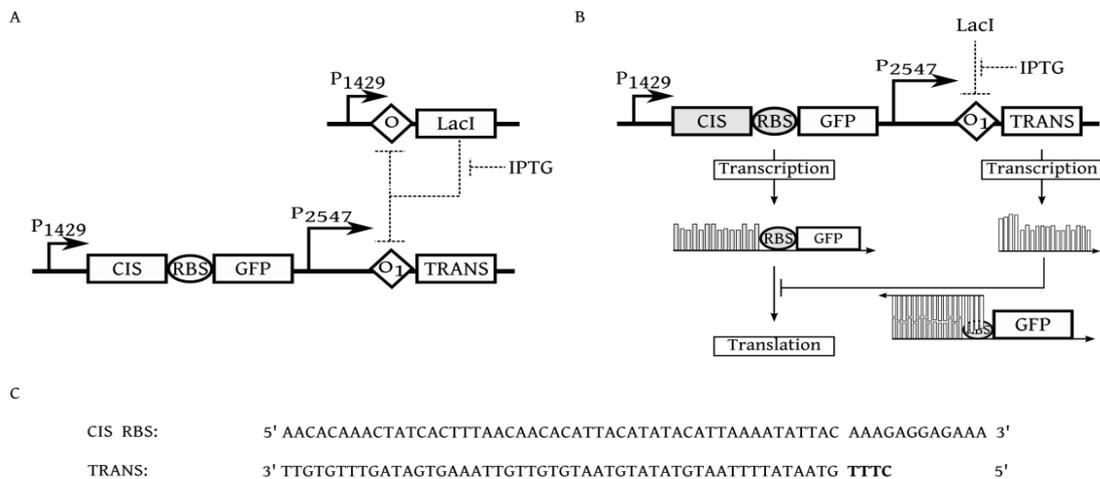


Figure 2.1: Post-transcriptional controller characterized in [5]. In panel A the gene circuits for the regulated expression of LacI repressor and the post-transcriptional controller originally developed in the ICM Lab is shown. Induction with IPTG releases the repression exerted by LacI on the operator sites O and O<sub>1</sub>, causing an increased transcription of the TRANS acting sequence. Panel B shows a functional scheme of the post-transcriptional controller: hybridization of the CIS-GFP and TRANS RNAs leads to a partial occlusion at the RBS, inhibiting ribosome docking. In panel C details of the CIS and TRANS nucleotide sequences are reported. CIS and TRANS are 50 base-pair long complementary sequences designed to prevent partial or improper annealing, acquisition of secondary structures impeding their interaction and limit crosstalk with endogenous RNAs. The CIS element ends with a RBS. The four nucleotides at the 5' end of the TRANS sequence, responsible for the partial RBS occlusion, are shown in bold.

In this chapter we will describe the development of a catalogue of synthetic gene circuits aimed at investigating the experimental feasibility of an independent control of the first and second order moments of the steady-state distribution for the expression levels of a fluorescent reporter gene. We will then present an initial population level characterization of elements isolated from this collection, implementing either a transcriptional (TC gene circuit) or a post-transcriptional (pTC gene circuit) control in the expression of the fluorescent reporter gene. As well as verifying the correct behaviour of the analysed circuits, the experimental data are used for the definition of deterministic models, providing a macroscopic description of the gene circuits' function. The parameters identified in the deterministic models are used in stochastic simulations aimed at assessing the differential effect of the cloning vector's copy number on stochasticity in gene expression.

## 2.1 Methods

### 2.1.1 Plasmid construction and strain

The synthetic gene circuits were assembled using biological parts which adhere to the BioBrick standard format [94]. Biological parts were taken from the Registry of Standard Biological parts or synthesized to achieve this requirement (CIS and TRANS acting sequences, Invitrogen). The genetic program of the circuits' catalogue elements comprises two transcriptional units, both ending with the same transcriptional terminator (BBa\_B0015). In the former, a Tet repressor protein (TetR) downregulates the cocistronic transcription of the CIS non-coding sequence, including the strong RBS based on Elowitz repressilator (BBa\_B0034) [10], and a green fluorescent reporter gene (GFP) labelled with an LVA degradation tag (BBa\_J04631). Six variants of the regulated promoter were assembled cloning one of three TetR operator sites - TetO (BBa\_K079036), TetO-4C (BBa\_K079037) or TetO-wt/4C5G (BBa\_K079038) - denoted by decreasing binding affinity for the repressor protein downstream of a synthetic, constitutive promoter,  $P_{2547}$  (BBa\_J23100) or  $P_{1429}$  (BBa\_J23118). In the second transcriptional unit a LacI repressible promoter controls the synthesis of the TRANS oligoribonucleotide. Two promoter variants were assembled placing one of the two natural LacI operator sites ( $O_1$  or  $O_2$ , the latter having a one order magnitude weaker affinity for the repressor protein [95]) in the  $P_{2547}$  proximal region. Control circuits, composed only of the first transcriptional unit, or the latter, modified through insertion of the GFP encoding gene in order to obtain a fluorescent readout proportional to the TRANS sequence abundance, were included in the catalogue. The genetic circuits were alternatively cloned at the multiple cloning site of a high copy number plasmid (pSB1A2) containing a pUC19-derived pMB1 origin of replication and selection marker for ampicillin resistance. The synthetic gene circuits, whose experimental characterization is described in this chapter, were additionally cloned in the pSB4A5 low copy number plasmid, containing pSC101 origin of replication and ampicillin antibiotic selection marker. TOP10F' competent *E. coli* cells (Invitrogen) were selected as host strain, having a genomic overexpression site for both repressor proteins (TetR and LacI) controlling the circuits' function. Heat shock transformation, performed according to manufacturer's guidelines, was used for plasmid insertion in the host strain.

### 2.1.2 Fluorescence assay

Single colonies of *E. coli* TOP10F' strain transformed with the desired plasmid were inoculated, from a freshly streaked LB-agar plate, in 5ml of selective (100  $\mu$ g/ml ampicillin) M9 minimal medium, completed with casamino acids, thiamine hydrochloride and glucose as carbon source (Sigma). After

an overnight growth (37°C, 220 rpm orbital shaking) in Erlenmeyer flasks, cell cultures were spun-down (10 min, 3,500 rpm) and resuspended in fresh pre-warmed medium to enable metabolites removal. Upon dilution to an initial optical density  $OD_{600} = 0.05$ , a volume equal to 200  $\mu\text{l}$  of each culture sample was transferred into a 96-well microplate along with 5  $\mu\text{l}$  of IPTG at the proper concentration (10, 25, 50, 100, 200  $\mu\text{M}$ ). The volume of samples not requiring induction was balanced with 5  $\mu\text{l}$  of M9 minimal medium. After being covered with a breath easy sealing membrane (Sigma), the microplate was loaded in a Infinite M200 microplate reader (Tecan). A time-course experiment aimed at following bacterial growth and fluorescence was then performed. The experiment, set through the i-control™ software (Tecan), consisted of 75 kinetic cycles, each of them including: 180 s linear shaking (3 mm amplitude), 10 s wait, optical density measurement (600 nm), bottom reading fluorescence measurement (excitation: 501 nm, emission: 535 nm, gain: 60), plate movement out, 180 s wait and plate movement in. For each genetic circuit, nine transformants' colonies, along with sterile medium and wild-type host strain for absorbance and fluorescence background correction, were assayed with this protocol in technical duplicate.

### 2.1.3 Data analysis

Raw absorbance and fluorescence time-series acquired through the microplate reader were analysed using custom code written in Python. Optical density measurements were corrected by subtracting the absorbance of sterile M9 medium, in order to infer the real bacterial growth over time. A linear regression of the log-transformed  $OD_{600}$  time-series was then performed to identify the temporal extremes of the log-phase in the growth curve. Similarly, background fluorescence correction was implemented through subtraction of the autofluorescence acquired on wild-type cell cultures. The steady-state fluorescence data reported are normalized by the  $OD_{600}$  value at which they were extracted, in order to provide a signal proportional to the mean cell fluorescence. Specifically, an  $OD_{600} = 0.3$ , reached approximately 3 hours after induction with IPTG and falling in the middle of the exponential growth phase (where the linear correlation between fluorescence and  $OD_{600}$  justifies the performed normalization), was adopted.

### 2.1.4 Numerical simulations

**Deterministic models.** All the parameters of deterministic models describing the function performed by the TC gene circuit were retrieved from [5] except for the Hill function describing the variation in TRANS-GFP transcription rate upon IPTG induction (equation 2.5). The Hill coefficient as well as the dissociation constant was fitted on the normalized fluorescence data acquired on the TC

gene circuit, cloned in pSB1A2 plasmid. It is worth noting that, due to the observed saturation phenomenon, the maximum transcriptional rate  $k_{r,TG}^{MAX}$  is lower than the value theoretically expected based on the ratio of the  $P_{2547}/P_{1429}$  promoters' transcriptional strengths. Assuming the presence of 80 plasmids for the pSB1A2 cloning vector, the ratio between the normalized fluorescence values acquired on the TC gene circuit, cloned in both plasmids, at IPTG = 200  $\mu$ M was used to fit the copy number for the pSB4A5. The Hill function defined on the TC gene circuit's dose-response curve was used to model the increase in TRANS sequence transcriptional rate, upon IPTG induction, in the synthetic circuit implementing post-transcriptional control in GFP expression. The effect of TRANS sequences on translation was modeled by an increase in the coding mRNA degradation rate induced by hybridization with the *trans*-acting oligoribonucleotide [96, 97]. The stoichiometry-dependent degradation rate describing the strength of interaction between the CIS-GFP mRNA and the TRANS sequence was determined by minimizing the difference between the experimental ratio of normalized fluorescence for the pTC gene circuit, cloned in the high copy number plasmid, at IPTG = 0 and at maximum induction and the numerical equivalent, computed by simulating the circuit in the same conditions. Assuming  $P_c = 80$  plasmids for the pSB1A2 cloning vector, the plasmids in the cell compartment when the pTC gene circuit is cloned in pSB4A5 was chosen to reproduce the experimental ratio of the pTC normalized fluorescence in absence of induction. Indeed, when we tried to adopt the same number of plasmids determined for the TC gene circuit, the simulated dose-response curve could not reproduce the experimental data.

**Stochastic models.** Stochastic simulations were performed using the Gillespie's direct method, implemented using custom code written in Python. At least 1000 trajectories were simulated for each configuration, using for the kinetic rates the parameters values fitted in the deterministic models. The steady-state mean and variance in GFP molecules was used to compute the dose-response curves and to quantify the stochasticity in the expression of the fluorescent reporter, expressed as the *coefficient of variation (CV)*.

Symbols and parameters used in the deterministic and stochastic simulations are summarized in Table 2.1.

Table 2.1: State variables and parameters values adopted in simulations. § Values obtained by experimental measurements. +Values defined through the fitting procedure of the dose-response curves. References are included for literature retrieved parameters.

Symbol	Definition	Value	Units
$P_c$	Number of plasmids per cell	80 in pSB1A2 60 for TC in pSB4A5 36 for pTC in pSB4A5	molecules cell <sup>-1</sup>
$M_{CG}$	CIS-GFP mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$M_T$	TRANS mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$M_{TG}$	TRANS-GFP mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$G$	GFP molecules per cell	-	molecules cell <sup>-1</sup>
$k_{r,CG}$	CIS-GFP transcription rate	$7.5 \cdot 10^{-3}$ §	s <sup>-1</sup>
$k_{r,T}^{MAX}$	Maximum transcription rate of the TRANS sequence	$6.8 \cdot 10^{-3}$ §	s <sup>-1</sup>
$k_{r,TG}^{MAX}$	Maximum transcription rate of the TRANS-GFP sequence	$6.8 \cdot 10^{-3}$ §	s <sup>-1</sup>
$IPTG_{50}$	Hill curve dissociation constant	70.7 <sup>+</sup>	μM
$n$	Hill coefficient	2.6 <sup>+</sup>	-
$k_p$	GFP translation rate	0.11 [98]	s <sup>-1</sup>
$\gamma_r$	mRNA degradation rate	$4.5 \cdot 10^{-3}$ [99-102]	s <sup>-1</sup>
$\gamma_h$	Stoichiometry dependent degradation rate	$1.5 \cdot 10^{-3}$ +	molecules <sup>-1</sup> s <sup>-1</sup>
$\gamma_p$	Protein degradation rate	$3.6 \cdot 10^{-4}$ [99-102]	s <sup>-1</sup>

## 2.2 Results and discussion

### 2.2.1 Definition of the gene-circuits catalogue

The simultaneous transcriptional and post-transcriptional regulation in the expression of the fluorescent reporter should enable the synthesis of a desired GFP concentration with different noise strengths. The topology of the designed noise tester circuits was hence derived from the post-transcriptional controller including a transcriptional regulation of the CIS-GFP sequence. In the resulting genetic program, summarized in Figure 2.2, the transcription of the CIS-GFP sequence occurs with different dynamics depending on the considered promoter variant and its kinetic rate can be tuned over a wide range upon induction with anhydrotetracycline (aTc). Similarly, IPTG administration, displacing endogenous LacI from the promoter driving transcription of the TRANS oligoribonucleotide, provides an external control on the cytosolic concentration of the silencing transcript. Regulating the ratio between the coding mRNA and TRANS sequence molecules number, the two inducers allow a simple control of the encoded noise tester circuits. Indeed, aTc concentration sets the number of GFP coding mRNA molecules, while IPTG levels, regulating the number of TRANS molecules, constrain the pool of coding mRNA competent for translation and should buffer the fluctuations in CIS-GFP mRNA counts through the annealing-dependent post-transcriptional control. Different desirable features were considered while cloning these synthetic circuits. Other than being simple to control, an ideal noise tester should be easy to implement and allow the wide tuning of any protein mean and variance, while limiting the metabolic burden its presence and function imposes on transformants. In order to fulfil these requirements, the synthetic circuits were built using biological parts whose physical standardization enables an easy assembly and manipulation. To further support functional modularity, useful in *a priori* predictions of synthetic circuits' variants behaviour, all the regulated promoters were assembled cloning an operator site in the proximal region of a constitutive promoter. In order to expand the investigated tuning regime, different inducible promoters were considered for each transcriptional unit of the genetic program. Finally, while the properties of CIS and TRANS elements should support the regulation of the protein's mean and variance independently of its encoding gene, the experimental investigation of the circuits' behaviour was performed in a host strain characterized by genomic overexpression sites for the TetR and LacI repressors. Compared to the original implementation of the post-transcriptional controller (Figure 2.1), this choice is advantageous as it requires the transformation of the host strain with a single plasmid harbouring the gene circuit and the use of a single antibiotic for transformants selection.

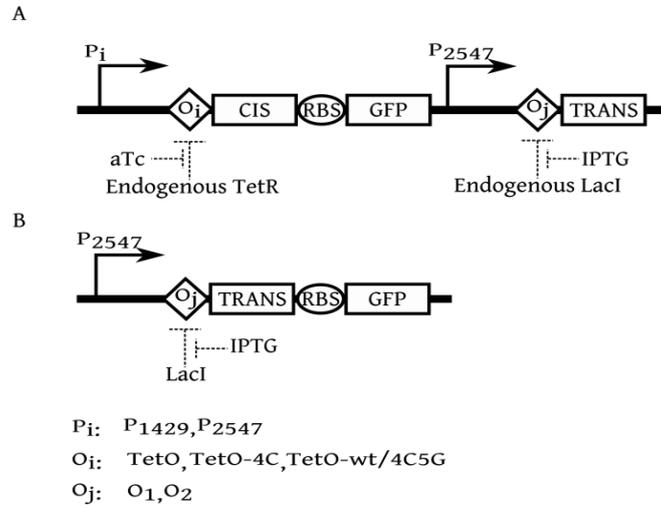


Figure 2.2: Gene circuits catalogue. Panel A shows the noise tester catalogue, where transcription of the CIS-GFP mRNA can proceed from two alternative promoters (P<sub>1429</sub> or the stronger P<sub>2547</sub>) belonging to the Anderson's promoters library. The cloning of an operator site for TetR repressor (O<sub>i</sub>) downstream of this promoter allows the transcription of CIS-GFP to be tuned upon aTc administration. Three operator site variants were used: TetO, TetO-4C and TetO-wt/4C5G. The former is the wild type operator sequence and is denoted by a high binding affinity for TetR repressor. TetO-4C, obtained by inserting a T-C mutation in the fourth nucleotide on both sides of the native consensus sequence; has a medium TetR binding affinity. In TetO-wt/4C5G, the double mutation of the fourth and fifth nucleotide in the right half of the consensus sequence compromises TetR binding: the resulting repression is approximately 50 fold less stringent than in the natural operator. In the second transcriptional unit shown in panel A, transcription of the TRANS element can be differently repressed depending on the O<sub>j</sub> operator site, for which only two natural sequences were considered. In panel B a gene circuit which differs from this transcriptional unit because of the cloning of the RBS and GFP encoding gene downstream of the TRANS acting sequence is shown. As indicated in the main text, investigating this synthetic circuit behaviour provides a fluorescent readout proportional to the transcription of the TRANS element in the noise tester.

In the next section we will present the results of the experimental, population-level characterization of the function implemented in a subset of the circuits' catalogue described so far. As well as providing a deeper understanding of the noise tester's basic components, the parameter identification performed on this experimental data will prove useful for numerical investigations of the mean and variance in protein levels achievable with other members of the circuits' library.

### 2.2.2 Experimental characterization of synthetic circuits implementing transcriptional or post-transcriptional control of GFP expression

As it can be seen in Figure 2.3, the considered subset is composed of two synthetic circuits exerting either transcriptional or post-transcriptional control in the expression of the fluorescent reporter gene. In the former, hereafter referred to as TC gene-circuit, induction with IPTG is expected to unbind endogenous LacI from its natural operator site  $O_1$ , leading to an increased transcription of the TRANS-GFP sequence. In this gene-circuit, the TRANS element, given the absence of its complementary CIS, does not perform any regulatory function. However, considering the observation of unexpected effects exerted by sequences flanking operator sites on their affinity for protein regulators, cloning TRANS upstream of the GFP encoding gene reproduces a genetic context similar to the one denoting the second transcriptional unit of the noise tester. As a consequence, fluorescence data acquired when characterizing the TC gene circuit was used to estimate the  $P_{2547}$  inducible promoter's transcriptional dynamic in the noise tester. The second gene circuit exerts the post-transcriptional regulation in GFP expression through hybridization of the CIS and TRANS complementary sequences and is therefore referred to as pTC. With the aim of assessing the synthetic devices' performance over a wide range of GFP concentrations and obtaining an experimental setup suitable for the investigation of plasmid copy number's effect on stochasticity in gene expression, both gene circuits were excised from pSB1A2 plasmid and cloned in a low copy number plasmid (pSB4A5).

The criteria motivating the selection of the synthetic circuits initially undergoing experimental characterization are the simple control of GFP expression and the dim intensity of the corresponding single cell fluorescent signal. Indeed, the regulation of the fluorescent reporter expression, implemented via alternative control mechanisms in the two circuits, relies on a single signal: the concentration of IPTG inducer. Furthermore, the pSB4A5 harbouring the pTC gene-circuit, in which the CIS-GFP sequence is transcribed from the weak  $P_{1429}$  promoter, provides a platform to test if the intensity of the emitted fluorescent signal, dampened by the post-transcriptional control mechanism, is sufficiently high to be detected in single cell measurements acquired by our fluorescence microscopy setup.

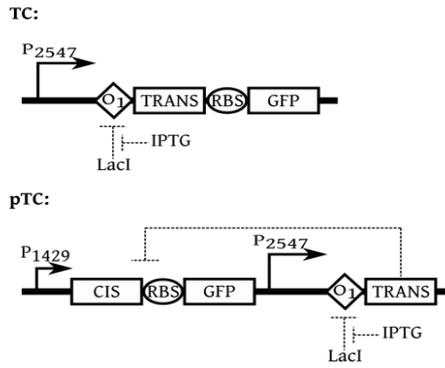


Figure 2.3: Genetic program for the gene circuits initially undergoing experimental characterization. The image shows the topology of the TC (top) and pTC (bottom) gene circuits selected for the experimental characterization. The circuits were cloned either in a high (pSB1A2) or low (pSB4A5) copy number plasmid, containing ampicillin resistance.

The results of experimental characterization of TC and pTC gene circuits performed with the microplate reader are shown in Figure 2.4, where the steady-state normalized fluorescence values are compared for the low (blue bars) and high (green bars) copy number plasmid. Coherent with our expectations, the data shows that induction with IPTG yields a marked increase in GFP expression levels for the TC gene circuit and a moderate decrease for the pTC. This indicates that the synthetic gene circuits operate as intended. Focusing on the pTC circuit, the comparison of the normalized fluorescence values measured in absence of IPTG, i.e. when transcription of the TRANs sequence is abolished and GFP is constitutively expressed, and at full induction, when the P<sub>2547</sub> promoter reaches its maximum transcriptional activity and the repressive effect of TRANs sequence is maximized, highlights a 30% reduction in GFP synthesis (preserved on both plasmids). This experimental evidence, suggesting that the post-transcriptional control mechanism allows tuning the GFP mean expression levels over only a limited range, might be unexpected considering the strength of the P<sub>1429</sub> promoter relative to P<sub>2547</sub>: quantified to be 0.56. Considering both the high fluorescence value measured on the TC gene circuit at IPTG = 200  $\mu$ M and the similar reduction in GFP synthesis observed in the original investigation of the pTC (performed in CSH126 *E. coli* cells transformed with a plasmid harbouring the *LacI* gene), the limited repressive action on GFP translation cannot be ascribed to the overabundance of LacI repressor in the cellular compartment. An alternative explanation for the limited efficiency of the post-transcriptional control mechanism relates to the probability of occurrence or the stability of the CIS-TRANs silencing complex. As the two regulatory RNAs were synthesized in order to be complementary over the whole sequence, the limited repression on GFP translation might arise from the low hybridization rate between CIS-GFP mRNA and *trans*-acting oligoribonucleotide or the ribosomes ability of processing annealed sequences. It is worth noting that the mild translational repression observed in the pTC gene circuit closely

replicates the weak repression of natural microRNAs (miRNA) on their target genes [103, 104]. Moreover, its coherent appearance in the two plasmid backbones considered does not hamper our investigation of plasmid copy number effect on the average value and dispersion of GFP expression levels.

A closer inspection of the data shown in Figure 2.4 reveals that the plasmid-dependent normalized fluorescence scales proportionally with IPTG induction in both gene circuits. While this observation is expected under the hypothesis that the number of plasmid molecules present in the cellular compartment does not saturate the endogenous machinery's processing capability, the value of ratio between the normalized fluorescence provided by the high and low copy, at each induction level, is not in line with theoretical expectations. Indeed, based on the properties of their origins of replication, the pSB1A2 high copy plasmid is expected to inhabit the cellular compartment with ~100-300 plasmid molecules, while the pSB4A5 should occur at ~5 copies per cell. In absence of saturating effects, the ratio of the normalized fluorescence measured, at each IPTG concentration, on the high and low copy number plasmids should be proportional to the ratio of their respective plasmid counts. As a consequence, the ratio of the normalized fluorescence measured on the high and low copy number plasmid is expected to be in the range 20– 60. When evaluated on the fluorescence values measured on the pTC gene circuit, the ratio is 2.2: approximately one order magnitude smaller than the minimal value predicted theoretically. This evidence, coupled with the high yields of miniprep solutions prepared from cultures of *E. coli* cells transformed with the pSB4A5 plasmid, would suggest that this is not a low copy number plasmid. This hypothesis was subsequently supported by experimental investigations performed by R. Shetty, who designed this plasmid [94].

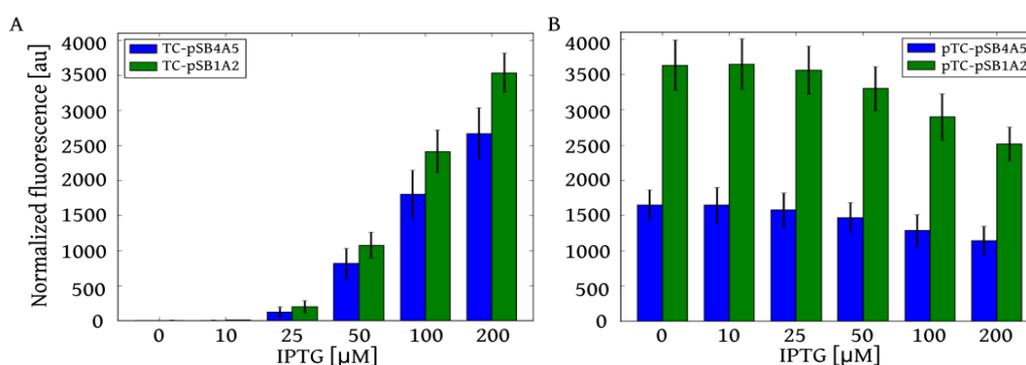


Figure 2.4: Normalized steady-state fluorescence for TC and pTC gene circuits. Panel A shows a bar plot of comparison between the normalized fluorescence values acquired, upon induction with five IPTG concentrations from 0 to 200 μM, on the TC gene circuit when cloned in the low (pSB4A5, blue bars) or high (pSB1A2, green bars) copy number plasmid. In panel B the monotonic decrease of the normalized fluorescence with IPTG in pTC gene circuit, harboured either in the low (pSB4A5, blue bars) or high copy number plasmid (pSB1A2, green bars), is shown. Experimental data is reported as the mean fluorescence value, computed over 9 colonies assayed in technical duplicate, with error bars representing the standard deviation. As can be seen, GFP mean expression levels measured on a single gene circuit scale proportionally with the plasmid copy number. However, the ratio of the fluorescent reporter concentrations observed in the high and low copy number plasmids at a given induction level proves incoherent with theoretical predictions based on plasmid counts expected for the two origins of replication and differs for the two synthetic circuits.

The medium copy number attributed to the pSB4A5 could hinder the experimental evaluation of plasmid counts effect on gene expression stochasticity in the subset of synthetic circuits under analysis.

Another unexpected feature in the experimental data is the plasmid-dependent difference in GFP expression between the TC and pTC gene circuits. As emphasized in Figure 2.5, when the data measured on pSB4A5 plasmid are shown as a function of their equivalent acquired on pSB1A2, the ratio of normalized fluorescence between the two plasmid backbones, proportional to the ratio between the respective plasmid molecules, appears dependent on the considered gene circuit. Specifically, the slope of the linear predictor computed by a least-square fit of the ‘low’ copy data as a function of the high copy ones is 0.75 for the TC and 0.45 for the pTC gene circuits. Considering the simple topology of the gene circuits under analysis, a dependency of plasmid counts on the synthetic device located in their multiple cloning site is unlikely. The comparison of the normalized fluorescence values for the TC gene circuit at IPTG = 200  $\mu$ M with the data measured for the pTC in absence of induction on the high copy number plasmid (Figure 2.4) would instead suggest the presence of a saturation phenomenon for the TC cloned on pSB1A2. Indeed, at these induction levels, the transcriptional activity of the  $P_{1429}$  promoter in pTC and  $P_{2547}$  in TC should approach the maximum value and, as a consequence, the GFP expression levels obtained with the transcriptional control mechanism should almost double the ones measured on the pTC gene circuit, as occurs for the pSB4A5 cloning vector. By characterizing the TC and pTC gene circuits with the plate reader, we observed the appearance of a saturation phenomenon occurring in the circuit implementing transcriptional control in the expression of the fluorescent reporter cloned in pSB1A2 plasmid. Furthermore, we outlined that the pSB4A5 cloning vector occurs in the cell compartment with a number of copies significantly higher than that expected for a low copy number plasmid.

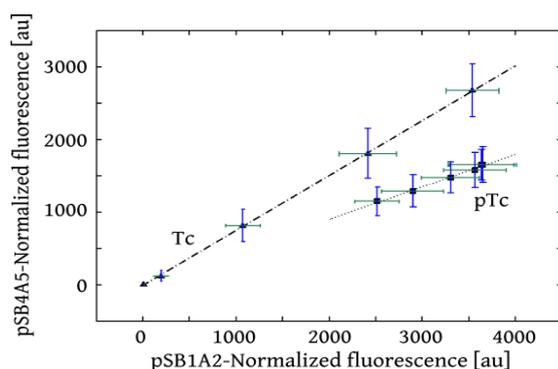


Figure 2.5: Normalized fluorescence values dependence on the cloning vector’s copy number for the TC and pTC gene circuits. The normalized fluorescence measured, at the tested inducer concentrations, when the two gene circuits are cloned in plasmid pSB4A5 are reported as a function of their respective data on pSB1A2. The average of the normalized fluorescence is shown for the TC and pTC gene circuits, with error bars representing the standard deviation in the high (green bars) and ‘low’ (blue bars) copy number plasmids. The linear fit to the data for the TC (dashed line) and pTC (dotted line) gene circuits shows that the fluorescence values scale proportionally in the two plasmid backbones, upon IPTG induction. The slope of the linear predictor ( $R^2 > 0.99$ ) differs between the two gene circuits.

We will now turn to use this data to parametrize deterministic models of the gene circuits' behaviour.

### 2.2.3 Macroscopic description of the TC and pTC gene circuits' function

In this section deterministic models describing the population-level GFP expression experimentally assessed for the TC and pTC gene circuits will be presented. The major goal was the identification of parameters describing the kinetic rates of biochemical reactions underpinning the two synthetic circuits' behaviour. While different strategies for the parameters fitting were investigated, we will report only the procedure yielding the best description of the experimental data. We were interested in identifying, using a chemical rate equations formulation, parameters to be adopted in the microscopic description of the system, provided by the CME.

The following set of biochemical reactions defines a possible model of the TC gene circuit:



In reactions 2.1-2.4  $M_{TG}$  and  $G$  are the number of mRNA and protein molecules per cell,  $P_c$  is the number of plasmid counts present in the cellular compartment,  $\gamma_r$  and  $\gamma_p$  are respectively the degradation rates of mRNA and protein molecules,  $k_{r,TG}(IPTG)$  is the transcription rate, and  $k_p$  is the translation rate per mRNA molecule. It is interesting to note that the formalism adopted in equation 2.1 provides an abstract description of the ongoing biological reality, subsuming that all the  $P_c$  plasmids are recruited and induction with IPTG increases the rate of transcription driven by the  $P_{2547}$  promoter. A more precise and biologically meaningful modelling of IPTG effect on transcription could be achieved considering that the inducer concentration regulates the fraction of active operator sites  $O_1$ , from which transcription occurs at a constant rate  $k_{r,TG}$ . Both the higher complexity of such a model and the equivalent results provided by the two alternatives, lead us to favour the version here reported.

In reactions 2.1- 2.4 the dependence of  $k_{r,TG}(IPTG)$  on the inducer concentration, assuming fast transcription factor/DNA binding, was modelled by a Hill equation:

$$k_{r,TG}(IPTG) = k_{r,TG}^{MAX} \frac{\left(\frac{IPTG}{IPTG_{50}}\right)^n}{1 + \left(\frac{IPTG}{IPTG_{50}}\right)^n} \quad 2.5$$

Where  $n$  is Hill coefficient and  $IPTG_{50}$  is the inducer concentration providing a transcriptional rate equal to half the maximum value. Under the hypothesis that the amounts of biochemical molecules of interest, homogeneously distributed within the cell compartment, evolve continuously through instantaneous reactions dependent on the actual state of the system, a macroscopic description of the gene circuit's function can be provided by the following set of linear ODEs:

$$\frac{dM_{TG}}{dt} = k_{r,TG}(IPTG)P_c - \gamma_r M_{TG} \quad 2.6$$

$$\frac{dG}{dt} = k_p M_{CG} - \gamma_p G \quad 2.7$$

The TC gene circuit's model defined by equations 2.6-2.7 was simulated in Matlab-Simulink (2013a, The MathWorks, Natick, MA), with the  $k_{r,TG}(IPTG)$  and  $P_c$  parameters values determined by fitting the experimental dose-response curves (see section 2.1.4 for more details). The steady-state solutions of the ODEs system, numerically obtained using the ode45 solver as implemented in Matlab, accurately reproduces the experimental dose-response curves acquired for the TC gene circuit.

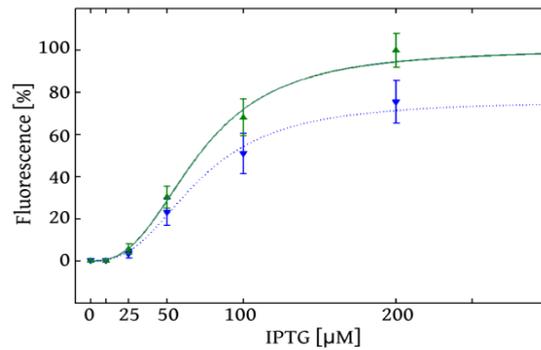
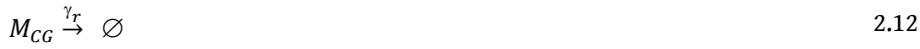


Figure 2.6: Dose-response curves for the TC gene circuit cloned in pSB4A5 and pSB1A2 plasmids. Fluorescence data is normalized by the average fluorescence measured in the TC circuit, cloned in pSB1A2, at maximum induction (IPTG = 200  $\mu$ M). Experimental data is reported as mean (triangle marks), with error bars representing the standard deviation, for both the pSB1A2 (green colour) and pSB4A5 (blue colour) cloning vectors. In the same panel, dose-response curves simulated for the TC gene circuit placed in the pSB1A2 (green, continuous line) and pSB4A5 (blue, dotted line) are shown.

A possible model of the gene circuit implementing post-transcriptional control in GFP expression is defined by the following set of chemical reactions:



In reactions 2.8-2.14  $M_{CG}$  and  $M_T$  are the number of CIS-GFP and TRANS mRNA molecules per cell,  $G$  is the number of protein molecules per cell,  $k_{r,CG}$  and  $k_{r,T}(IPTG)$  are the transcription rates of  $M_{CG}$  and  $M_T$ ; and  $\gamma_h$  the stoichiometry-dependent degradation rate of mRNA molecules due to the hybridization of the CIS-TRANS sequences. The remaining symbols have the same meaning adopted in reactions 2.1-2.4. It is worth noting that the bimolecular reaction 2.11 describes the silencing mechanism due to the direct, irreversible interaction between the  $M_{CG}$  and  $M_T$  RNA molecules. The partial occlusion at the RBS due to the hybridization process reduces the amount of  $M_{CG}$  molecules competent for translation, effecting the post-transcriptional control, and leads to degradation of the annealed molecules. This representation was preferred to one describing separately the RNA-RNA interaction, resulting in the naissance of an hybridized mRNA, which is subsequently degraded. Despite being conceptually equivalent and yielding the same results, the description adopted is simpler and allows reproducing the circuit's behaviour using only three state variables. The modulation of  $k_{r,T}(IPTG)$  by the inducer concentration was modelled using an equation analogous to 2.5, adopting for parameters  $n$  and  $IPTG_{50}$  the values determined by the fitting of TC dose-response curve.

The deterministic model derived from reactions 2.8-2.14 is composed of the following set of ODEs:

$$\frac{dM_{CG}}{dt} = k_{r,CG}P_c - (\gamma_r + \gamma_h M_T)M_{CG} \quad 2.15$$

$$\frac{dM_T}{dt} = k_{r,T}(IPTG)P_c - (\gamma_r + \gamma_h M_{CG})M_T \quad 2.16$$

$$\frac{dG}{dt} = k_p M_{CG} - \gamma_p G \quad 2.17$$

As can be seen in equation 2.15, the hybridization reaction depicting the post-transcriptional control introduces a non-linearity in the system, impeding the determination of an analytical solution for the steady state number of GFP molecules. Upon fitting the parameter  $\gamma_h$  (the detailed procedure is described in section 2.1.4), the steady-state solutions of the set of ordinary equations was numerically computed using the ode15s solver implemented in Matlab- Simulink.

The simulated dose-response curves for the pTC gene circuit, cloned in pSB1A2 and pSB4A5 plasmid backbones, faithfully reproduce the fluorescence decrease with IPTG induction experimentally observed (Figure 2.7).

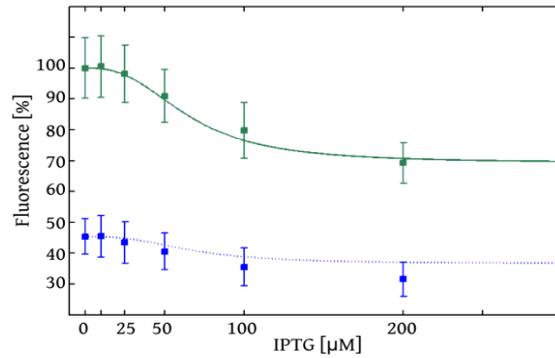


Figure 2.7: Dose-response curves for the pTC gene circuit cloned in pSB4A5 and pSB1A2 plasmids. Fluorescence data is normalized by the average fluorescence measured in the pTC circuit, cloned in pSB1A2, in absence of induction (IPTG = 0 μM). Experimental data is reported as mean (square marks), with error bars representing the standard deviation, for both the pSB1A2 (green colour) and pSB4A5 (blue colour) cloning vectors. In the same panel, dose-response curves simulated for the pTC gene circuit placed in the pSB1A2 (green, continuous line) and pSB4A5 (blue, dotted line) are shown.

The deterministic models presented so far provide a macroscopic, simplified description of the synthetic circuit's behaviour. Such mathematical formalism proves inadequate whenever the counts of interacting molecules is low, as occurs for natural genes, and prevents the quantification of stochasticity in gene expression process.

In the next section we will abandon the kinetic rate equations model and we will instead consider the probabilistic occurrence of the biochemical reactions underpinning the analysed gene circuits' function, yielding a stochastic evolution of the system over time.

#### **2.2.4 Copy number effect on stochasticity in GFP expression under transcriptional and post-transcriptional control: a numerical investigation**

Stochastic models were used to numerically predict how the number of plasmids in the cellular compartment affects the variability in GFP expression levels among an isogenic population of transformants. While the gene circuit's average behaviour computed with deterministic and stochastic models is expected to coincide in linear systems, i.e. those described by zeroth and first order biochemical reactions, transitioning from a macroscopic to a microscopic picture of the biological model usually requires a proper scaling of the kinetic rates governing the system's evolution over time. Indeed, state variables are conventionally expressed as molecules concentrations when writing rate equations, while in the microscopic description provided by the CME they appear as particles numbers. Having expressed the deterministic model's state variables as molecules counts, stochastic models of the investigated gene circuits used the same parameters values identified in the fitting procedure previously outlined.

Using the Gillespie algorithm, trajectories of the CME providing a probabilistic picture of the TC circuit, cloned in both the pSB1A2 and pSB4A5 plasmid, were simulated using the set of reactions 2.1-2.4. The steady-state mean and variance in GFP expression, averaged over 1000 trajectories, was computed varying the transcription rate of the TRANS-GFP sequence according to equation 2.5. The simulated dose-response curves, shown in panel A of Figure 2.8, correctly reproduce the experimental data for both plasmid contexts. In addition, the perfect match between the numerical and analytical dose-response curves ensured the correctness of the implemented algorithm. When considering the stochasticity in GFP expression, quantified by the *coefficient of variation*, the model describes the expected decrease in biological noise upon induction with IPTG. Administering the inducer causes an increase in TRANS-GFP transcription, leading to higher expression of the fluorescent reporter: a condition which limits the relevance of finite number effect. In addition, the numerical results suggest a higher stochasticity in GFP expression for the pSB4A5 plasmid. While this observation is in line with theoretical considerations, the obtained curves are not statistically different. A plausible explanation is the limited fractional change in counts between the two plasmids, which was revealed by the deterministic model.

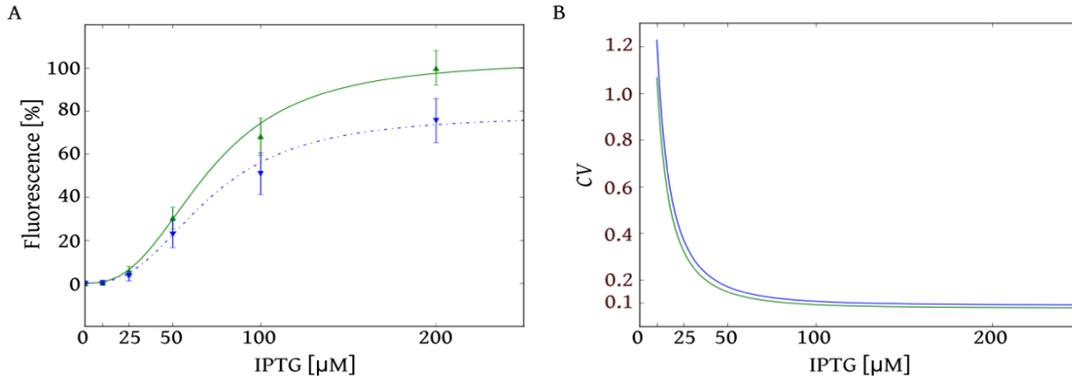


Figure 2.8: Results of the stochastic model for the TC gene circuit. In panel A the agreement between the dose-response curves simulated by stochastic simulations and the experimental data acquired on the TC gene circuit is shown. Experimental data are reported as mean  $\pm$  standard deviation, using green upper triangles for the circuit cloned in pSB1A2 and blue lower triangles for the low copy number cloning vector. Experimental values are normalized by the average fluorescence measured in TC gene circuit, cloned in pSB1A2, at maximum induction (IPTG = 200  $\mu\text{M}$ ). Panel B shows the trend of the numerical *coefficient of variation* (CV) with increasing IPTG for the pSB1A2 (green line) and pSB4A5 (blue line) plasmid backbones. In line with our expectations, the curves reproduce a reduction in GFP expression stochasticity upon induction: under this condition the higher synthesis of fluorescent reporter molecules constrains the relevance of finite number effect. When comparing the noise strengths originating from cloning vectors with different copy number, the numerical predictions suggest a higher variability in GFP expression levels for the TC gene circuit cloned in pSB4A5 plasmid.

With analogous procedure, stochastic models based on the set of reactions 2.8-2.14 were developed for the pTC gene circuit. It is worth noting that the bimolecular nature of reaction 2.11, which embodies the post-transcriptional control mechanism, could lead to discrepancies in the dose-response curves simulated with the deterministic and stochastic version of the pTC gene circuit's model. The good agreement between simulated dose-response curves and the experimental decrease in normalized fluorescence with IPTG indicates that the mean obtained with stochastic simulations equals the steady-state solution of the rate equations. When considering GFP expression's variability, the model, in contrast with the results for the TC gene circuit, provides a coefficient of variation which is almost constant upon induction with IPTG. The difference in noise amplitude computed for the pSB4A5 and pSB1A2 cloning vectors is higher than the one estimated numerically for the TC circuit. While this supports the hypothesis that the limited difference observed for the gene circuit implementing transcriptional control in the expression of the fluorescent reporter is due to the similar number of plasmids estimated in the fitting procedure based on the deterministic model, this difference is relatively low and might not be detectable in experimental measurements.

Finally, a qualitative comparison among the CV estimated for the TC and pTC gene circuits, cloned in the high copy number plasmid, at maximum (IPTG = 200  $\mu\text{M}$ ) and in absence of induction

respectively –condition under which the mean normalized fluorescence has similar values (Figure 2.4) – would suggest that the post-transcriptional regulation in GFP expression produces a lower stochasticity than the one provided by the transcriptional control.

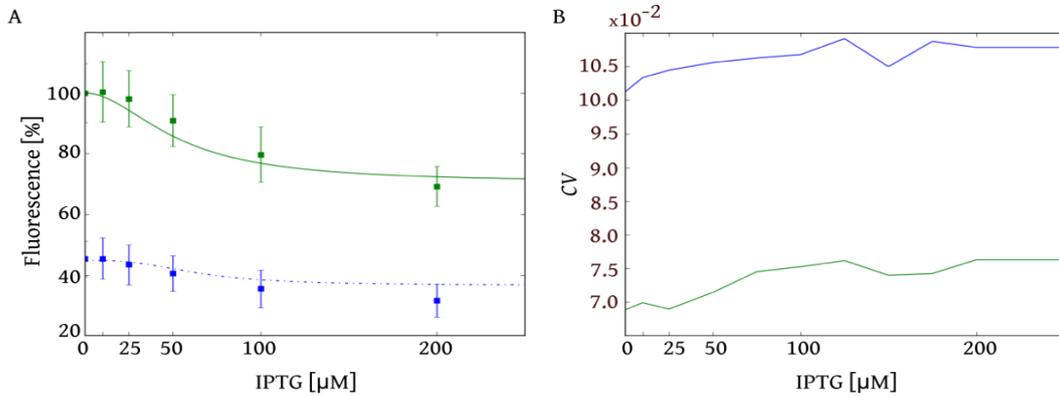


Figure 2.9: Results of the stochastic model for the pTC gene circuit. In panel A the agreement between the dose-response curves simulated by stochastic simulations and the experimental data acquired on the pTC gene circuit is shown. Experimental data are reported as mean  $\pm$  standard deviation, using green squares for the circuit cloned in pSB1A2 and blue squares for the pSB4A5 cloning vector. Experimental values are normalized by the average fluorescence measured in pTC gene circuit, cloned in pSB1A2, in absence of induction. Panel B shows the trend of the numerical *coefficient of variation* (CV) with increasing IPTG for the pSB1A2 (green line) and pSB4A5 (blue line) plasmid backbones. The limited fractional change in GFP expression upon IPTG induction is probably responsible for a constant value of the CV with increasing inducer concentrations. When comparing the noise strengths originating from cloning vectors with different copy number, the numerical predictions suggest a higher variability in GFP expression levels for the pTC gene circuit cloned in pSB4A5 plasmid.

## 2.3 Conclusions

In this chapter we described the development of a catalogue of gene circuits which could act as a noise tester, providing a tool for *a priori* predictions inherent the robustness of newly designed synthetic devices or for testing existent molecular widgets. A subset of these gene circuits, cloned in plasmid backbones with different copy number, was characterized at the population level. The experimental measurements confirmed that the TC and pTC gene circuits operate as expected: upon IPTG induction the fluorescent reporter synthesis increases in the gene circuit implementing transcriptional control and decreases when the regulation of GFP expression is effected at the post-transcriptional level. Analysing the normalized fluorescence data we outlined both the presence of a saturation phenomenon occurring in the TC gene circuit cloned in the high copy number plasmid, where GFP expression at maximum induction is comparable with the one provided by the uninduced pTC gene circuit, and the Parts Registry's misleading classification of pSB4A5 as a low copy number plasmid. These experimental evidences forced us to favour a parameters fitting procedure which disregards an accurate description of the ongoing biological reality but faithfully reproduces the measured dose-response curves. The identified parameters were subsequently used in stochastic simulations aimed at evaluating the effect of plasmid copy number on stochasticity in gene expression under the alternative control mechanisms. The results suggest a reduced biological noise for the high copy number cloning vector in both gene circuits and indicate that noise strength can be tuned administering IPTG in the TC gene circuit, while remains almost constant in the circuit implementing post-transcriptional control. As the discrepancies in the expected differences in plasmid counts associated with the adopted cloning vectors could hinder the experimental evaluation of a differential stochasticity in GFP expression, in the next chapter we will elaborate the presented methodology by making experimental measurements of the noise profile arising from the alternative control mechanisms encoded in the TC and pTC gene circuits cloned in pSB4A5. The experimental stochasticity in GFP expression will be compared with model predictions, in order to develop a theoretical framework to deepen our understanding of the transcriptional and post-transcriptional contributions to the variability in protein levels.



### 3 Experimental measurements and mathematical modelling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits<sup>7</sup>

As already mentioned, investigating how phenotypical noise is affected by the different regulatory mechanisms that control gene expression would pave the way for a better understanding of biological processes and an efficient design of more robust synthetic circuits. In this chapter, the noise in protein concentration will be compared between two synthetic networks, previously characterized at the macroscopic population level, implementing either a transcriptional or a post-transcriptional control of gene expression.

While transcriptional control has long been considered the most widespread gene regulatory mechanism in nature, post-transcriptional regulation has attracted increasing interest with the discovery of a significant number of non-coding RNAs.

The simplest reason to undertake the comparison here presented is the investigation of the differential effects of transcriptional and translational processes on protein variability. Indeed, translation acts as a natural amplifier of mRNA fluctuations, thought to be the prime source of intrinsic noise. Moreover, due to proteins' high molecular stability, the resulting stochasticity persists long after the degradation of the coding mRNA molecules. These considerations suggest that a direct, translational control would allow an efficient buffering of unavoidable mRNA fluctuations tuning intrinsic noise. Such a strategy, already exploited in natural system where essential genes requiring tight control of expression levels are often translated from low efficient RBSs, might be adopted to reduce noise in the design of synthetic gene-circuits.

The control of noise by mutations in the RBS sequence, explored in the landmark study of Ozbudak et al. [6], represents an optimal strategy when noise level is a static characteristic of the circuit, i.e. gene expression stochasticity does not require adaptation in fluctuating environments. However, in other cases, an external control on noise strength might be necessary. As an example, the concentration of a protein might be critical in static environmental conditions, while a more

---

<sup>7</sup> Most of the content of this chapter is published as Bandiera, L., et al., *Experimental measurements and mathematical modeling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits*. J Theor Biol, 2016. **395**: p. 153-60.

dispersed distribution of expression levels would ensure a faster response to fluctuating environments [41]. The intriguing phenomenon of stochastic resonance mirrors how a noise level sensitive to an external signal might be useful to implement complex cellular behaviours. Due to stochastic resonance, a system might oscillate in response to a weak periodic input signal, if associated with a specific level of noise. Once noise is abolished, the outcome of the system ceases to be periodic. Thus, contrary to intuition, the signal-to-noise ratio increases for increasing noise levels [105]. Stochastic resonance has been proposed as a plausible mechanism for explaining circadian oscillations in biological systems [106, 107], which are indeed caused by a weak periodic signal (the day-night cycle) in a noisy environment. A mechanism similar to stochastic resonance contributes to stochastic focusing, where noise improves the sensitivity of a detector [108]. Under the perspective of controlling noise by means of an external signal, circuits based on stochastic resonance (or stochastic focusing) might be turned on/off in response to changes in the environmental conditions. This possibility gives rise to two exciting corollaries. First, tunable noise could be used to implement complex functionalities in synthetic gene-circuits. Second, thanks to an external control on noise strength, it would be possible to directly test the role of mechanisms as stochastic resonance and stochastic focusing, or more generally of noise itself, on cellular processes.

It is thus apparent how both the inherent features of gene expression process and the ambition of tuning biological noise by means of an external signal converged to a renewed interest in post-transcriptional control mechanisms. Moreover, quantifying the poorly investigated effects of mRNA-based translational regulation on gene expression stochasticity would support the use of this intriguing control mechanism in the design of novel synthetic circuits.

Post-transcriptional mechanisms are known to exert an important control on gene expression both in eukaryotes and in prokaryotes. In bacteria, such as *Escherichia coli*, post-transcriptional control by small RNA (sRNA) molecules seems predominant in stress response pathways and in virulence genes regulation [109]. Theoretical analyses revealed that a gene downregulated by a trans-acting sRNA, the term used to identify a riboregulator transcribed from a genomic locus distant from the chromosomal gene it controls, might exhibit three regimes of expression [80]. When sRNA molecules outnumber the pool of coding mRNA, expression is silenced. At the other side of the spectrum, i.e. in presence of a surplus of coding mRNA, the concentration of protein increases linearly with the rate of transcription of the target gene. In between these extreme conditions, there is a crossover regime, where variance on protein concentration is maximized [110]. Stochastic simulations proved that in the silenced-regime, the post-transcriptional control mechanism has minimal noise, and that this noise level is lower than the one exhibited when the same average concentration of protein is synthesized through a transcriptional control mechanism [110, 111]. The low-level of noise in the silenced-regime might be an explanation for the post-transcriptional control of critical genes, like

the ones responsible for the response to oxidative stress [112]. However, experimental analyses on the iron homeostasis network of *E. coli* have not revealed any reduction in noise levels related to post-transcriptional control [113]. In this case, transcriptional control turned out to be less noisy, even at low protein concentrations [114]. This lack of experimental evidence for a decrease in noise due to post-transcriptional control might be explained by two factors. First, extrinsic noise might be predominant, thus masking any effect of the post-transcriptional control mechanism on intrinsic noise. Second, it is plausible that the post-transcriptional control mechanism of the iron homeostasis network cannot reach the silenced-regime, as a consequence of toxic effects related to iron deprivation.

The comparative, both experimental and numerical, analysis of protein variability originating from alternative regulatory mechanisms here presented addresses the selection of network topologies useful to minimize noise in synthetic devices. Furthermore, this study could provide insights on the mentioned contradicting conclusions drawn on post-transcriptional control noisiness. Finally, dealing with the necessary characterization of single cell behaviour, the presented investigation constitutes, to the best of our knowledge, the first quantification of phenotypic variability associated with a synthetic circuit implementing post-transcriptional control of gene expression.

We begin with a description of the experimental methods used in the project and a related lab project that adds validity to the presented data. We will then move to the experimental and theoretical analysis by which we compare protein variability originating from transcriptional and post-transcriptional regulation of the fluorescent reporter expression.

## 3.1 Methods

### 3.1.1 Single cell fluorescence assays

Single colonies of *E. coli* TOP10F' strain from a freshly streaked LB-agar plate were inoculated in 5 ml of selective (100 µg/ml ampicillin) M9 minimal medium, completed with casamino acids, thiamine hydrochloride and glucose as a main carbon source (Sigma). After an overnight growth (37°C, 220 rpm orbital shaking) in Erlenmeyer flasks, cell cultures were spun-down (10 min, 3,500 rpm) and resuspended in fresh pre-warmed medium to enable metabolites removal. Cell cultures were then diluted to an initial OD<sub>600</sub> = 0.05. After induction with IPTG concentrations ranging from 0 to 200 µM, cell cultures were incubated at 37°C with orbital shaking until they reached the mid log phase of growth.

**Flow cytometric measurements.** Upon a 3-fold dilution in sterile PBS, fluorescence distributions of recombinant bacteria were analysed by means of a Partec PAS II flow cytometer equipped with an argon ion laser using the 488 nm blue line for excitation, while fluorescence emission was acquired in FL1 through a 515-545 nm band pass filter. At least three colonies for each circuit and induction level were sampled. For each sample, at least 150,000 events (over the signal threshold) were acquired. A non-fluorescent TOP10F' culture was always included to measure the background fluorescence.

**Microscopy measurements.** Prior to image acquisition, samples were concentrated with a fivefold volume reduction and resuspended in sterile PBS, in order to maximize the cardinality of the sampled population while preserving an optimal field of view coverage and minimizing the background autofluorescence during the experiment. Each of the biological triplicates was assayed at the 5 IPTG concentrations adopted in the flow cytometer acquisition: the null induction was discarded due to the inability of identifying the cells outline in absence of fluorescence signal. At least 70 images, split on 6 slides, were acquired for each sample. Images were collected through an inverted microscope Eclipse TE2000-U (Nikon) equipped with a DS-Qi1 monochrome digital, cooled camera (Nikon) using an S-Fluor 40x objective. Once acquired with the Nis Elements Documentation v 4.20 software, images were exported in 8 bits tiff format and encoded in the RGB colour space for the subsequent digital image processing, performed with a custom made segmentation algorithm, coded in Python language.

### 3.1.2 Data analysis

**Flow cytometric data.** Flow cytometric data was processed with the MATLAB 2013b suite (MathWorks, Natick, MA), adopting the *fca\_readfcs* script [115], an FCS data reader package (Laszlo Balkay), and ad-hoc routines to perform a proper gating, through removal of undesired fluorescent events [116]. The log-binned values of forward scatter (FSC), side scatter (SSC) and fluorescence (FL1) are integer values in the range 0-4095. Events having FL1 equal to zero were removed [115]. Linear-scale values of FSC, SSC and FL1 were obtained from log-binned values [99]. In all experiments, more than 30,000 events remained after the described processing and the fluorescent distributions were all unimodal. The FL1 arithmetic mean of the non-fluorescent culture was computed and subtracted from the FL1 values of fluorescent cultures to remove background fluorescence [117]. The mean value and variability of GFP expression levels obtained from the fluorescence distributions were hence adopted for the experimental evaluation of the *Fano factor* and the *squared coefficient of variation*.

**Image analysis.** Raw images were initially corrected for background fluorescence, which limits the SNR by reducing the assessed dynamic range, and camera response function (CRF) aberration, which impairs the empirically observed fluorescence distributions [118]. In order to avoid hypothesis on background features, the former was estimated through a morphological opening greyscale to the analysed image in which the structuring element was selected to be of the same size or bigger than a cell's projection. The result was pixel-wise subtracted from the analysed image. The CRF distortion was compensated by evaluating at the 256 grey levels of the input image the inverted third degree polynomial relating image intensity and scene radiance, as fitted on preliminary experiments. Upon application of a Gaussian filter to the so obtained images, segmentation was performed through an algorithm structured in zero-crossing detection method, hole filling procedure and double BW erosion. Individuals' fluorescence was hence evaluated by averaging the intensities of the pixels within the identified cell outline. Finally, the single cell fluorescence was scaled by a correction factor accounting for the different exposure times adopted while acquiring samples with various fluorescence intensities, corresponding to increasing inducer concentration. The mean expression level and its dispersion within the cell population, extracted from the experimental distributions, were used to compute biological noise, as quantified by the *squared coefficient of variation*. The described post-processing software was developed by Cortesi and is detailed in a dedicated manuscript (in preparation).

### 3.1.3 Stochastic simulations

Stochastic simulations were performed using the Gillespie's direct method, implemented using custom code written in Python. More than 10,000 trajectories were simulated for each configuration, with each trajectory sampling more than 100 cell division events. A division event was simulated each time the cell volume ( $V$ ), exponentially increasing with a rate  $\alpha$ , became higher than 100 arbitrary volume elements. Results did not change upon modification of this threshold (data not shown). The cell growth rate,  $\alpha$ , was defined using the time constant of the exponential phase determined experimentally. The mRNA and protein degradation rates ( $\gamma_r$  and  $\gamma_p$ ) were chosen to produce a half-life of 5 min and 40 min respectively [98, 100-102]. The translation rate,  $k_p$ , was defined in order to achieve an average of 300 proteins per mRNA molecule in the transcriptional control (TC) gene circuit [119]. The dependence of transcription rate in the TC gene-circuit ( $k_{r,TC}$ ) on the IPTG concentration was modelled using a Hill equation (equation 3.5). The asymptotic value  $k_{r,TC}^{MAX}$  was chosen to reproduce the experimental value of  $CV^2$  at IPTG concentration equal to 200  $\mu$ M for the TC gene circuit.  $IPTG_{50}$  and  $n$  were determined by fitting the dose response curve of the TC gene circuit. The transcription rate of the CIS-GFP sequence in the post-transcriptional control (pTC) circuit ( $k_{r,CIS}$ ) was defined as  $0.65 * k_{r,TC}^{MAX}$ , where the multiplicative factor corresponds to the strength of the promoter  $P_{1429}$  relative to the promoter  $P_{2547}$ , as determined experimentally (data not shown). The transcription rate  $k_{r,T}$  was modelled by an equation analogous to equation 3.5, with  $k_{r,T}^{MAX}$  equal to  $(835/58) * k_{r,TC}^{MAX}$ , where 835 and 58 are respectively the number of nucleotides in the mRNA molecules  $M_{TC}$  and  $M_T$ . The values of  $IPTG_{50}$  and of  $\gamma_h$  were determined by fitting the dose-response curve of the pTC gene circuit. Symbols and parameters used in the stochastic simulations are summarized in Table 3.1.

Table 3.1: State variables and parameters values adopted in stochastic simulations. §Values obtained by experimental measurements. +Values defined through the fitting procedure of the dose-response curves. References are included for literature retrieved parameters.

Symbol	Definition	Value	Units
$V$	Cell volume	-	a. u.
$\alpha$	Cell growth rate	$7.84 \cdot 10^{-5} \text{ §}$ in TC $7.23 \cdot 10^{-5} \text{ §}$ in pTC	$\text{s}^{-1}$
$M_{CG}$	CIS-GFP mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$M_T$	TRANS mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$M_{TG}$	TRANS-GFP mRNA molecules per cell	-	molecules cell <sup>-1</sup>
$G$	GFP molecules per cell	-	molecules cell <sup>-1</sup>
$k_{r,CG}$	CIS-GFP transcription rate	$1.95 \cdot 10^{-3} \text{ §}$	molecules s <sup>-1</sup>
$k_{r,T}^{MAX}$	Maximum transcription rate of the TRANS sequence	$4.3 \cdot 10^{-2} \text{ §}$	molecules s <sup>-1</sup>
$k_{r,TG}^{MAX}$	Maximum transcription rate of the TRANS-GFP sequence	$3.0 \cdot 10^{-3} \text{ §}$	molecules s <sup>-1</sup>
$IPTG_{50}$	Hill curve dissociation constant	42.0' in TC 67.3' in pTC	$\mu\text{M}$
$n$	Hill coefficient	1.5'	-
$k_p$	GFP translation rate	0.11 [98]	$\text{s}^{-1}$
$\gamma_r$	mRNA degradation rate	$2.31 \cdot 10^{-3}$ [99-102]	molecules <sup>-1</sup> s <sup>-1</sup>
$\gamma_h$	Stoichiometry dependent degradation rate	$1.1 \cdot 10^{-4} \text{ +}$	$\text{s}^{-1}$
$\gamma_p$	Protein degradation rate	$2.89 \cdot 10^{-4}$ [99-102]	$\text{s}^{-1}$

## 3.2 Results and discussion

The noise that affects protein expression at steady state was studied in two synthetic gene-circuits (Figure 3.3): i) one in which protein synthesis is controlled by a transcriptional mechanism (TC); and ii) one implementing a post-transcriptional control of gene expression (pTC). As mentioned in Chapter 2, the design of both circuits derives from previous works carried out in the ICM Lab to explore the modular design of gene circuits [5, 16]. To briefly recapitulate, the transcriptional control in the TC gene-circuit was achieved by cloning an operator site ( $O_1$ ) for the lactose repressor protein (LacI) downstream of a constitutive synthetic promoter ( $P_{2547}$ ). Exogenous Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) can act as an inducer to regulate the transcriptional rate of the resulting promoter. In order to preserve coherence between the structures of the two gene-circuits, a non-coding DNA sequence (TRANS), playing a pivotal role in the post-transcriptional control mechanism implemented by the pTC gene-circuit, was cloned upstream of the ribosome binding site (RBS) of the green fluorescence protein (GFP) gene also in the TC circuit. The pTC gene-circuit implements a post-transcriptional control mechanism: the regulated promoter described above drives the transcription of a TRANS-acting oligoribonucleotide able to hybridize its complementary sequence (CIS) placed upstream of the GFP gene. CIS-GFP mRNA molecules are transcribed from a constitutive promoter ( $P_{1429}$ ). TRANS mRNAs compete with ribosomes for binding to the CIS-GFP mRNA molecules. Therefore, an increase in the concentration of TRANS sequences is associated with a decrease in GFP translation efficiency. The TC and pTC gene-circuits were cloned in low-copy number pSB4A5 plasmid. TOP10F' *Escherichia coli* strain, overexpressing the LacI repressor, was used as host. Expression levels of the GFP gene in TC and pTC at different IPTG concentrations were evaluated by flow cytometry.

Before dealing with the outlined comparison, we will summarize the results of the experiments performed to validate and characterize the optical microscopy set-up, available in the ICM Lab, for quantifying the single cell and population fluorescence signal from *E. coli* transformants expressing GFP. To this end, measurements of the steady-state expression distributions for the TC gene circuit were compared for the flow cytometric and microscopy acquisitions.

### 3.2.1 TC gene circuit: flow cytometry and optical microscopy acquisitions

The TC gene circuit was selected as a benchmark for validating the usability of our microscopy set-up for quantifying single cell fluorescence. This choice was based on the wide range of fluorescent

signal intensities, inducible with the administration of the gratuitous inducer IPTG. The average and dispersion of the dose-response curve acquired with both instruments under homogeneous genetic and environmental conditions could therefore be used to infer their technical properties. It is worth noting that the use of fluorescence microscopy for single cell acquisitions, while being advantageous as it supports the direct integration of fluorescence intensity and morphological information, is generally more time-demanding than flow cytometry measurements. In addition, the microscopy sampled population has a lower cardinality than the one interrogated in a flow cytometry acquisition. Since small variations in imaged cells counts could introduce a bias in the experimental characterization of populations with low cardinality, the microscopy fluorescence distributions were computed over an equal number of individuals at each induction level. Specifically, the cardinality was chosen to be the number of cells in the least numerous sample (~12,000 individuals). Upon normalization by the respective average fluorescence intensity acquired at maximum induction (200  $\mu\text{M}$ ), the dose response curves obtained with the flow cytometer and the microscopy set-up showed good agreement (Figure 3.1, panel A). In particular, while the experimental mean values are almost superimposed, the linearity of the analytical relation between the fluorescence intensities acquired with the two instruments (data not shown) and the comparable standard errors prove the ability of the microscopy set-up to capture the fluorescent signal's dispersion within the cell population. Despite the closely matching results, the flow cytometer's wider dynamic range ensures a more reliable discrimination of fluorescence intensities at low expression levels, where the lower sensitivity of the microscopy set-up might be critical. An analogous observation holds when the  $CV^2$  quantified with the two instruments are considered. In this case, the limited dynamic range of the microscopy set-up is witnessed by the slight underestimation of the *squared coefficient of variation* it provides (Figure 3.1, panel B).

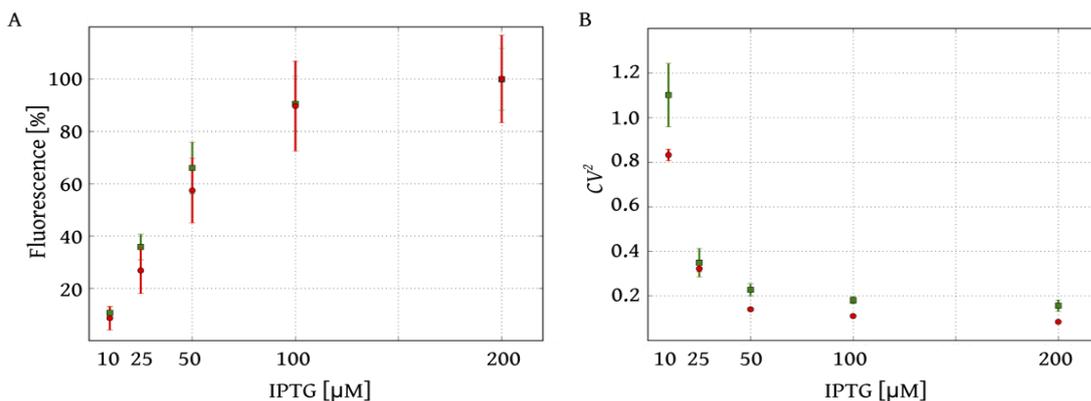


Figure 3.1: Mean expression level and *squared coefficient of variation* measured with the microscope set-up. In panel A the mean expression level extracted from the microscope fluorescence distributions (red dots) at the tested IPTG concentrations is compared with the flow cytometric equivalent (green squares). Microscope and flow cytometry average fluorescence values, shown with their standard error, are almost superimposed (Pearson correlation coefficient higher than 0.99). In panel B the decrease in  $CV^2$ , reported as mean plus/ minus standard error, for increasing induction levels is shown for microscopy (red dots) and flow cytometry (green squares) acquisitions. The underestimation in  $CV^2$  provided by the microscopy set-up, which is not consistent across IPTG concentrations, appears to be more significant for dimmer fluorescence distribution.

To further explore the usability of the characterized set-up, we investigated the impact of the number of imaged cells on the validating data. To this aim, sets of size  $n$  were sampled without replacement from the total population observed. The average fluorescence and  $CV^2$  at increasing  $n$  were computed and compared with the corresponding flow cytometric values through the evaluation of the Pearson correlation coefficient ( $R^2$ ).

In Figure 3.2 the results, indicating that a population composed by few hundreds cells is sufficient to achieve reliable and stable statistics, allow the identification of a lower threshold on the sufficient number of imaged cells. Remarkably, the assessed threshold is in-line with the reference often considered in time-series microscopy acquisitions based on microfluidic platforms (see Chapter 4).

The reported data suggests that the characterized microscopy set-up is an adequate instrument for quantifying the single cell behaviour of *E. coli* transformants.

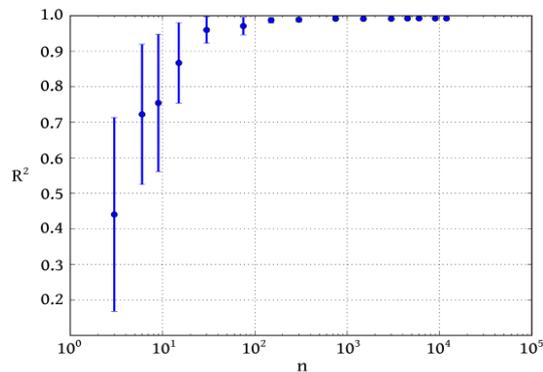


Figure 3.2: Results of the characterization of population size impact on the deviations of microscopy  $CV^2$  from the value measured by flow cytometry. The panel shows the Pearson correlation coefficient evaluated over the experimental *squared coefficients of variation* acquired with both the flow cytometer and the microscope, as a function of the number of cells composing the considered imaged sample ( $n$ ). Mean  $R^2$  values (blue dots) are reported together with the standard error. As it can be seen, the microscope fluorescence distribution corresponding to a population size of 100 cells is sufficient for achieving results in-line with the flow cytometric acquisitions. Similar conclusions were drawn when the  $R^2$  was evaluated over the mean expression levels, analyzed with the same procedure (data not shown).

We will now turn to the analysis of topology dependent steady-state variability in protein synthesis. As stated earlier, this comparison was performed using experimental fluorescence distributions acquired with the flow cytometer.

### 3.2.2 TC gene circuit

The average experimental value of the single-cell fluorescence extracted from the flow cytometric distributions increases monotonically with IPTG concentration in cells transformed with the TC gene-circuit (Figure 3.3).

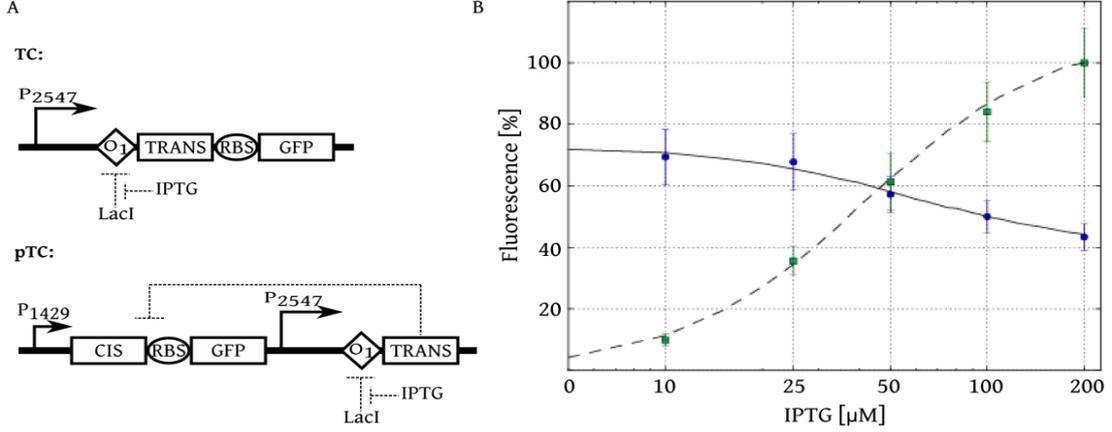


Figure 3.3: Genetic program and dose-response curves for TC and pTC gene circuits. Panel A shows a scheme of the circuits implementing transcriptional (TC) and post-transcriptional (pTC) regulation in the expression of the fluorescent reporter. Dose-response curves of the two gene circuits are shown in panel B. Mean fluorescence, extracted from the experimental distributions at different inducer concentrations, is reported for TC (green squares) and pTC (blue circles) together with the standard error. Experimental fluorescence values are normalized by the average fluorescence measured in the TC circuit at maximum induction (IPTG = 200  $\mu$ M). In the same panel, dose-response curves simulated with a mathematical model that includes cell division are shown for TC (dashed line) and pTC (continuous line) gene circuits.

The following set of reactions defines a possible model of the gene-circuit:



In reactions 3.1-3.4,  $M_{TG}$  and  $G$  are the number of mRNA and protein molecules per cell,  $\gamma_r$  and  $\gamma_p$  are respectively the degradation rates of mRNA and protein molecules,  $k_{r,TG}(IPTG)$  is the transcription

rate, and  $k_p$  is the translation rate per mRNA molecule. IPTG modifies the binding affinity of LacI repressor for the binding site  $O_1$ , thereby modifying the transcription rate of the TRANS-GFP sequence. Thus, in reactions 3.1-3.4, the transcription rate  $k_{r,TG}(IPTG)$  is the only parameter affected by the inducer concentration. Assuming that the binding-reactions between IPTG and LacI and between LacI and  $O_1$  are at equilibrium, the effect of IPTG on the transcription rate  $k_{r,TG}$  might be modelled by a Hill equation:

$$k_{r,TG}(IPTG) = k_{r,TG}^{MAX} \frac{\left(\frac{IPTG}{IPTG_{50}}\right)^n}{1 + \left(\frac{IPTG}{IPTG_{50}}\right)^n} \quad 3.5$$

In equation 3.5,  $n$  is the cooperativity coefficient,  $k_{r,TG}^{MAX}$  is the transcription rate when the blockage of LacI on transcription is completely released, and  $IPTG_{50}$  is the concentration of IPTG displacing half of the LacI molecules from the operator site  $O_1$ . The experimental dose-response curve of the gene-circuit TC can be perfectly reproduced by a model based on reactions 3.1-3.4 and equation 3.5, using  $n$  and  $IPTG_{50}$  as fitting parameters (data not shown). However, this model is not able to describe the dispersion affecting the number of proteins within an isogenic population of transformed cells. In the simple case of the mathematical model based on reactions 3.1-3.4, the steady-state value of the indexes commonly used to quantify gene expression stochasticity, i.e. the *squared coefficient of variation* ( $CV^2$ ) and the *Fano factor* ( $F$ ) [6, 36, 51], can be derived analytically using the moment generating function (see appendix for further details). These indexes can be formalized as:

$$CV^2 = \frac{\langle G^2 \rangle - \langle G \rangle^2}{\langle G \rangle^2} = \frac{\gamma_p \gamma_r}{k_{r,TG}(IPTG)} \left( \frac{1}{k_p} + \frac{1}{\gamma_p + \gamma_r} \right) \quad 3.6$$

$$F = \frac{\langle G^2 \rangle - \langle G \rangle^2}{\langle G \rangle} = 1 + \frac{\frac{k_p}{\gamma_r}}{1 + \frac{\gamma_p}{\gamma_r}}. \quad 3.7$$

The discrepancy between experimental values and model predictions emerges clearly from a comparison between equation 3.7 and experimental data. The  $F$  predicted by the mathematical model does not depend on the transcription rate constant  $k_{r,TG}(IPTG)$ , and as a consequence, it should not be modified by the concentration of IPTG. This prediction is contradicted by experimental measurements, which show a marked increase of the  $F$  upon IPTG induction (Figure 3.4, green marks).

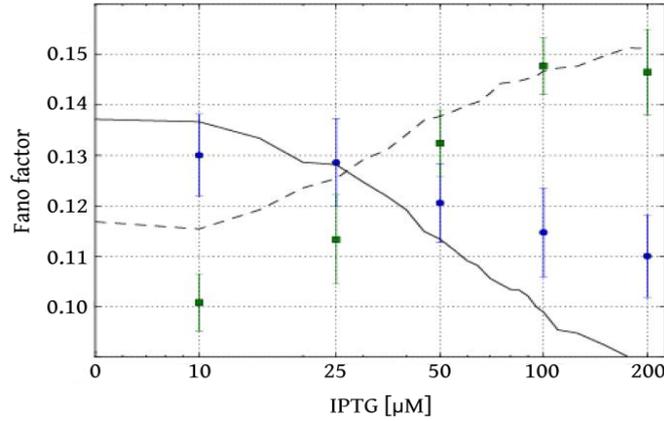


Figure 3.4: *Fano factors* in gene circuits TC and pTC. Both experimental and numerical *Fano factors* are expressed in arbitrary units. Experimental data is reported for TC (green squares) and pTC (blue circles) together with the standard error. In the same panel, numerical  $F_s$ , computed by stochastic simulations including cell division, are shown for TC (dashed line) and pTC (continuous line) gene circuit. As stated in the main text, once cell division is explicitly accounted for in the stochastic simulations, the trend in numerical  $F_s$  qualitatively reproduces the experimental one.

Cell growth and division, which is not considered in reactions 3.1-3.4, is expected to contribute to the variability of protein concentration via the random partitioning of molecules among the daughter cells [120]. Its effect was included in the mathematical model of the TC gene-circuit, combining reactions 3.1-3.4 with a reaction that describes the increase in cell volume ( $V$ ):



The model comprising reactions 3.1-3.4 and 3.8, with the transcription rate  $k_{r,TC}$  defined by equation 3.5, was simulated using the Gillespie's direct method [89]. In order to implement cell division in the Gillespie algorithm, a division event was simulated every time the cell volume reached a pre-defined upper boundary. The choice of modelling a size-based control of cell division, supported by the availability of the average growth rate ( $\alpha$ ) fitted on bulk measurements, mirrors the traditional hypothesis, in which the attainment of a critical cell size triggers replication events, leading to bacterial division [121-123]. As well as being reasonable for the expected periodicity of cell division events in exponentially growing individuals, the size-based control has been recently identified as the best candidate for a simple mathematical formalization of bacterial growth. Indeed, contrary to time-based control models, it is robust to variability in single-cell growth rate and septal ring localization [124]. At each division event, all the model variables were divided between the two

daughter cells according to a binomial distribution [35, 120, 125]. Such hypothesis holds for homogeneous cytoplasm and independent molecules segregation at cell division. Upon partition, the stochastic simulation continued tracking only one of the two daughter cells. More details about the algorithm adopted for the stochastic simulations and the procedures used to define the parameters of the mathematical model are provided in section 3.1.3. This mathematical model, with explicit treatment of cell division, successfully reproduced the average protein concentration upon IPTG induction (Figure 3.3, dashed line), as well as the experimental data on  $CV^2$  (Figure 3.5, dashed line).

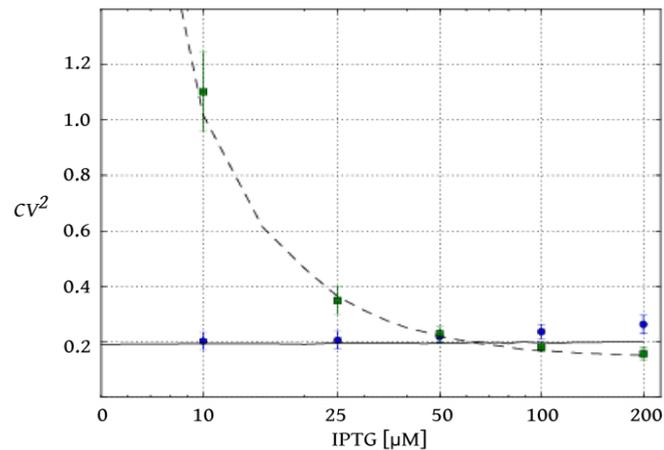


Figure 3.5: Squared coefficients of variation in gene circuits TC and pTC. Experimental data is reported for TC (green squares) and pTC (blue dots) together with the standard error. In the same panel, numerical  $CV^2$ , computed by stochastic simulations including cell division, are shown for TC (dashed line) and pTC (continuous line) gene circuit.

The importance of considering cell division in the mathematical model clearly emerges when the *Fano factor* is considered. Indeed, while in absence of cell division the  $F$  was not affected by IPTG concentration, when cell division is explicitly taken into account the *Fano factor* increases monotonically with IPTG concentration, as experimentally observed (Figure 3.4, dashed line). It is worth noting that, contrary to  $CV^2$ , the numeric value of the  $F$  depends on the measurement unit adopted for protein concentrations. When protein concentration is measured as number of molecules per cell, the  $F$  is equal to 1 if protein distribution obeys a Poisson statistics. In the present case protein concentration is measured in arbitrary fluorescence units normalized to the average fluorescence of the TC gene-circuit at maximum induction via 200  $\mu\text{M}$  IPTG. An analogous measurement unit was used for the results of stochastic simulations, i.e. protein counts are normalized to the average number of molecules in the TC gene-circuit synthesized at maximum induction.

Thanks to this definition of the measurement units, it is possible to compare the average fluorescence and  $F_s$  between experiments and simulations, even if the conversion factor between arbitrary fluorescence units of flow cytometric experiments and number of molecules is unknown. However, as a consequence of this normalization procedure, the numeric value of  $F$  does not represent the deviation from Poisson statistics. Nevertheless, the observed trend (increase in the  $F$  with IPTG concentration) is conserved regardless of the adopted measurement unit. In the case of the model with explicit treatment of cell division, the Pearson correlation coefficient between experimental and simulated  $F_s$  is above 0.99. The high correlation between simulated and experimental  $F_s$ , together with the agreement between simulations and experiments for  $CV^2$ , suggests that the mathematical model with explicit treatment of cell division correctly reproduces how the variability in protein concentration is tuned by IPTG concentration. These results alone are not sufficient to prove that cell division events are really responsible for the dependency of noise upon IPTG concentration. Other biological processes, such as interactions between mRNA and ribosomes, RNA polymerase binding events, or transcriptional bursts, could introduce noise in the system. Therefore more detailed mathematical models that include these events could reproduce the same experimental data of the TC gene-circuit. The aim of this study was to compare transcriptional and post-transcriptional control mechanisms. In this context, the model of the TC circuit represents only an effective description, which is used to evaluate the differences between the two gene circuits, and hence a simple mathematical model is certainly preferable than a more complicated one.

### 3.2.3 pTC gene circuit

In the pTC circuit, an increase in IPTG concentration induces an increase in the transcription rate of the TRANS sequence, thereby decreasing the number of GFP coding mRNA molecules competent for translation. Accordingly, the experimental single cell fluorescence decreases with increasing IPTG concentration (Figure 3.3, blue marks). The following set of reactions defines a possible model of the pTC gene-circuit:





$M_{CG}$  and  $M_T$  correspond to the number of mRNA molecules of CIS-GFP and TRANS, respectively;  $k_{r,CG}$  and  $k_{r,T}(IPTG)$  are the transcription rates of  $M_{CG}$  and  $M_T$ ; and  $\gamma_h$  the stoichiometry-dependent degradation rate of mRNA molecules due to the hybridization of the CIS-TRANS sequences [126, 127]. Coherently with the known biological features of the post-transcriptional regulation exerted by *trans*-acting sRNA, of which the TRANS sequence constitutes a synthetic analogue, the RNAs base-pairing (reaction 3.11) is formalized as an irreversible process [80, 110, 128]. Specifically, it conveys the dual role attributed to naturally occurring sRNA: inhibition of translation and enhanced degradation of both the coding transcript (CIS-GFP) and the synthetic sRNA (TRANS) [96, 97]. Indeed, upon hybridization, the biochemical nature of the CIS and TRANS sequences prevents ribosome docking to the RBS, thereby reducing the pool of coding mRNA translated. At the same time, the RNAs base-pairing leads to degradation of the interacting transcripts. The remaining symbols in reactions 3.8-3.14 have the same meaning adopted in reactions 3.1-3.4. Cell division was included in the model of the pTC circuit using the same method adopted for the TC circuit. As in the TC circuit, the only parameter affected by IPTG concentration is the transcription rate of the TRANS sequence, which was modelled by an equation analogous to equation 3.5. All the parameters of this model were based on the ones adopted for the TC gene-circuit, as described in the Methods section. The only exceptions were  $IPTG_{50}$  and  $\gamma_h$  that were used to fit the experimental data on the average fluorescence per cell. The usage of a different  $IPTG_{50}$  value in the TC and pTC gene-circuits is justified by the fact that different DNA backgrounds might modify LacI binding affinity for the operator site [5].

The model based on reactions 3.8-3.14 correctly reproduces the average fluorescence upon IPTG induction (Figure 3.3, continuous line). The experimental data on  $CV^2$  are almost constant at the different IPTG concentrations (Figure 3.5, blue marks), and the  $F$  exhibits a moderate decrease upon IPTG induction (Figure 3.4, blue circles). Although stochastic simulations underestimated protein variability at IPTG concentrations above 100  $\mu$ M, the mathematical model reasonably reproduces the experimentally observed behaviours, i.e. increase in  $F$  and constant  $CV^2$ . Experimental and simulated  $F$ s were strongly correlated as observed for the TC circuit (Pearson correlation coefficient equal to 0.97).

These results suggest that the mathematical models with cell division correctly reproduce how the  $F$  and the  $CV^2$  are tuned by IPTG concentration in both gene circuits.

### 3.2.4 Differences between TC and pTC gene circuits

A meaningful comparison between the protein variance observed in the TC and pTC gene circuits should be performed at the same average concentration of the fluorescent reporter. However, exerting a direct control on the average protein concentration by means of an experimental set-up is not trivial. Thus, in order to facilitate the comparison between experimental data of TC and pTC, the dependency of protein variance on the mean expression level was fitted by a power law in both gene-circuits (Figure 3.6, panel A) [117, 129]. The experimental value of  $CV^2$  in the pTC gene circuit is identical to the one estimated for the TC gene circuit when fluorescence values are high, corresponding to IPTG concentrations below 50  $\mu\text{M}$  (Figure 3.6, arrow). Under this condition, TRANS sequences are transcribed at a minimal rate, the effects of the post-transcriptional control are minimized, and therefore the difference between TC and pTC circuits is minimal. Instead, when the average protein concentration decreases, the difference in protein variance between the TC and pTC gene circuits increases. At higher IPTG concentrations, the TRANS sequence exerts its post-transcriptional control on gene expression, leading to a reduction in protein variability. An intuitive interpretation of this process implies that as the concentration of TRANS molecules increases, GFP translation becomes less efficient because of sequestration of coding mRNA molecules by the CIS-TRANS hybridization process, which makes protein production less noisy. It is important to remark that the noise of the pTC gene circuit never exceeds the level measured for the TC circuit at the same average protein concentration. These results are not in line with previous experimental observations about the iron homeostasis system in *E. coli*, where transcription of the sRNA RyhB was associated with a noisier expression of the main target gene *sodB* [114]. It is worth noting that a feature shared between the two experimental systems is their inability of reaching the high-silenced regime, where post-transcriptional control is expected to exhibit a remarkably lower noise than the transcriptional one. Indeed, the above iron homeostasis system was tested within a range of iron deprivation conditions that do not impair physiological growth, while the GFP expression level of our pTC circuit at maximum induction is only about 40% of the reference value. The observed discrepancy might be attributed to differences in transcription between the two experimental systems. Previous theoretical analyses showed that in the presence of transcriptional bursts, the noise is generally higher for a post-transcriptional control mechanism than for a transcriptional one, and that only at very low protein concentrations, as reached in the silenced-regime, the post-transcriptional control mechanism becomes less noisy than the transcriptional one [110]. On the other hand, when transcriptional bursts are absent, the difference at high protein concentration between the two mechanisms is minimal, and the transition to the silenced – less noisy - regime takes place at higher protein concentrations. Therefore, the difference between the iron homeostasis network and the pTC/TC gene circuits could be attributed to an absence or a limited extent of transcriptional bursts

in these latter synthetic circuits. Indeed, in agreement with our wetlab results, stochastic simulations based on a mathematical model that did not include any description of transcriptional bursts, see equations 3.8-3.14, qualitatively reproduce the experimental behaviours (Figure 3.6, A vs B).

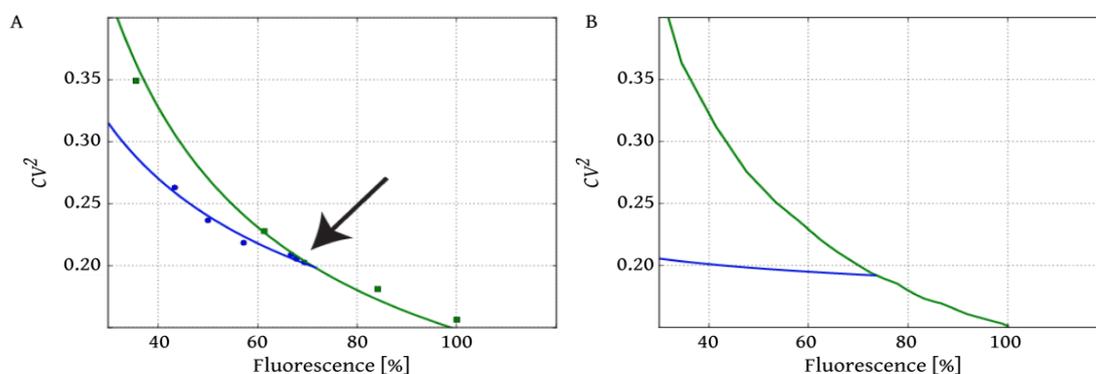


Figure 3.6: *Squared coefficients of variation as a function of the average expression level.* In panel A the experimental  $CV^2$  are shown for TC (green squares) and pTC (blue circles) gene circuits. Regression curves of the experimental data fitted by a power-law are reported adopting the same colour code. The arrow points at an equivalent value of  $CV^2$  for TC (green line) and pTC (blue line) reached when high fluorescence values, corresponding to IPTG concentrations below 50  $\mu\text{M}$ , are considered. In panel B the numerical  $CV^2$ , computed by stochastic simulations, are reported for the TC (green line) and pTC (blue line) gene circuits.

### 3.3 Conclusions

Extensive research effort has been devoted to investigate the interplay between gene network architecture and stochasticity in gene expression, under the dual perspective of shedding light into how naturally occurring pathway evolved to counteract or amplify noise amplitude, and defining design requirements for synthetic gene circuits with reliable function. In this study, we compared phenotypic noise within an isogenic population of cells transformed with elementary synthetic circuits implementing either a transcriptional or a post-transcriptional control of gene expression. The circuits were designed in order to compare transcriptional and post-transcriptional control mechanisms in two circuits that were as similar as possible, i.e. the mechanism adopted for modulating gene-expression is shared between the two circuits. The main result of this study highlights that noise is lower for the gene-circuit with post-transcriptional control, and that the difference in noise between the two circuits increases when the post-transcriptional control on gene-expression is more efficient. The experimental data on protein variability provides a measure of the entire noise affecting protein synthesis, i.e. the sum of intrinsic and extrinsic noise. Therefore, we are not able to exclude that the difference between the two gene circuits is related to changes in extrinsic noise. However, the circuits were compared using identical experimental protocols, and the circuits themselves are similar. It is therefore unlikely that induction by IPTG has a different effect on the extrinsic noise in the two gene-circuits. In our opinion, it is plausible that IPTG modulates the intrinsic noise by the post-transcriptional control mechanism, which acts on translation efficiency. This hypothesis is supported by the results of stochastic simulations, in which IPTG has only intrinsic effects, acting on the transcription of GFP gene or the TRANS sequence. It is interesting to note that a similar reduction of noise related to post-transcriptional control has recently been observed in eukaryotic cells and related to changes in intrinsic noise [71]. The data presented in this study supports the hypothesis that post-transcriptional control might be used to decrease the noise on protein expression. However, since in other systems post-transcriptional control has been shown to increase noise [114], or to act on the correlation among different genes [113], post-transcriptional control might represent a common strategy used both by prokaryotes and eukaryotes to tune protein concentration variability. The pTC circuit analysed in this study represents a possible strategy to modulate noise by means of external signals. This possibility could be useful to develop novel applications in synthetic biology, and to examine the effects of noise on cell behaviours.



## **4 Investigation of *GAL10*-lncRNA effects on *GAL1* transcriptional activation using fluorescence microscopy and microfluidics**

This chapter outlines the research activity conducted during a six months period staying at the “Centre for Synthetic and System Biology” (SynthSys) of the University of Edinburgh, under the supervision of Professor Peter Swain.

The implemented study, belonging to Elco Bakker’s research activity, represents an extension of my doctoral project on several aspects.

First, focusing on phenotypic variability in a eukaryotic organism, it allowed the analysis of a biological model denoted by increased complexity, where cellular processes result from the interaction of molecular players operating in distinct cell compartments.

Second, as single cell fluorescence measurements, aimed at tracking gene-specific induction, were performed in time-lapse microscopy adopting a microfluidic system, they allowed the assessment of cellular behaviour over a wide temporal window and exposure to controlled environmental conditions.

Finally the project aimed to analyse the regulatory function of an antisense, non-coding transcript on the expression of an inducible gene assessed, on a single cell basis, by quantitative fluorescence microscopy and in a microfluidic device. In this, the focus of the study reasonably represents a natural conjugate of the mRNA interference regulation implemented in the synthetic system described in the previous chapters.

## 4.1 The GAL Network in *Saccharomyces cerevisiae*

*Saccharomyces cerevisiae* has been widely investigated in scientific research, being an organism exhibiting fast growth in laboratory conditions and genetic tractability. Beyond its simple use, the baker's yeast retains the biological complexity inherent to eukaryotic cells, and its proteins, as well as metabolic/regulatory pathways [130](e.g. the *Leloir* pathway [131]), prove to be highly conserved in mammals. These characteristics designate it as candidate model to unravel complex biological processes and the evolutionary optimization of gene networks underpinning them [132]. Among those, the galactose network (*GAL* network)– encoding *S. cerevisiae* ability to prioritize the carbon source to metabolize and achieve optimal growth in varied environmental conditions – stands as a model for decision making and gene regulation [132, 133].

*S. cerevisiae* is able to retrieve carbon, and therefore energy, from both sugars and unconventional substrates. Among sugars, the preference for glucose metabolism, when compared to the consumption of its epimer galactose, is due to the greater energy cost associated with breaking-down the latter. In addition to a halved stoichiometric ratio of the high energy compound adenosine triphosphate (ATP) synthesized for each sugar molecule, galactose metabolism requires the expression of a minimum of eight specific genes belonging to the *GAL* network. This, together with the additional observation that under fully induced conditions the amount of *GAL* enzymes is estimated to constitute about 5% of the overall cell protein content [131], justifies the tight down-regulation on *GAL* network activation and the associated metabolism reprogramming in glucose.

The genes constituting the *GAL* network encode for regulatory and catalytic proteins. The first cluster includes proteins responsible for cellular internalization (*Gal2*), sensing of cytosolic galactose concentration (*Gal3*) and coherent regulation of the transcription of the other structural genes in the pathway (*GAL4*, the activator, and *GAL80*, the repressor). The remaining molecular players (*Gal1*, *Gal5*, *Gal7* and *Gal10*) constitute the *Leloir* pathway [133], which converts  $\beta$ -D-galactose into glucose-6-phosphate which then enter glycolysis.

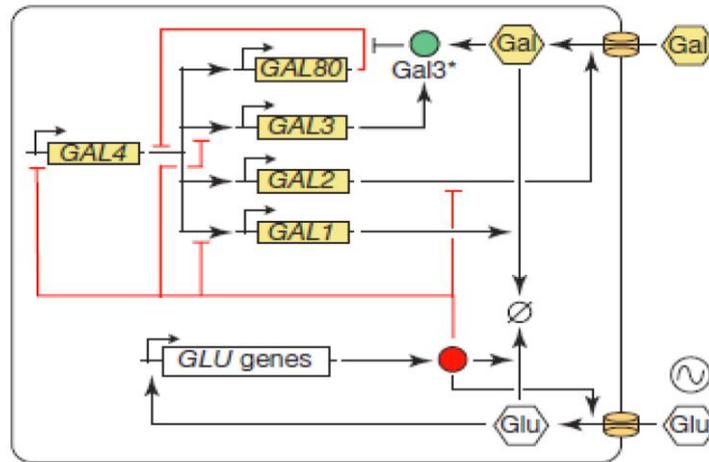


Figure 4.1: Functional scheme of the nested feedback loops characterizing the GAL network and the major interactions between the proteins encoded in it. Red and black are adopted to represent repression and induction. Reproduced with permission from Bennett et al. [5].

Due to the presence of multiple feedback loops, the network can be found in three different states: uninduced, induced and repressed, depending on the environmentally predominant nutrient. When the cell experiences glycerol or raffinose, carbon sources that are not observed to induce or repress the network, the transcriptional activator Gal4 binds as a homodimer the instances of upstream activating sequences (UAS) located in the promoter region of the other GAL genes [134]. However, the simultaneous presence of the repressor Gal80, bound as a dimer to Gal4 activation domain, prevents the pathway induction. Galactose is imported in the cytoplasm by facilitated diffusion through glucose hexose transporters (HXT) or the galactose permease Gal2. Cytoplasmic galactose, in association with ATP, causes the activation of the ligand sensor Gal3. Its subsequent interaction with Gal80 releases the repression on Gal4, leading to the switch in transcriptional program as identified by the appearance of GAL mRNAs within minutes [135].

The repressed state is associated with the availability of environmental glucose. This preferred monosaccharides downregulates galactose pathway through multiple mechanisms: it lowers cellular galactose levels, both reducing the fraction of Gal2 and repressing transcription of the inducer Gal3, it promotes Mig1 shuttling to the nucleus where it represses transcription of GAL4 and it hastens degradation of GAL1 and GAL3 transcripts [136]. As shown by Johnston et al. [137], the time scale of these mechanisms implements a tight control on the GAL network and their synergy accounts for the 1000 fold repression of Gal1, assumed as an indicator of the overall induction level. It is worth noting that, while the GAL utilization network is supposed to be activated when depletion brings glucose below a predefined threshold, Escalante-Chong et al. [138] recently showed that, in analogy with observations in bacteria [139, 140], induction of GAL genes in budding yeast is dependent on the ratio of sensed glucose and galactose. The ratio sensor behaviour was not limited to laboratory strains, despite the dependency of induced fraction on ratio value and genetic context, and was found to

break down and be replaced by threshold galactose sensor when extracellular glucose concentration fell below 0.006%. These observations were consistent with a location of the ratio sensing upstream of the *GAL* pathway and its attribution to a competition of extracellular sugars for the hexose transporters.

#### 4.1.1 The *GAL* gene cluster

The *GAL* gene cluster is a group of three genes, placed in proximity to the centromeric region of chromosome II [132], extensively studied over the past thirty years to identify the timing of biochemical events leading to nutrients dependent induction of genes. More recently, the discovery of non-coding RNAs transcribed from this locus suggested its use as a tool for achieving new insights into the functions performed by these transcripts within euchromatic regions.

The *GAL* cluster encodes three genes belonging to the *Leloir* pathway: *GAL7*, *GAL10* and *GAL1*. *GAL1* and *GAL10* constitute a bidirectional gene pair, since the two nucleotidic sequences lie, with opposite polarization, on complementary DNA strands and their transcription is driven by the shared bidirectional promoter *GAL1-10*.

Gal7, Galactose-1-phosphate uridyl transferase, acts as a dimer in the *Leloir* pathway, catalysing the reaction of UDP-D-glucose and  $\alpha$ -D-galactose-1-phosphate in glucose-1-phosphate and UDP-galactose [141, 142]. Even though it is defined as a non-essential gene, null mutants prove unable to grow in galactose only containing media and their growth rate is reduced when this sugar is present. This observation is consistent with the accumulation of galactose-1-phosphate and the toxic effect related to its high levels [143].

Gal10, as a mutarotase, catalyses the first step of the reactions chain responsible for the conversion of  $\beta$ -D-galactose in glucose-6-phosphate. In addition, this protein acts at a later stage of the pathway as an epimerase, leading to the availability of UDP-glucose. Like Gal7, its deletion prevents yeast growth in galactose [144].

In 2001 Ideker et al. [145], adopting the *GAL* network as a model to test a methodological approach of integration between experimental data (DNA microarrays and quantitative proteomics) and numerical simulations to extract information on pathway behaviour, highlighted a global decrease in *GAL* enzymes levels, upon induction, in *GAL7* $\Delta$  and *GAL10* $\Delta$  strains. Either the detrimental accumulation in galactose-1-phosphate or a related metabolic derivative [143], or a transcriptional interference mechanism impairing *GAL7* and *GAL1* expression at the *GAL1-10* gene locus could explain the observation. Subsequently, the analogy in expression levels between a double mutant strain *GAL1* $\Delta$ *GAL10* $\Delta$ , in which the *Leloir* pathway was corrupted upstream of Gal7 function, and *GAL1* $\Delta$

supported the idea of a cellular repression of *GAL* network activity, aimed at constraining the accumulation of harmful metabolites.

Gal1 is a monomer phosphorylating  $\alpha$ -D-galactose. In spite of a high level identity (~70%, ~90% similarity) in amino acids sequence with Gal3, which supports the recovery of a Gal3's enzymatic activity upon insertion of two residues at position 164 [146], the deletion of this Leloir enzyme impedes cellular growth when galactose is the only available carbon source.

Gal1 and Gal3 are paralogues resulting from asymmetric evolution, subsequent to a genomic duplication event, of a common ancestor which implemented both an enzymatic and signal transduction function. Their evolutionary asymmetry is reflected both by their regulatory regions, namely the number of UAS in the promoter regions, and actual functions. Indeed, while Gal3 acts exclusively as a ligand-sensor, Gal1's enzymatic role is complemented by the capacity of performing as a weak inducer of the *GAL* network [147]. This ability has been highlighted both in the presence and absence of Gal3. While exploring the dynamic and cellular localization of the interaction between Gal4, Gal80 and Gal3, Reece and co-workers [148] identified in fluorescent images, and later confirmed using Chromatin Immunoprecipitation assay (ChIP), the initial formation of a tripartite complex involving Gal4-Gal80-Gal3. This was followed by Gal4-Gal80-Gal1, when cells were grown in galactose containing media for more than two hours. The observation suggested the partial replacement of Gal3 by Gal1 while the metabolic reprogramming proceeds, an idea supported by the evaluation of the different expression levels following induction (3x for Gal3, 1000x for Gal1 [149]). In *GAL3* mutants (*GAL3* $\Delta$ ), Gal1 can induce the network albeit over a time-scale of 48-72 hours rather than few hours. This time difference was probably due to its reduced binding affinity for the transcriptional repressor and lower basal expression level. The phenomenon, termed long term adaptation (LTA), has been linked with phenotypic heterogeneity observed in a population of *GAL3* $\Delta$  mutants grown in neither non-inducing nor repressive conditions, upon switch to galactose media. The bias in cellular response to galactose was ascribed to differential levels in Gal4 [150].

Considering the dependency of *GAL1* expression on active Gal3, the mechanism that should ensure the enzymatic activity of Gal1 in absence of the ligand sensor has not been clarified yet [151, 152].

Being the most expressed among the *GAL* genes, Gal1 represents an appealing reporter of the overall induction level. Furthermore, Gal1's high expression, together with its stability, is a determinant in identifying it as a cytoplasmic inheritance factor facilitating transcriptional memory. This term, in general, defines the cell's ability to remember a past environment to which it was exposed. So far, two different types of transcriptional memory have been distinguished: reinduction and persistent memory. Reinduction memory arises as a faster and more homogeneous induction pattern when cells, after an initial exposure to galactose followed by a period (<12 hours, corresponding to 6-7 cell divisions) of repressive growth, face again the activating sugar. In particular, the first induction is

denoted by a three hours lag and requires up to eight hours for Gal1 to reach its steady state. In this condition the fraction of OFF cells - namely cells with a low level of Gal1 - decreases with time, leading to a unimodal distribution by twelve hours. In the second induction *GAL1* expression occurs at a uniform rate across the cell population and without lag time. In both cases the same steady-state expression level is reached. Heterokaryon experiments, aimed at identifying the mechanisms underlying the memory phenomenon, revealed that the progeny inheritance of Gal1, acting as an inducer despite being diluted at each mitotic event, ensure the faster kinetics of the second induction [153]. The relevance of other plausible candidates, such as the inheritance of an active chromatin state due to the activity of SWI/SNF complex, the tethering of *GAL* genes at the nuclear periphery after transcription or the influence of the ligand-sensor Gal3 has been constrained to the initial phase of glucose repression, when Gal1's role is less relevant [150]. The relevance of different molecular players depending on the duration of repressive growth has led to the additional distinction of short and long term memory.

The experiments performed to exploit mechanisms behind transcriptional memory, and their time of activation, focused on cells grown in media containing a single sugar (either glucose or galactose) at high concentrations (typically 2% w/v). The bimodal induction pattern frequently observed when cells sense a weakly inducing environment (high galactose concentration combined with glucose traces or low galactose titration) is a touchstone of persistent memory. Used to indicate the maintenance of memory pertaining a previous environment, it was first documented by Biggar et al. [154, 155] who observed the distinction between ON and OFF cells for up to 14 hours after exposure to fully inducing and partially repressing media, following growth in glucose or raffinose. The carbon source of initial exposure determined the pattern of induction: glucose grown cells display a bimodal *GAL1* promoter activity, while raffinose produces a homogeneous response with steady-state Gal1 levels dependent on the concentration of glucose in the second media. In contrast, cells grown in raffinose and switched to a condition of low galactose concentration remained bimodal for up to 27 hours. In an analogous experiment, the historical exposure to galactose produced a unimodal distribution. So far, the mechanisms underlying persistent memory have been related to the bistability of the *GAL* network.

#### **4.1.2 *GAL10*-lncRNA: evidences from scientific literature**

As mentioned above, the identification of antisense long non-coding RNAs within the *GAL1-10* gene cluster has recently attracted scientific interest, providing a model for unveiling the regulatory functions of these motifs on the expression of inducible genes. Long non-coding RNAs (lncRNAs), mainly identified in the last decade following the advent of genomic approaches, are molecules



0.001-0.1% galactose. The lowest galactose concentration able to induce a detectable activation of *GAL1* (0.01%), when combined with 0.02% glucose, produced a highly different behaviour and was hence selected for analysing the dynamics of activation. This showed different kinetics over 6 hours of induction, establishing a repressive role of *GAL10*-lncRNA in limited sugar environments. To test the mode of action of the lncRNA and discriminate between *cis* and *trans*, the authors cloned heterozygous diploids having a wild-type *GAL1-10* locus in the  $\alpha$  allele and either a wild-type *GAL1-10* locus or *Reb1BS* $\Delta$  *GAL1-10* locus in the  $a$  allele. Methylation and acetylation patterns (H3 K36 trimethylation and H3 K27 acetylation) associated with the *GAL10*-lncRNA proved to behave independently in each allele, suggesting a *cis* activity of the antisense transcript. This idea was moreover supported by the low transcriptional frequency of *GAL10*-lncRNA, estimated in 1 nascent transcript every 50 min.

The altered methylation and acetylation profiles, together with a *cis* activity, lead the authors search for histone deacetylases (HDACs) underlying the repressive action of *GAL10*-lncRNA. Among HDACs, the Rpd3S complex acts at a late stage of transcription, recognizing H3 K36me3 deposited by Set2 at the 3' end of active genes and catalysing H3 and H4 deacetylation. This is Eaf3 dependent and prevents an incorrect displacement of nucleosomes at the 3' end of genes, ensuring the dormancy of cryptic promoters populating intergenic regions. The measurements of *GAL1* mRNA levels 3 hours after induction in wild-type and *Reb1BS*-silent mutant with selective HDAC knockouts indicated that, in *Eaf4* mutants, de-repression of *GAL1* was not enhanced in the absence of *GAL10*-lncRNA. Other than revealing a concomitant repressive action of the lncRNA and the Rpd3S complex, this evidence supported a model in which the act of transcription of the *GAL10*-lncRNA causes H3 K36me3, thereby leading to the recruitment of the repressive Rpd3S complex.

In 2009, Pinskaya et al. [4] analysed the effect of histone modifying enzymes, particularly Set1 and Set2, on the kinetics of *GAL1* activation. Measuring, both in laboratory and naturally occurring strains, the concentration of *GAL1* mRNA after 1 hour of induction in 2% galactose, the authors observed that *Set1* $\Delta$ , differently from *Set2* $\Delta$ , significantly increased induction when compared to wild-type cells. This supported a role for H3 K4 methylation in *GAL1* repression, as confirmed by an increased activation of this gene detected in strains carrying H3 K4A mutation. Reasoning on the possible mechanisms behind Set1 repression, Pinskaya and co-workers performed measurement of mRNA stability and RNA pol II occupancy at *GAL1* locus, revealing that the effect was limited to the initiation of transcription. In attempt to evaluate a regulatory role of Set1 on *GAL1* transcription through RNA-dependent mechanisms, the authors probed total RNA in strains defective for either the cytoplasmic 5'-3' decay (*xrn1* $\Delta$ ) or the nuclear surveillance 3'-5' pathway (*trf4* $\Delta$ ). They hence identified the three transcripts independently observed by Houseley et al., and proved their transcription dependency on Reb1 protein by means of a *Reb1-1* thermosensitive degron strain. The

stabilization of *GAL10*-lncRNA in *xrn1Δ* and *trf4Δ* mutants seemed to have no effects on *GAL1* activation. Furthermore, the similarities in the assessed methylation profile in *Reb1-1* and *Set1Δ* strains proved their synergistic role in altering H3 K4 methylation pattern.

The role of *Dcp2* and *Xrn1* in the regulation of *GAL1-10* gene cluster was assessed via deletion of the genes encoding these enzymes, and hence stabilization of *Gal10*-lncRNAs, by Geisler et al. [159]. Analysing *Dcp2Δ* and *Xrn1Δ* mutants upon switch from 2% raffinose to 2% galactose, the authors observed increased levels of *GAL10*-lncRNA and concomitant delay of *GAL1* induction. The delayed activation proved to be consistent with an active chromatin state observed over the whole locus in the first 30 min of induction. This was identified by the spreading of H3 K18 acetylation, in *Dcp2Δ* strain.

Cloutier et al. [2] adopted Houseley's strains to investigate the effects of *GAL10*-lncRNAs stabilization – through impairment of two decay pathway (*Xrn1Δ* and *Dcp2Δ*) and deletion of RNA helicase encoding gene *Dbp2* – on the induction of the genes within the *GAL1-10* cluster. As in Houseley et al., cells cultured overnight in YEP media containing either 2% glucose or 2% raffinose were induced with 2% galactose, once they reached an optical density ( $OD_{600}$ ) equal to 0.4. After three hours, the authors measured *GAL1* mRNA levels and the lag time – time from induction required for a detectable signal of *GAL1* activation. *Dbp2Δ* strain showed significantly reduced lag times compared to wild-type cells, when induced from glucose. The lag time difference was abrogated in *Dbp2Δ/Reb1BSΔ*. Comparable changes were observed in *Dcp2Δ* and *Xrn1Δ* strains, though in the first case a variation in steady state *GAL* mRNAs levels after 5 hours of induction was also detected. Once an acquisition with higher density time-points was adopted, a statistically significant difference in *GAL10* and *GAL7* lag times was revealed between wild-type and *Reb1BSΔ* strains upon switch from 2% glucose to 2% galactose. Using CHIP, the mechanism of action of *GAL10*-lncRNA was attributed to the displacement of *Cyc8* repressive complex, which was observed over the promoter and 5' end of the *GAL* genes in *Dcp2Δ* and *Xrn1Δ* strains under repressive conditions. Additionally, the authors investigated the effect of *Dcp2* on *GAL1* induction when *GAL10*-lncRNA was knocked out. In this case, *GAL10* was deleted and the cells transfected with a plasmid harbouring a *GAL10* gene deprived of *Reb1* binding sites. Inducing these strains from raffinose, Geisler et al. evidenced that *GAL10*-lncRNA deletion counteracts the effect due to *Dcp2* knockout, hypothesizing a dependency of *Dcp2* repression on the antisense transcript. The authors justified the surprising observation of the modulation of *GAL10*-lncRNA's repressive effect by the degradation machinery by proposing two explanatory models: a repression mediated by R loop formation at the locus or an attenuation of *GAL10*-lncRNA induced repression by *Dcp2* and *Xrn1*.

The scientific literature review presented so far highlights discrepant opinions concerning the nutrients regime under which *GAL10*-lncRNA transcription occurs and the effect of the antisense transcript on *GAL* genes expression. In particular, Houseley et al. observed a significant repressive effect of *GAL10*-lncRNA on *GAL1* activation only when cells faced mixes of low concentrations nutrients. By contrast, other groups saw an impact of the antisense transcript even in standard laboratory conditions. In the experiments in which Pinskaya et al. [4] observed the strong repressive effect after induction with 2% galactose, a much larger genetic perturbation was introduced than that underlying the experiments of Houseley et al. This difference might justify the discrepancies in conclusions drawn by the two groups. Similar reasoning applies to the methodology adopted by Geisler et al. [159], with the additional note that the comparison between wild-type cells and *GAL10*-lncRNA mutants upon induction, despite not stressed by the authors, appears not significant.

The ascription, by Cloutier et al. [2], of a *GAL10*-lncRNA activating role upon induction from repressive conditions appears in conflict with the discussed published results. Coherence may be found analysing the same strains under both experimental conditions. Indeed, while Houseley et al. and Geisler et al. drew their conclusions from *GAL1* acquiring steady state mRNA levels in glucose and 40 minutes time-series for 2% galactose induction from 2% raffinose, Cloutier et al. opted for an activating effect inferred on high density time-series data of *GAL10* and *GAL7* transcripts measured over 5 hours on *Reb1BS* $\Delta$  strains induced from glucose to galactose.

All the published results shared the assessment of mRNA levels encoding *GAL* gene cluster on batch cultures. Inspired by the idea that dissection of single cell behaviour through microfluidics may convey deeper understanding of *GAL10*-lncRNA mode of action, eventually reconciling the aforementioned observations, we requested the strains characterized in Houseley et al.: wild-type and *Reb1BS*-silent mutants, hereafter referred to as *GAL10*-lncRNA $\Delta$  with abuse of notation, with *GAL1*-GFP fusion. Though the use of a reporter protein for assessing *GAL10*-lncRNA-dependent *GAL1* transcriptional activation might be technically challenging, the microfluidics based analyses conveys the advantage of a potential correlation between cellular fluorescence and morphological traits.

As a proof of concept, we first aspired to collect results in line with those of Houseley et al.

## 4.2 *GAL10*-lncRNA effect on *GAL1*-GFP activation: single cell microscopy time-series data in microfluidic device

In attempt to replicate Houseley et al. results, we retrieved experimental conditions analogous to theirs: wild-type and *GAL10*-lncRNA $\Delta$  cells were cultured overnight in YEP media with 2% raffinose, diluted in the morning in fresh pre-warmed media to an OD<sub>600</sub> of 0.05, and loaded into the microfluidic device at the beginning of the log-mid phase growth (OD<sub>600</sub> ~ 0.2). The adopted device was a modified version of ALCATRAS - A Long-term Culturing And TRapping System [160] - which, consisting of three chambers, allows the simultaneous monitoring of up to three strains subjected to the same media conditions. Once loaded, cells experienced YEP with 2% raffinose, 0.01% galactose and 0.02% glucose over the 20 hours of imaging. While the loading protocol, set by the device structure, impeded the precise identification of the instant at which cells first sensed the partially inducing media, efforts were taken to constrain it to the 20 minutes preceding the acquisition start.

The results, shown in Figure 4.3, are in qualitative agreement with those published by Houseley et al., reproducing a higher fluorescence for the *GAL10*-lncRNA $\Delta$  strain over the first 5 hours of induction. Considering the constant supply of fresh media during the experiment, the comparable fluorescent levels observed in the following 4 hours could not be attributed to the authors paved hypothesis of a decrease in batch cultures cellular growth rate at high OD<sub>600</sub>, where the nutrients content might become a limiting factor. Finally, the data revealed the appearance of a tardive difference in *GAL1* expression among the tested strains, which could reflect the *GAL* pathway activation. In that case, the unusually long lag time could be ascribed to both the inability to detect faint initial differences in fluorescent signal, due to the high autofluorescence of YEP media, and the low galactose concentration, whose activating effect is moreover dampen by the presence of glucose traces.

The late fluorescent boost was never observed in subsequent experiments, although it might reflect a cellular behaviour liable to investigation with acquisition of extended duration. Such experiments present technical challenges in ensuring cell fitness.

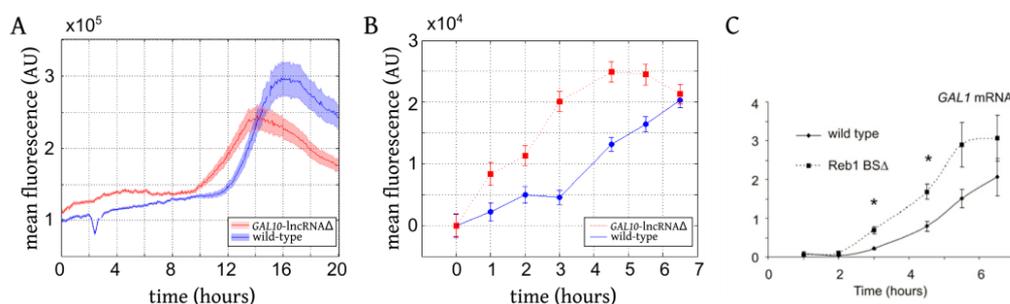


Figure 4.3: Results of wild-type and *GAL10-lncRNAΔ* cells, induced in accordance with Houseley et al. experimental conditions (SC with 2% raffinose, 0.01% galactose and 0.02 %) using the microfluidic device. The results refer to the 433 wild-type and 488 *GAL10-lncRNAΔ* cells that remained present for more than 220 time points. In panel A the mean fluorescence over the whole acquisition is shown for wild-type (blue) and *GAL10-lncRNAΔ* (red) population, with shaded area representing the standard error on the mean. The sharp depression in wild-type mean fluorescence 2 hours after the beginning of the acquisition is due to a loss of focus. As detailed in the main text, the comparison of populations behaviour qualitatively indicate a faster and stronger activation for *GAL10-lncRNAΔ* strain. Ten hours after the beginning of induction the high fluorescence boost, more prominent for wild-type cells, is observed. Panel B shows experimental data (mean  $\pm$  standard error) extracted, for both strains, at the same time-points at which RNA samples were harvested in Houseley et al. paper. In this case the mean fluorescence value at the first time point was subtracted from all data. The comparison of panels B and C, where Houseley et al. [3] results are reproduced with permission, denotes a general agreement in the two set of experiments.

Although the results of the above experiment were in line with our expectations, subsequent repeats revealed poor reproducibility, providing data set in which *GAL1* expression levels of the analysed strains were hardly distinguishable. We reasoned that the originally revealed qualitative agreement indicated our proximity to a nutrients environment in which strain-dependent differences in *GAL1* expression could be detected.

We therefore employed flow cytometry analysis to screen transformants behaviour when exposed to slightly different mixes of sugars in Synthetic Complete (SC) media, usually selected for the preparation of liquid yeast cultures committed to fluorescent measurements because of its reduced autofluorescence compared with other richer media, such as XY or YPD.

We grew overnight wild-type cells and *GAL10-lncRNAΔ* mutants in SC media with 2% raffinose (30°C, 200rpm). In the morning, cell cultures were diluted to an  $OD_{600}$  of 0.05 in fresh pre-warmed raffinose media and further incubated until they reached an  $OD_{600}$  of 0.2, at the beginning of log phase growth. Samples harvested from these were hence centrifuged and the pellet resuspended in an equal volume of SC containing 2% raffinose, 0.02% glucose and galactose concentrations equal to 0.01%, 0.02%, 0.04%, 0.06%, 0.08%, 0.1%. Two hours later, samples were assayed at the flow cytometer to quickly measure the fluorescent distribution ( $\sim 2 \times 10^4$  fluorescent events) of the two transformants. Each sample was acquired twice, and, given that the cells were not fixed, the order adopted for

samples test was reversed among the two runs to compensate for distortions in results due to temporal bias during the transient activation.

Among the assayed conditions, 0.04% and 0.06% galactose showed a robust and statistically significant difference in *GAL1* expression between wild-type and *GAL10-lncRNAΔ* strains. The first one (Figure 4.4), being closer to the 0.01% galactose originally adopted by Houseley et al. [3], was selected for the subsequent experiments.

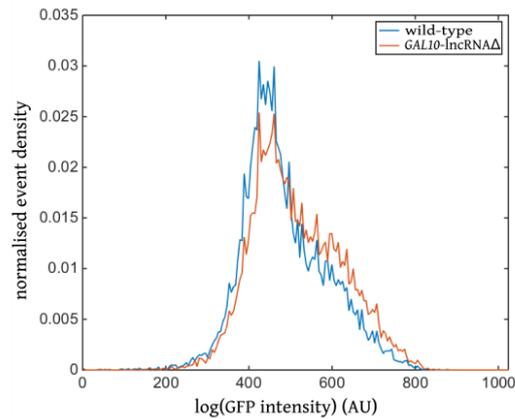


Figure 4.4: Comparison of fluorescent histograms plot of wild-type (blue line) and *GAL10-lncRNAΔ* (red line) cells exposed to SC containing 2% raffinose, 0.02% glucose and 0.04% galactose. On the y axes the number of fluorescent events is normalized to the whole set of acquired ones. The histograms were found to be statistically different ( $p\text{-value} < 10^{-5}$ ) by the two samples KS test ( $\alpha = 5\%$ ). Among the tested inducing media providing robust and discernible expression, this one was selected as having the galactose concentration closest to the one adopted in Houseley et al.

### 4.3 Induction of wild-type and *GAL10-lncRNA* $\Delta$ strains with SC 2% raffinose, 0.02% glucose and 0.04% galactose

The newly identified sugar concentrations characterized the inducing media in further microfluidics investigations of wild-type and *GAL10-lncRNA* $\Delta$  strains. Being aware of the gap that the modified inducing media would have introduced with respect to our literature reference, we adjusted the experimental protocol to include the use of bovine serum albumin (BSA) and cyanine 5 (Cy5). BSA, added in the media filling device and in the inducing one, exerts a lubricating action which facilitates the loading of cells into the device and the daughter cells removal at the low flow rates of the inducing media. Cy5 is a fluorescent dye, frequently used in biomedical imaging, which allows us to track the presence of glucose, providing a signal useful to ascertain the arrival of the inducing media. Furthermore, we decided to make use of the third chamber of the device to monitor control cells (BY4741) which do not express fluorescent reporter. Upon acquisition, imaged cells were identified, tracked, segmented and the relative data extracted using automated software routines developed in the Lab. Data analysis was limited to the cells that were imaged for more than 220 time-points, corresponding to approximately 80% of the acquisition duration. The data extracted for the loading control strain were used to implement an autofluorescence correction aimed at facilitating the discrimination of slight differences in *GAL1* expression levels at early stages of induction. The autofluorescence subtraction was structured in i) normalization by the mean fluorescence evaluated, for each strain, over the first 6 time-points and ii) subtraction of a time-varying autofluorescent component. The scaling for the multiplicative constant, representative of the mean strain-dependent fluorescence in a phase in which we did not expect *GAL1* expression, was applied to compensate plausible differences in focus among the three chambers composing the device. Such scaling factor resulted equal to  $2 \times 10^4$ ,  $2.1 \times 10^4$  and  $2.9 \times 10^4$  for wild-type, *GAL10-lncRNA* $\Delta$  and loading control respectively. The time-varying autofluorescence component was estimated as the fluorescence value determined by the cell size over time, when adopting a linear fitting of autofluorescence versus size for the data-set acquired on the loading control population (Figure 4.5).

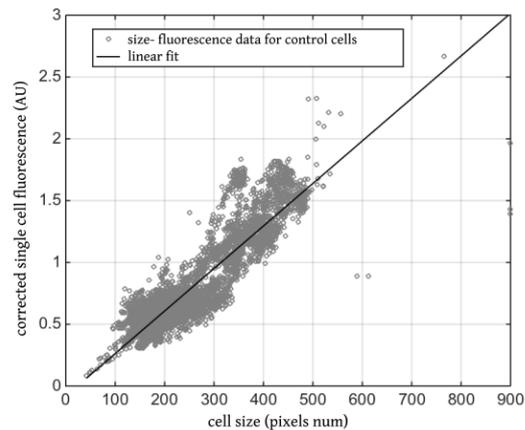


Figure 4.5: Time-varying autofluorescence estimate. The plot shows the linear fitting of the normalized fluorescence dependency upon size for the loading control strain. Each point coordinates represent size (x axis), measured as the number of pixel within the identified cell outline, and normalized fluorescence value for the loading control individuals, at each time point. The coefficient of determination,  $R^2$ , of the first degree polynomial fitted on this data is 0.8. This rather low parameter value indicates that the assumption of a linear relationship is probably not the most precise one. Upon scaling by the normalization constant, the autofluorescence estimated through this fit was subtracted, at each time point, from single cell fluorescence time-series.

The average population results, shown in Figure 4.6, highlighted a faster and higher activation of *GAL1*-GFP expression for the *GAL10*-lncRNA $\Delta$ . The difference in mean fluorescence is statistically significant ( $p$ -value  $< 10^{-4}$ ) over the whole induction.

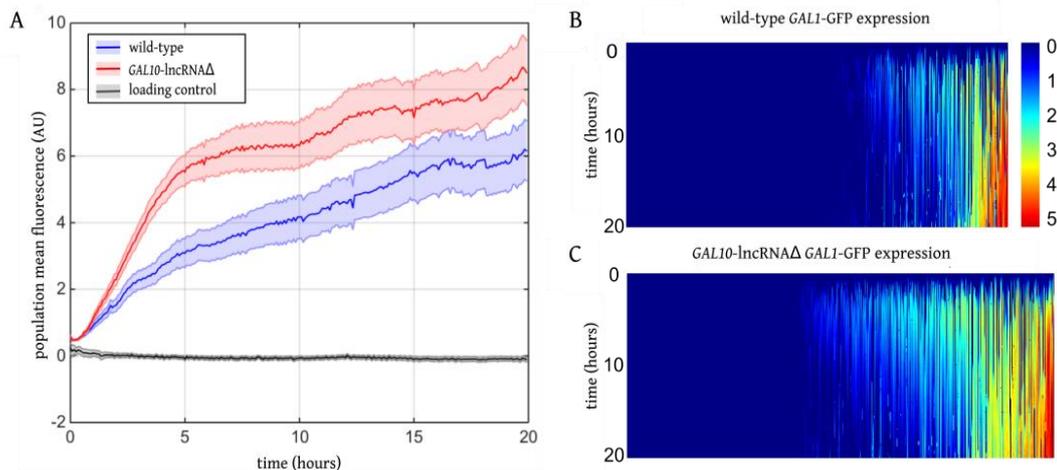


Figure 4.6: Results of wild-type and *GAL10*-lncRNA $\Delta$  induction with SC 2% raffinose, 0.02% glucose and 0.04% galactose. In panel A the average population behaviour is shown for wild-type (blue line), *GAL10*-lncRNA $\Delta$  (red line) and loading control (black line) strains, with shaded area representing the standard error on the mean. The results pertain 514 wild-type, 515 *GAL10*-lncRNA $\Delta$  and 28 loading control cells which were present for more than 220 time-points. The expression pattern of the activated strains is clearly discernible, with *GAL10*-lncRNA $\Delta$  cells responding to the inducing media earlier and stronger. In panels B and C the kymographs of the logarithm of single cell fluorescence over time for wild type and *GAL10*-lncRNA $\Delta$  strain is reported. Time is reported on y axis, while each x coordinate corresponds to a cell. The pixels encode with colours the logarithm of corrected fluorescence. The kymographs clearly indicate the presence of unresponsive individuals in each population.

A closer inspection of single cell fluorescence intensity for wild-type and *GAL10-lncRNAΔ* pointed out the existence of a fraction, whose proportion seems to be strain-dependent, of the population which appears unresponsive to the inducing media.

To further investigate this apparent phenotypic heterogeneity, a heuristic method was adopted to classify each strain in subpopulations of unresponsive (OFF) and responsive (ON) cells. In particular, a cell was labelled as ON if its corrected fluorescence exceeded, for more than 50 time-points (250 minutes), an intensity threshold defined as three standard deviations of the residuals determined with the linear fitting procedure used for the autofluorescence quantification. The pertinence of the classification criteria was evidenced by the absence of ON cells within the loading control strain.

Performing the ON/OFF distinction, we first confirmed that the percentage of responsive cells was higher for *GAL10-lncRNAΔ* (Figure 4.7).

Furthermore, restricting the previous analysis to the ON subpopulation of wild-type and mutant strain we noted that their discernible behaviour is transient, leading to the same steady-state expression levels.

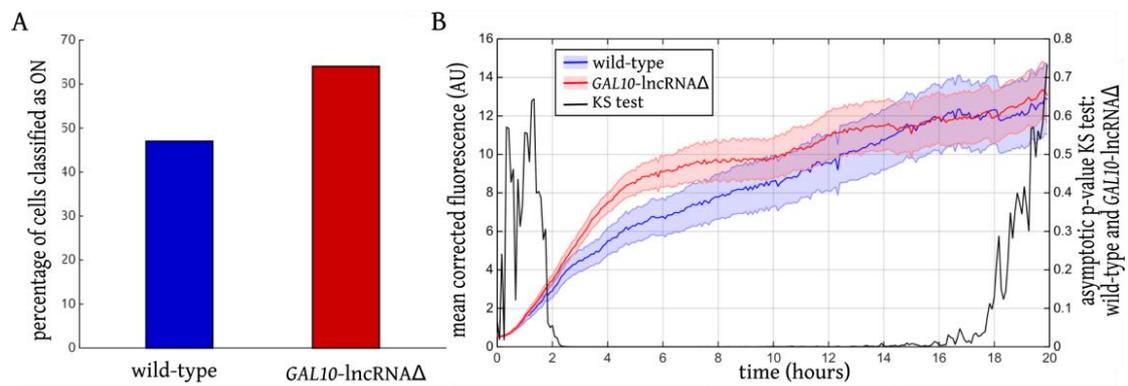


Figure 4.7: Comparison of wild-type and *GAL10-lncRNAΔ* responsive (ON) cells. In panel A bar plot is used to represent the fraction of wild-type (46%, blue) and *GAL10-lncRNAΔ* (64%, red) cells classified as responsive to the inducing media. These percentages proved statistically different at 1% significance level in a chi square test. Panel B shows the mean *GAL1*-GFP expression assessed over the ON subpopulations of wild-type (blue line) and *GAL10-lncRNAΔ* (red line) cells, with shaded area representing the standard error on the mean. The black line is the asymptotic p-value as evaluated with the KS test ( $\alpha = 5\%$ ) applied to the fluorescent data of ON cells. Its high (low) value indicates periods in which the two subpopulations are (not) considered as statistically different by the test.

These results suggest that mutations in *Reb1BS*, causing the abrogation of *GAL10-lncRNA* transcription, might exert a major effect on the early kinetics of *GAL* network activation. We tested the hypothesis that the emergence of distinct ON/OFF behaviour depends on the variability of single cell induction kinetics by evaluating kinetic statistics for each fluorescent trace of the ON subpopulations. The definition of these statistics, retrieved from Cloutier et al. [2] who applied them

to detail divergence in population level behaviour upon induction, is described in Figure 4.8. Among the kinetic statistics, the lag time and initial accumulation velocity were found to be statistically different (KS test,  $\alpha=5\%$ ) between wild-type and *GAL10-lncRNA $\Delta$* .

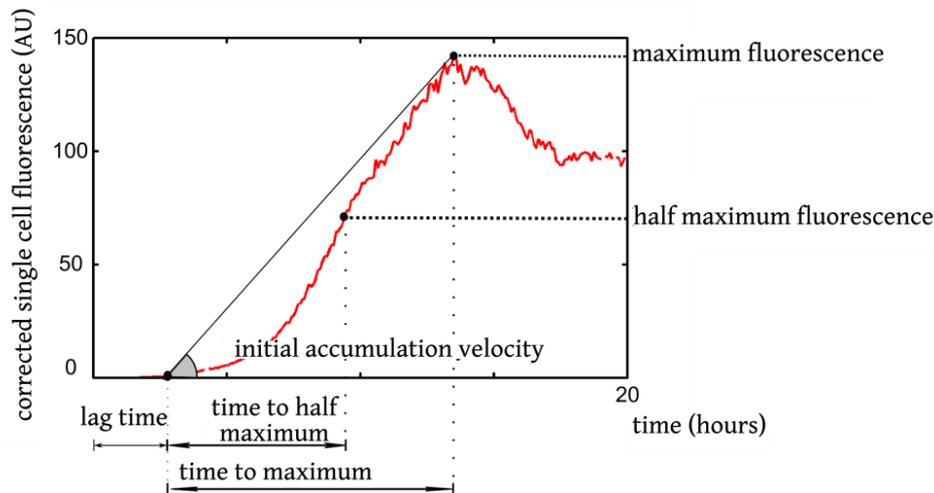


Figure 4.8: Visual description of kinetics statistics adopted to explore the divergent induction of wild-type and *GAL10-lncRNA $\Delta$*  ON cells. The six computed parameters, whose definition was retrieved from [2], are labelled within the figure using an acquired single cell trace as an example. These statistics relate fluorescence expression levels (y axes) with the time required for their attainment (x axes). The initial accumulation velocity was defined as the first derivative of the straight line connecting the activation threshold (fluorescence at the lag time) to the maximum fluorescence. While these statistics were originally applied to population level measurements, here they were evaluated on single cell fluorescence traces of ON individuals for wild-type and *GAL10-lncRNA $\Delta$*  strains. The inherent stochasticity of single cell behaviour probably played a role in limiting the statistical differences to lag time and initial accumulation velocity.

The data presented up to now revealed, when the population observational level is adopted, a faster and higher expression of *GAL1-GFP* in *GAL10-lncRNA $\Delta$*  strain over the whole acquisition. A closer look at single cell behaviour evidenced phenotypic heterogeneity, supporting the classification in responsive (ON) and unresponsive (OFF) individuals. The percentage of ON cells is higher for *GAL10-lncRNA $\Delta$*  and proved statistically different between the two strains. The fluorescence analysis for the ON subpopulations restricted their differences to the early phase of activation, as confirmed by the evaluation of kinetic statistics.

The more rapid induction in *GAL10-lncRNA $\Delta$*  cells proved coherent with a repressive role attributed by Houseley et al. [3] to the antisense transcript. However, while Houseley et al. and Cloutier et al. reported a temporary difference in the strains behaviour, restricting the effects of *GAL10-lncRNA* to the early stage of induction, in our population level data the discrepancy, albeit more relevant in the initial phase, is preserved over the whole acquisition. When focusing on the subset of ON cells our

results mirror the ones presented by Houseley et al. as the activated individuals of both strains, more abundant in the mutant case, reached the same steady state expression levels.

Multiple factors pertaining the measurement methodology might justify the population level discrepancies, among which the different inducing media and composition of cell populations analysed in our and Houseley et al. experiments. Indeed, while batch cultures retain both mothers and newly born cells, our signals are collected over the same mother cells, since daughters are removed by the flow media.

Given the plausible association, outlined by Houseley et al. [3], between attainment of low time-varying expression levels and limiting culture conditions in concealing late differences in wild-type cells and *GAL10-lncRNA* $\Delta$  mutants, we decided to juxtapose to the analysis presented so far an investigation of growth rate trend over time. Beyond extending the preliminary results reported on the kinetics of activation, the assessment of this measure of fitness underlines the potential of combining imaging and microfluidics to provide morphological information that can be integrated in a quantitative understanding of cell behaviour. Considering the manual counting of birth events a time-consuming option, we made use of an automated daughter identification script developed by M. Crane in the Lab. Though a more thorough characterization is required, preliminary applications indicated its results suitable to draw general conclusions.

The results are reported in Figure 4.9 and show the emergence of a significant difference in birth rate between ON and OFF subpopulations of wild-type and *GAL10-lncRNA* $\Delta$ , three hours after the induction start. In particular, ON cells are denoted by a higher birth, independently of the examined strain. This indicates that the antisense transcript, tuning the percentage of responsive cells, might provide a fitness advantage in environments denoted by particular mixes of sugars. Such hypothesis has never been proposed, to our knowledge, in scientific literature. However, our data do not support the definition of an exciting causality between cells responsiveness and fitness. Indeed, our observations could indicate either that the activation of the GAL pathway provides a fitness advantage in an environment with sugars mixtures or that growing, healthy cells are more likely to respond to induction. These considerations could reconcile the divergent results inherent to fluorescence data: the higher growth rate for ON cells, coupled with the inheritance of the activated state and the independence of birth events number on the compared genetic contexts, could justify the indistinguishable tardive expression in culture experiment.

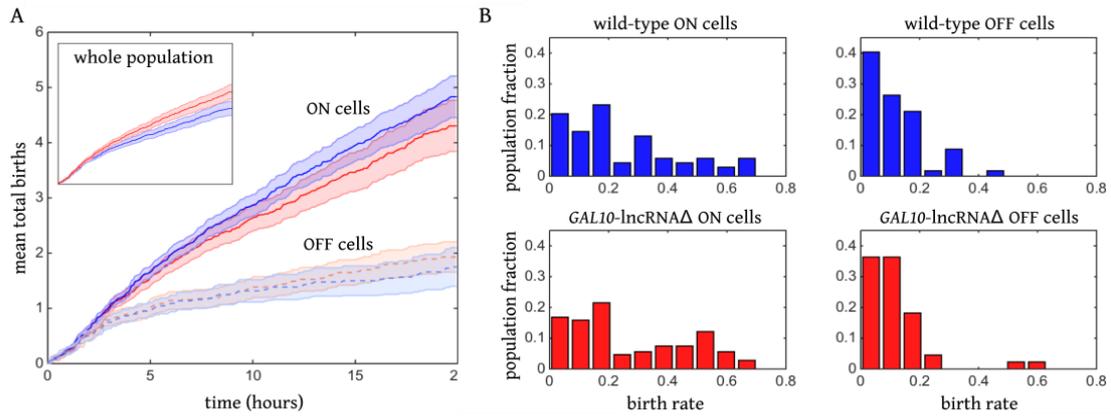


Figure 4.9: Results of birth events detection performed using the automated daughter counts identification script. Panel A shows the mean of the total number of budding events, as a function of time, of the ON (dark color) and OFF (light color) subpopulations for wild-type (blue) and *GAL10-lncRNAΔ* (red) strains. The corresponding colored shaded area represents the standard error on the mean. In the inset, mean total births is reported for the whole population of the analyzed strains. As can be seen, the birth rate trend does not appear to depend on *GAL10-lncRNA* transcript. In both strains, the similar population birth rate masks the higher fitness of the ON subpopulation compared to the OFF one. In panel B histograms of birth rates, defined as the number of budding events occurring in the acquisition period of each mother divided by the time the mother is present for, are compared among ON and OFF cells of wild-type and *GAL10-lncRNAΔ* strains. Birth rates of ON and OFF cells of each strain were found to be significantly different by a KS test ( $\alpha = 5\%$ ), while the statistical analysis failed to discriminate between responsive (unresponsive) clusters of wild-type and *GAL10-lncRNAΔ*.

Though potentially relevant, the aforementioned interpretations rely upon a birth rate statistics assessed on a limited number of mother cells, and hence these results might be biased by the low cardinality of the samples. A strategy to strengthen our data, identifying the direction of causality between *GAL1*-GFP expression and higher ON cells growth rate, was identified in the application of similar analysis on new strains in which *GAL1* was knocked out and replaced with a suitable reporter protein. These strains, originally conceived as a model to explore the feasibility of employing fluorescent reporters with selected properties to infer the activation dynamics of a promoter, will be the focus of the next section.

## 4.4 Analysis of transcriptional dynamics using the UBI-M $\Delta$ k-GFP $\gamma$ reporter

To extend the analysis of transcriptional dynamics we require the fluorescent protein used as readout of promoter activation to provide an easily detectable and fast signal, being denoted by high brightness, maturation and degradation rate.

For these reasons, considering its previously performed characterization with the microscopy set-up available in the Swain Lab, we chose to adopt as a fluorescent reporter GFP $\gamma$  with N terminal degron tag. GFP $\gamma$  originated from combined mutations of an enhanced GFP variant directed to improve its brightness. The term N degron tag ( $\Delta$ k) refers to an N terminal degradation tag, first described in 2012 by Houser et al. [161], that marks proteins for fast recruitment of the degradation machinery. The degradation rate of the resulting tagged protein depends on the amino acid residue exposed at the N terminal upon proteolytic removal an ubiquitin sequence.

In the construction procedure of these strains, *GAL1*-GFP was knocked out in the previously analysed ones and replaced by UBI-M $\Delta$ k-GFP $\gamma$ , followed by histidine selection marker.

Although designed for the acquisition of time-series data for inference analysis of *GAL1-10* promoter, the deletion of *GAL1* gene in BY4741 UBI-M $\Delta$ k-GFP $\gamma$  (wild-type\*) and *GAL10*-lncRNA $\Delta$ : UBI-M $\Delta$ k-GFP $\gamma$  (*GAL10*-lncRNA $\Delta$ \*) should make them unable to metabolize galactose [144]. Hence they were experimentally characterized to gain insights in galactose metabolism related growth rate in wild-type and Gal10-lncRNA $\Delta$ .

The behaviour of wild-type\* and *GAL10*-lncRNA $\Delta$ \* strains was assessed imaging the cells within the microfluidic device over a 20 hours acquisition performed under the same experimental conditions previously defined.

As before, we restricted the analysis to the fraction of the cells that was present for more than 200 time points of the acquisition before assessing the growth rate of wild-type\* and Gal10-lncRNA $\Delta$ \*. As can be seen in Figure 4.10, even in this experiment each strain revealed the presence of an ON and OFF cell cluster, although the criteria on which they were defined differs from the one adopted above and will be detailed later.

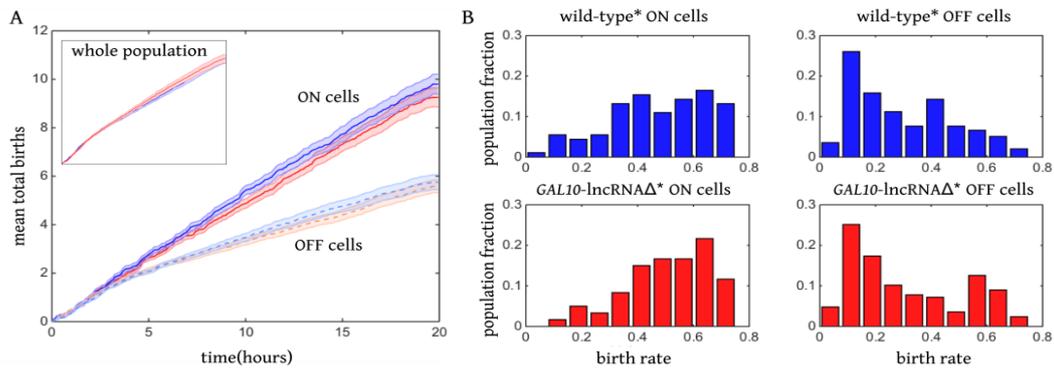


Figure 4.10: Birth events detection of wild-type\* and *GAL10-lncRNAΔ*\* strain. In analogy with the results shown in Figure 4.9, panel A depicts the mean of the total number of budding events over time for the ON (dark colour) and OFF (light colour) for wild-type\* (blue) and *GAL10-lncRNAΔ*\* (red) strains. Though the *GAL1* deletion performed in these strains should prevent galactose metabolism, ON cells are denoted by a higher growth rate, which seems to be independent of the presence of the antisense transcript. Birth rate statistics were computed on 287 wild-type and 227 *GAL10-lncRNAΔ*\* identified mother cells. In panel B histograms of birth rate for the ON and OFF subpopulations of each strain are shown. As before, a KS test revealed statistical differences in ON and OFF clusters within each strain, but did not support differences ascribable to the antisense transcript.

The birth events investigation proved relevant for performing a comparison with the growth rate trend observed in wild-type and *Gal10-lncRNAΔ* strains. Despite the similar genetic background of the strains, the knockout of *GAL1* gene in those presented in this section makes them unable to metabolize galactose [144]. Although further investigations and experimental repeats are required to confer solidity to the assessed behaviour, the analogy in mean total budding events trend over time – clear on either the whole population or the ON and OFF cluster – suggests that the higher growth rate of ON cells in the original strains is likely ascribable to an improved fitness promoting the activation of the *GAL* gene cluster upon induction, and not an advantage coming from the metabolism of the inducing sugar. It is worth noting that in the new strains all ON cells were dividing, and 3% of OFF cells were quiescent. A comparison with inline evaluations performed on wild-type and *GAL10-lncRNAΔ* (data not shown) cells indicates an overall higher fitness of the strains lacking *GAL1* gene, attributable to the metabolic burden of activating the *GAL* network when galactose is present at low levels and mixed with preferred carbon sources.

While the reasons underlying the partition of the wild-type\* and *GAL10-lncRNAΔ*\* populations, both unable to metabolize galactose, in subgroups of ON and OFF cells remain unclear, the similar proportion of responsive individuals among the two strains (29% and 24% respectively) would exclude a consistent effect of the antisense transcript.

Coupling the results of the experiments on the whole set of strains, it seems likely that the distinction in responsive and non-responsive cells is the result of a previous population state. New experiments, in which cells are induced with a media containing a subset of the tested sugars, could provide insights into this partitioned commitment.

Results on wild-type\* and *GAL10-lncRNAΔ*\* fluorescence analysis are shown in Figure 4.11. The population averaged fluorescence revealed an oscillatory trend, more marked for the wild-type\*, and a similar GFP expression pattern, in contrast with the data acquired on wild-type and *GAL10-lncRNAΔ*. An inspection of the single cell traces highlighted the presence of bursts in the fluorescent reporter expression, as opposed to a clear continuous activation. These bursts were identified as periods, lasting more than 5 time points (25 minutes), in which the fluorescence exceeded a threshold set to the maximum of the mean autofluorescence plus 3 standard deviations. Applying the same analysis with 1 or 2 standard deviations in the activation threshold preserved the ratio between ON and OFF subpopulations. The number of responsive (ON) cells, defined as individuals with at least one burst in GFP expression, is low and comparable between the two strains (31% and 29% for wild-type\* and *GAL10-lncRNAΔ*\* respectively).

The presence of a pulsatile activation might reflect a partial re-establishment of a repressed state, due to the presence of glucose. This might occur when galactose appears in trace concentrations, is sensed by cells which hence activate the network, but cannot be metabolized because of *GAL1* knockout. Furthermore, the absence of Gal1, which seems to perform Gal3 function at a later stage of induction, could coincide with the loss of a signal able to sustain activation. The pulsatile expression observed in wild-type\* and *GAL10-lncRNAΔ*\* has not been observed in the induction of other *GAL1Δ* mutants investigated in the Lab (data not shown), leading us to believe that this effect might be amplified by the properties of the inducing media here adopted.

So far, our data does not support the exclusion of a similar *GAL1* expression pattern for wild-type and *GAL10-lncRNAΔ*, where the readout of *GAL1-10* bidirectional promoter activation provided by the slow decaying eGFP could mask the pulsatile dynamics. As for the single cell fluorescence, the computation of parameters aimed at exploring bursts statistics (peaks amplitude and duration other than the number of bursts for each cell) failed in revealing statistical differences between the *GAL1* knockout strains (KS test,  $\alpha = 5\%$ ).

To further explore the activation state of *GAL1-10* bidirectional promoter in wild-type\* and *GAL10-lncRNAΔ*\* strains, eventually providing an *in silico* tool to support the formulated hypothesis on the mechanisms that cause the divergent behaviour observed, the application of a Bayesian inference scheme on this dataset is ongoing in the Lab.

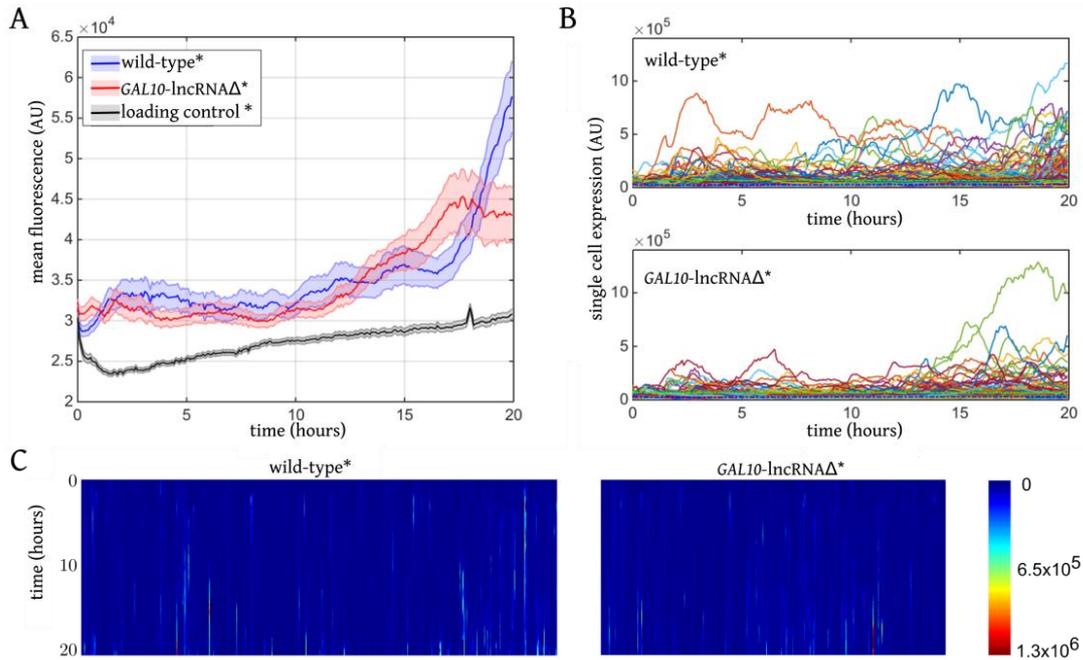


Figure 4.11: Results of the induction experiment (SC 2% raffinose, 0.02% glucose and 0.04% galactose) performed on wild-type\* and *GAL10-lncRNAΔ\** strains. In panel A the mean single cell fluorescence against time is shown for wild-type\* (blue), *GAL10-lncRNAΔ\** (red) and loading control\* (black) cells, with shaded bars indicating the standard error on the mean. In the loading control\* *GAL1* was knocked out and replaced with an unobserved fluorescent reporter. The mean was computed on 645 (wild-type\*), 469 (*GAL10-lncRNAΔ\**) and 477 (loading control\*) individuals. As it can be seen, the population behaviour differs from the one assessed in wild-type and *GAL10-lncRNA* (Figure 4.6): here a comparable and slightly pulsatile expression leads to a late activation. Panel B shows single cell fluorescence traces for the analysed strains. In panel C, kymographs of single cell fluorescence are reported. Panels B and C highlight the pulsatile activation of *GAL1-10* promoter in both strains, with temporary increases in GFP signal not leading to a steady state level.

## 4.5 Conclusions

The results presented in this chapter prove the feasibility of performing analysis of long non coding RNA regulation, on a single cell basis, using quantitative fluorescence microscopy and suitable transcriptional reporters.

The evidence of detectable differences among the strains suggests that the selected strategy is suitable to disclose the features of the antisense-transcript regulation on *GAL1* expression. Such investigations, easily extendable to other molecular players operating in the GAL network, could provide deeper understanding of the multiple regulatory mechanisms exemplified in the extensively studied metabolic pathway.

Up to now, technical complications impeded the acquisition of a significant number of repeats and controls of these experiments, which are required to provide solidity to the presented data and conclusions.

Once we will accomplish this goal, the experimental protocol could be adapted to include the assessment of cellular growth rate before induction (when cells are exposed to raffinose) and upon switch to the original environment after the attainment of a sustained activation. This, and an altered composition of the inducing media, could reveal insights into the hypothesized presence of a previously defined fate commitment and the benefits of the observed phenotypic variability. Preliminary test have been performed using an external mixer, namely a T-junction whose entries are connected to two syringe pumps (containing different media) while the output branch feeds the device with media flow. The use of such a system, other than basically reinforcing the precision with which we are able to define the time point of initial induction, expands the range of possible experiments through inclusion of behaviour analysis under dynamically changing environments.

Finally, the application of inference techniques to the acquired time-series data could allow the identification of parameters and mathematical models able to discern alterations in *GAL1-10* promoter states due to the regulatory function encoded in *GAL10*-lncRNA. The so defined *in silico* models, to our knowledge not available up to now, would allow fast investigation of the hypothesis inherent to the ongoing mechanisms underlying sense- antisense transcriptional regulation.

## 5 Final Remarks

In this thesis different experimental and mathematical modelling techniques have been adopted to address questions pertaining the selection of gene expression control mechanisms and network topologies in the design of synthetic devices able to reliably operate in the stochastic cellular context. Indeed, only when stochasticity effects will be included as a specific design criterion will Synthetic Biology fulfil the promises of providing engineered tools to solve biotechnological, medical and environmental challenges.

In Chapter 2, prompted by published numerical results elucidating the possibility of exerting an independent control on the mean expression level and noise profile of a target gene by regulating the two steps of gene expression, we described the development of a synthetic gene circuits' catalogue which could act as a noise tester. The gene circuits of this collection, whose topology derives from previous investigations carried out in the Laboratory of Cellular and Molecular Engineering (ICM Lab) of the University of Bologna, allow the post-transcriptional silencing of any gene of interest *via* the hybridization of two regulatory sequences. Specifically, a *cis*-acting element (CIS), composed of a 50 base-pair non-coding sequence and a ribosome binding site (RBS), is cloned upstream of the gene (in our case the nucleotidic sequence encoding the green fluorescent reporter) to be silenced. Upon transcription, the molecular annealing of the CIS-GFP sequence with its complementary *trans*-acting oligoribonucleotide (TRANS) causes a partial occlusion at the RBS, preventing ribosome docking. Though it is still necessary to verify the experimental feasibility of an independent control of the first and second order moments characterizing the expression levels' distribution of the fluorescent reporter gene over a wide regime, preliminary results presented indicate that this circuits' library will provide a useful tool for testing the robustness of available molecular widgets to biological noise.

Two gene circuits, implementing either a transcriptional (TC gene circuit) or post-transcriptional control (pTC gene circuit) in the expression of the fluorescent reporter, were selected from the catalogue for detailed investigation. The TC and pTC gene circuits, upon cloning in plasmids with different copy number (pSB1A2 and pSB4A5), were transformed in TOP10F' *E. coli* cells. The dynamic protein expression was measured in populations of transformants growing in the microplate reader. The steady-state normalized fluorescence values confirmed the gene circuits operate as expected: following induction with exogenous IPTG, GFP expression increases in TC and decreases in pTC gene circuits. Furthermore, the experimental data made it possible to detect both the appearance of a

saturation phenomenon, occurring in the TC gene circuit cloned in the pSB1A2 high copy number plasmid at maximum induction, and the Parts Registry's misleading classification of pSB4A5 as a low copy number cloning vector. Indeed, the fluorescence values measured on both circuits cloned in this plasmid are in line with those ascribable to a medium copy number cloning vector.

Deterministic models providing a macroscopic description of the TC and pTC circuits' function were developed considering as few species and biochemical reactions as possible, in order to minimize the models' features that could not be constrained by the data. Due to the aforementioned experimental evidences, the fitting procedure used to define the deterministic models disregarded a precise picture of the ongoing biological reality in favour of a faithful description of the measured dose-response curves. The identified parameters were subsequently adopted in stochastic simulations, using the Gillespie algorithm, aimed at investigating variability in gene expression levels in both gene circuits at different plasmid copy numbers. Simulation results conformed to theoretical principals concerning the scaling of biological noise strength with the number of involved molecules, indicating a lower stochasticity in GFP expression, independently of the exerted control mechanism, for the high copy number plasmid. The unexpected medium copy number attributed to the pSB4A5 was probably responsible for the limited fractional change in protein variability computed for the two plasmids. Simulation results indicate that noise strength can be tuned over a wide range by administering IPTG in the TC gene circuits, while it remains almost constant and at a lower level in the circuit implementing post-transcriptional control in GFP expression.

In Chapter 3 we summarized the results of the experiments performed to validate and characterize the optical microscopy set-up, available in the ICM Lab, for quantifying the single cell and population fluorescence signal from *E. coli* transformants expressing GFP. To this end, measurements of the steady-state expression distributions for the TC gene circuit, adopted as a benchmark for its ability to provide a wide range of fluorescence intensities upon induction with IPTG, were compared for the flow cytometric and microscopy acquisition. Both the linearity of the analytical relation between the fluorescence intensities acquired with the two instruments and the comparable standard errors proved the ability of the microscopy set-up to capture the fluorescent signal's dispersion within the cell population. Despite the closely matching results, the flow cytometer's wider dynamic range ensures a more reliable discrimination of fluorescence intensities at low expression levels, where the lower sensitivity of the microscopy set-up might be critical. In the analysed conditions, the reduced dynamic range of the microscopy set-up caused a slight underestimation of stochasticity in GFP expression, quantified as the *squared coefficient of variation* ( $CV^2$ ). We further explored how the number of imaged cells affects the measured variability in protein expression. The results indicated that a population composed by few hundreds cells is sufficient to achieve robust and reliable statistics, leading to the identification of a lower threshold on the sample size to be acquired for

collecting quantitative data. This analysis greatly improved the usability of our microscopy set-up, as the time required for the acquisition of a cells number comparable to the cardinality of fluorescent events detected in a typical flow cytometer experiment would be prohibitively time-consuming. The reported data suggests that the characterized microscopy set-up is an adequate instrument for quantifying the single cell behaviour of *E. coli* transformants.

In the second half of the chapter we presented an experimental and theoretical comparison of phenotypic noise within an isogenic population of bacterial cells transformed with either the TC and pTC gene circuits, cloned in pSB4A5 plasmid. In the study, fluorescence distributions of the expression of GFP subject to transcriptional or post-transcriptional control, exerted in gene circuits with similar topology, were acquired by flow cytometry. The results highlights that noise is lower for the gene-circuit with post-transcriptional control, and that the difference in noise between the two circuits increases when the post-transcriptional control on gene-expression is more efficient. Interestingly, only when cell division events were included in the stochastic models of the gene circuits' function, were simulation results in qualitative agreement with the experimental  $CV^2$ . As the experimental data on protein variability relates to a single fluorescent reporter, we were not able to exclude that the differential stochasticity of the two control mechanisms is related to extrinsic noise. Considering the experimental protocol and the similar genetic contexts in which the control mechanisms were compared, we think that the inducer IPTG acts on the intrinsic component via tuning of GFP gene (TC gene circuit) or TRANS sequence (pTC gene circuit) transcription. It is interesting to note that a similar reduction in gene expression stochasticity related to post-transcriptional control has been recently observed in eukaryotic cells, where it was related to changes in intrinsic noise. The numerical evaluation of the correlation in the expression of two fluorescent reporters, easily modelled considering the presence of two copies of the genetic circuits in the cell context, could provide an additional validation of this hypothesis. The data presented in this chapter constitutes, to the best of our knowledge the first single-cell characterization of a synthetic circuit implementing post-transcriptional control in gene expression. Globally, our results provides an experimental validation of theoretical studies attributing to post-transcriptional control a role in minimizing noise on protein expression. As a consequence, the pTC gene circuit represents a possible strategy to modulate noise by means of external signals: an intriguing possibility for the design of novel application in synthetic biology and the investigation of noise effects in the control of cellular behaviour.

In Chapter 4 we turned to investigating the role of a natural long non-coding RNA (*GAL10*-lncRNA) in the expression of a metabolites-induced gene (*GAL1*), adopting as a biological model the bakers' yeast *Saccharomyces Cerevisiae*. *S. cerevisiae*, widely studied over the past 60 years as an example of transcriptional regulation in eukaryotes, has recently attracted a renewed scientific interest due to

the identification of antisense transcripts encoded in the open reading frame of genes belonging to the galactose metabolic pathway. In the study, quantitative fluorescence microscopy and microfluidic devices were used to investigate, on a single-cell basis, the debated role exerted by GAL10-lncRNA in the regulation of *GAL1* gene. Both the long term monitoring of cells behaviour in the homogeneous environment provided by the microfluidic device and the extraction of a large amounts of statistical fluorescence and morphological data performed through automated analysis allowed us to support the repressive effect exerted by the *GAL10*-lncRNA on *GAL1* activation. Relating the lncRNA transcription to phenotypic variability, we provided an experimental evidence of its hypothesized effect on the heterogeneous cellular response to the inducing media. Using an automated daughter identification script developed in the laboratory headed by Professor Peter Swain (Swain Lab), we run a preliminary analysis of single-cell birth rate events and related them to the observed induction dynamic. This data allow reconciliation of the apparent discrepancies between the microfluidic and published batch-cultures results. To further explore the variability in single cell birth rate events, eventually revealing its dependence on galactose utilization, we built strains in which the *GAL1*-GFP open reading frame was replaced by the UBI-MΔk-GFP $\gamma$  fluorescent reporter, optimized for the inference of transcriptional dynamics. The lack of statistical differences in growth rates computed between the original and the new strains, which should be unable to metabolize galactose, led us to exclude the hypothesis that the positive correlation between the fluorescent reporter expression and birth rate emerges as a consequence of the beneficial galactose utilization, but likely witnesses the existence of a previously defined fate commitment. Looking at the fluorescence data of the *GAL1*-knocked out strains, we observed a pulsatile activation as opposed to the sustained one characterizing the original strains. This data might be explained considering the impairment of the *gal1* mediated positive feedback loop operating in the galactose network.

Once a significant number of repeats and controls of these experiments, required to provide solidity to the presented data and conclusions, will be available, the outlined hypothesis could be tested by adapting the protocol to include the assessment of cellular growth rate before induction (when cells are exposed to raffinose) and upon switch to the original environment after the attainment of a sustained activation. Furthermore, by altering the composition of the inducing media, we could gain insights into the benefits of the observed phenotypic variability.

The single-cell characterization of the TC and pTC gene circuits, cloned in the high copy number plasmid, is ongoing in the ICM Lab. Other than providing more sensitive dataset for the investigation of the unexpected saturation phenomenon occurring in the gene circuit implementing transcriptional control in GFP expression, this data will be used to experimentally quantify the scaling in noise strength with the cloning vectors' copy number. Once a proper concentration range

of the inducer regulating transcription of the CIS-GFP sequence has been identified, we will expand the analysis to other members of the implemented circuit catalogue in order to prove the usability of the system as a noise tester.

Looking forward, it would be interesting to explore the dynamic behaviour of these genetic circuits outside the steady-state. In this perspective, coupling the tested usability of our fluorescence microscopy set-up with the features of the more sophisticated analysis ongoing in the Swain Lab, the acquisition of single-cell time-series data would enable a more thorough characterization of the circuits' functional properties, thereby informing their potential use in more sophisticated circuit design and mathematical models.

The theoretical and the experimental results presented in this thesis demonstrate that the topology of the gene circuits is a major determinant of its noise properties, and that it is possible to control the variability in gene expression by external signals. The gene circuits developed during this PhD program might be useful to evaluate the role of noise in other biological systems, and in general they represent a first example of synthetic circuits designed to control noise characteristics.

In addition to the theoretical value of this issue, it should be noticed that noise control is expected to boost the potential of synthetic biology in biotechnological and biomedical applications. More than increasing the yield of chemicals at low environmental impact (e.g. see DuPont 'Sorona' textiles, Amyris 'Artemisinin' antimalarial drug or Joule biofuels), robustness of synthetic gene circuits allows us to anticipate that the future will see smart cells able to perform diagnostic and therapeutic tasks [162]. As the non-linearity inherent to living matter complicates the analysis – and therefore the design – of functional synthetic biological tools in an industrial perspective, addressing this hurdle is expected to strengthen the economic feasibility of synthetic biology as an industrial business model and an innovation platform in a global market estimated to grow up to € 35 billion value in 2020 [<https://www.alliedmarketresearch.com/synthetic-biology-market>].



# Bibliography

1. Elowitz, M.B., et al., *Stochastic gene expression in a single cell*. Science, 2002. **297**(5584): p. 1183-6.
2. Cloutier, S.C., et al., *Long noncoding RNAs promote transcriptional poisoning of inducible genes*. PLoS Biol, 2013. **11**(11): p. e1001715.
3. Houseley, J., et al., *A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster*. Mol Cell, 2008. **32**(5): p. 685-95.
4. Pinskaya, M., S. Gourvenec, and A. Morillon, *H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation*. EMBO J, 2009. **28**(12): p. 1697-707.
5. Ceroni, F., et al., *A synthetic post-transcriptional controller to explore the modular design of gene circuits*. ACS Synth Biol, 2012. **1**(5): p. 163-71.
6. Ozbudak, E.M., et al., *Regulation of noise in the expression of a single gene*. Nat Genet, 2002. **31**(1): p. 69-73.
7. Voigt, C.A. and J.D. Keasling, *Programming cellular function*. Nat Chem Biol, 2005. **1**(6): p. 304-7.
8. Hasty, J., D. McMillen, and J.J. Collins, *Engineered gene circuits*. Nature, 2002. **420**(6912): p. 224-30.
9. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline*. Mol Syst Biol, 2006. **2**: p. 2006 0028.
10. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators*. Nature, 2000. **403**(6767): p. 335-8.
11. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli*. Nature, 2000. **403**(6767): p. 339-42.
12. Rinaudo, K., et al., *A universal RNAi-based logic evaluator that operates in mammalian cells*. Nat Biotechnol, 2007. **25**(7): p. 795-801.
13. Ham, T.S., et al., *Design and construction of a double inversion recombination switch for heritable sequential genetic memory*. PLoS One, 2008. **3**(7): p. e2815.
14. Win, M.N. and C.D. Smolke, *Higher-order cellular information processing with synthetic RNA devices*. Science, 2008. **322**(5900): p. 456-60.
15. Friedland, A.E., et al., *Synthetic gene networks that count*. Science, 2009. **324**(5931): p. 1199-202.
16. Ceroni, F., et al., *Rational design of modular circuits for gene transcription: A test of the bottom-up approach*. J Biol Eng, 2010. **4**: p. 14.
17. Lee, S.K., et al., *Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels*. Curr Opin Biotechnol, 2008. **19**(6): p. 556-63.
18. Weber, W., et al., *A synthetic mammalian gene circuit reveals antituberculosis compounds*. Proc Natl Acad Sci U S A, 2008. **105**(29): p. 9994-8.

19. Lu, T.K. and J.J. Collins, *Dispersing biofilms with engineered enzymatic bacteriophage*. Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11197-202.
20. Marguet, P., et al., *Biology by design: reduction and synthesis of cellular components and behaviour*. J R Soc Interface, 2007. **4**(15): p. 607-23.
21. Novick, A. and M. Weiner, *Enzyme Induction as an All-or-None Phenomenon*. Proc Natl Acad Sci U S A, 1957. **43**(7): p. 553-66.
22. Shahrezaei, V. and P.S. Swain, *The stochastic nature of biochemical networks*. Curr Opin Biotechnol, 2008. **19**(4): p. 369-74.
23. Arias, A.M. and P. Hayward, *Filtering transcriptional noise during development: concepts and mechanisms*. Nat Rev Genet, 2006. **7**(1): p. 34-44.
24. Kussell, E. and S. Leibler, *Phenotypic diversity, population growth, and information in fluctuating environments*. Science, 2005. **309**(5743): p. 2075-8.
25. Arkin, A., J. Ross, and H.H. McAdams, *Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells*. Genetics, 1998. **149**(4): p. 1633-48.
26. Dodd, I.B., K.E. Shearwin, and J.B. Egan, *Revisited gene regulation in bacteriophage lambda*. Curr Opin Genet Dev, 2005. **15**(2): p. 145-52.
27. Balazsi, G., A. van Oudenaarden, and J.J. Collins, *Cellular decision making and biological noise: from microbes to mammals*. Cell, 2011. **144**(6): p. 910-25.
28. Losick, R. and C. Desplan, *Stochasticity and cell fate*. Science, 2008. **320**(5872): p. 65-8.
29. Suel, G.M., et al., *Tunability and noise dependence in differentiation dynamics*. Science, 2007. **315**(5819): p. 1716-9.
30. Maamar, H., A. Raj, and D. Dubnau, *Noise in gene expression determines cell fate in Bacillus subtilis*. Science, 2007. **317**(5837): p. 526-9.
31. Suel, G.M., et al., *An excitable gene regulatory circuit induces transient cellular differentiation*. Nature, 2006. **440**(7083): p. 545-50.
32. Davidson, C.J. and M.G. Surette, *Individuality in bacteria*. Annu Rev Genet, 2008. **42**: p. 253-68.
33. Maloney, P.C. and B. Rotman, *Distribution of suboptimally induced -D-galactosidase in Escherichia coli. The enzyme content of individual cells*. J Mol Biol, 1973. **73**(1): p. 77-91.
34. Spudich, J.L. and D.E. Koshland, Jr., *Non-genetic individuality: chance in the single cell*. Nature, 1976. **262**(5568): p. 467-71.
35. Berg, O.G., *A model for the statistical fluctuations of protein numbers in a microbial population*. J Theor Biol, 1978. **71**(4): p. 587-603.
36. Swain, P.S., M.B. Elowitz, and E.D. Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12795-800.
37. Paulsson, J. and M. Ehrenberg, *Noise in a minimal regulatory network: plasmid copy number control*. Q Rev Biophys, 2001. **34**(1): p. 1-59.

38. Boyd, D., et al., *Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple Escherichia coli plasmid-chromosome shuttle system*. J Bacteriol, 2000. **182**(3): p. 842-7.
39. Rosenfeld, N., et al., *Gene regulation at the single-cell level*. Science, 2005. **307**(5717): p. 1962-5.
40. Raser, J.M. and E.K. O'Shea, *Noise in gene expression: origins, consequences, and control*. Science, 2005. **309**(5743): p. 2010-3.
41. Chapon, C., *Expression of maltT, the regulator gene of the maltose region in Escherichia coli, is limited both at transcription and translation*. EMBO J, 1982. **1**(3): p. 369-74.
42. Baumeister, R., et al., *Lack of a 5' non-coding region in Tn1721 encoded tetR mRNA is associated with a low efficiency of translation and a short half-life in Escherichia coli*. Nucleic Acids Res, 1991. **19**(17): p. 4595-600.
43. Trotot, P., et al., *Comparative analysis of the cya locus in enterobacteria and related gram-negative facultative anaerobes*. Biochimie, 1996. **78**(4): p. 277-87.
44. Fraser, H.B., et al., *Noise minimization in eukaryotic gene expression*. PLoS Biol, 2004. **2**(6): p. e137.
45. Blake, W.J., et al., *Noise in eukaryotic gene expression*. Nature, 2003. **422**(6932): p. 633-7.
46. Thattai, M. and A. van Oudenaarden, *Intrinsic noise in gene regulatory networks*. Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8614-9.
47. Kaern, M., et al., *Stochasticity in gene expression: from theories to phenotypes*. Nat Rev Genet, 2005. **6**(6): p. 451-64.
48. Raj, A., et al., *Stochastic mRNA synthesis in mammalian cells*. PLoS Biol, 2006. **4**(10): p. e309.
49. Kafri, R., M. Springer, and Y. Pilpel, *Genetic redundancy: new tricks for old genes*. Cell, 2009. **136**(3): p. 389-92.
50. Guido, N.J., et al., *A bottom-up approach to gene regulation*. Nature, 2006. **439**(7078): p. 856-60.
51. Paulsson, J., *Summing up the noise in gene networks*. Nature, 2004. **427**(6973): p. 415-8.
52. Pedraza, J.M. and J. Paulsson, *Effects of molecular memory and bursting on fluctuations in gene expression*. Science, 2008. **319**(5861): p. 339-43.
53. Lestas, I., G. Vinnicombe, and J. Paulsson, *Fundamental limits on the suppression of molecular fluctuations*. Nature, 2010. **467**(7312): p. 174-8.
54. Becskei, A., B. Seraphin, and L. Serrano, *Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion*. EMBO J, 2001. **20**(10): p. 2528-35.
55. Isaacs, F.J., et al., *Prediction and measurement of an autoregulatory genetic module*. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7714-9.
56. Thieffry, D., et al., *From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli*. Bioessays, 1998. **20**(5): p. 433-40.
57. Schedl, A., et al., *Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities*. Cell, 1996. **86**(1): p. 71-82.

58. Rao, C.V., D.M. Wolf, and A.P. Arkin, *Control, exploitation and tolerance of intracellular noise*. Nature, 2002. **420**(6912): p. 231-7.
59. Simpson, M.L., C.D. Cox, and G.S. Sayler, *Frequency domain analysis of noise in autoregulated gene circuits*. Proc Natl Acad Sci U S A, 2003. **100**(8): p. 4551-6.
60. Orrell, D. and H. Bolouri, *Control of internal and external noise in genetic regulatory networks*. J Theor Biol, 2004. **230**(3): p. 301-12.
61. Becskei, A. and L. Serrano, *Engineering stability in gene networks by autoregulation*. Nature, 2000. **405**(6786): p. 590-3.
62. Dublanche, Y., et al., *Noise in transcription negative feedback loops: simulation and experimental analysis*. Mol Syst Biol, 2006. **2**: p. 41.
63. Austin, D.W., et al., *Gene network shaping of inherent noise spectra*. Nature, 2006. **439**(7076): p. 608-11.
64. Alon, U., *Network motifs: theory and experimental approaches*. Nat Rev Genet, 2007. **8**(6): p. 450-61.
65. Balazsi, G., et al., *The temporal response of the Mycobacterium tuberculosis gene regulatory network during growth arrest*. Mol Syst Biol, 2008. **4**: p. 225.
66. Nevozhay, D., et al., *Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression*. Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5123-8.
67. Singh, A. and J.P. Hespanha, *Optimal feedback strength for noise suppression in autoregulatory gene networks*. Biophys J, 2009. **96**(10): p. 4013-23.
68. Zhang, M.X., et al., *Regulation of endothelial nitric oxide synthase by small RNA*. Proc Natl Acad Sci U S A, 2005. **102**(47): p. 16967-72.
69. Zhang, M.X., et al., *Biogenesis of short intronic repeat 27-nucleotide small RNA from endothelial nitric-oxide synthase gene*. J Biol Chem, 2008. **283**(21): p. 14685-93.
70. Hinske, L.C., et al., *A potential role for intragenic miRNAs on their hosts' interactome*. BMC Genomics, 2010. **11**: p. 533.
71. Schmiedel, J.M., et al., *Gene expression. MicroRNA control of protein expression noise*. Science, 2015. **348**(6230): p. 128-32.
72. Hooshangi, S., S. Thiberge, and R. Weiss, *Ultrasensitivity and noise propagation in a synthetic transcriptional cascade*. Proc Natl Acad Sci U S A, 2005. **102**(10): p. 3581-6.
73. Pedraza, J.M. and A. van Oudenaarden, *Noise propagation in gene networks*. Science, 2005. **307**(5717): p. 1965-9.
74. Powell, E.O., *An outline of the pattern of bacterial generation times*. J Gen Microbiol, 1958. **18**(2): p. 382-417.
75. Rosenfeld, N. and U. Alon, *Response delays and the structure of transcription networks*. J Mol Biol, 2003. **329**(4): p. 645-54.
76. Swain, P.S., *Efficient attenuation of stochasticity in gene expression through post-transcriptional control*. J Mol Biol, 2004. **344**(4): p. 965-76.

77. Muffler, A., et al., *The RNA-binding protein HF-I plays a global regulatory role which is largely, but not exclusively, due to its role in expression of the sigmaS subunit of RNA polymerase in Escherichia coli*. J Bacteriol, 1997. **179**(1): p. 297-300.
78. Tsui, H.C., G. Feng, and M.E. Winkler, *Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of Escherichia coli K-12*. J Bacteriol, 1997. **179**(23): p. 7476-87.
79. Bardwell, J.C., et al., *Autoregulation of RNase III operon by mRNA processing*. EMBO J, 1989. **8**(11): p. 3401-7.
80. Levine, E., et al., *Quantitative characteristics of gene regulation by small RNA*. PLoS Biol, 2007. **5**(9): p. e229.
81. Komorowski, M., J. Miekisz, and A.M. Kierzek, *Translational repression contributes greater noise to gene expression than transcriptional repression*. Biophys J, 2009. **96**(2): p. 372-84.
82. Lu, T.K., A.S. Khalil, and J.J. Collins, *Next-generation synthetic gene networks*. Nat Biotechnol, 2009. **27**(12): p. 1139-50.
83. Davidson, E.A. and A.D. Ellington, *Engineering regulatory RNAs*. Trends Biotechnol, 2005. **23**(3): p. 109-12.
84. Bayer, T.S. and C.D. Smolke, *Programmable ligand-controlled riboregulators of eukaryotic gene expression*. Nat Biotechnol, 2005. **23**(3): p. 337-43.
85. Isaacs, F.J., et al., *Engineered riboregulators enable post-transcriptional control of gene expression*. Nat Biotechnol, 2004. **22**(7): p. 841-7.
86. Isaacs, F.J., D.J. Dwyer, and J.J. Collins, *RNA synthetic biology*. Nat Biotechnol, 2006. **24**(5): p. 545-54.
87. Balagadde, F.K., et al., *Long-term monitoring of bacteria undergoing programmed population control in a microchemostat*. Science, 2005. **309**(5731): p. 137-40.
88. Basu, S., et al., *A synthetic multicellular system for programmed pattern formation*. Nature, 2005. **434**(7037): p. 1130-4.
89. Gillespie, D.T., *Exact stochastic simulation of coupled chemical reactions*. The Journal of Physical Chemistry, 1977. **81**(25): p. 2340-2361.
90. Burrage, K., T. Tian, and P. Burrage, *A multi-scaled approach for simulating chemical reaction systems*. Prog Biophys Mol Biol, 2004. **85**(2-3): p. 217-34.
91. Cao, Y., D.T. Gillespie, and L.R. Petzold, *The slow-scale stochastic simulation algorithm*. J Chem Phys, 2005. **122**(1): p. 14116.
92. Zhou, W., et al., *Accelerated stochastic simulation algorithm for coupled chemical reactions with delays*. Comput Biol Chem, 2008. **32**(4): p. 240-2.
93. Lu, T., et al., *A molecular noise generator*. Phys Biol, 2008. **5**(3): p. 036006.
94. Shetty, R.P., D. Endy, and T.F. Knight, Jr., *Engineering BioBrick vectors from BioBrick parts*. J Biol Eng, 2008. **2**: p. 5.
95. Marians, K.J. and J.D. Brooker, *Structure of the lactose operator*. Nature, 1976. **260**(5549): p. 360-3.

96. Gottesman, S., *The small RNA regulators of Escherichia coli: roles and mechanisms\**. Annu Rev Microbiol, 2004. **58**: p. 303-28.
97. Storz, G., J. Vogel, and K.M. Wassarman, *Regulation by small RNAs in bacteria: expanding frontiers*. Mol Cell, 2011. **43**(6): p. 880-91.
98. Andersen, J.B., et al., *New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria*. Appl Environ Microbiol, 1998. **64**(6): p. 2240-6.
99. Zucca, S., et al., *Multi-Faceted Characterization of a Novel LuxR-Repressible Promoter Library for Escherichia coli*. PLoS One, 2015. **10**(5): p. e0126264.
100. Lim, H.N., Y. Lee, and R. Hussein, *Fundamental relationship between operon organization and gene expression*. Proc Natl Acad Sci U S A, 2011. **108**(26): p. 10626-31.
101. Bernstein, J.A., et al., *Global analysis of Escherichia coli RNA degradosome function using DNA microarrays*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2758-63.
102. Yang, E., et al., *Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes*. Genome Res, 2003. **13**(8): p. 1863-72.
103. Baek, D., et al., *The impact of microRNAs on protein output*. Nature, 2008. **455**(7209): p. 64-71.
104. Selbach, M., et al., *Widespread changes in protein synthesis induced by microRNAs*. Nature, 2008. **455**(7209): p. 58-63.
105. Wiesenfeld, K. and F. Jaramillo, *Minireview of stochastic resonance*. Chaos, 1998. **8**(3): p. 539-548.
106. Guantes, R. and J.F. Poyatos, *Dynamical principles of two-component genetic oscillators*. PLoS Comput Biol, 2006. **2**(3): p. e30.
107. Goldbeter, A., *Computational approaches to cellular rhythms*. Nature, 2002. **420**(6912): p. 238-45.
108. Paulsson, J., O.G. Berg, and M. Ehrenberg, *Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation*. Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7148-53.
109. Beisel, C.L. and G. Storz, *Base pairing small RNAs and their roles in global regulatory networks*. FEMS Microbiol Rev, 2010. **34**(5): p. 866-82.
110. Mehta, P., S. Goyal, and N.S. Wingreen, *A quantitative comparison of sRNA-based and protein-based gene regulation*. Mol Syst Biol, 2008. **4**: p. 221.
111. Jost, D., A. Nowojewski, and E. Levine, *Small RNA biology is systems biology*. BMB Rep, 2011. **44**(1): p. 11-21.
112. Zhang, A., et al., *The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs*. Mol Cell, 2002. **9**(1): p. 11-22.
113. Arbel-Goren, R., et al., *Effects of post-transcriptional regulation on phenotypic noise in Escherichia coli*. Nucleic Acids Res, 2013. **41**(9): p. 4825-34.
114. Lavi-Itzkovitz, A., et al., *Quantitative effect of target translation on small RNA efficacy reveals a novel mode of interaction*. Nucleic Acids Res, 2014. **42**(19): p. 12200-11.
115. Salis, H.M., *The ribosome binding site calculator*. Methods Enzymol, 2011. **498**: p. 19-42.

116. Zucca, S., et al., *Characterization of an inducible promoter in different DNA copy number conditions*. BMC Bioinformatics, 2012. **13 Suppl 4**: p. S11.
117. Bar-Even, A., et al., *Noise in protein expression scales with natural protein abundance*. Nat Genet, 2006. **38**(6): p. 636-43.
118. Waters, J.C., *Accuracy and precision in quantitative fluorescence microscopy*. J Cell Biol, 2009. **185**(7): p. 1135-48.
119. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation*. Nat Biotechnol, 2007. **25**(1): p. 117-24.
120. Huh, D. and J. Paulsson, *Non-genetic heterogeneity from stochastic partitioning at cell division*. Nat Genet, 2011. **43**(2): p. 95-100.
121. Ullman, G., et al., *High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking*. Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1611): p. 20120025.
122. Donachie, W.D., *Relationship between cell size and time of initiation of DNA replication*. Nature, 1968. **219**(5158): p. 1077-9.
123. Cooper, S. and C.E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coli B/r*. J Mol Biol, 1968. **31**(3): p. 519-40.
124. Robert, L., et al., *Division in Escherichia coli is triggered by a size-sensing rather than a timing mechanism*. BMC Biol, 2014. **12**: p. 17.
125. Rigney, D.R., *Stochastic model of constitutive protein levels in growing and dividing bacterial cells*. J Theor Biol, 1979. **76**(4): p. 453-80.
126. Aiba, H., *Mechanism of RNA silencing by Hfq-binding small RNAs*. Curr Opin Microbiol, 2007. **10**(2): p. 134-9.
127. Masse, E., F.E. Escorcia, and S. Gottesman, *Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli*. Genes Dev, 2003. **17**(19): p. 2374-83.
128. Wang, R.S., et al., *Modeling post-transcriptional regulation activity of small non-coding RNAs in Escherichia coli*. BMC Bioinformatics, 2009. **10 Suppl 4**: p. S6.
129. Vallania, F.L., et al., *Origin and consequences of the relationship between protein mean and variance*. PLoS One, 2014. **9**(7): p. e102202.
130. Stockwell, S.R., C.R. Landry, and S.A. Rifkin, *The yeast galactose network as a quantitative model for cellular memory*. Mol Biosyst, 2015. **11**(1): p. 28-37.
131. Bhat, P.J. and R.S. Iyer, *Epigenetics of the yeast galactose genetic switch*. J Biosci, 2009. **34**(4): p. 513-22.
132. Barnett, J.A., *A history of research on yeasts 7: enzymic adaptation and regulation*. Yeast, 2004. **21**(9): p. 703-46.
133. Thoden, J.B., et al., *Molecular structure of Saccharomyces cerevisiae Gal1p, a bifunctional galactokinase and transcriptional inducer*. J Biol Chem, 2005. **280**(44): p. 36905-11.

134. Lohr, D., P. Venkov, and J. Zlatanova, *Transcriptional regulation in the yeast GAL gene family: a complex genetic network*. FASEB J, 1995. **9**(9): p. 777-87.
135. Sellick, C.A., R.N. Campbell, and R.J. Reece, *Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them*. Int Rev Cell Mol Biol, 2008. **269**: p. 111-50.
136. Bennett, M.R., et al., *Metabolic gene regulation in a dynamically changing environment*. Nature, 2008. **454**(7208): p. 1119-22.
137. Johnston, M., J.S. Flick, and T. Pexton, *Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae*. Mol Cell Biol, 1994. **14**(6): p. 3834-41.
138. Escalante-Chong, R., et al., *Galactose metabolic genes in yeast respond to a ratio of galactose and glucose*. Proc Natl Acad Sci U S A, 2015. **112**(5): p. 1636-41.
139. Kaplan, S., et al., *Diverse two-dimensional input functions control bacterial sugar genes*. Mol Cell, 2008. **29**(6): p. 786-92.
140. Setty, Y., et al., *Detailed map of a cis-regulatory input function*. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7702-7.
141. Christacos, N.C., et al., *Subcellular localization of galactose-1-phosphate uridylyltransferase in the yeast Saccharomyces cerevisiae*. Mol Genet Metab, 2000. **70**(4): p. 272-80.
142. Lai, K. and L.J. Elsas, *Overexpression of human UDP-glucose pyrophosphorylase rescues galactose-1-phosphate uridylyltransferase-deficient yeast*. Biochem Biophys Res Commun, 2000. **271**(2): p. 392-400.
143. De Robichon-Szulmajster, H., *Induction of enzymes of the galactose pathway in mutants of Saccharomyces cerevisiae*. Science, 1958. **127**(3288): p. 28-9.
144. Douglas, H.C. and D.C. Hawthorne, *Enzymatic Expression and Genetic Linkage of Genes Controlling Galactose Utilization in Saccharomyces*. Genetics, 1964. **49**: p. 837-44.
145. Ideker, T., et al., *Integrated genomic and proteomic analyses of a systematically perturbed metabolic network*. Science, 2001. **292**(5518): p. 929-34.
146. Sellick, C.A. and R.J. Reece, *Eukaryotic transcription factors as direct nutrient sensors*. Trends Biochem Sci, 2005. **30**(7): p. 405-12.
147. Rubio-Teixeira, M., *A comparative analysis of the GAL genetic switch between not-so-distant cousins: Saccharomyces cerevisiae versus Kluyveromyces lactis*. FEMS Yeast Res, 2005. **5**(12): p. 1115-28.
148. Abramczyk, D., et al., *Interplay of a ligand sensor and an enzyme in controlling expression of the Saccharomyces cerevisiae GAL genes*. Eukaryot Cell, 2012. **11**(3): p. 334-42.
149. Kundu, S. and C.L. Peterson, *Dominant role for signal transduction in the transcriptional memory of yeast GAL genes*. Mol Cell Biol, 2010. **30**(10): p. 2330-40.
150. Kar, R.K., et al., *Stochastic galactokinase expression underlies GAL gene induction in a GAL3 mutant of Saccharomyces cerevisiae*. FEBS J, 2014. **281**(7): p. 1798-817.
151. Nogi, Y., *GAL3 gene product is required for maintenance of the induced state of the GAL cluster genes in Saccharomyces cerevisiae*. J Bacteriol, 1986. **165**(1): p. 101-6.

152. Torchia, T.E. and J.E. Hopper, *Genetic and molecular analysis of the GAL3 gene in the expression of the galactose/melibiose regulon of Saccharomyces cerevisiae*. Genetics, 1986. **113**(2): p. 229-46.
153. Zacharioudakis, I., T. Gligoris, and D. Tzamarias, *A yeast catabolic enzyme controls transcriptional memory*. Curr Biol, 2007. **17**(23): p. 2041-6.
154. Acar, M., A. Becskei, and A. van Oudenaarden, *Enhancement of cellular memory by reducing stochastic transitions*. Nature, 2005. **435**(7039): p. 228-32.
155. Biggar, S.R. and G.R. Crabtree, *Cell signaling can direct either binary or graded transcriptional responses*. EMBO J, 2001. **20**(12): p. 3167-76.
156. Wang, S. and E.J. Tran, *Unexpected functions of lncRNAs in gene regulation*. Commun Integr Biol, 2013. **6**(6): p. e27610.
157. Pelechano, V. and L.M. Steinmetz, *Gene regulation by antisense transcription*. Nat Rev Genet, 2013. **14**(12): p. 880-93.
158. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome*. Science, 2005. **309**(5740): p. 1564-6.
159. Geisler, S., et al., *Decapping of long noncoding RNAs regulates inducible genes*. Mol Cell, 2012. **45**(3): p. 279-91.
160. Crane, M.M., et al., *A microfluidic system for studying ageing and dynamic single-cell responses in budding yeast*. PLoS One, 2014. **9**(6): p. e100042.
161. Houser, J.R., et al., *An improved short-lived fluorescent protein transcriptional reporter for Saccharomyces cerevisiae*. Yeast, 2012. **29**(12): p. 519-30.
162. Kis, Z., et al., *Mammalian synthetic biology: emerging medical applications*. J R Soc Interface, 2015. **12**(106).



# 1 Appendix to experimental measurements and mathematical modelling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits

## 1.1 Using the moment generating function to derive analytical expressions for protein mean and variance

As highlighted in the introduction, the probability distribution of the system being in state  $\mathbf{x}$  at time  $t$ , given the state  $\mathbf{x} = \mathbf{x}_0$  at the initial time  $t_0$ ,  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$  constitutes the exact solution of the CME. While the full probability distribution can be analytically determined only in rare cases, most of physically accessible information it conveys might be summarized by its first and second order moments. If the propensity functions of the reaction channels occurring within the system are constant or linear in the chemical species, as is the case for zeroth- and first-order reactions, the moment generating function, e.g. the z-transform of  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ , can be used to derive analytical expression for the steady-state mean and variance of the interacting molecules.

In this paragraph we will use the moment generating function to derive analytical expressions of the indexes quantifying the stochasticity in the expression of the green fluorescent reporter gene in the TC gene circuit.

The TC gene circuit's behaviour can be described by the following set of reactions:



Reactions 1.1-1.4 represent synthesis and degradation of TRANS-GFP mRNA ( $M_{TG}$ ) and GFP ( $G$ ) molecules. In these reactions, the dependence of the transcription rate on the inducer concentration,  $k_{r,TG}(IPTG) = k_{r,TG}$ , has been neglected for the sake of a simpler representation.

At each time  $t$ , the state of the system is defined by the number of molecules counts of each chemical species:

$$\mathbf{x} = \begin{pmatrix} M_{TG} \\ G \end{pmatrix} \quad 1.5$$

With  $P(\mathbf{x}, t | \mathbf{x}_0, t_0) = P(m_{TG}, g)$ , the CME for this model reads:

$$\begin{aligned} \frac{\partial P(m_{TG}, g)}{\partial t} = & k_{r,TG}P(m_{TG} - 1, g) + \gamma_r(m_{TG} + 1)P(m_{TG} + 1, g) \\ & + k_p m_{TG} P(m_{TG}, g - 1) + \gamma_p(g + 1)P(m_{TG}, g + 1) \\ & - k_{r,TG}P(m_{TG}, g) - \gamma_r m_{TG} P(m_{TG}, g) \\ & - k_p m_{TG} P(m_{TG}, g) - \gamma_p g P(m_{TG}, g) \end{aligned} \quad 1.6$$

The first two rows composing the right hand side of equation 1.6, depict the occurrence of reactions of synthesis and degradation, for  $m_{TG}$  and  $G$  respectively, towards state  $(m_{TG}, g)$ . Analogously, the remaining rows describe synthesis and degradation of the chemical species from the state  $(m_{TG}, g)$ .

For this two-dimensional system, the moment generating function is defined by:

$$F(z_1, z_2) \equiv \sum_{m_{TG}=0}^{+\infty} \sum_{g=0}^{+\infty} z_1^{m_{TG}} z_2^g P(m_{TG}, g). \quad 1.7$$

Applying the time-derivative to equation 1.7 and substituting for equation 1.6 yields:

$$\begin{aligned} \frac{\partial F(z_1, z_2)}{\partial t} = & k_{r,TG} z_1 F(z_1, z_2) - k_{r,TG} F(z_1, z_2) \\ & + \gamma_r \frac{\partial F(z_1, z_2)}{\partial z_1} + k_p z_1 z_2 \frac{\partial F(z_1, z_2)}{\partial z_1} - \gamma_r z_1 \frac{\partial F(z_1, z_2)}{\partial z_1} - k_p z_1 \frac{\partial F(z_1, z_2)}{\partial z_1} \\ & + \gamma_p \frac{\partial F(z_1, z_2)}{\partial z_2} - \gamma_p z_2 \frac{\partial F(z_1, z_2)}{\partial z_2} \end{aligned} \quad 1.8$$

The steady-state mean and variance for  $m_{TG}$  and  $g$  can be derived as:

$$\frac{\partial F(1,1)}{\partial z_1} = \langle m_{TG} \rangle = \frac{k_{r,TG}}{\gamma_r} \quad 1.9$$

$$\frac{\partial^2 F(1,1)}{\partial z_1^2} = \langle m_{TG}^2 \rangle - \langle m_{TG} \rangle^2 = \frac{k_{r,TG}}{\gamma_r} \quad 1.10$$

$$\frac{\partial F(1,1)}{\partial z_2} = \langle g \rangle = \frac{k_p k_{r,TG}}{\gamma_p \gamma_r} \quad 1.11$$

$$\frac{\partial^2 F(1,1)}{\partial z_2^2} = \langle g^2 \rangle - \langle g \rangle^2 = \frac{k_p k_{r,TG}}{\gamma_p \gamma_r} \left( 1 + \frac{\frac{k_p}{\gamma_r}}{1 + \frac{\gamma_p}{\gamma_r}} \right) \quad 1.12$$

The evaluation of the *squared coefficient of variation* and *Fano factor* quantifying the stochasticity in the fluorescent reporter expression hence provides:

$$CV^2 = \frac{\langle g^2 \rangle - \langle g \rangle^2}{\langle g \rangle^2} = \frac{\gamma_p \gamma_r}{k_{r,TG}} \left( \frac{1}{k_p} + \frac{1}{\gamma_p + \gamma_r} \right) \quad 1.13$$

$$F = \frac{\langle g^2 \rangle - \langle g \rangle^2}{\langle g \rangle} = 1 + \frac{\frac{k_p}{\gamma_r}}{1 + \frac{\gamma_p}{\gamma_r}}. \quad 1.14$$



# Scientific writing

## Publications in international journals

1. L. Bandiera, S. Furini and E. Giordano, “*Phenotypic variability in synthetic biology applications: dealing with noise in microbial gene expression*”, *Frontiers in Microbiology- Microbiology, Ecotoxicology and Bioremediation*, 2016 (provisionally accepted for publication).
2. L. Bandiera, A. Pasini, L. Pasotti, S. Zucca, G. Mazzini, P. Magni, E. Giordano and S. Furini, “*Experimental measurements and mathematical modeling of biological noise arising from transcriptional and translational regulation of basic synthetic gene circuits*”, *J Theor Biol*, 2016. **395**: p. 153-60.

## Abstracts in proceedings of international conferences

1. L. Bandiera, A. Pasini, M. Cortesi, E. Giordano and S. Furini., ‘*Reducing phenotypic variability in Synthetic Biology applications via post-transcriptional control of noise in gene expression*’, CAS Synthetic Biology II, Biocenter Ludwig-Maximilians university, 26-30 July 2015, Munich, Germany.
2. L. Bandiera, A. Pasini, M. Cortesi, E. Giordano and S. Furini., ‘*In vitro/in silico analysis of phenotypic noise under transcriptional and post-transcriptional control in elementary synthetic gene-circuits*’, *Network Models in Cellular Regulation, International School and Conference on Network Science*, 1-3 June 2015, Zaragoza, Spain.
3. L. Bandiera, S. Furini, A. Pasini, M. Cortesi and E. Giordano, “*A synthetic circuit to control noise in gene expression*”, *Proceedings of the Fourth Bioengineering National Congress*, 25-27 July 2014; Pavia, Italy.
4. M. Cortesi, L. Bandiera, A. Pasini and E. Giordano, “*A computational approach for the automatic classification of mesenchymal vs epithelial cancer cell phenotype*”, *Third International Conference “Translational Research in Oncology”*, 7-9 May 2014; Forlì, Italy.

5. L. Bandiera, S. Furini and E. Giordano, “A *genetic circuit to modulate noise and analyse its effects in synthetic devices*”, Proceedings of the 6<sup>th</sup> International Meeting on Synthetic Biology, Imperial College, 9-11 July 2013, London, UK.