Noise propagation in *Escherichia coli*'s regulatory network

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Arantxa Urchueguía Fornes
aus Spanien

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Erik van Nimwegen
Prof. Dr. Urs Jenal

Basel, den 23.04.2019

Prof. Dr. Martin Spiess
Dekan

A mis padres y a mi madrina, Marina.

Abstract

The ability to regulate gene expression allows bacteria to grow under diverse conditions, often involving large regulatory networks. As gene expression is an inherently stochastic process, accurate regulation will only be achieved if the molecules involved in the process adapt perfectly to the different conditions and show low noise themselves. In *Escherichia coli* it has been reported that high noise promoters are characterized by containing a large number of regulatory binding sites in their sequences, suggesting that regulation and noise may be intimately coupled. However, little is known about the extent by which this coupling determines high noise levels in bacteria or even how individual promoters vary their noise in response to changes in the environment.

The work presented in this thesis aims to elucidate the main determinants of condition-dependent noise in *Escherichia coli*, and to systematically investigate how noise properties vary across environments. To do so, we have quantified the genome-wide variation in transcriptional noise across 8 diverse growth conditions using a combination of flow cytometry, high-throughput microscopy, and computational modeling. We find that a promoter's noise can be decomposed into two components: a promoter-independent component, linearly decreasing as a function of growth rate, and equally affecting all promoters in a given condition, and a noise propagation component resulting from noise propagation from transcription factors. Using a simple linear model we identify a set of TFs that contribute to condition-specific and condition-independent noise propagation. Moreover, the overall correlation structure of genome-wide expression properties uncovers that genes are organized along two principal axes, with the first one sorting genes by their mean expression and evolutionary rate, and the second one by their expression noise, number of regulatory inputs and expression plasticity. Overall, the results of the thesis show that noise and regulation are intimately coupled, and that this coupling is explained by noise propagation which has evolved independently of a promoter's expression level or evolutionary rate in its coding region. This work provides the first systematic investigation of genome-wide noise properties across several growth conditions and reveals that noise propagation is the main mechanism underlying the tight coupling between gene expression noise and gene regulation.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Mechanisms of gene expression in bacteria

Gene expression is the most fundamental processes in biology, as it allows to 'read' the genetic code stored as DNA (deoxyribonucleic acid) and translate it into proteins, which will ultimately perform all functions that cells require to grow, survive and replicate. Since the discovery of DNA by Friedrich Miescher in 1869 [1], keystone discoveries set the grounds for our current understanding of the way genes are structured and their expression mechanisms. James Watson and Francis Crick with contributions from Rosalind Franklin and Maurice Wilkins solved the DNA structure in 1953 [2], and in 1958 Crick postulated the central dogma [3, 4] in biology, which describes the flow of genetic information in a cell. But it was not until 1963 that the genetic code (how to 'read' DNA and translate it into the aminoacids that conform the proteins) was solved by Nirenberg [5, 6]. Much time has passed since these early studies and today we have a clearer picture of the mechanisms and molecules involved in gene expression. This section focuses on the current understanding of the molecular machinery that allows bacteria, specifically *Escherichia coli*, express their genes and synthesize proteins.

The first step in the synthesis of proteins is the initiation of transcription, which is the key process undergoing regulation and is believed to be one of the major determinants of gene expression levels in bacteria [7]. The central molecule involved is the multi-subunit DNA-dependent RNA polymerase (RNAP) [8]. The RNAP is a protein complex that consists of a core enzyme of five sub-units $\beta\beta'\alpha_2\omega$ [9] that is capable of binding to the DNA and initiate transcription, but not to initiate a

'controlled' or promoter-mediated transcription initiation (i.e. under regulation) [10]. The sub-units involved in the binding to the DNA and the conformational structural changes they adopt have been extensively described (see, for example, [11, 12]). However, for the core enzyme to be able to start promoter mediated transcription initiation, it must form a 'holoenzyme' by interacting with small sub-units named sigma factors, σ [13]. Bacteria contain different classes of sigma factors each allowing the regulation of a different set of genes. In *Escherichia coli*, seven types of sigma factors control the response to distinct environmental stimuli [14, 15], although the main factor, σ^{70}, allows the polymerase to recognize most promoters. Other mechanisms, such as the binding from transcription factors (TFs), also contribute to the control of the specific binding of the RNAP to different promoters and the transcription of the genes they regulate. These mechanisms are described in more detail in the section 1.3 of this chapter.

For the recognition between the holoenzyme complex and the DNA, four sequence elements are necessary: an UP element (at position -37 and -58 from the transcription start site +1), the -35 element (positions -35 to - 30), the - 10 element (positions -17 to -14) and a discrimination element (-6 to -4), all of which are recognized by the sigma factors, excepting the UP element which interacts with the C-terminal domains of the RNAP α subunit [10]. The activity of a promoter (i.e. gene expression level) is often determined by these elements and the 'strength' by which the holoenzyme complex binds to them. A classical way in which the strength of the binding is determined is by describing the interaction between RNAP and DNA in terms of sequence-dependent binding energies, in which each individual nucleotide along the binding site contributes to the total energy independently [16–18]. In Kinney *et al.* 2010 [17], for example, the sequence-dependent energy binding matrix between the σ70-RNAP complex was quantitatively estimated using a combination of computational and experimental techniques.

An important aspect of transcription initiation is that the number of free polymerase molecules is limited ($\sim$ 2000-3000 molecules per cell, depending on the growth rate [15, 19]), as well as the supply of sigma factors. These numbers are small compared to the number of genes a typical bacteria contains (4694 in *E.coli* as documented in RegulonDB [20] at the moment), and therefore promoters must compete in order to attract the holoenzyme complex to their sequences [21]. It has been proposed that promoters have evolved a *pick-and-mix* mechanism in their sequences that results in different promoter hierarchies when it comes to transcription

initiation frequencies [22, 23]. Once a RNAP-holoenzyme is bound to a promoter, an open complex is formed in which the double stranded DNA around the transcription start site is unwounded. This allows the synthesis of a DNA-templated RNA chain (mRNA) in a process termed elongation. Since promoters can only be occupied by one RNAP at a time, the transition to elongation is rate limiting. For example, it's been observed that RNAPs are stalled in many promoters [24]. Around 28% of RNAPs are non-specifically bound to the DNA (sliding through the DNA in a three dimensional space), whereas only 12% of them are free [19].

Elongation is the final stage in transcription and it is well established that its termination is critical for the generation of different functional classes of mRNA [25]. In bacteria, two main mechanisms have been described to achieve transcription termination: intrinsic sequence elements (GC-rich followed by a T stretch [26]) that dissociate the RNAP-holoenzyme complex through the formation of sequence-specific RNA secondary structures [27, 28] and rho-dependent elements that dissociate the complex via a RNA helicase (termed Rho) [25]. Proteins are then synthesized via translation from the produced mRNA chains through large macromolecular complexes, called ribosomes, in intervals of few seconds [29]. In [30] all components and processes involved in translation and protein synthesis have been reviewed. Ribosomes are typically ensemble into polysomes, which consist of two or more ribosomes together with an average spacing of 77 nucleotides between them [29]. The polysome-ensemble allows fast production of proteins from the same mRNA molecule, although it has been proposed that queuing and collisions can have a detrimental effect on translation efficiency [31]. Changes in translation efficiency (i.e. how many proteins can be synthesized from one mRNA) are determined by various factors. The ribosomal binding site (RBS) which is placed upstream of the translation start site is a key element for translation efficiency [32], as secondary structures and folding of the RNA around it are known to lower translation rates [33]. Stability of mRNA [34], translation elongation speed or codon usage [35, 36] also influence translation efficiency. The process of synthesis terminates with one of the three stop codons (UAG, UGA, UAA) that are recognized by release factors (RF1 and RF2), which induce peptide release. Ribosomes are then recycled for new rounds of synthesis [37]. Apart from transcription and translation rates, expression levels are also influenced by the number of proteins and mRNAs present in the cell in a given moment, which is determined by their half-lives and degradation rates. Proteins are stable and have, typically, longer half-lives (around 20h [38]) than the

generation time in *E.coli* (around 20 minutes in nutrient rich environments [39]). Therefore, protein levels are affected by dilution from growth rather than degradation. On the contrary, mRNAs have very short half-lives (between 3 and 8 minutes [40]) and their levels are affected by degradation processes.

In the past years, great efforts have been made in the development of experimental quantitative techniques to measure protein and mRNA levels, as well as gene expression dynamics in bacteria [41–45]. These methods have stimulated the development of thermodynamic and rate-equation models that aim to mathematically model the distributions of mRNA and protein levels in individual cells [46, 47].

1.2 The importance of gene regulation

Bacteria are often exposed to drastic changes in the environment they live in. These can comprise changes in nutrient quality or availability, temperature changes, oxidative stress, etc. In order to survive, it is essential for them to be able to change their transcriptional response according to the environmental stimuli they receive and synthesize the proteins that are important under their current circumstances. As explained in the previous section, the number of free RNAP molecules available for transcription is small compared to the total number of genes. It is, therefore, essential for bacteria to have regulatory mechanisms that attract the RNAP to the genes that need to be expressed in a given moment. In bacteria, genes are organized in operons, allowing genes involved in a specific function to be co-regulated. The most well studied one in *E.coli* is the *lac* operon (discovered by Jacob and Monod in 1961 [48]), which coordinates the response to the switch from glucose to lactose as main carbon source. Gene expression adaption to environmental changes is mainly achieved by regulating transcription initiation and, specially, by guiding the binding of the RNAP to the promoters of the important genes in each condition. In order to achieve this, external stimuli need to be integrated by external sensors and transmitted internally by response regulators, which will modify the transcriptional response accordingly [49]. The next section focuses on the response regulators in *E.coli* that allow cells to change their transcriptional program as a response to environmental changes.

1.3 *Escherichia coli*'s regulatory network

Transcription initiation can be controlled at different levels: either by regulating the formation of the RNAP holoenzyme complex, facilitating/impeding promoter recognition by the holoenzyme or by changing a promoter's activity [10]. The factors that interact with the RNAP to form the holoenzyme are, mainly, the sigma factors. In *E.coli*, there are seven distinct types and each guide the RNAP to different promoters, constituting the simplest layer of regulation [14]. The different factors compete with each other for free RNAP, although under most conditions, the housekeeping factor (σ^{70}) outcompetes the others. σ^{70} regulates the vast majority of promoters, including those of all essential genes [15]. However, under specific conditions, the level of one or the other alternative σ factor increases and displaces σ^{70} [21]. These alternative factors include: σ^{54} (rpoN), σ^{38} (rpoS), σ^{32} (rpoH), σ^{28} (rpoF; fliA), σ^{24} (rpoE), σ^{19} (fecI); and they play a role in different scenarios [20]. σ^{38} is essential for transcription of stationary phase and general stress genes, σ^{54} for the response to nitrogen deficiency, σ^{32} for the heat-shock response, σ^{28} for flagellar and chemotaxis genes, σ^{24} for genes involved in the correction of misfolded proteins that may be caused by different stresses and σ^{19} is involved in the activation of the fec operon, which regulates the expression of genes for ferric citrate transport [15]. All factors, except σ^{54}, share a common ancestor [50]. The best well studied example of sigma factors exchange is between the housekeeping factor σ^{70} and σ^{38} [51] and is known to be facilitated by small molecules (Crl, Rsd and ppGpp), which respond to changes in cellular state [10].

There are other small factors, apart from the sigma factors, that are known to also interact with RNAP. In *E.coli*, DksA stabilizes or destabilizes the RNAP-DNA complexes in cooperation with ppGpp as a response to certain metabolic stresses [52]. Some factors, like 6S RNA, sequester the holoenzyme complex lowering the number of available molecules to initiate transcription [53]. Another factor, SoxS, is known to guide the RNAP to specific promoters by preventing the holoenzyme to interact with the promoters UP elements, but instead with Sox-box sequences only present in the promoters of genes involved in oxidative stress [54]. Moreover, changes in NTP levels can also modulate the activity of RNAP. This modulation has been shown to affect genes involved in pyrimidine biosynthesis, specially [55].

The second most important layer of regulation is achieved by transcription factors that directly target promoter sequences and activate or repress transcription initiation. Repression is mainly achieved by blocking the access of the RNAP-holoenzyme to the -10 and -35 elements. Activation, which facilitates the recruitment of RNAP to the promoter, is typically separated in two distinct classes: in class I activation, the activator binds to the promoter and recruits the RNAP through protein-protein interactions, and in class II activation, the activator binds to the -35 element and recruits the RNAP. Another class of activation involves the binding of the regulator to sequence elements between the -10 and -35 elements to bring them closer and facilitate the binding of the holoenzyme [10]. It should be noted that, in fact, many factors can act both as repressors and activators (dual factors) [20]. Transcription factors respond to external stimuli sensed by transporters and signal transduction mechanisms [49]. Typically, transcription factors contain two domains: one that senses external stimuli and another that interacts with the DNA through structural motifs that recognize specific sequences (being the helix-turn-helix motif the most common element for DNA binding in bacteria) [56]. In many cases, binding is achieved after oligomerization of the transcription factor and bounding of an effector molecule [57]. Other strategies for transcription initiation control include: DNA methylation by Dam, DNA supercoiling control or transient DNA inversions which reverse the orientation of promoters [10], but they will not be discussed in the scope of this thesis.

Transcription factors regulate more than one promoter and promoters are regulated by more than one transcription factor. Moreover, the expression of many transcription factors is regulated by other transcription factors, as well as by themselves [56], generating very complex regulatory dynamics. In *E.coli*, seven regulatory proteins control the expression of around 51% of all genes (CRP, FNR, IHF, Fis, ArcA, LrP and H-NS) and their control is amplified by addition of other transcription factors in hierarchical layers [56, 58], adding more complexity to the regulatory interactions within the network (Figure 1.1 illustrates the hierarchical structure of *E.coli*'s regulatory network). Even though *E.coli*'s regulatory network is complex, it has been shown that transcription networks contain a set of recurring regulation patterns (network motifs) from which all interactions are built on. This network motifs have been classified in [59] and cover most of the known regulatory interactions in *E.coli*. Extensive experimental [60–62] and computational methods [18, 63–65] have been developed to identify transcription factors binding sites in genome

sequences and infer regulatory interactions. Every year, more and more information is available thanks to advances in the development of these methods. Databases, such as RegulonDB [20] or Ecocyc [66], provide curated up-to-date knowledge of identified binding sites, as well as information on the specific roles of the different transcription factors (activator, repressor or dual).

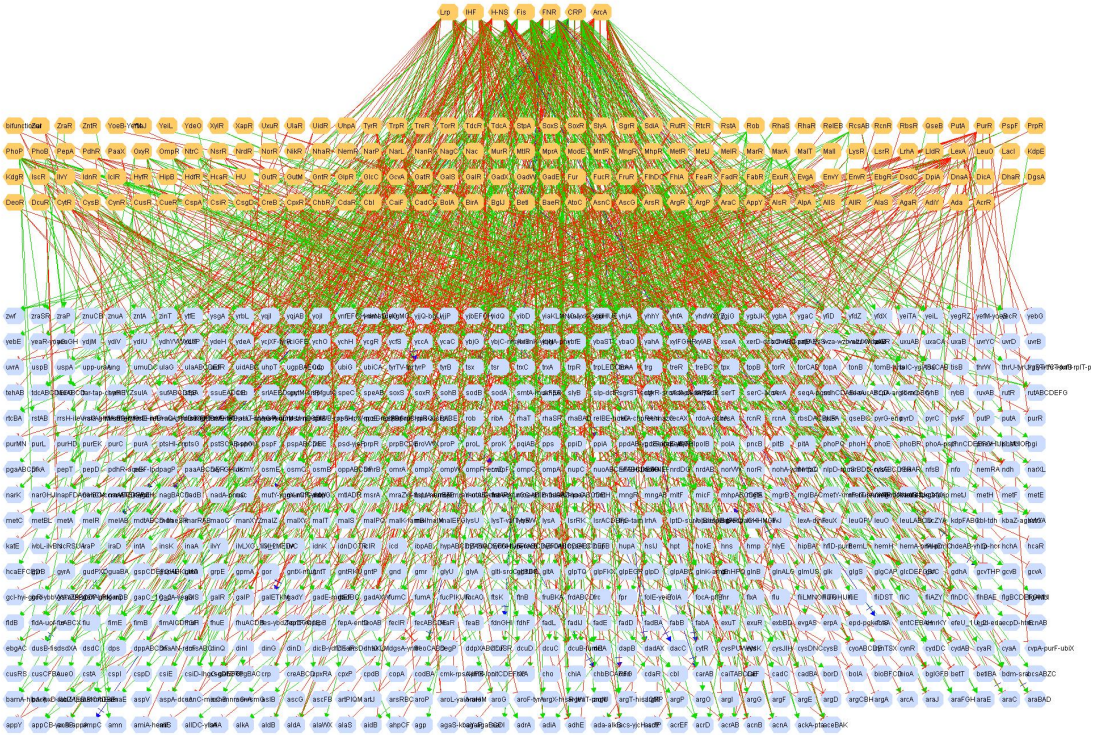


Figure 1.1: TF-operon network downloaded from RegulonDB [20]. Illustration of the hierarchy structure of *E.coli*'s regulatory network. As cited in RegulonDB [20]: "In this transcriptional network, the nodes represent the transcription factors (TFs) and their regulated genes; the edges represent the relation of the transcription regulation and the arrows indicate a relation directed from the TFs to the operons they regulate".

A fundamental aspiration in systems biology is the development of computational methods that will ultimately allow the prediction of gene expression levels or regulatory interactions from any given input [65]. An example of such computational methods are simple linear models [67, 68], which can be used to infer the effect of reg-

ulatory binding sites in the promoter sequence of different genes on their expression level. However, a limitation for the use of these models is the lack of understanding of how some of the regulatory interactions act *in vivo*. At the moment, there are roughly 215 transcription factors, 3355 regulatory interactions and 4694 genes annotated in RegulonDB [20] for *E.coli*. Nevertheless, more than half of the genes do not contain any regulatory annotations, implying that there is an underestimation of all possible regulatory interactions in *E.coli*'s genome. Moreover, recent studies also seem to indicate that our current understanding of *E.coli*'s regulatory network is not enough to explain experimentally measured gene expression levels [69]. Although recent efforts have been made in trying to experimentally quantify the levels of different transcription factors across different conditions [43, 70] and their genome-wide binding sites (both computationally and experimentally [18, 61], to cite some examples), there might be major challenges in accurately doing so. One of these challenges is the fact that for many of the different transcription factors we do not really know in which conditions they may be active or bound to the genes they regulate.

1.4 Noise in gene expression

In an ideal world, organisms would be able to precisely regulate the levels of gene expression required at any given time and in any given condition. Being able to do so would, nevertheless, require having a large arsenal of diverse molecular machineries for each possible condition. This is of course impossible, as genome sizes are limited and the energetic cost of maintaining the different machineries would be too high and detrimental for an organism fitness. It is since long known now that gene expression is highly stochastic (i.e. noisy), implying that identical cells growing in homogeneous environments will show cell-to-cell differences in gene expression levels [71, 72]. These differences arise from the inherent randomness and the low copy numbers associated to the molecules involved in gene expression, unavoidably limiting its precision [73]. Noise in gene expression has been proposed to help organisms display a variety of responses in different environments without the cost of having to express, in each environment, the precise molecular response to it [74–76]. This section focuses on the current understanding of the origins of gene expression noise and its functional role.

The first quantitative experimental measurements of gene expression fluctuations across single-cells by the seminal work of [77] and [78] led to the development of a new field in which researchers have combined experimental and theoretical efforts to disentangle the origins and consequences of gene expression noise [79–81]. Since then, it has been now experimentally [77], and theoretically [82] proved that variations in mRNA and protein levels arise from two independent noise sources: intrinsic and extrinsic; both appearing across different taxa [41, 79, 83–86]. Intrinsic noise, which is the result of the inherent stochasticity of the molecules involved in gene expression, can be modelled by assuming Poissonian statistics in transcription and translation rates (i.e. constant rates of expression), where the squared of coefficient of variation in protein levels ($CV^2 = \frac{\sigma_p^2}{\mu_p^2}$) scales with $\frac{1}{\mu_p}$ [41, 87]. This model of gene expression, however, is able to explain noise levels of low expressed genes only (<10 protein molecules per cell, according to [41]). The deviations in CV^2 that highly expressed (>10 protein molecules per cell [41]) genes show, which scale with the square of the mean, are more consistent with a model in which promoters switch between on and off states in stochastic bursts [45, 88, 89]. Extrinsic noise sources, such as cell-to-cell differences in copy numbers of the different molecules involved in expression (transcription factors, ribosomes, polymerases, etc.), cell cycle dependent changes or changes in DNA supercoiling have been attributed to affect the frequency of bursts between different cells [45, 90–94]. However, making a clear distinction between all possible extrinsic noise sources remains challenging, mainly due to the difficulty in isolating each of them from each other. Although both scaling entities will contribute to some degree to the noise of promoters expressing at different levels, there is now clear evidence that different genes exhibit different levels of gene expression noise and that theses differences are to some extent highly dependent on the specific features of their promoter sequences [90, 95–98], implying that noise is an evolvable trait. If noise is subject to selection, what determines the noise level of a promoter and how is it exploited by natural selection?

Noise has been classically seen as an undesirable effect of gene regulation, with some studies arguing that natural selection may have acted to minimize noise levels [99–101]. In *E.coli*, for example, essential genes tend to have lower noise levels [95]. Still, there is now increasing experimental [74, 102–104] and theoretical [75, 76, 96] evidence showing that noise is another mechanism by which phenotypic variability can arise across genetically identical individuals, increasing the chances of survival under fluctuating environments via 'bet-hedging' or division of labor.

This functional role is, thus, favorable for positive selection. In a recent study, a set of synthetic *E.coli* promoters were evolved *de novo* based on their mean, but not noise properties [96]. Surprisingly, all evolved sequences showed noise levels close to the minimal noise displayed by some native promoters, implying that low noise is the default state of promoters and that high noise promoters have been positively selected by natural selection.

1.5 The relationship between gene regulation and noise

Strikingly, in the same study, it was also shown that the major feature of native high noise promoters in *E.coli* is that they are highly regulated (i.e. contain a large number of binding sites for known transcription factors in their promoter sequences). The observed link between regulation and noise was argued to emerge from noise propagation from the transcription factors to the genes they regulate [96]. Since transcription factors are themselves subject to experience noise (e.g. through cell-to-cell differences in copy numbers [105] or binding dynamics [86, 97, 106]), their noise will unavoidably be transmitted through the regulatory cascade to their targets [107]. Moreover, noise propagation also implies that noise is highly condition-dependent, as transcription factors will also likely change their noise levels in different conditions. In the context of a general theory, it was proposed that noise propagation may have promoted the evolution of early regulatory mechanisms from states without regulation [96]. The theory predicts that, whenever an unregulated promoter couples to a transcription factor, both its mean and noise will become correlated with the mean activity and noise of the transcription factor in a condition-dependent manner. In the study, they investigated how the mean condition-response and noise propagation coupling affected fitness. It was found that coupling to a noisy regulator can improve fitness whenever regulation is imprecise, as some genes will benefit from its noise propagation. Under some circumstances, the coupling may even be more beneficial than evolving a highly precise regulation, what may result in positive selection to keep certain transcription factor's noise levels relatively high. However, to what extent noise propagation is shaping condition-dependent genome-wide noise levels, and which transcription factors are most responsible of

noise propagation across different conditions in *E.coli*'s regulatory network remains largely unknown.

1.5.1 Noise propagation in regulatory networks

Considering the complexity of *E.coli*'s regulatory network (Section 1.3), noise propagation from transcription factors is likely playing an important role in shaping genome-wide noise levels. For example, the fact that co-regulated genes show correlated gene expression fluctuations in yeast [108], demonstrates that noise propagation is an important source of variability. It has also been established that transcription factor binding and unbinding constitutes another important source of variability in eukaryotes [86, 97, 106]. The seminal work of Pedraza *et al.* 2005 [109], where the expression of each gene of a synthetic regulatory cascade was monitored in single-cells in *E.coli*, demonstrated that fluctuations in the expression of one gene will affect the noise levels of all the other downstream genes as a consequence of noise propagating through the regulatory cascade. Growth, for example, has also been shown to be affected by noise propagation, as fluctuations in growth rate have been linked to fluctuations in the concentration of metabolic enzymes across single-cells in *E.coli* [110]. There has been extensive theoretical work analyzing how noise can be propagated through regulatory networks [107, 111]. Notably, it has been mathematically proved that there are fundamental molecular constraints in the ability to suppress noise propagation in genetic networks as the amount of proteins required to reduce it are very large [112]. Experimentally, it has been measured that in synthetic regulatory networks, long regulatory cascades tend to amplify noise [113]. The amount of noise that will be propagated by a network is, also, strongly dependent on its topology: positive regulatory feedback loops allow better noise buffering [114] than negative feedback loops [115]. Some topologies favour noise propagation to happen through the entire network, while others use molecular circuits to 'isolate' the effects of noise propagation in sub-networks [116]. Other theoretical studies have also investigated how noise propagation can help revealing underlying regulation structures and dynamics [117, 118]. From a functional perspective, it has been proposed that noise propagation through regulatory networks could, for example, lead to differentiate states within homogeneous populations (such as the well known two stable states in bacteria) via feedback loops or excitable systems

[119] or, as previously discussed, serve as a rudimentary strategy for the evolution of gene regulation from states without regulation [96].

1.6 Outline of the thesis

The work presented in this thesis explores the interplay between gene regulation and gene expression noise. The main hypothesis the thesis is based on is that noise propagation from transcription factors is a major source of noise explaining elevated noise levels in *E.coli*'s promoters, and that it is highly condition-dependent. The hypothesis is supported by evidence discussed in the previous sections, however, it has not yet been systematically investigated. The main predictions of the noise propagation hypothesis are: a) that highly regulated promoters will tend to be more noisy, independently of the condition they are growing in, b) that individual promoters will change their noise across different conditions, as a result of condition-dependent noise propagation, and c) that the noise of a promoter can be explained by its regulatory inputs. To test the noise propagation hypothesis, we've investigated the following key questions during my PhD: How do individual genes in *E.coli* change their noise levels in different conditions? Do individual genes display a certain noise level by default across all conditions or do they change their noise, as a result of noise propagation being highly condition-dependent? Does the previously observed general association between high noise and regulation by transcription factors [96] hold in different conditions? If so, which transcription factors are most responsible for noise propagation in different conditions? And, finally, what is the default noise level of genes that are not regulated and express with different means in different conditions? Do they show systematic changes in noise across the conditions?

In Chapter 2, the systematic study of single-cell noise properties using flow cytometry of a high fraction of native *E.coli* promoters is described in a set of 8 conditions. Strong evidence for a general association between noise and regulation is presented. As predicted, a strong positive correlation between high noise promoters and those that contain a large number of regulatory sites is observed in all conditions, but individual promoters show complex behaviours in terms of changes in their noise levels. The condition-dependent noise level of each promoter is then modelled in terms of regulatory binding sites and noise propagation strengths of the transcription factors

that regulate them. The computational approach used [68], has already been successfully employed to identify key regulators driving gene expression in mammalian cells [67] and here, it is extended for the first time to identify key transcription factors driving gene expression fluctuations (i.e. 'noisy' transcription factors) in specific conditions and on average in all 8 conditions. A genome-wide correlation between sequence and expression properties is uncovered, in which *E.coli* genes are organized along two principal axes: one sorting genes by mean expression and a second one by noise properties. Finally, data showing that slow growth conditions tend to display overall increased noise levels than fast growth conditions is presented, suggesting that growth rate sets a minimal noise level which all promoters will display 'by default'.

In Chapter 3, the dependency between mean expression and transcriptional noise across different conditions is analyzed in a set of constitutive synthetic and native promoters expressing at different levels. Studying constitutive promoters is relevant, as it allows to quantify the 'default' effect of general extrinsic sources to their noise and mean levels. Physiological variations due to environmental changes are known to affect these sources in different ways [120, 121], which will affect the expression dynamics of constitutive promoters. Understanding how unregulated promoters respond in mean and noise to these changes can help in disentangling the specific effects of noise propagation from regulation on noise from more general physiological effects. Two independent high-throughput methods are employed: flow cytometry, as in Chapter 2, and a custom developed microscopy method that uses convolutional neural networks for image analysis. Both approaches confirmed that unregulated promoters are among the lowest noise promoters in *E.coli*. All exhibit no deviations from a minimal noise floor, likely arising from general condition-dependent extrinsic noise sources that affect all promoters by the same amount. Remarkably, measurements done using microscopy do not show a lower bound on variance as a function of mean that measurements using flow cytometry do, indicating that flow cytometers introduce a non-negligible shot-noise component affecting variance estimations [122].

Both chapters are presented as individual stand-alone publications, with some cross-references between them.

2

Noise propagation shapes condition-dependent gene expression noise in *Escherichia coli*

Arantxa Urchueguía[1], Luca Galbusera[1], Gwendoline Bellement[1], Thomas Julou[1,*], Erik van Nimwegen[1,*]

[1] Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Basel, Switzerland.
[*] to whom correspondence should be addressed: thomas.julou@normalesup.org, erik.vannimwegen@unibas.ch

A preprint version of this chapter is available in bioRxiv:
https://doi.org/10.1101/795369

2.1 Abstract

Although it is well appreciated that gene expression is inherently noisy and that transcriptional noise is encoded in a promoter's sequence, little is known about how promoters vary their transcriptional noise across growth conditions. Using flow cytometry we have quantified transcriptional noise genome-wide across 8 growth conditions in *Escherichia coli*, and find that noise and gene regulation are intimately coupled. Apart from a growth-rate dependent lower bound on noise, we find that individual promoters show highly condition-dependent noise levels and that this condition-dependent expression noise is shaped by noise propagation from regulators to their targets. A simple model of noise propagation identifies transcription factors that most contribute to noise propagation in specific and across all measured growth conditions. The set of regulators that appear to consistently contribute to noise propagation across all environments include H.NS, Sigma38, CRP, PhoB, GadX and GadW. Moreover, the overall correlation structure of sequence and expression properties of *E. coli* genes uncovers that genes are organized along two principal axes, with the first axis sorting genes by their mean expression and evolutionary rate of their coding regions, and the second axis sorting genes by their expression noise, the number of regulatory inputs in their promoter, and their expression plasticity. This study provides the first systematic investigation of genome-wide noise properties across several growth conditions and demonstrates that noise propagation is the main mechanism underlying the tight coupling between gene expression noise and gene regulation.

2.2 Introduction

It is well established that isogenic cells growing in homogeneous environments show cell-to-cell fluctuations in gene expression [77]. This is not surprising from a biophysical point of view due to the small number of molecules involved in these reactions and their inherent stochasticity [73]. Since the first experimental observation of expression heterogeneity in 1957 by Novick and Weiner [71], seminal studies both theoretically and experimentally have shed light upon the mechanisms by which gene expression noise arises and its consequences for cellular physiology and survival (see, for example, [79, 80]). These studies have linked this heterogeneity to

various sources including transcriptional and translational regulation [78, 89, 106], chromatin state [98, 123, 124], promoter architecture [90, 97] or cellular growth [125, 126], among others.

In bacteria, several researchers have proved that different genes exhibit various levels of expression noise [41, 90, 95, 96] and that part of these differences are to some extent encoded in their promoter sequence. The role natural selection may have had in shaping noise levels across different genes is an ongoing field of research with some researchers arguing that expression noise has been minimized to ensure optimal cell functioning [101] and others, that under fluctuating environments noise has been positively selected as a bet-hedging strategy [75, 127].

A recent study in our group showed that in *Escherichia coli* genes with high noise levels correspond to those that are highly regulated (i.e. contain a large number of regulatory binding sites in their promoters) [96], suggesting that part of their noise is coming from noise propagation of their corresponding transcription factors. Noise propagation in biological networks is a well established phenomenon [107, 109, 111, 114, 115] and it has been theoretically proved that molecular control needed to suppress it is limited [112]. *E.coli*'s regulatory network is large and complex, with 215 transcription factors and 3355 regulatory interactions documented in Regulon DB [20] that can give rise to a mixture of simple regulatory mechanisms (e.g. negative and positive regulation) and feed-forward loops which generate more complex regulation processes [59]. This complexity entails that noise propagation as a consequence of regulation should be considered an important source of gene expression variability. In eukaryotes, for example, it has been demonstrated that transcription factor binding and unbinding amplifies transcriptional noise [86, 97, 106] and that co-regulated genes show correlated gene expression fluctuations [108].

Under a noise propagation scenario, optimal gene regulation will only be achieved if transcription factors activity adapt perfectly to each environment and show low expression variability themselves. On the other hand transcription factor fluctuations, which can arise from either the difference in their concentration among single cells [105] or their bound/unbound dynamics [86, 90, 97, 106], will unavoidably be transmitted to the genes they regulate. In turn, increasing expression noise as a function of their noise propagation strength (i.e. what fraction of a promoter noise can be explained by it). Therefore, a gene that is regulated by many transcription factors will respond to many noise sources and will with a higher probability see its variability increased.

As the variability of individual transcription factors is likely to change across environments [43], noise propagation should be considered a condition-dependent mechanism. Following this assumption all targets of a given transcription factor will exhibit noisier gene expression in a condition where its activity is variable and its noise propagation is strong, while in a different condition their noise may be lower or uncorrelated, as the transcription factor displays a less variable activity across individual cells. This will result in a large noise plasticity between conditions across its targets (Figure 2.1). However, so far there has been no systematic investigation into how the noise properties of genes in *E. coli* vary across conditions.

We here systematically quantify how genome-wide gene expression noise in *E. coli* varies across conditions by using flow cytometry in combination with a library of fluorescent transcriptional reporters [128] in 8 different growth conditions, including different nutrients, stresses, and in stationary phase. We investigate how global noise properties vary across conditions and quantify noise propagation by modeling the condition-dependent transcriptional noise of each gene in terms of annotated regulatory inputs in their promoters. Using this modeling we infer which TFs are contributing most to expression noise in each condition, and identify several TFs that consistently contribute to noise propagation in all growth conditions. Our analysis shows that gene expression noise and gene regulation are intimately coupled. In particular, the number of regulatory inputs of a gene, its expression plasticity, its gene expression noise, and also the plasticity in its gene expression noise, are all highly positively correlated.

2.3 Results

Understanding how noise propagation from TFs shapes patterns of gene expression noise is experimentally challenging as it is difficult to manipulate the activity of TFs in a predictable way. However, since the state of gene regulatory networks is known to change between conditions, TFs activity must vary in different conditions. Therefore, if noise propagation is a key determinant of gene expression noise, we expect noise levels across different conditions to exhibit certain general features, as illustrated in Figure 2.1.

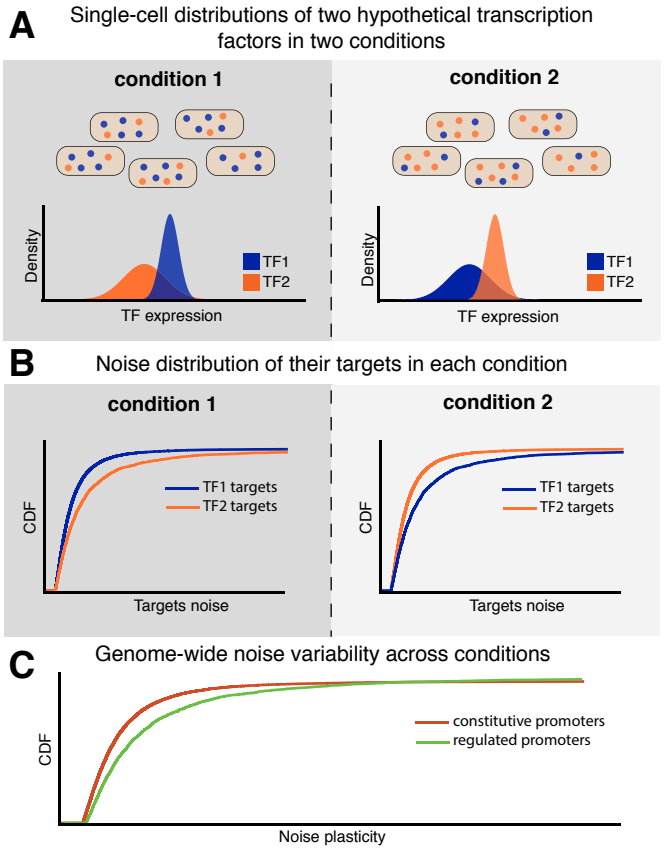


Figure 2.1: Condition-dependent noise propagation. One mechanism by which noise propagation can arise is from heterogeneous expression of transcription factors. **A:** Two independent transcription factors (TF) represented each by an orange and a blue dot show differential single-cell expression distributions across two conditions. TF1 (blue dots) is homogeneously expressed in condition 1, but not in condition 2, whereas TF2 (orange dots) shows the opposite behaviour. **B:** As noise propagation is a condition-dependent mechanism, in the condition where a given transcription displays a heterogeneous expression, its targets will also show increased variability. If now, in a different condition its expression is more homogeneous, its targets will also show a decrease in variability. As shown in the illustration, since TF2 is noisier than TF1 in condition 1, its targets show also higher noise than those of TF1, whereas in condition 2 the opposite occurs. **C:** Given the condition-dependent nature of noise propagation, regulated promoters will show larger plasticity in noise across conditions than constitutive ones, as the latter are not affected by condition-dependent noise propagation. Here we define noise plasticity as the variance of relative noise levels of a given promoter across conditions.

Let's consider a simple case scenario where two individual transcription factors show variable activities between cells and between conditions (Figure 2.1A). If in a given condition (condition 1, Figure 2.1A, dark grey panel), transcription factor 1 shows little variation in activity compared to transcription factor 2; due to unavoidable noise propagation, targets of transcription factor 2 will show higher

variability than those regulated by transcription factor 1 (Figure 2.1B, dark grey panel). Now, if in a different condition (Condition 2, Figure 2.1A, light grey panel), the heterogeneity of the transcription factor activities is reversed and transcription factor 1 shows larger variability in activity, its targets will be noisier than those of transcription factor 2 (Figure 2.1B, light grey panel) in the new condition. This condition-dependent noise propagation entails that highly regulated promoters will display larger noise plasticity across conditions (Figure 2.1C, green line) than unregulated ones, as the latter are not affected by noise propagation coming from regulation (Figure 2.1C, red line).

To prove this assumption we measured a fluorescent library comprising about 75% of all native *E.coli* promoters [128] under a set of 8 different conditions (Figure 2.2A). The library consists of each intergenic region inserted upstream a strong ribosomal binding site and a fast-folding GFP in a low copy plasmid and has already been extensively used to study noise properties in *E.coli* [95, 96, 129]. By using this library we were able to measure differences in protein levels across promoters, which we believe to originate mostly from changes in transcriptional regulation, as the mRNAs of the different reporters are almost identical.

The conditions were chosen to span a wide range of growth rates and maximize diversity in regulatory states (Figure 2.2B; Supplementary Figure S2.1). They consist of M9 minimal media with three different carbon sources (0.2% glucose, 0.2% glycerol and 0.2% lactose), two stresses (sub-MIC antibiotic: Ciprofloxacin 1.5 ng/ml + 0.2% glucose and osmotic: 0.4M NaCl + 0.2% glucose), two time points in stationary phase (16h and 30h of growth in 0.2% glucose) and a MOPS based synthetic rich media that contains all necessary metabolites (supplied with 0.2% glucose, Supplementary Table S2.1). Cell sizes changed accordingly with the growth rate (Supplementary Figure S2.2; Supplementary Figure S2.3) [130].

In each of these conditions, we measured GFP levels of single-cells during mid exponential growth (except in both stationary phases) using high-throughput flow cytometry (Figure 2.2A; Supplementary Table S2.1; Supplementary Figure S2.1). We characterized each promoter by its mean and variance in log-fluorescence with a method that identifies debris as outliers in the forward and side scattering and fits the log-fluorescence distribution of the remaining cells with a mixture of a Gaussian and a uniform distribution to further remove possible outliers (e.g. contaminants, non-growing cells) [122]. Noise propagation can arise from two main mechanisms, which are in turn binding/unbinding dynamics or fluctuations in single-cell concentrations

(Figure 2.1A). However, given our experimental setup we do not distinguish in which regime of noise propagation the cells are.

We found the measurements to be highly reproducible between replicates measured in different days (Supplementary Figure S2.4). Moreover, we performed a time-course experiment in one of the conditions to determine whether the estimates of gene expression noise were dominated by systematic measurement variations introduced by the experimental setup (Supplementary Figure S2.5; Supplementary Figure S2.6). During the time-course experiment, we measured GFP levels at different time-points of 95 individual strains growing in bulk and found that the correlation between consecutive time-points both for the mean and variance was higher than the correlation between replicates measured in different days. This indicates that the differences we observed among replicates mainly reflect biological variability and not technical measurement errors produced by the flow cytometer, validating our estimations.

Figure 2.2B shows the variance as a function of mean for each promoter measured in M9 minimal media + 0.2% lactose (see Supplementary Figure S2.7 for all conditions). Note that the variance in log-fluorescence is equal to the square of the coefficient of variation (CV^2) whenever fluctuations are small relative to the mean [96]. This approximation applies in our data, as the majority of promoters ($\sim$75% across all conditions) have a variance smaller than 0.3 (Supplementary Figure S2.7).

As it has been observed in previous studies [41, 85, 86, 96] there is a clear lower bound on noise as a function of the mean expression level of the promoter (Figure 2.2B). As explained in the Material and Methods section, we have previously shown [96] that the functional form of the noise floor, equation (2.3), can be derived assuming that GFP variance is the sum of two terms: one 'multiplicative' contribution with variance proportional to the square of the mean expression, and one 'Poissonian' contribution with variance proportional to mean expression. The former term, whose magnitude we denote by a_c corresponds to the minimal variance of very high expressed promoters and likely results from global fluctuations in transcription, translation, mRNA decay, and growth [41, 77]. This term is often referred to as an 'extrinsic noise' contribution. The latter term, whose magnitude we denote by b_c and is often referred to as the 'intrinsic noise' term, could in principle derive from intrinsic expression noise whose magnitude scales proportional to mean expression [41, 131]. However, by comparing microscopy and flow cytometry measurements we have recently shown that, at these expression levels, the component b_c

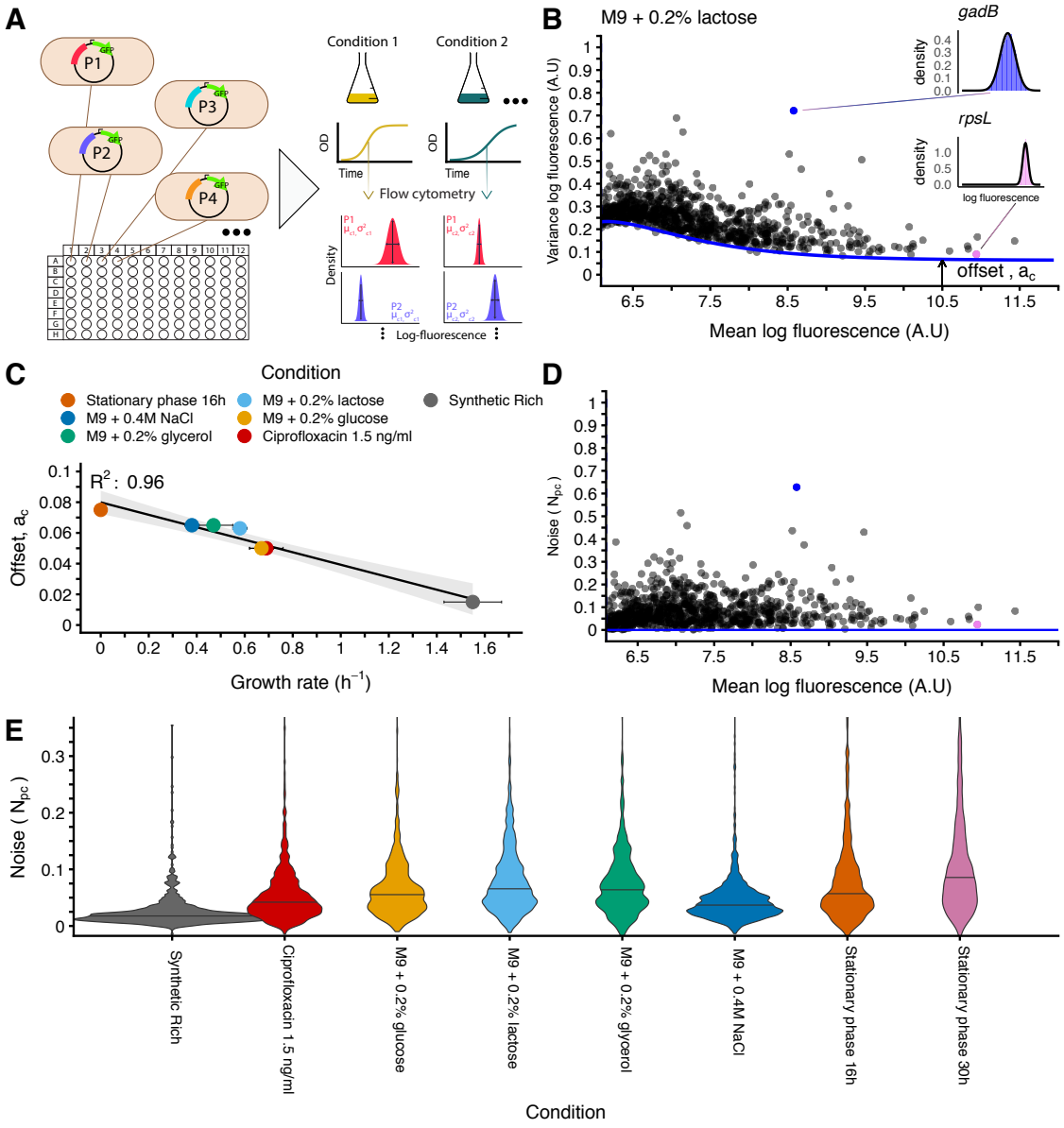


Figure 2.2: Expression noise of native *E. coli* promoters under different growth conditions. A: For each growth condition and *E. coli* promoter, we used flow cytometry to measure the distribution of GFP levels across single cells of the corresponding fluorescent reporter. The 8 growth conditions comprised synthetic rich media, minimal media with different carbon sources, an osmotic and DNA damage stress, and two time points in stationary phase. **B:** Mean (x-axis) and variance (y-axis) of log GFP levels for all promoters with expression above a background level for growth in M9 0.2% lactose (see Supplementary Figure S2.7 for results in all conditions). The blue line shows the fitted minimal variance as a function of mean expression and the corresponding minimal extrinsic noise a_c is indicated with an arrow. The insets show distributions of log-GFP levels for two example promoters. **C:** Minimal extrinsic noise a_c as a function of the growth rate in the respective condition (*stationary phase at 30h not shown*). The line indicates a linear fit (0.95 confidence interval in grey) and annotated is the R^2 Pearson correlation coefficient. **D:** To compare noise of promoters with different means, we defined the noise level of a promoter as the difference between its variance and the fitted minimal variance at its mean expression. Shown are noise levels versus mean for promoters in M9 0.2% lactose. **E:** Noise distributions of the full library in each of the measured conditions. The horizontal lines indicate the medians. The vertical scale is clipped at 0.35 for better visibility (Supplementary Figure S2.8 has the full distributions).

derives almost entirely from the measurement noise of the flow cytometer [122]. As shown in Supplementary Figure S2.7, the same functional form describes the noise floor in all conditions and we estimated the noise floor a_c at high expression in each condition.

Remarkably, we observed that the noise floor a_c is an almost perfectly decreasing linear function of growth-rate ($R^2 = 0.96$, Figure 2.2C). Thus, the slower cells grow, the higher the minimal cell-to-cell variability in gene expression. A similar anti-correlation between noise and growth-rate has been previously observed in eukaryotes, but was proposed to derive from heterogeneity in cell cycle stage [126]. However, our results show that this general anti-correlation between noise and growth-rate also occurs in prokaryotes that do not have analogous cell cycle stages.

In order to have a noise measure for each gene that does not systematically depend on its mean expression, we defined the noise N_{pc} of promoter p in condition c as the difference between its variance in log-fluorescence and the minimal variance at its mean expression level (Materials and Methods equation (2.4) and Figure 2.2D). Figure 2.2E shows the distribution of noise levels N_{pc} in each of the conditions, sorted from high to low growth-rate. We see that not only the noise floor, but also the distribution of noise on levels on top of this noise floor vary substantially across conditions. Moreover, like the noise floor, both the median of the noise levels N_{pc} as well as the variability in noise levels increase as the growth-rate decreases, e.g. the noise levels are lowest in synthetic rich conditions ($p = 3 \times 10^{-30}$, Wilcoxon rank-sum test) and highest at 30h of stationary phase ($p = 5 \times 10^{-68}$, Wilcoxon rank sum test). That is, not only do minimal noise levels increase as growth-rate decreases, the variability in noise levels across genes increases as well. The only exception to this general trend is the osmotic stress condition M9 + 0.4M NaCl, which has relatively low variability in noise levels N_{pc} compared to other conditions with similar growth-rate (Figure 2.2E), even though its noise floor is *not* deviating from the general dependence on growth-rate. These results show that growth-rate, and more generally the physiological state of the cell determines the distribution of noise levels. However, in this work we will focus on how the *relative* levels of different promoters vary across conditions.

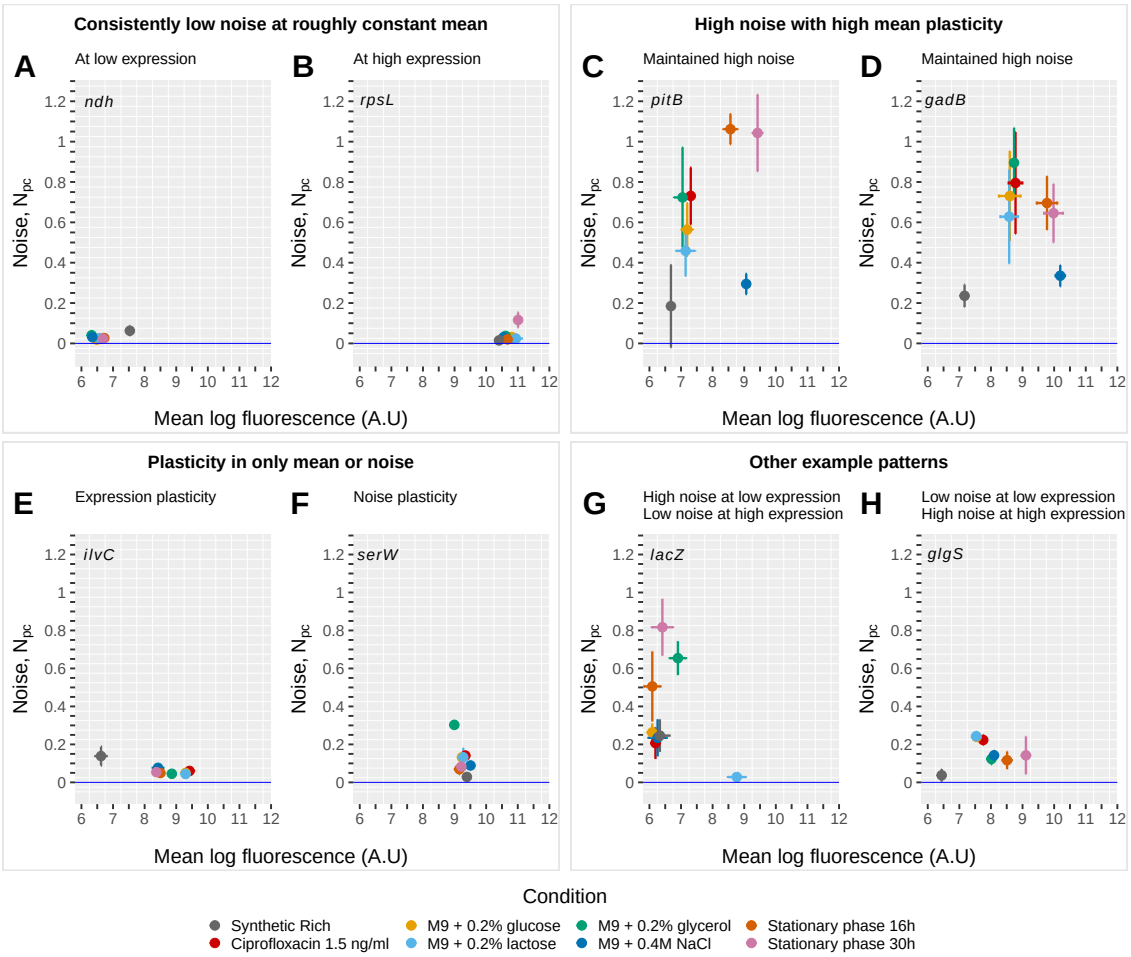


Figure 2.3: Individual promoters show diverse patterns of changes in expression noise across conditions. Each panel shows the noise as a function of mean across conditions (different colors) for an individual promoter (in each panel the name of the immediately downstream gene is annotated). Error bars denote standard-errors of the estimates. Each of the 4 pairs of panels indicate different examples of behavior in mean and noise across conditions, as described on top of each pair of panels.

If changes in noise levels across conditions were mostly driven by the overall physiology of the cells, then we would expect different genes to exhibit coherent changes in noise across conditions. For example, noise levels might rescale across conditions as a function of the mean expression of the gene in the condition. In contrast, we observe that different promoters show highly diverse changes in their noise levels across conditions (Figure 2.3). Some promoters show consistently low noise at either low or high mean expression (Figure 2.3A and B), some promoters show consistently high noise that also strongly varies across conditions in a manner not correlated with

mean expression (Figure 2.3C and D), some promoters show only plasticity in mean (Figure 2.3E) or only plasticity in noise (Figure 2.3F), but many other patterns of behavior were observed. For example, there are also promoters that show only low noise when the promoter has high mean (Figure 2.3G), or only low noise when the promoter has low mean (Figure 2.3H).

The growth media was not a predictor of how individual genes were going to change their mean and noise. For example, while overall the whole library is shifted towards lower noises in synthetic rich, individual genes can show higher noise in this condition compared to other conditions (e.g. Figure 2.3A, E, and G). We highlighted this particular condition as an example, but the same principle can be extrapolated to others. These observations indicate that global changes in the cell physiology or in the expression level only cannot explain how the noise of a promoter varies across conditions. This implies that there is a promoter specific source of noise shaping gene expression variability across the environments.

To confirm whether noise propagation coming from regulation could explain part of their additional variance, we investigated if previously reported noise propagation features still applied in our data. In particular, it has been demonstrated that there is a correlation between high noise and large number of regulatory binding sites. Applying a similar analysis as in Wolf et al. 2015 [96], we found that this relationship still holds in all 8 conditions (Figure 2.4A; Supplementary Figure S2.9; Material and Methods), confirming the hypothesis that transcription factor regulation constitutes a significant source of noise.

Under a condition-dependent noise propagation scenario (Figure 2.1C), as the noise propagation strength of individual factors is likely to change between environments, highly regulated promoters will display higher plasticity in noise between conditions than constitutive ones. In Figure 2.4B we show that, as expected, highly regulated promoters showed increased noise plasticity compared to non-regulated ones, supporting the condition-dependent nature of noise propagation ($p < 3.7 \times 10^{-10}$, two-sided Welch's t-test).

The presented data clearly demonstrates the strong association between noise and regulation. The fact that high noise promoters show high noise plasticity between conditions, indicates that the noise propagation strength of the individual transcription factors varies across conditions and therefore, their noises are being transmitted accordingly to their targets.

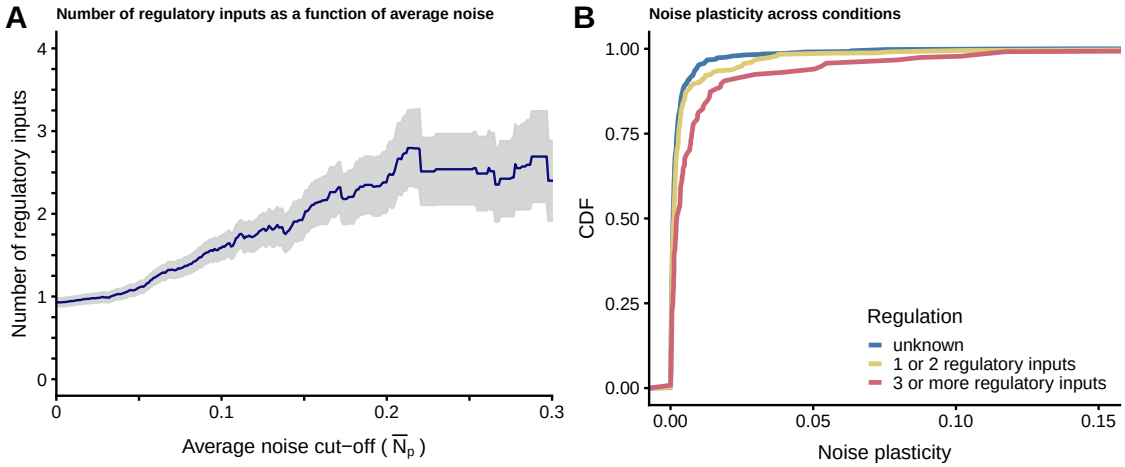


Figure 2.4: Condition-dependent noise propagation features. A: There is a positive association between number of regulatory inputs and noise level. We sorted promoters by their average noise $\bar{N}_p$ across the 8 conditions and calculated the mean (y-axis) and standard error (grey area) of the average number of known unique regulatory inputs of all promoters with noise above $\bar{N}_p$, as a function of $\bar{N}_p$ (x-axis). **B:** Promoters with many regulatory inputs show larger noise plasticity than unregulated ones. Shown is the cumulative distribution of the variance in noise of each promoter across the 8 conditions for promoters without known regulatory interactions (blue), 1 or 2 known regulators (yellow), and 3 or more known regulators (red).

If noise propagation is a key determinant of the condition-dependent changes in the noise levels of promoters, then it should be possible to explain some of these changes in terms of the regulatory sites occurring in the promoter sequences. We have previously developed a model, called Motif Activity Response Analysis [67, 68] that models gene expression patterns in terms of computationally predicted regulatory sites in promoters genome-wide and 'activities' of regulatory motifs. Here we adapted this approach to model the condition-dependent noise levels of promoters in terms of known regulatory inputs and 'noise propagating activities' of regulators. In particular, we model the noise N_{pc} of each promoter p in each condition c as a linear function of its known regulatory inputs S_{pr}, and the unknown noise propagating activities A_{rc} of each regulator r in each condition c:

$$(N_{pc} - \bar{N}_c) = \epsilon + \sum_r (S_{pr} - \bar{S}_r) A_{rc}, \tag{2.1}$$

where $\bar{N}_c$ is the average noise level of all promoters in condition c, ϵ is a noise term that is assumed Gaussian distributed with mean 0 and unknown variance. We used the RegulonDB database [20] to set a binary matrix of known regulatory inputs, i.e. S_{pr} is 1 when promoter p is known to be regulated by TF r and 0 otherwise. In

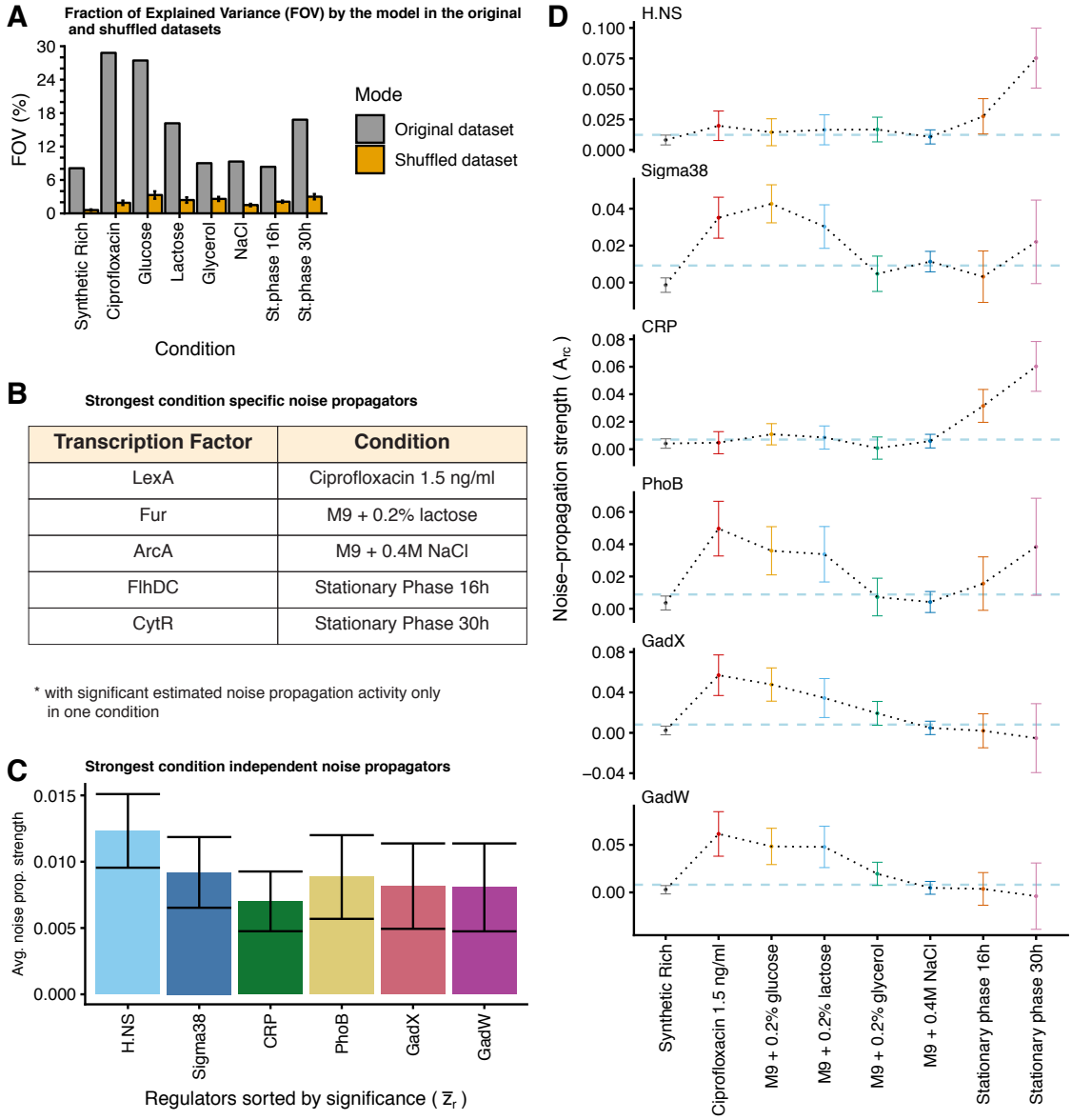


Figure 2.5: Strongest noise propagators in *E.coli's* regulatory network. A: Fraction of Explained Variance (FOV, %) by the adapted Motif Activity Response Analysis model (y-axis) in each of the 8 conditions (x-axis) after running it in two modes: in the original dataset (grey bars) and in a randomly shuffled dataset (yellow bars). The randomized data was shuffled multiple times and shown is the average FOV value obtained +/- se. **B:** Table of transcription factors predicted by the model as significant condition-specific noise propagators (with $A_{rc} > \delta A_{rc}$). **C:** Average noise propagation strengths ($\bar{A}_r$, y-axis) and their error bars ($\delta\bar{A}_r$, vertical lines) of the strongest 6 noise-propagators (with $\bar{A}_r > \delta\bar{A}_r$), sorted by significance ($\bar{z}_r$, x-axis), that consistently contribute to explain noise levels in all 8 conditions. **D:** Condition-dependent noise propagation strengths (A_{rc}) and error bars (δA_{rc}) of the strongest 6 noise propagators. The dashed blue line indicates the average noise propagation strength, $\bar{A}_{rc}$.

addition $\bar{S}_r$ is the average of S_{pr} across all promoters, i.e. the fraction of promoters targeted by regulator r.

For each condition c we infer the noise propagating activities A_{rc} by fitting the model (2.1) using a Gaussian prior of the activities A_{rc} to avoid overfitting, which allows us to calculate a full posterior probability distribution over the activities A_{rc} [68]. It should be noted that this extremely simple linear model of course only provides a caricature of the complex interactions between TFs and promoters, i.e. we ignore the number, positioning, and affinities of the binding sites, the potential interactions between binding sites for different TFs, the non-linear dependence on TF concentrations, and so on. We thus do not expect that this simple model will accurately predict the expression noise of each promoter across the conditions. Rather, the main aim is to test whether noise propagation can explain a significant fraction of the variation in noise levels across promoters, and to identify which TFs are most responsible for noise propagation in each condition.

As shown in Figure 2.5A (grey bars), a substantial fraction of the variance in noise levels in each condition, i.e. between 9% and 29%, can be explained by the simple model of equation (2.1). To confirm the significance of these fits we fitted the same model to data in which rows of the noise matrix N_{pc} were randomly shuffled, i.e. the association between regulatory inputs and noise levels were randomized, and observed that the fraction of explained variance on the randomized data was always much lower (Figure 2.5A, yellow bars). By running the model assigning different values of the Gaussian prior manually (in the range of the optimal one), we found the inference to be almost identical in all conditions (Supplementary Figure S2.12)

Apart from estimating the noise propagating activities A_{rc} of each regulator r in each condition c, the model (2.1) calculates an error bar δA_{rc} for each of these activities and Supplementary Figure S2.10 shows, for each condition, all TFs for which the noise propagating activity was larger than its error-bar, i.e. $A_{rc} > \delta A_{rc}$. We first focused on TFs that contributed to noise-propagation in a highly condition-specific manner. As shown in Figure 2.5B, there were 5 TFs that had significant noise propagating activity in only 1 condition. For example, the TF LexA contributed to noise propagation only in the sub-MIC ciprofloxacin condition. LexA is a repressor of the SOS response genes which responds to DNA damage by auto-cleaving in recA polymers that form on DNA double-strand breaks [132]. Indeed, it is known ciprofloxacin can induce DNA damage and induce the SOS response [133]. In particular, since we employed ciprofloxacin at a concentration well below the minimal

inhibitory concentration, DNA damage likely only occurred in a subset of the cells, leading to heterogeneity in lexA activity across the cells. A second example of a condition-specific noise propagating TF is ArcA, whose activity was only significant in M9 + 0.4M NaCl. ArcA is a general regulator that controls the aerobic/anaerobic expression of respiratory proteins and diverges metabolism into fermentation [134]. It is known that under salt stress major adaptations in metabolism occur and fermentation products increase [135], which is consistent with heterogeneous activity of ArcA across these conditions. Third, we found that FlhDC, the master regulator of the flagellar biosynthesis [136], was contributing to noise propagation only in early stationary phase, i.e after 16h of growth in 0.2% glucose. It is known that flagellar synthesis has a peak in expression during late exponential phase and decreases shortly after entry in stationary phase [137]. Since the 16h condition is a transition between late exponential growth and entry into stationary phase, it seems plausible that some cells had entered growth arrest and were no longer expressing components of the flagellar machinery, while others had not yet transitioned. Fourth, the TF CytR was found to contribute to noise propagation in late stationary phase. CytR regulates genes involved in nucleoside uptake and utilization [138] and it was recently found that mutations in CytR have a fitness advantage during long term stationary phase [139], which was hypothesized to result from an increased ability to import and use nucleosides that occur in the stationary phase environment due to cell death. Heterogeneity in CytR activation late in stationary phase is consistent with this functional role. Finally, the TF Fur, which regulates genes involved in iron homeostasis [140], had significant noise propagating activity only in the M9 + 0.2% lactose condition. In contrast to the other four cases, we do not have a concrete biological hypothesis for which Fur activity might be especially heterogeneous during growth on lactose.

In addition to these condition-specific noise propagators, it is noteworthy that many of the most significant noise propagators were found in multiple conditions (Supplementary Figure S2.10). To identify regulators that were consistently contributing to noise propagation in all conditions we calculated, for each regulator r, its average noise propagating activity $\bar{A}_r$ averaged over all conditions (Figure 2.5C and Figure 2.5D and Material and Methods). The most significant noise propagating factor was H.NS, a general transcriptional repressor that regulates around 5% of all *E.coli* promoters. It belongs to the family of 'nucleoid associated' proteins, acting as a histone-like molecule, by binding to curved DNA and inhibiting transcription

[141]. It is noteworthy that, in eukaryotes, histone positioning plays an important role in determining noise levels [123, 124], although the gene regulatory mechanisms are too different between prokaroytes and eukaryotes to imply a direct mechanistic link between these observations.

The second most significant noise propagating TF is Sigma38(*rpoS*), which is considered the central regulator of gene expression in stationary phase and under environmental stress, as it interacts with the RNA polymerase and activates genes involved in the overall stress response and genes required to survive long periods of food starvation and growth arrest [142, 143]. It has been established that, in contrast to rich media, *rpoS* levels in minimal media (the basis in 7 of our 8 conditions) are also high during exponential phase, although the molecular mechanisms behind these differences in *rpoS* activity are unclear [144]. Interestingly, it has been reported that in a mutant strain unable to produce ppGpp, noise levels in a set of synthetic genes were significantly reduced in stationary phase compared to a WT strain [145]. This observation is in line with the prediction of Sigma38 playing an important role in shaping genome-wide expression noise, as ppGpp is an 'alarmone' responsible for regulating genes upon entry into stationary phase [146] and promoters regulated by Sigma38 require ppGpp for induction [147]. Moreover, it was recently shown that *rpoS* activity is heterogeneous among single cells in M9 glucose [103].

Two further significant noise propagators are CRP and PhoB. CRP is a global regulator of genes involved in carbon source catabolism and its activity has been proposed to reflect carbon source influx [148]. PhoB regulates the response to inorganic phosphate (Pi) starvation and binds to the DNA as a dimer after being phosphorylated by a histidine kinase (PhoR) under Pi limited conditions [149]. We hypothesize that, in our growth conditions, both carbon source influx and Pi concentration are sufficiently limiting that there are significant cell-to-cell fluctuations, leading to fluctuations in the activities of CRP and PhoB. We also note that it has previously been observed that promoters associated with carbon metabolism regulation are over represented among high noise promoters [95].

Finally, the two last factors we find to be significantly contributing to noise propagation across all conditions (GadX and GadW), belong to a family of regulators involved in the response to acid stress [150]. The appearance of these factors may also be explained by our experimental setup. Oxygen levels in microtiter plates can easily become limiting and this oxygen deprivation leads to production of fermentation products [134], even when oxygen is still present [151]. Fermentation products

are known to acidify the medium [152], which can activate the response to acid stress in some cells. Moreover, the fact that we find GadX and GadW as noise propagators is in accordance with a recent publication, where it was shown that heterogeneous expression of the gadBC operon (heavily regulated by GadX and GadW) correlated with single-cell survival to high acid induced by an antibiotic [104].

Together our results show that noise propagation by TFs plays a major role in shaping noise levels across genes, and that TFs have different noise propagating activities in each condition, leading to highly condition-dependent noise levels across genes.

Above we have shown that, through noise propagation, gene regulation and gene expression noise are intimately coupled, such that highly regulated genes tend to be more noisy and also vary their noise levels more across conditions. We next set out to understand how regulation and expression relate to other properties of genes on a genome-wide scale. For example, previous analysis of gene features have uncovered that, on a genome-wide scale, genes are organized along a one-dimensional axis that relates evolutionary rates, codon bias, and gene expression [153–156], i.e. highly expressed genes tend to have strong codon bias and slowly evolving coding regions, whereas lowly expressed genes tend to have weak codon bias and evolve more rapidly. To investigate how gene regulatory and expression noise properties relate to other gene features we collected a set of features for *E. coli* genes on a genome-wide scale from the literature including the absolute expression levels at both the RNA [41] and protein level [157], sequence properties such as codon bias and evolutionary rates at both synonymous and nonsynonymous sites [155] and the number of regulatory inputs of each gene [20]. We then complemented these features with gene regulatory annotations and gene expression levels that we measured here including mean expression level, expression plasticity across the 8 growth conditions, mean expression noise, and noise plasticity across the 8 growth conditions.

In total we gathered 10 different gene features and then calculated an overall normalized covariance matrix C of correlations between these features, i.e. with C_{ij} the squared Pearson correlation between features i and j. We then performed Principal Component Analysis (PCA) of the matrix C to characterize the overall genome-wide correlation structure of these gene features. As shown in Supplementary Figure S2.11, the first two principal components capture significantly more of the total variance than the other 8 components, and more than 50% of the variance. That is, genes are mainly organized along two PCA major axes in the the 10-dimensional

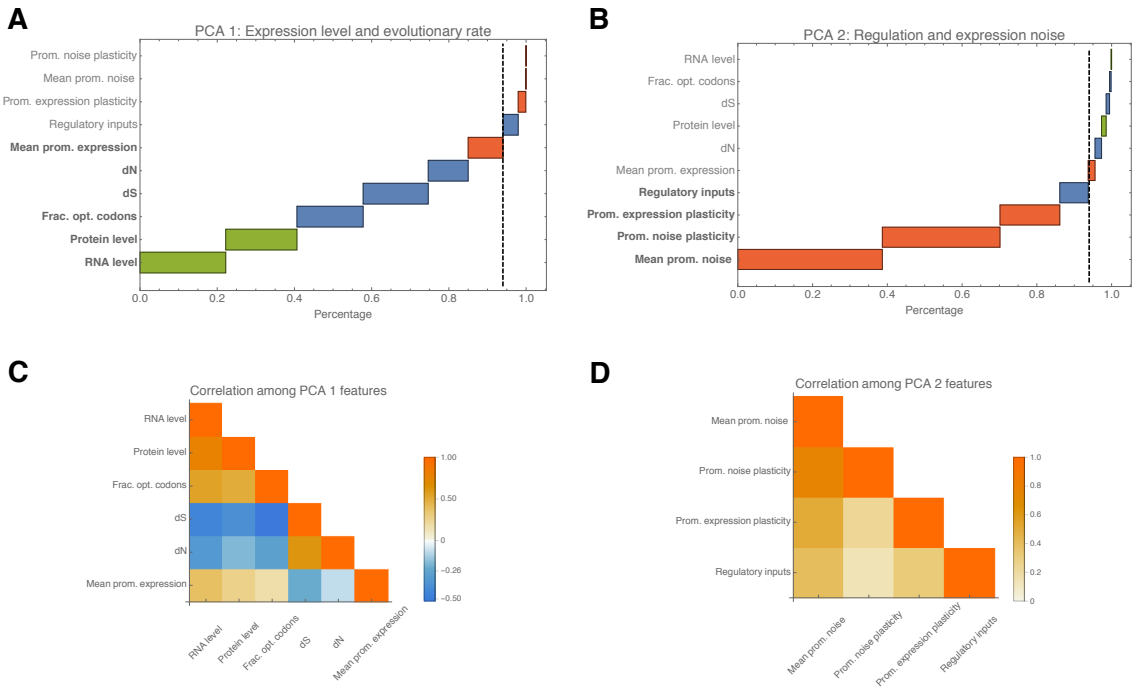


Figure 2.6: PCA analysis of the genome-wide structure of gene features. A: Relative contribution of 10 gene features to the first PCA component, sorted from top to bottom. The features in bold together account for 94% of the vector. In green are expression measurements obtained from previous studies, sequence features are in blue, and features measured in this study are in red. **B:** As in panel A, but now for the second PCA component. **C:** Correlation structure of the features contributing to the first PCA component. Negative correlations are in blue and positive correlations in orange. **D:** As in panel C but now for the second PCA component.

space of gene features. The first PCA axis sorts genes by their absolute gene expression and evolutionary rate (Figure 2.6A). That is, 94% of the weight along this first PCA component is accounted by mean RNA and protein levels, codon bias, and evolutionary rates at synonymous and nonsynonymous sites (Figure 2.6A) and, whereas the absolute expression levels and codon bias are all positively correlated with each other, and negatively correlated with the evolutionary rates (Figure 2.6C). That is, this first PCA axis corresponds to the general organization of genes by their absolute expression levels and evolutionary rates observed previously [153–156].

Strikingly, the second PCA axis is almost entirely oriented along features associated with gene regulation and gene expression noise. That is, 94% of the vector's weight is accounted for by gene expression noise, noise plasticity, plasticity in mean expression, and number of regulatory inputs (Figure 2.6B). Moreover, these four features are all positively correlated with each other (Figure 2.6D). That is, this second PCA axis organizes genes by their regulation and expression noise. That is on one end of this axis are constitutively expressed genes that do not change their mean expression level across conditions, and have low noise in all conditions, whereas on the other end of the axis are highly regulated genes that are highly plastic in expression, have high and varying expression noise across conditions. This result not only further confirms that gene regulation and expression noise are intimately coupled on a genome-wide scale, it also shows that these gene regulatory features are varying *independently* of other principal axis that organizes genes by absolute expression level and the evolutionary rate of their coding region.

2.4 Discussion

Although it is now well-established that gene expression is an inherently noisy process, so far little is known in bacteria about how noise levels of genes vary across growth conditions. Here we used high-throughput flow cytometry in combination with a library of fluorescent transcriptional reporters to quantify expression noise of *E. coli* promoters genome-wide. The general picture that emerges from our study is that the expression noise of a given gene in a given condition is the sum of two separate contributions: a minimal amount of noise that derives from global physiological fluctuations and that is approximately equal for all genes, and a highly gene- and condition-specific component that is due to noise propagation from regulators

to their targets. Constitutively expressed promoters only exhibit the physiological 'noise floor' in each condition, and the more regulated a gene is, the more additional noise from noise propagation it exhibits, and the more variable this additional noise is across conditions.

We observed that the noise floor is itself significantly varying across conditions. In particular, the noise floor systematically decreases with the growth-rate of the cells, and is highest in stationary phase (Figure 2.2). Both its dependence on growth-rate, and the fact that this noise floor appears to equally affect all genes, strongly suggest that the noise floor is driven by global physiological fluctuations. However, it is currently unknown what physiological fluctuations most contribute to this noise floor. It is plausible that fluctuations in chromosome copy number, polymerase concentration, ribosome and charged tRNA concentrations, mRNA decay rates, and fluctuations in growth-rate itself, may all contribute to determining the noise floor. To gain further understanding which fluctuations set the lower noise floor, and why the noise floor decreases with growth-rate will likely require quantitative time course data, for example from approaches that combine microfluidics with time-lapse microscopy [158, 159].

Our results show that, in addition to this noise floor, each gene exhibits additional expression noise due to noise propagation. The additional noise is not only highly condition-dependent, but different genes show highly diverse behaviors of noise across conditions (Figure 2.3). In addition, the more regulatory inputs a gene has, the higher its noise levels tend to be, and the more variable its noise levels across conditions. The intimate relationship between expression noise and gene regulation was further underscored by a global analysis of the correlation structure of a diverse set of gene features. We found that *E. coli* genes are broadly organized along two independent axes in the space of sequence, evolutionary, and gene expression features. While the first axis organizes genes by evolutionary rate and absolute expression level, with low expression and fast evolving genes on one end, and high expression and slow evolving genes on the other, the second axis organizes genes by gene expression noise. Here constitutively expressed genes with consistently low expression noise occur on one end of the axis, while highly regulated genes with high expression plasticity, high noise, and high noise plasticity, occur on the other end of the scale.

To identify which TFs are most responsible for noise propagation in each condition, we adapted a simple linear model that we previously developed for modeling gene

expression in terms of regulatory sites in promoters [68], to model gene expression noise in terms of known regulatory inputs and noise propagation activities of TFs. This analysis showed that, in spite of the simplicity of the model, a significant fraction of the variance in expression noise can be explained by noise propagation and we identified both a number of TFs that propagate noise in only a specific condition, and a number of TFs that appear to significantly contribute to noise propagation in all conditions. Among these latter ubiquitously noise propagating TFs are the histone-like TF H.NS, the stationary phase sigma factor Sigma38, the global carbon and phosphate regulators CRP and PhoB, and the Gad TFs involved in acid stress. It is likely that these ubiquitous noise propagating TFs reflect aspects that were shared between all our growth-conditions, i.e. batch culture growth in microtiter plates.

Although our simple linear noise propagation model captures a significant amount of the variation in noise levels, it only captures a modest fraction of the total variance in absolute terms. This is not surprising. In order to make quantitatively accurate predictions of the expression noise of promoters much more realistic models would be needed that take into account that different TFs compete for binding at promoters, that binding rates depend on TF concentrations in a non-linear manner, that interaction between bound TFs and RNA polymerase depends on the relative positioning of sites, and so on. To develop such quantitative models one would likely need more detailed data on the expression dynamics of different promoter architectures. A particularly interesting question that such more detailed data might answer is whether the noise propagation results mainly from fluctuations in TF concentrations across cells, or whether the main source of noise propagation is the stochastic binding and unbinding of the TFs to the promoters.

2.5 Material and Methods

Strains

All strains used in this study have been previously described [95, 128]: each strain carries a transcriptional fusion of a given native *E.coli* promoter followed by a strong ribosomal binding site and *gfp-mut2* (a fast maturing GFP) on a low copy-number plasmid (pUA66 or pUA139 with pSC101 origin, $\sim$ 5 copies per cell). The library

contains a construct for ∼75% of all intergenic regions longer than 40bp in *E.coli*'s genome flanked by 50 (resp. 150bp) of the downstream (resp. upstream) sequence in order to include most regulatory interactions found on the chromosome.

Growth conditions

The strains library was stored at -80°C in LB + 7.5% glycerol in microtiter plates. Individual plates were inoculated into fresh media of interest (200 μl) and incubated for two overnights in the same condition before fluorescence measurements. Dilutions (∼1/2000) between overnights were done using a 96 Solid Pin Replicator (V&P, 409). The library was grown in a total of 8 different conditions: minimal media, M9 (0.1mM $CaCl_2$, 1mM $MgSo_4$, 1 x M9 salts [Sigma M6030]) supplemented with either 0.2% glucose (w/v), 0.2% glycerol (v/v), 0.2% lactose (w/v), 0.4M NaCl (+ 0.2% glucose [w/v]) or 1.5 ng/ml ciprofloxacin (+ 0.2% glucose [w/v]); a MOPS based synthetic rich media (Teknova, M2105) supplemented with 0.2% glucose, and two stationary phase conditions, where plates were grown for either 16h or 30h in M9 minimal media + 0.2% glucose (w/v). Note that optical density typically saturates after about 10 hours of growth in these conditions (Supplementary Figure S2.1).

All media, except the one containing ciprofloxacin, were supplemented with 50 μg/ml kanamycin. The overnights for the sub-MIC ciprofloxacin condition were done in M9 glucose 0.2%, and only at the day of quantification ciprofloxacin was added. On the quantification day, cells were diluted between 200 and 1000-fold depending on the condition (Supplementary Table S2.1) and grown until mid-exponential phase at 37°, shaken at 600rpm. Growth rates were estimated independently for individual strains in each condition by monitoring the optical density (OD_{600}) every 90s during 15-25 hours at 37°C in a plate reader (Biotek Synergy 2). We defined the growth rate as α as the slope of a straight-line fit of $\log(OD_{600})$ against time.

To estimate cell sizes, a strain of the library containing a plasmid without promoter was selected and grown as described. Cells were then placed on a 1% agarose pad and phase contrast images were obtained with a Nikon Ti-E microscope using a 100× Ph3 objective (NA 1.45) and an Hamamatsu Orca-Flash 4.0 v2 camera. Cell outlines were identified using a custom MATLAB pipeline.

Flow cytometry quantification of fluorescence

We measured the distribution of GFP fluorescence levels in single cells using a FACSCanto II (BD Biosciences) with a high-throughput sampler (HTS), fluorescence excitation at 488 nm and a 530/30 nm filter for emission. For each strain we collected 5×10^4 events. Distributions were individually inspected and promoters showing clear signs of contamination were discarded. For the remaining measurements, we used a Bayesian procedure that removes outliers to extract the mean and variance of the log-fluorescence distributions as described in [122]. Briefly, we first fitted the 4-dimensional signal distribution of forward and side scatter heights and widths by a mixture of a multi-variate Gaussian and a uniform 'background' distribution. For each event, we then calculated the posterior probability that it derives from the central multi-variate Gaussians, and all events with lower than 50% posterior probability were removed. For the remaining cells, the logarithms of the fluorescence signals (logarithm of the height of the peak) were fitted to a mixture of a Gaussian and a uniform background distribution. That is, the probability of observing log-fluorescence y has the form:

$$P(y|\mu,\sigma,\rho) = \frac{\rho}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} + \frac{1-\rho}{\Delta}, \tag{2.2}$$

, where μ and σ^2 are the mean and variance of the log-fluorescence distribution, ρ is the fraction of cells deriving from the Gaussian, and $\Delta = y_{\max} - y_{\min}$ is the range of observed log-fluorescence values. Given n single-cell log-fluorescence measurements y_1, y_2, ..., y_n for a given promoter in a given condition, the likelihood is simply given by $L(\mu,\sigma,\rho) = \prod_{i=1}^{n} P(y_i|\mu,\sigma,\rho)$ and we fit μ, σ, and ρ by maximizing this likelihood. The data processing method of [122] is available as an R package at (`https://github.com/vanNimwegenLab/vngFCM.git`). In order to assess reproducibility of the measurements, we measured a subset of the library on multiple days and estimated means and variances of each promoter separately for each day (Supplementary Figure S2.4). We defined the mean and variance of each promoter that was measure more than once as the average over its replicates. For the individual promoters shown in Figure 2.3, 6 independent measurements of each were taken.

Time-course quantification of fluorescence

One of the plates of the library (95 strains) was grown in M9 + 0.4M NaCl during two overnights. At the day of the quantification a 1/200 dilution was done in 1ml of fresh media in a 96 deep-well plate (with 1 glass-bead per plate for better shaking). The plate was covered with a breathable sealing film and grown at 37°, shaken at 600 rpm. At 9 consecutive time points after dilution (after 0h, 1h, 2h, 3h, 5h, 6.5h, 8.5h, 10h and 11h), 100 μl of the culture were transferred into a 96-well plate and used for fluorescence quantification.

Minimal variance as a function of mean and noise estimation

We observed a clear lower bound on noise levels (variance of log-fluorescence) that depends on the mean of expression. In previous work [96] we derived a functional form for the noise minimum as a function of mean expression which takes into account that total fluorescence is a sum of background fluorescence and fluorescence deriving from GFP, and that the variance in GFP levels is a sum of a 'Poissonian' term that is proportional to mean fluorescence, and a term proportional to mean fluorescence squared. If we denote the background fluorescence in condition c by $f_{bg,c}$ and the average fluorescence of promoter p in condition c by $\langle f_{p,c} \rangle$, the minimal variance in log-fluorescence takes the form

$$\sigma^2_{\min}\left(\langle f_{p,c}\rangle\right) = a_c \left(1 - \frac{f_{bg,c}}{\langle f_{p,c}\rangle}\right)^2 + \frac{b_c}{\langle f_{p,c}\rangle}\left(1 - \frac{f_{bg,c}}{\langle f_{p,c}\rangle}\right), \tag{2.3}$$

where b_c is the prefactor of the component of the noise proportional to the mean, and a_c is the prefactor of the noise proportional to the square of the mean.

We estimated the average background fluorescence in each condition from plasmids without a promoter upstream of *gfp-mut2* that were included in each individual plate. As the model breaks down in the regime where promoters display fluorescence levels close to background fluorescence, we only considered promoters with mean larger than $2 \times f_{bg,c}$ for further analysis. We fitted the following parameters for the minimal variance in each condition:

Condition	a_c	b_c	$f_{bg,c}$	% of promoters above background
Synthetic Rich	0.015	410	180	41.0
Ciprofloxacin 1.5ng/ml	0.05	570	230	59.2
M9 glucose	0.05	530	220	55.3
M9 lactose	0.063	550	220	57.9
M9 glycerol	0.065	580	205	59.7
M9 0.4M NaCl	0.065	500	205	52.0
Stationary phase 16h	0.075	600	190	59.3
Stationary phase 30h	0.075	600	190	60.1

To obtain a noise level for each promoter that does not systematically depend on mean, we defined the noise N_{pc} of promoter p in condition c as the difference between the measured variance and the fitted minimal variance:

$$N_{pc} = \sigma^2_{pc} - \sigma^2_{\min}\left(\langle f_{p,c}\rangle\right). \tag{2.4}$$

Noise propagation features

We used the same promoter annotation as in Wolf et al. 2015, where the promoter fragments had been re-annotated by mapping the primer pairs used to construct the library to the *E.coli* K12 MG1655 genome. From all promoters in the library we were able to annotate 94% unambiguously to an immediately downstream gene. We obtained all gene-TF regulation annotations from RegulonDB [20] and counted for each gene the number of unique transcription factors known to regulate it.

We sorted all annotated genes by their average noise across all conditions ($\bar{N}_p$) and as a function of a cut-off in $\bar{N}_p$, we calculated the mean and standard-error of the number of regulatory inputs of all genes with $\bar{N}_p$ values above the cut-off. We performed this analysis separately in each of the conditions (Supplementary Figure S2.9). As a measure of noise plasticity of each promoter p, we calculated the variance of the noise levels N_{pc} across conditions.

Fitting noise in terms of regulatory inputs

To model noise in terms of regulatory inputs we adapted a method that was previously developed in our group [67, 68], called Motif Activity Response Analysis, which models gene expression levels in terms of computationally predicted regulatory sites in promoters and condition-dependent activities regulators using a linear model. As explained in the main text, we model the noise N_{pc} of each promoter p in each condition c as a linear function of the condition-dependent noise-propagating activities A_{rc} of the regulators known to regulate promoter p, i.e. equation (2.1).

The binary matrix of regulatory interactions S_{pr} was constructed using the RegulonDB data [20], where $S_{pr} = 1$ when transcription factor r is known to target promoter p, and $S_{pr} = 0$, otherwise. N_{pc} is normalized by subtracting the average noise $\bar{N}_c$ in the condition and S_{pr} is normalized by subtracting the average of S_{pr} over all promoters, i.e. the fraction of promoters targeted by regulator r.

The noise term, which reflects the deviation between the measurements and our simple model, is assumed to be Gaussian distributed with unknown variance. To avoid overfitting, the model also includes a Gaussian prior over noise-propagation activities A_{rc} that has mean zero and a variance that is set using cross-validation. In particular, maximal posterior noise-propagation activities A_{rc} are inferred on 80% of the promoters, and the variance of the prior is set so as to minimize the squared-error of the predictions on the remaining 20% of the promoters. Thus, a different prior is fitted for each condition. As a simple measure of the quality of the fit, we used the fraction of the total variance in the data that is explained by the model (FOV).

For each regulator and condition, we obtain the full posterior distribution over the noise-propagation activity A_{rc} and use the standard-deviation δA_{rc} of this posterior as an error-bar for the inferred activity A_{rc}. In addition, we use the z-like statistic $z_{rc} = A_{rc}/\delta A_{rc}$ as a measure of significance of regulator r in condition c. Note that, roughly speaking, z_r corresponds to the number of standard-deviations the activity of regulator r is away from zero on average.

We defined the average noise-propagating strength $\bar{A}_r$ of each regulator r as a weighted average over the 8 conditions:

$$\bar{A}_r = \frac{\sum_c \frac{A_{rc}}{\delta A_{rc}^2}}{\sum_c \frac{1}{\delta A_{rc}^2}}, \tag{2.5}$$

and the corresponding error-bar $\delta\bar{A}_r$ as

$$\delta\bar{A}_r = \frac{1}{\sqrt{\sum_c \frac{1}{\delta A_{rc}^2}}}. \tag{2.6}$$

Finally, the average significance $\bar{z}_r$ of motifs over all conditions was then estimated as:

$$\bar{z}_r = \frac{\bar{A}_r}{\delta\bar{A}_r} \tag{2.7}$$

To test the robustness of the model in relation to the regulatory matrix, we runned it with three different versions of S_{pr}. One full version, containing all regulators with at least one target in the dataset (197 TF's in total), another version where all transcription factors with less than 3 targets had been removed (75 TF's in total) and one version containing the top 6 inferred transcription factors with the highest significance ($\bar{z}_r$) across all conditions, together with Sigma70 and 17 randomly chosen ones (24 TF's in total). In addition, we runned the model assigning different priors. All versions led to very similar results (Supplementary Figure S2.12, Supplementary Figure S2.13).

To confirm the significance of our fits, we performed tests in which we randomly shuffled the rows of the noise-level matrix N_{pc}, thereby randomizing the association between noise levels and regulatory inputs. We fitted the model to this randomized data and found consistently low FOVs (Figure 2.5C, yellow bars).

Principal component analysis

For each promoter we gathered a list of 10 features associated with the immediately downstream gene using both the measurement in this study as well as previously published data. In particular we obtained for each promoter:

1. Average RNA level (data taken from [41]).
2. Average protein level (data taken from [157]).
3. Fraction of optimal codons (data taken from [155]).
4. Substitution rate at synonymous sites dS (data taken from [155]).
5. Substitution rate at non=synonymous sites dN (data taken from [155]).
6. Average of the mean in log-expression across conditions (this study).
7. Expression plasticity, i.e. variance of the mean in log-expression across conditions (this study).
8. Average of the promoter noise across conditions (this study).
9. Noise plasticity, i.e. variance of the promoter noise across conditions (this study).
10. Number of regulatory inputs (data taken from [20]).

Using these measurement, we calculated a covariance matrix containing all the variances of each of these features across genes, and the covariances of each pair of features. Note that not all features were available for all genes so that, for each pair of features, we estimated the covariance from the set of genes for which both features were available. We then normalized the covariance matrix by dividing each entry C_{ij} by the square-root of the product of variances, i.e. $C_{ij} \to R_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$, turning it into a matrix of squared correlation coefficients. We then performed PCA on this normalized covariance matrix. Finally, for the first two principal components we calculated what fraction of the principal vector's length was accounted for by each feature.

2.6 Acknowledgements

We would like to thank our FACS Core Facility, specially Janine Bögli, for assistance with flow cytometry, Guillaume Witz for support with microscopy analysis and Mikhail Pachkov for help with setting up MARA. Thanks to Urs Jenal, Gasper Tkačik and Olin Silander for useful discussions on the project; and to Dany Chauvin

and Đorđe Relic for comments on the manuscript. This work was partly funded by the Werner Siemens Stiftung through a fellowhip to AU, the SystemsX.ch StoNets grant, and the SNF grant 31003A_159673 to EvN.

2.7 Author contributions

EvN and TJ designed the study. AU performed all experiments, except fluorescence quantification with flow cytometry in the Ciprofloxacin condition, which was done by GB. LG wrote the R package for the estimation of mean and variances from flow cytometry data. AU, TJ and EvN analyzed the data. AU, TJ, and EvN wrote the paper.

2.8 Conflict of interest

The authors declare no conflict of interest.

2.9 Supplementary information

Table S2.1: List of conditions. Description of the 8 environmental conditions in which the library of native *E.coli* promoters [128] has been grown. The chosen environmental conditions comprised a MOPS based synthetic rich media, minimal media (M9) with different carbon sources (glucose, lactose and glycerol), an osmotic and DNA damage stress (0.4M NaCl and Ciprofloxacin 1.5 ng/ml both supplemented with glucose), as well as two time points in stationary phase (16 and 30 hours of growth in M9 glucose).

Condition	Condition during overnights	Hours of growth before FACS measurements	Dilution after ON	Growth rate +/- sd (h^{-1})	Mean area +/- sd (in μm^2)
Synthetic Rich*	same	2h	1/1000	1.59 +/- 0.28	4.3 +/- 1.4
M9** + Ciprofloxacin 1.5 ng/ml (+ 0.2% glucose)	M9 + 0.2% glucose	4h	1/200	0.69 +/- 0.07	2.7 +/- 1
M9** + 0.2% glucose	same	4h	1/200	0.67 +/- 0.05	2+/- 0.6
M9** + 0.2% lactose	same	4h	1/200	0.58 +/- 0.03	2.1+/- 0.6
M9** + 0.2% glycerol	same	5h	1/200	0.5 +/- 0.11	1.6 +/- 0.5
M9** + 0.4M NaCl (+ 0.2% glucose)	same	15h	1/500	0.37 +/- 0.03	1.4 +/- 0.3
Stationary phase 16h (M9** 0.2% glucose)	same	16h	1/200	x	1.3 +/- 0.4
Stationary phase 30h (M9** 0.2% glucose)	same	30h	1/200	x	x

* MOPS based. Commercially available (Teknova M2105)

** Prepared as follows: 0.1mM $CaCl_2$, 1mM $MgSo_4$, 1 x M9 salts (Sigma M6030), 50 µg/ml Kanamycin, *d*H20

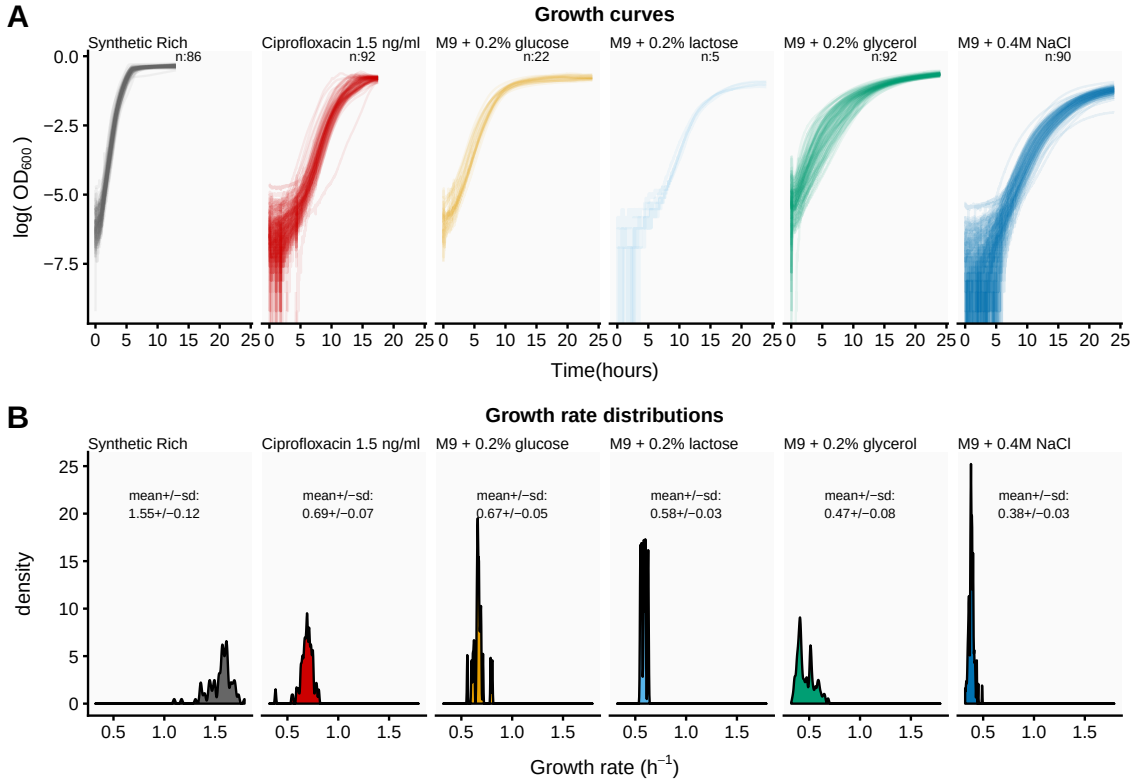


Figure S2.1: Growth curves across conditions. A: OD_{600} (log-scale, y-axis) as a function of time (in hours, x-axis). We measured OD_{600} in individual strains growing in bulk at intervals of 90 seconds during 15 to 25 hours. The number of strains used per condition is indicated in each panel. **B:** Density distribution of the estimated growth rates in each condition. The growth-rate α was defined as the slope of a linear fit of $\log(OD_{600})$ against time.

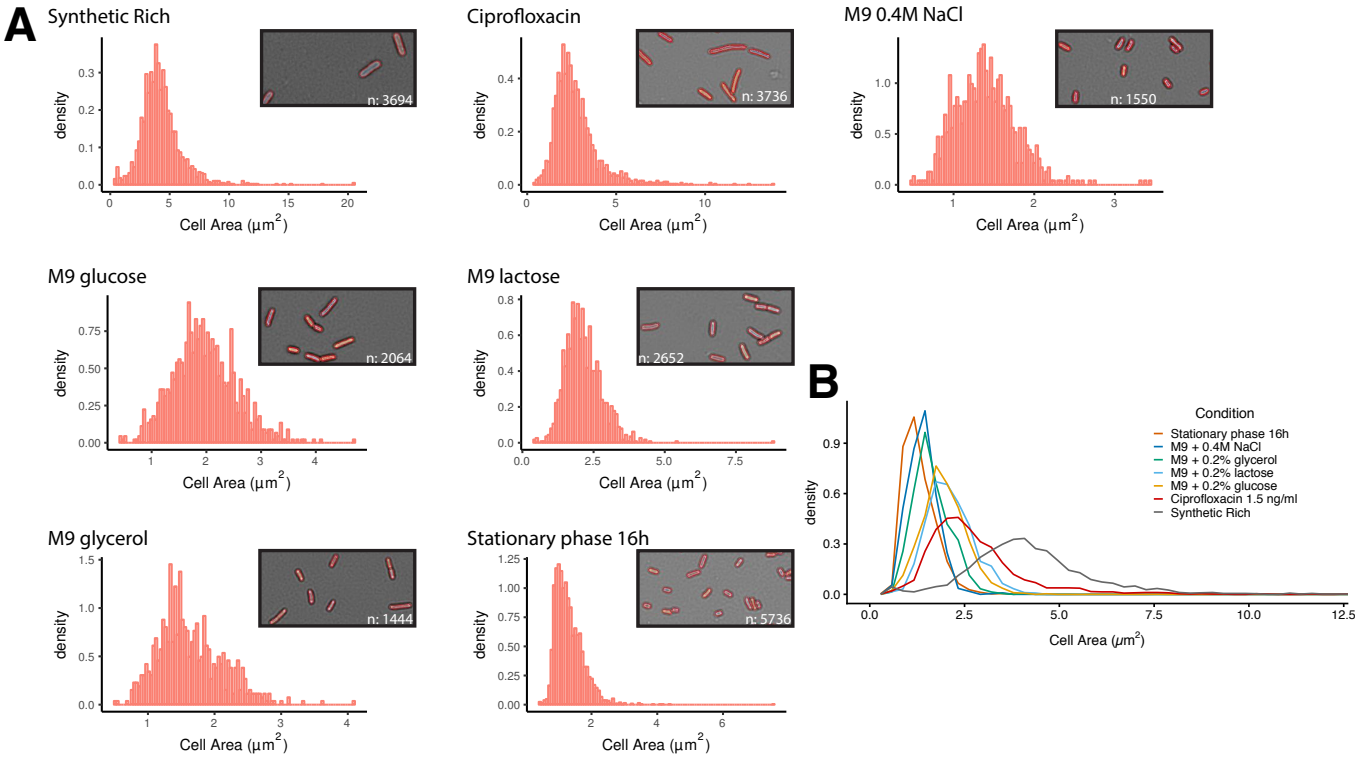


Figure S2.2: Cell sizes distributions. A: Histograms of the distribution of single-cell areas (μm^2, x-axis) in each condition. The insets in each condition show segmentation examples together with the number of cells used to estimate the mean and standard-deviation of the areas. **B:** Kernel-density estimates of the distribution of areas across all conditions (Areas bigger than 12.5 μm^2 are not shown).

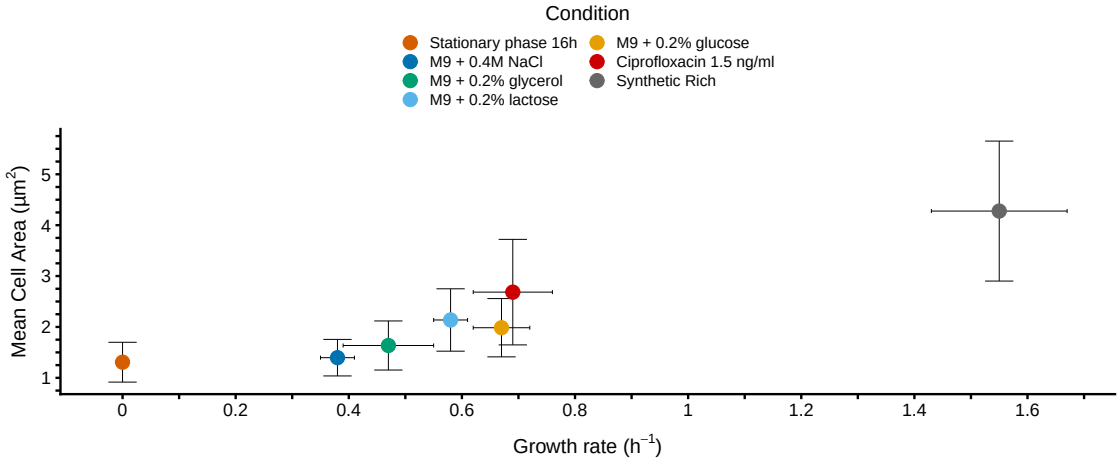


Figure S2.3: Cell area as a function of growth rate. Mean cell area (in μm^2) as a function of the growth rate (h^{-1}) in all conditions except stationary phase 30h. Cross-hairs indicate standard-deviations.

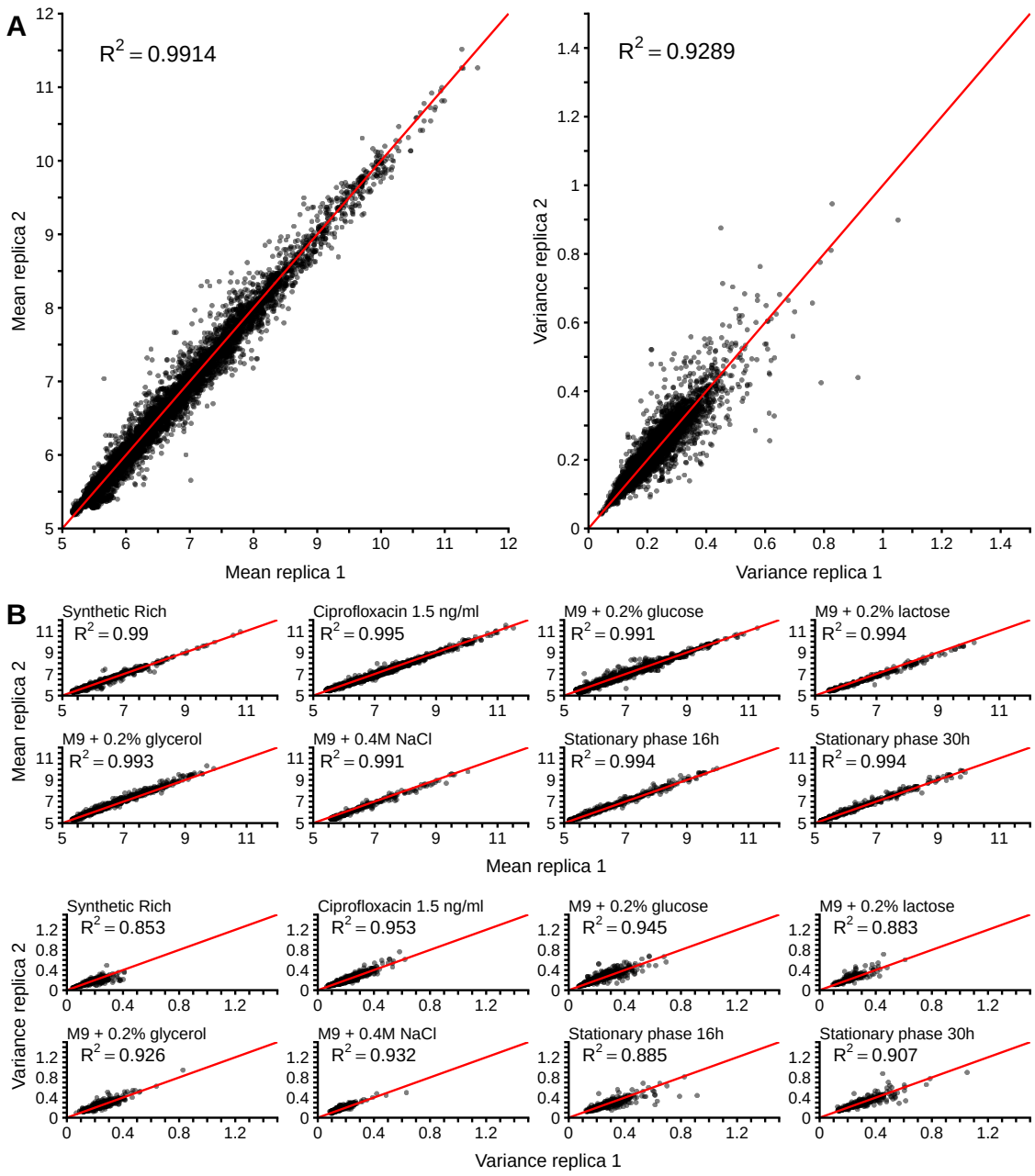


Figure S2.4: Reproducibility of measured means and variances. A: Means (left-panel) and variances (right-panel) of promoters (each represented by a black dot) measured on different days. The Pearson squared-correlations are indicated in each panel. **B:** Reproducibility of means (top panel) and variances (bottom panel) separately for each condition. Pearson squared correlations are indicated in each panel.

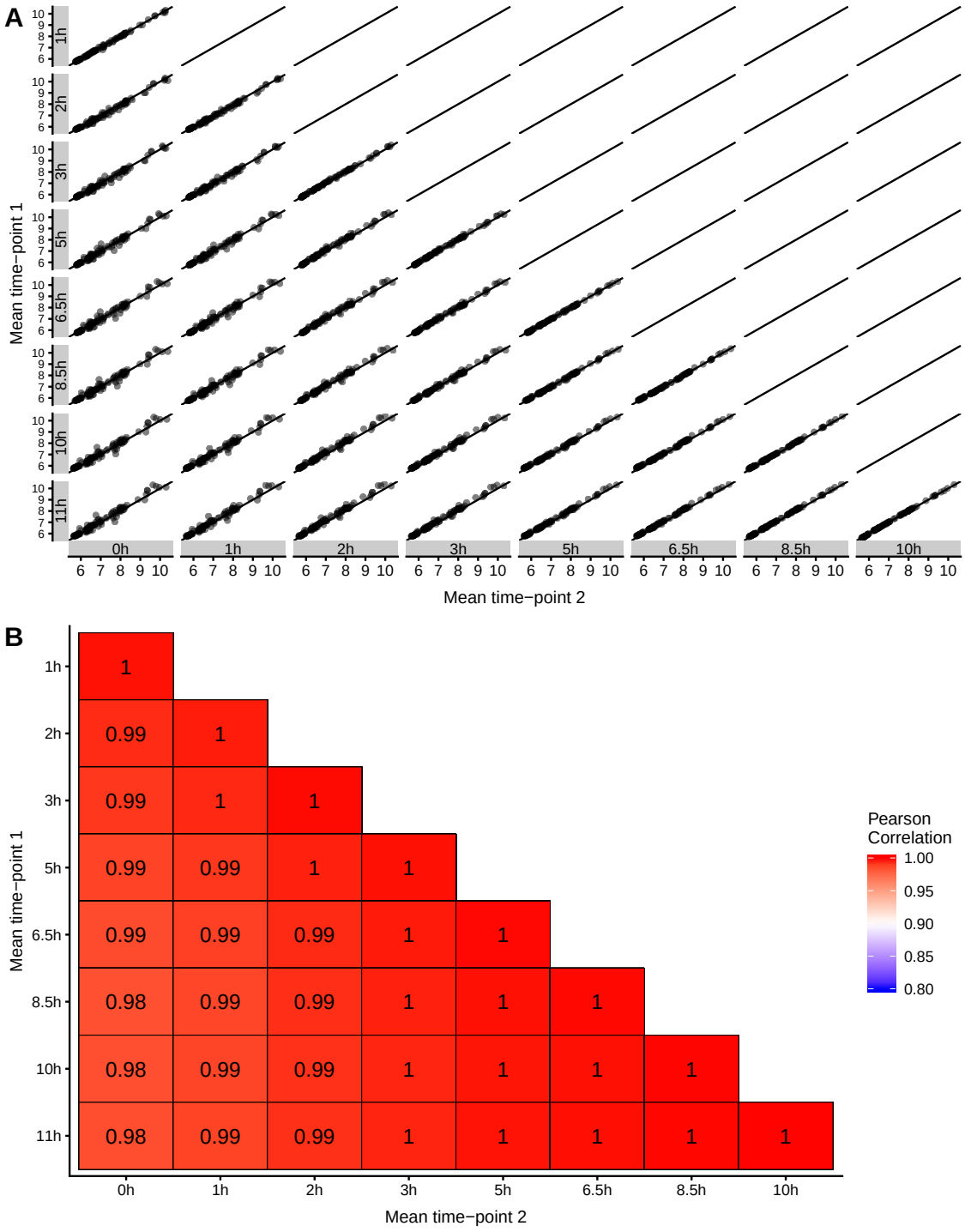


Figure S2.5: Reproducibility of measured mean fluorescences at different timepoints during growth. **A:** Correlations of mean expression levels for 95 promoters from the library, measured at consecutive time points during growth in M9 + 0.4M NaCl (+ 0.2% glucose). The time points range between 0h (freshly diluted culture) and 11 hours. The grey boxes on the axes indicate the time points that are being compared. **B:** R^2 Pearson correlation coefficients of measured mean expression levels for all pairs of timepoints.

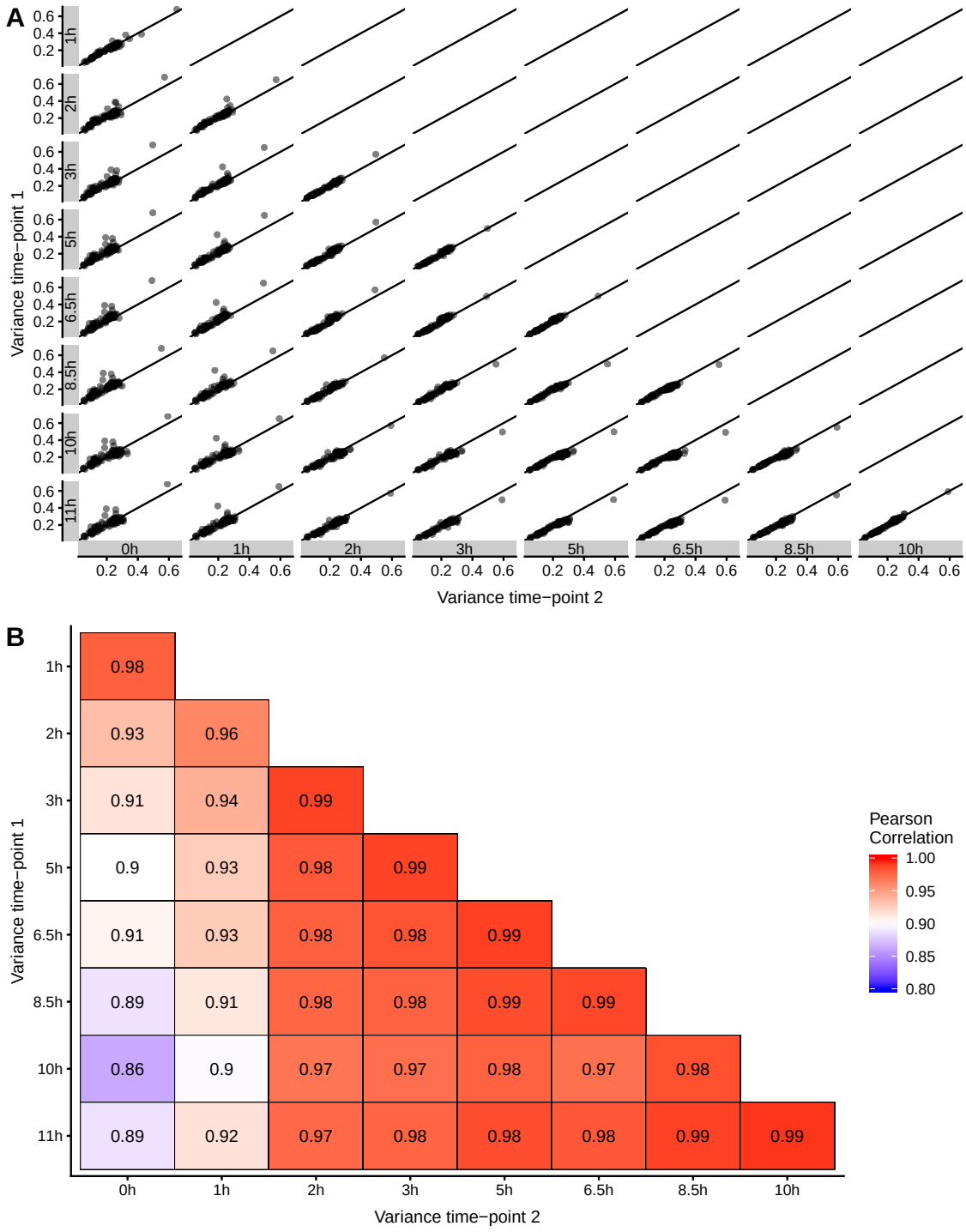


Figure S2.6: Reproducibility of measured fluorescence variances at different time-points during growth. **A:** Correlations of variance in expression levels for 95 promoters from the library, measured at consecutive time points during growth in M9 + 0.4M NaCl (+ 0.2% glucose). The time points range between 0h (freshly diluted culture) and 11 hours. The grey boxes on the axes indicate the time points that are being compared. **B:** R^2 Pearson correlation coefficients of measured variances in expression levels for all pairs of timepoints.

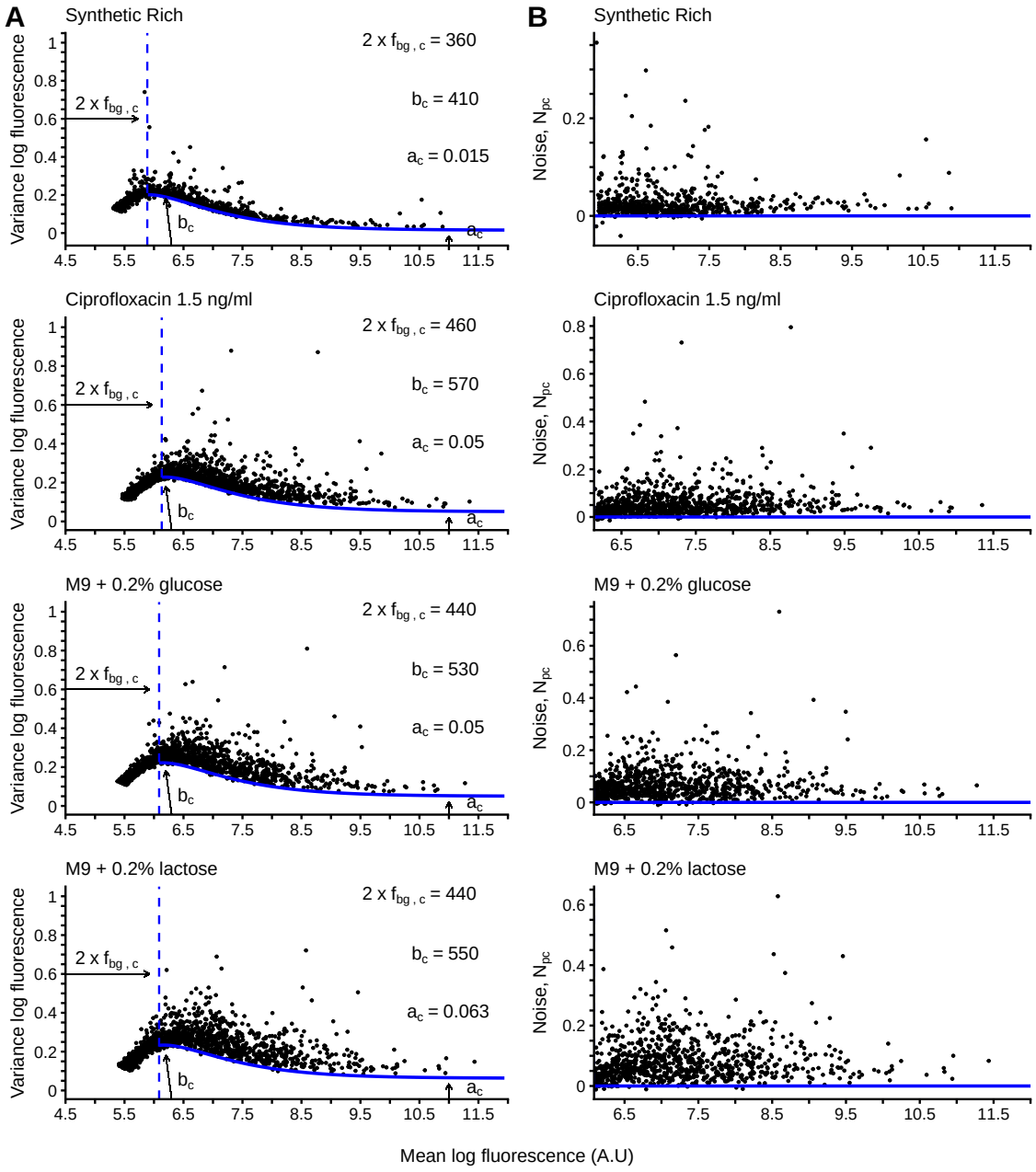


Figure S2.7: Means and variances of promoters from the library in all conditions. A: Variance as a function of mean for all promoters measured in each condition. Each promoter is represented by a black dot. The blue line indicates the predicted minimal variance as a function of mean . The model breaks at fluorescence levels close to background (left of the vertical blue dashed line), thus we only considered promoters above it. The number of promoters measured per condition is annotated inside each panel. **B:** Noise-level N_{pc} as a function of mean after correcting for the mean-dependent noise floor, i.e. differences between measured variance and minimal variance (Figure continued on next page).

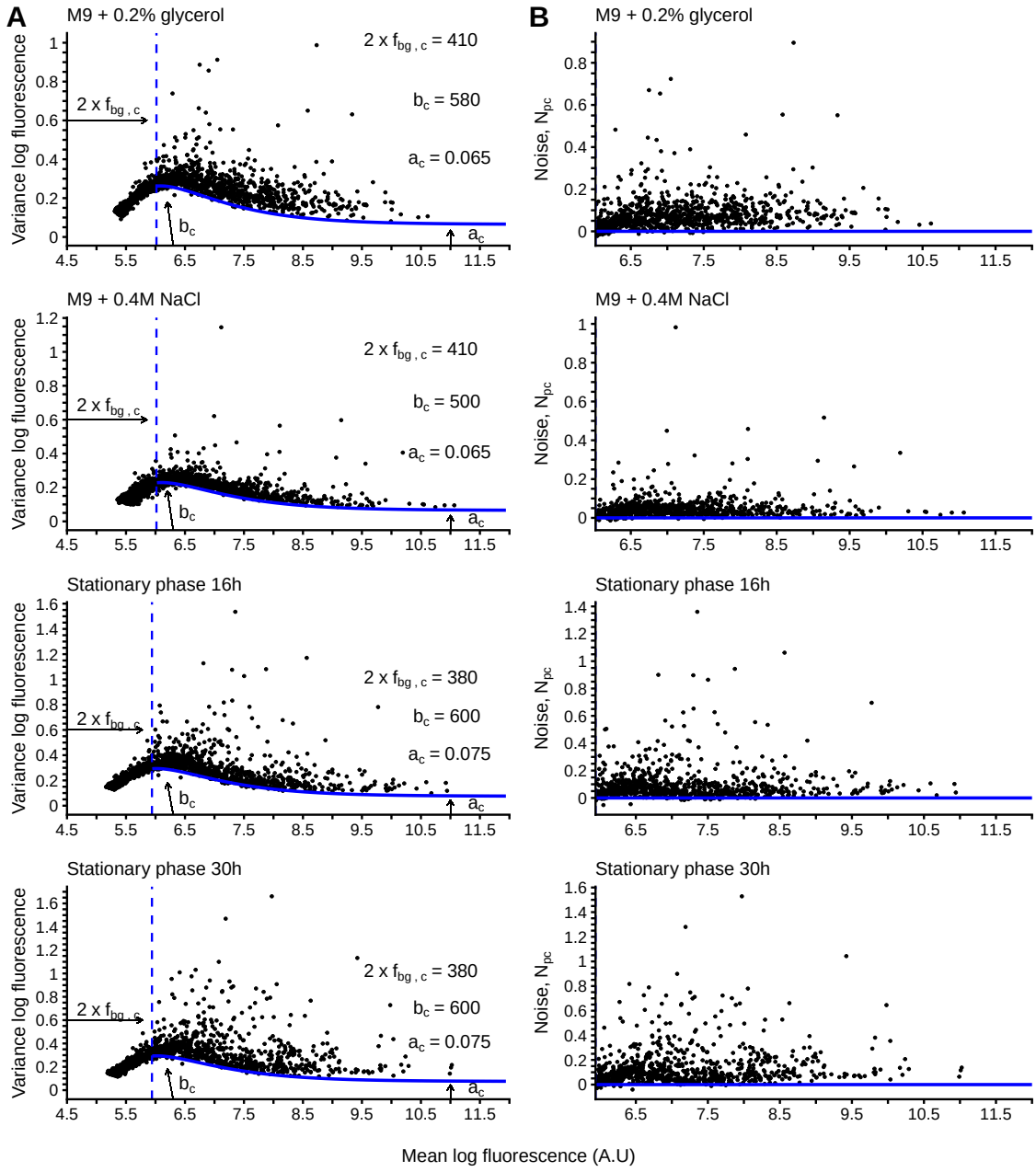


Figure S2.7: Means and variances of promoters from the library in all conditions. A: Variance as a function of mean for all promoters measured in each condition. Each promoter is represented by a black dot. The blue line indicates the predicted minimal variance as a function of mean . The model breaks at fluorescence levels close to background (left of the vertical blue dashed line), thus we only considered promoters above it. The number of promoters measured per condition is annotated inside each panel. **B:** Noise-level N_{pc} as a function of mean after correcting for the mean-dependent noise floor, i.e. differences between measured variance and minimal variance.

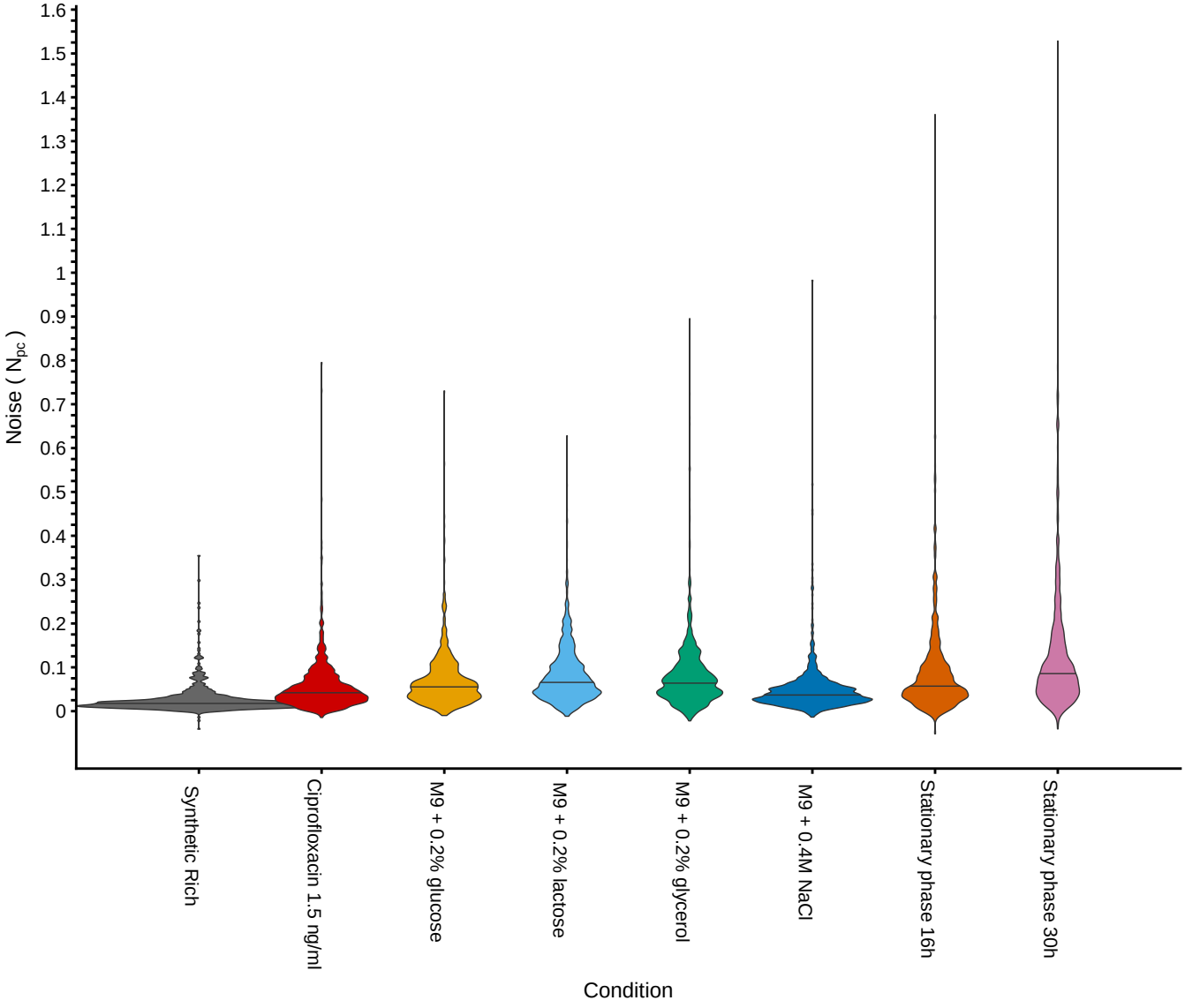


Figure S2.8: Noise levels across all conditions. The violin plots show the Noise distributions of all promoters in each of the measured conditions. The horizontal lines indicate the medians.

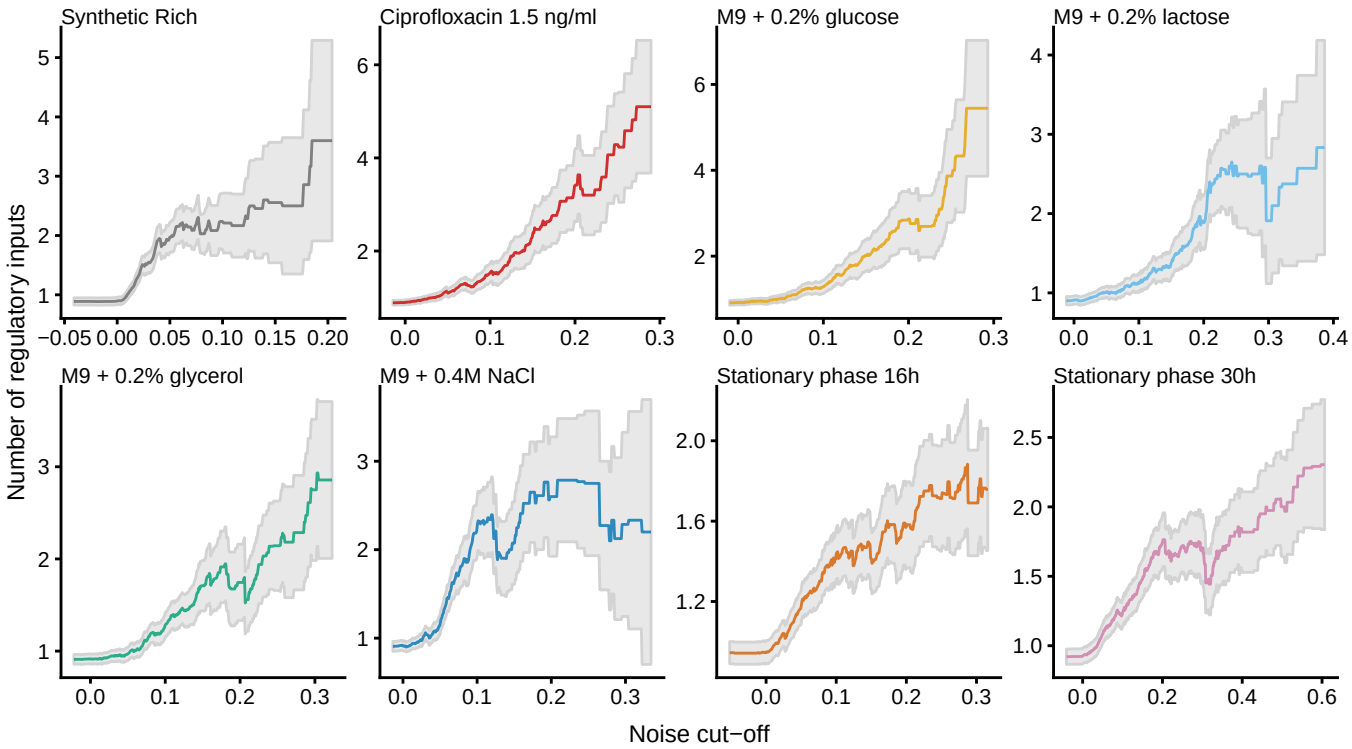


Figure S2.9: Noise as a function of number of regulatory inputs. In each condition we sorted promoters by their noise level N_{pc}, and calculated the mean and standard-error of the number of known regulatory inputs (y-axis) of all promoters above a cut-off in N_{pc} (x-axis). Regulatory interactions were annotated from Regulon DB [20].

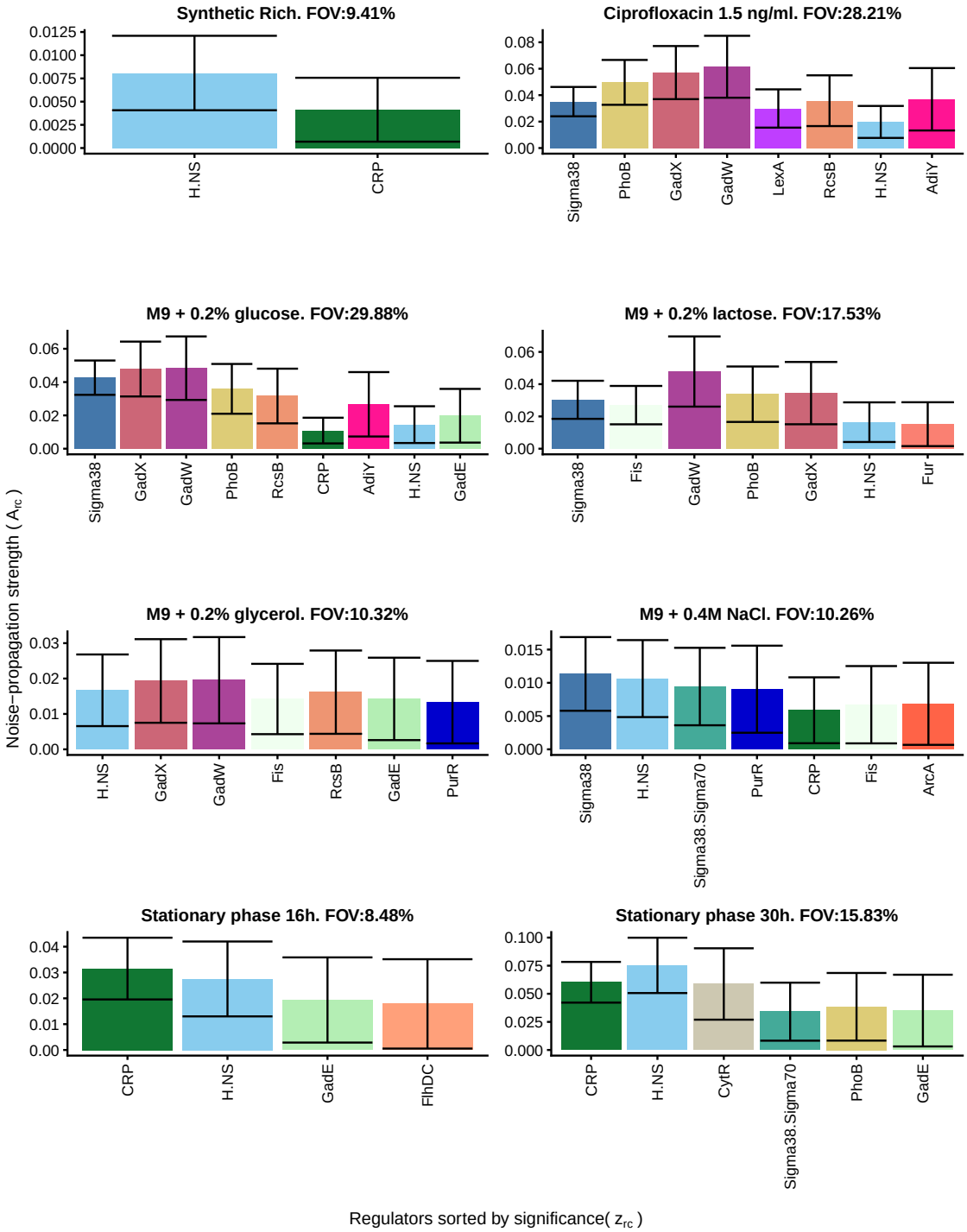


Figure S2.10: Strongest noise-propagators in each condition. Each panel corresponds to one growth condition and shows the inferred noise propagation strengths A_{rc} for the transcription factors for which $A_{rc} > \delta A_{rc}$ in that condition. The TFs are sorted by their overall significance z_r. The condition is indicated above each panel together with the fraction of variance (FOV) explained by the model.

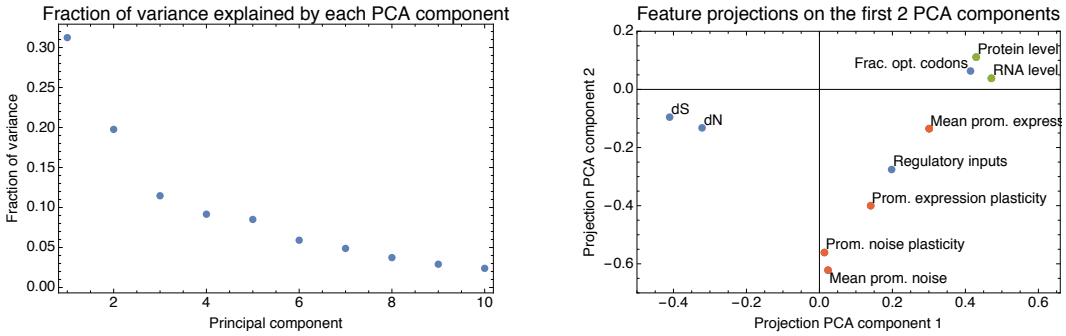


Figure S2.11: Principal component analysis of the 10 **gene features**. **A:** Fractions of the total variance in gene features captured by each of the PCA components. Note that the first two components together capture more than 50% of the variance. **B:** Projection of each of the 10 features on the first two PCA components. Expression levels from the literature are shown in green, sequence features are shown in blue, and gene expression features measured in this study are shown in red.

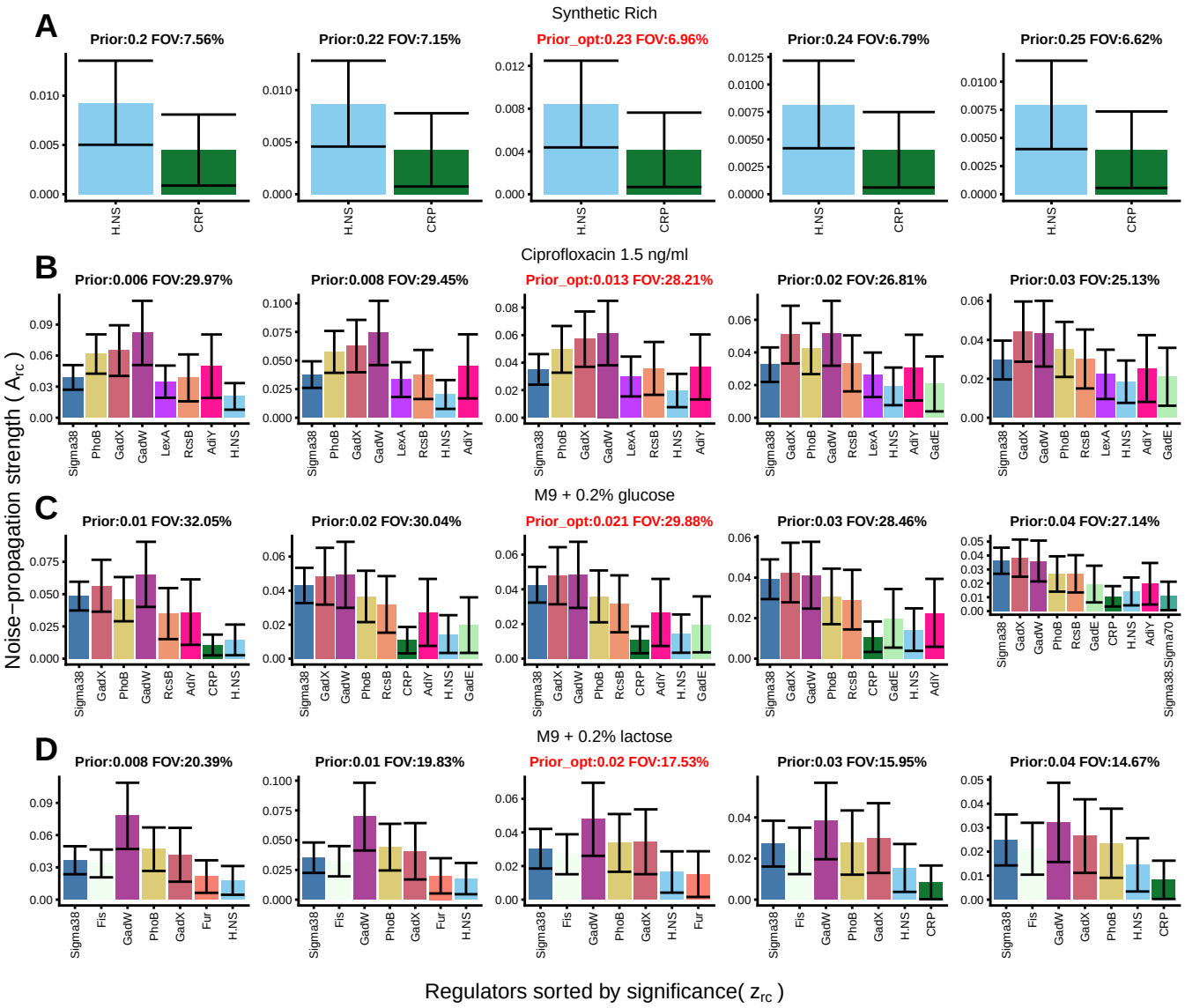


Figure S2.12: Inference with different Gaussian priors. **A:** Each panel shows the inferred transcription factors (with $A_{rc} > \delta A_{rc}$) contributing to explain noise using different values of the Gaussian prior used for cross-validation in Synthetic Rich. The plot in the middle with the title coloured in red shows the results with the optimal prior chosen automatically by the model and used in this study. **B:** Results in Ciprofloxacin 1.5 ng/ml. **C:** Results in M9 glucose 0.2%. **D:** Results in M9 lactose 0.2%.

The figure continues in the next page.

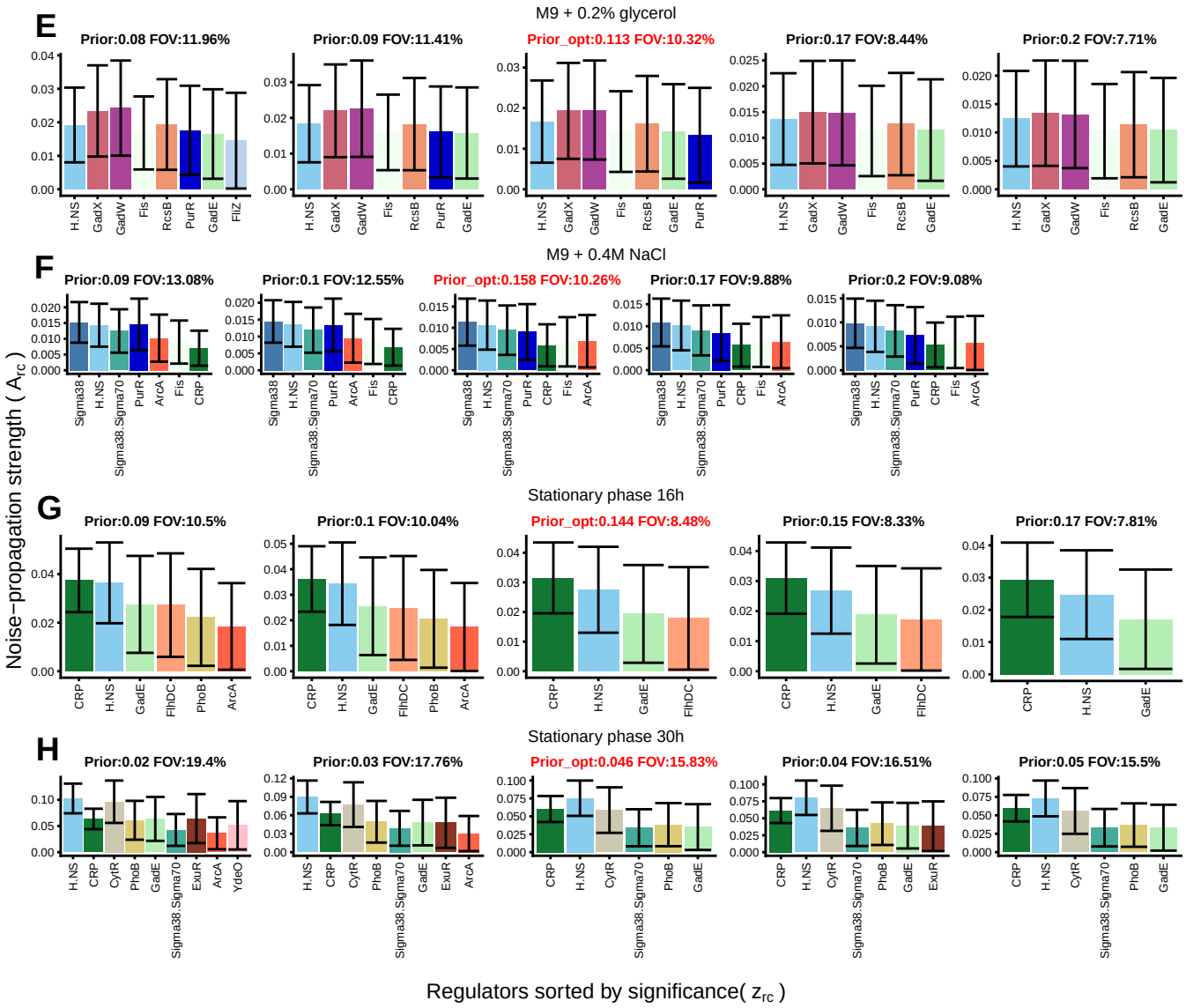


Figure S2.12: Inference with different Gaussian priors. **A:** Each panel shows the inferred transcription factors (with $A_{rc} > \delta A_{rc}$) contributing to high noise using different values of the Gaussian prior used for cross-validation in M9 glycerol 0.2%. The plot in the middle with the title coloured in red shows the results with the optimal prior chosen automatically by the model and used in this study. **B:** Results in M9 salt 0.4M NaCl. **C:** Results in Stationary phase 16h. **D:** Results in Stationary phase 30h.

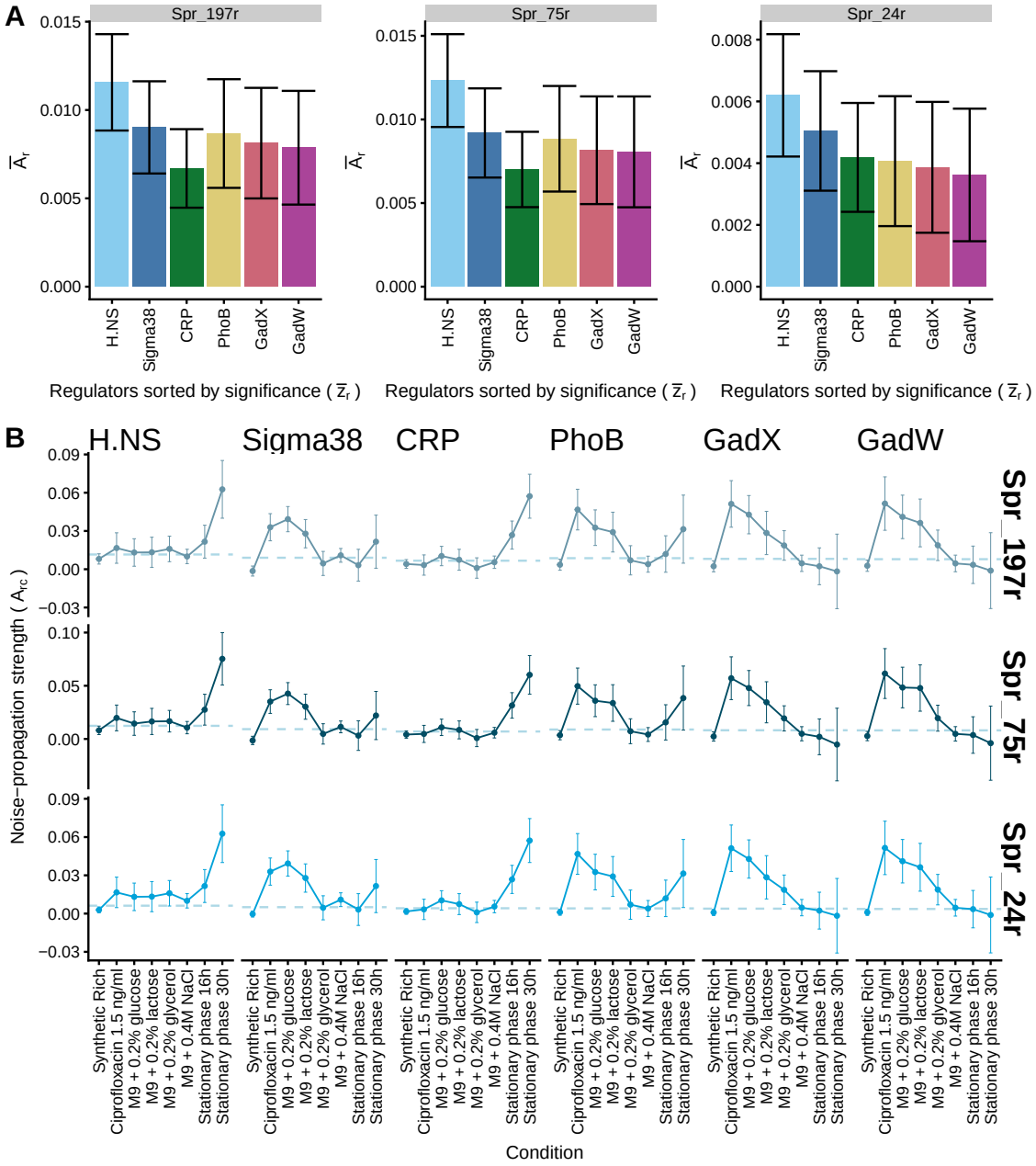


Figure S2.13: Inference with different versions of the binary regulatory matrix A: In each panel, a different version of the regulatory binary matrix has been used and shown are the average noise propagation strengths ($\bar{A}_r$, y-axis) and the error bars ($\delta\bar{A}_r$, vertical lines) of the top noise-propagators (with $\bar{A}_r > \delta\bar{A}_r$), consistently contributing to high noise levels in all conditions (sorted by significance, $\bar{z}_r$, x-axis). The different versions consist of a matrix with 197 transcription factor's (all transcription factors with at least 1 target in the library, left-panel), 75 transcription factors (TF's with more than 3 targets, panel in the middle) and a version with the top 6 inferred propagators on average, Sigma70 and 17 other randomly chosen ones (24 TF's in total, right panel). **B:** For each of the versions (vertical annotations , left) we show the condition-dependent (x-axis) noise propagation strengths, (A_{rc}, y-axis), and error bars, (δA_{rc}), of the top noise-propagators. The dashed blue line indicates the average noise propagation activity, $\bar{A}_{rc}$.

3

Low noise is the default state of constitutive promoters

Arantxa Urchueguía[1], Guillaume Witz[2], Thomas Julou[1], Erik van Nimwegen[1]

[1] Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Basel, Switzerland

[2] Mathematisches Institut, University of Bern, Bern, Switzerland

3.1 Introduction

Measurements of gene expression across homogeneous populations of cells have shown that in prokaryotes and eukaryotes different genes exhibit different levels of gene expression variability among individual cells [41, 86, 95, 96]. Theoretical models, which assume constant rates of transcription, translation and decay rates predict that the variance in expression should be proportional to its mean [87]. This source of variability, inherent to the mechanisms of gene expression, is typically referred to as intrinsic noise [77] and is believed to be dominated by Poisson fluctuations [160]. Under steady-state Poisson statistics, the Fano factor (defined as the variance divided to the mean) of protein and mRNA number distributions equals 1. Indeed, mean and variance measurements of single-cell mRNA and protein levels of individual promoters have shown that expression variance tends to scale with mean both in prokaryotes [41, 42] and eukaryotes [85, 86]. Highly expressed genes, however, display a variance much larger than the one predicted by a model that assumes Poissonian statistics in transcription and translation rates. Rather, their variances are better fitted with a model in which their promoters randomly switch between *on* and *off* states in stochastic bursts [41, 84, 88]. Extensive experimental evidence has confirmed this model of gene expression in prokaryotes [41, 42, 45, 89, 161] and eukaryotes [83, 162]. These sources are commonly referred to as extrinsic noise [77], and can affect both mRNA and protein production. Here we will mainly focus on the sources affecting bursty mRNA production in bacteria, which have been proposed to include: transcription factors concentration and their binding/unbinding dynamics [45, 90], intracellular changes in DNA supercoiling [163], differences in free RNA polymerase concentration [91], changes in gene copy number [164] or noise propagation from upstream components of regulatory networks [96, 109, 111], and extensively discussed in Chapter 2. The coupling between gene expression and growth [120, 121] has also been attributed to affect noise due, for instance, cell-to-cell differences in cell-cycle dependent changes [92], differences in single-cell growth rates [93, 165] or the random partitioning of molecules during cell division [94]. However, because all these different sources can lead to very similar results, making a clear distinction between them remains challenging [166].

The fact that two independent noise sources dominate at different expression levels has led to a widely accepted belief that gene expression variance follows a universal scaling law dictated by mean expression [41, 42, 166], with intrinsic noise dominating

at low expression and extrinsic noise at high expression. A set of recent publications in bacteria suggest, however, that there is no evidence to assume a universal scaling between mean and variance, but rather that the noise level of a promoter is determined by the specific features of its sequence. In a study by [90], it was found that the variability in mRNA production of constitutive promoters in *E.coli* could be explained by only assuming Poisson statistics in gene expression (i.e. constant rates of mRNA production and degradation). Deviations from the fluctuations expected by a Poisson process were achieved by just finely tuning promoter features, such as transcription factor binding strength, copy numbers and polymerase binding strength. In the work by [96], an evolved set of constitutive synthetic promoters in *E.coli*, expressing at two different levels, showed noise levels close to the minimal one that native promoters can exhibit, suggesting that noise is an evolvable trait that has been positively selected for in noisy promoters. Furthermore, as presented in Chapter 2, we have shown strong evidence of regulation being the most important feature explaining high noise in native *E.coli* promoters. Altogether, these studies deviate from the previously introduced idea of a universal scaling law between variance and mean and place the focus on the individual characteristics of promoter sequences. Nevertheless, as shown in Chapter 2, we still observed a strict lower bound on variance as a function of mean when measuring protein noise levels in terms of the CV^2 (equivalent to the variance in log-fluorescence) using flow cytometry. Since we recently discovered that flow cytometers introduce a noise component (shot-noise) that affects the variance estimates by an amount that scales inversely proportional to the mean expression level [122], we hypothesized that the observed lower bound in previous flow cytometry measurements is a result of shot-noise.

Here, we aim to understand to what extent shot-noise is dominating variance measurements acquired using flow cytometry, but most importantly, how unregulated promoters expressing at different levels change their variance relative to their mean in different conditions. We present protein mean and variance estimations from single-cell fluorescence distributions in a set of *E.coli* constitutive synthetic promoters [96] and low regulated native promoters [128]. The study of constitutive (i.e. unregulated) promoters is relevant, as they are not affected by noise propagation from regulation, thus making them ideal candidates to study how variance scales with mean. Moreover, measuring them across different conditions allows to systematically compare how regulated versus unregulated promoters change their noise across conditions and to quantify the extent by which differences in noise across en-

vironments are caused by general condition-dependent physiological changes, such as fluctuations in growth rate, cell size, ribosome copy-number, etc. To measure single-cell fluorescence levels of the different promoters, we used flow cytometry and an innovative high-throughput microscopy method we developed to overcome the possible biases in estimating variances with flow cytometry. The microscopy procedure consists of a unique way to measure and analyse 96-well plates images in an automated manner and in a reasonably short amount of time (mimicking a flow cytometry-like approach). A great advantage of the method is that it is easily extendable to other applications in which single-cell statistics are pivotal. In short, we found that the lower bound on variance as a function of mean observed using flow cytometry was not reproducible under the microscope. All measured promoters, at different expression levels, showed very similar variances, resulting in a very clear floor in variance in all conditions. The observed minimal variance floor is only dominated by extrinsic noise sources, as the chosen promoters express at a level at which only extrinsic fluctuations dominate (>10 proteins per cell [41]). We discuss the possible sources contributing to the floor in a condition-dependent context.

3.2 Results

In Figure 3.1A we illustrate the typical lower bound on variance as a function of mean expression that native *E.coli* promoters [128] display when measured using flow cytometry (FCM). Shown are promoters growing in minimal media supplemented with 0.2% glucose, albeit the lower bound consistently appears across different conditions (Figure 3.1B and Supplementary Figure S3.2A). As previously described (see Chapter 2) [128], the measured promoters consist of a library of transcriptional fusions of ∼75% of all *E.coli* intergenic regions. The constructs were fused to a fast-folding GFP protein downstream of a strong ribosomal binding site and were designed specifically to measure transcription. Although fluorescence measurements only allow direct estimations of protein levels, the observed differences in variance between promoters mainly reflect transcriptional variability. We believe this assumption to be accurate, as all mRNA's of the different constructs are identical (except in some fusions in which operon specific 5' untranslated regions may be present). Nonetheless, it should not be despised that part of the differences can also arise from fluctuations in translation rates. In Figure 3.1A each mean and variance

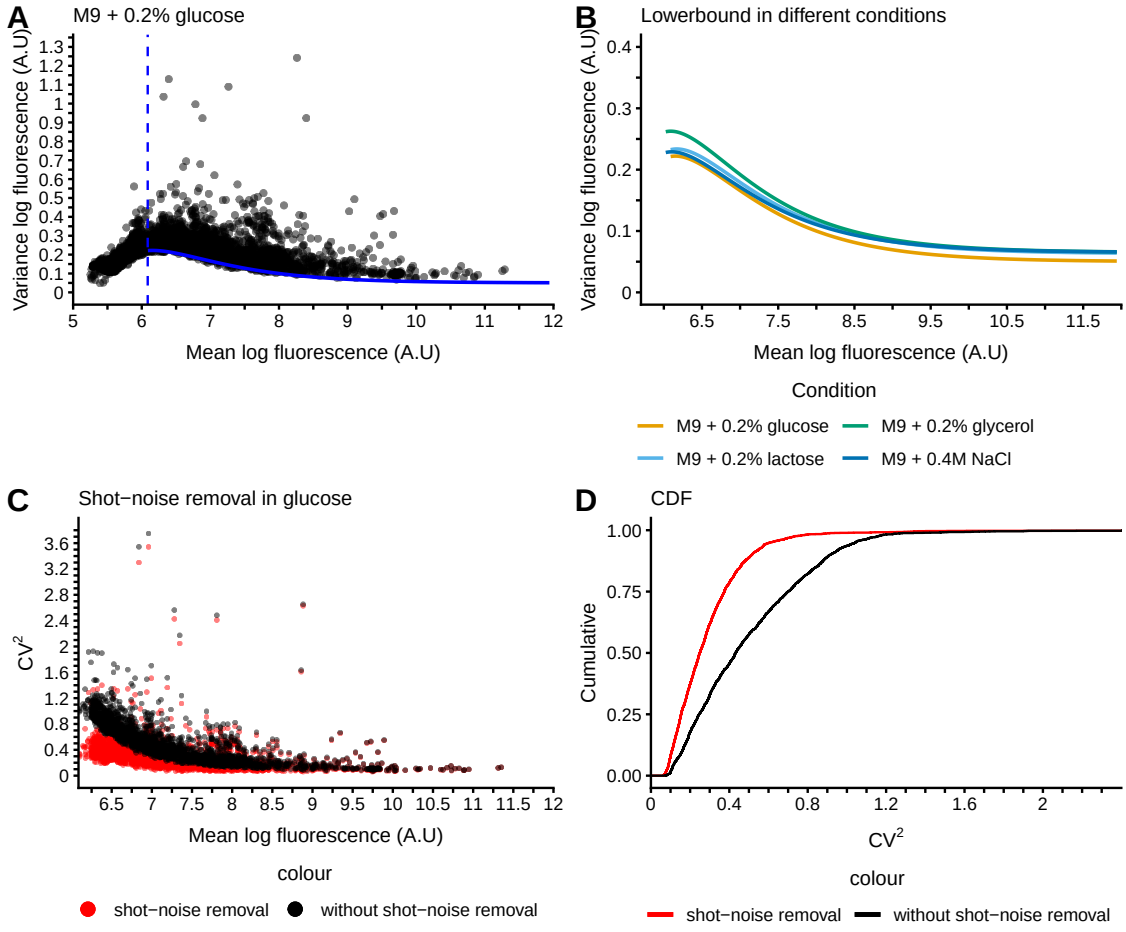


Figure 3.1: Flow cytometers introduce shot-noise. A: Shown are the mean (x-axis) and variance (y-axis) in log-fluorescence of a set of *E.coli* native promoters characterized with flow cytometry, and grown in minimal media + 0.2% glucose (each point represents one promoter). Clearly, there is a strict lower bound on variance as a function of mean. The blue line indicates the theoretical predicted minimal variance, which breaks for promoters below a background threshold (dashed vertical blue line). **B:** The lower bound (represented by the theoretical predicted minimal variance) consistently appears across different conditions (different colours, Supplementary Figure S3.2A). The displayed conditions consist of minimal media supplemented with either 0.2% glucose, 0.2% lactose, 0.2% glycerol and 0.4M NaCl (+0.2% glucose as carbon source). **C:** Each point indicates the CV^2 (y-axis, which is equivalent to variance of the log-fluorescence whenever fluctuations are small relative to the mean) as a function of mean expression (x-axis) of each promoter shown in panel A (above a background threshold, indicated by the vertical blue-dashed line) before (black dots) and after (red dots) shot-noise removal (See Supplementary Figure S3.2B for the other conditions). In panel A, the auto-fluorescence in each promoter was not removed, whereas here the auto-fluorescence in each promoter has been subtracted. **D:** Cumulative distribution of the CV^2 of the set of promoters represented in panel A, before (black line) and after (red line) shot-noise removal. See Supplementary Figure S3.2C for the other conditions.

was estimated from single-cell log-fluorescence protein distributions. The method has been described in detail in Chapter 2, as well as in other studies [95, 96].

Typically, in order to compare the variance of promoters with different mean, we correct for the dependency illustrated in Figure 3.1A by fitting a simple model that takes into account information on the background fluorescence and two parameters, one that scales with the square of the mean and another proportional to the mean [96] (Figure 3.1A and Supplementary Figure S3.2A, blue line; Figure 3.1B, various colours). We then define the corrected variance of a promoter as its variance, in log-space, above and beyond this lower bound (the term is equivalent to the CV^2 and we sometimes refer to it simply as noise). The physical interpretation of the two parameters, which determine the shape of the lower bound has been, classically, attributed to reflect both intrinsic and extrinsic noise sources, each dominating at different expression levels [41, 85, 86, 96]. Intrinsic fluctuations scale proportionally to the mean expression level, whereas extrinsic fluctuations scale with the square of the mean, dominating at high expression levels [41].

Recently, our group has put great efforts in trying to quantitatively estimate the accuracy and sensitivity of gene expression measurements in bacteria using flow cytometry [122]. We discovered that flow cytometers propagate a noise component (shot-noise), which affects variances by an amount inversely proportional in log-space to the mean level. Figure 3.1C and D illustrate how shot-noise affects the CV^2 estimations. In Figure 3.1C each promoter's variance is plotted against its mean and the colours indicate whether the shot-noise has been removed from the estimations or not. Since the variance is affected by an amount that scales inversely with the mean, the CV^2 of promoters expressing at high levels is not affected. In contrast, low expressed promoters show significant differences in CV^2 after removing the shot-noise (Supplementary Figure S3.1). Figure 3.1D shows the cumulative distribution of the CV^2 of all promoters before and after shot-noise removal. On average, promoters decreased 37% their CV^2 after removing shot-noise (with the highest decrease being 73.6% and the lowest 1.3%; p-value $<$ 2.2e-16, two-sample Kolmogorov-Smirnov test). Supplementary Figures S3.1, S3.2C and S3.2D show how, in different conditions, shot-noise also dominates the CV^2 estimations of low expressed promoters. These results were striking and opened the question of whether a big fraction of the observed mean-variance dependency was an artifact

introduced by flow cytometers, motivating us to develop an alternative method to flow cytometry.

Flow cytometry has numerous advantages, mainly its high throughput and ease of use, which allows fast and accurate measurements of a large number of samples [167]. We aimed to mimic its usability by developing a high-throughput method based on microscopy, instead. The goal was to ultimately be able to measure a large number of promoters and take *snapshots* of gene expression at specific time points in a FCM-like approach. The main advantage in using microscopy is that it allows direct visualization of single-cells. However, classical microscopy experiments, based on agar patches, involve manual strain-by-strain curation, which constitutes a major bottleneck when having a large number of environments or strains to investigate. Other microscopy approaches, such as Mother machine experiments [159], are great to study single-cell expression dynamics, but at the moment they do also present important limitations in their throughput. The method we describe next, allowed us to overcome the throughput bottleneck by automating microscopy control [168].

In Figure 3.2A we summarize the outline of the method. A step-by-step description of the protocol with all critical points is available as Supplementary text. In short, we used poly-L-lysine 0.01% to attach exponentially growing single-cells to the base of bottom-glass 96-well plates, followed by centrifugation and washing steps. We used a wide-field inverted fluorescence microscope for imaging (Nikon Ti), which we automatized in order to run 96-well plates without human interaction. An essential aspect of the pipeline was the generation of an image analysis routine to extract single-cell statistics from the acquired fluorescent images. We developed an automated high-throughput segmentation tool that uses convolutional neural networks for cell recognition, available as a stand-alone package in Python. In summary, in order to perform cell recognition, we generated a training set by segmenting fluorescent images of cells growing in minimal media + 0.2% glucose (highly expressing GFP) and applying the masks on phase correlation images to create weight-maps. The weights learned in the training were then used to segment all images and obtain total fluorescence values (in pixels) from individual cells.

In Supplementary Figure S3.3 we show different segmentation examples. The package uses OME-TIFF images [169] as input, the standard file format in which images are automatically saved after acquisition. Each tiff image is processed in parallel (speeding up the analysis time) and the output is stored in different formats.

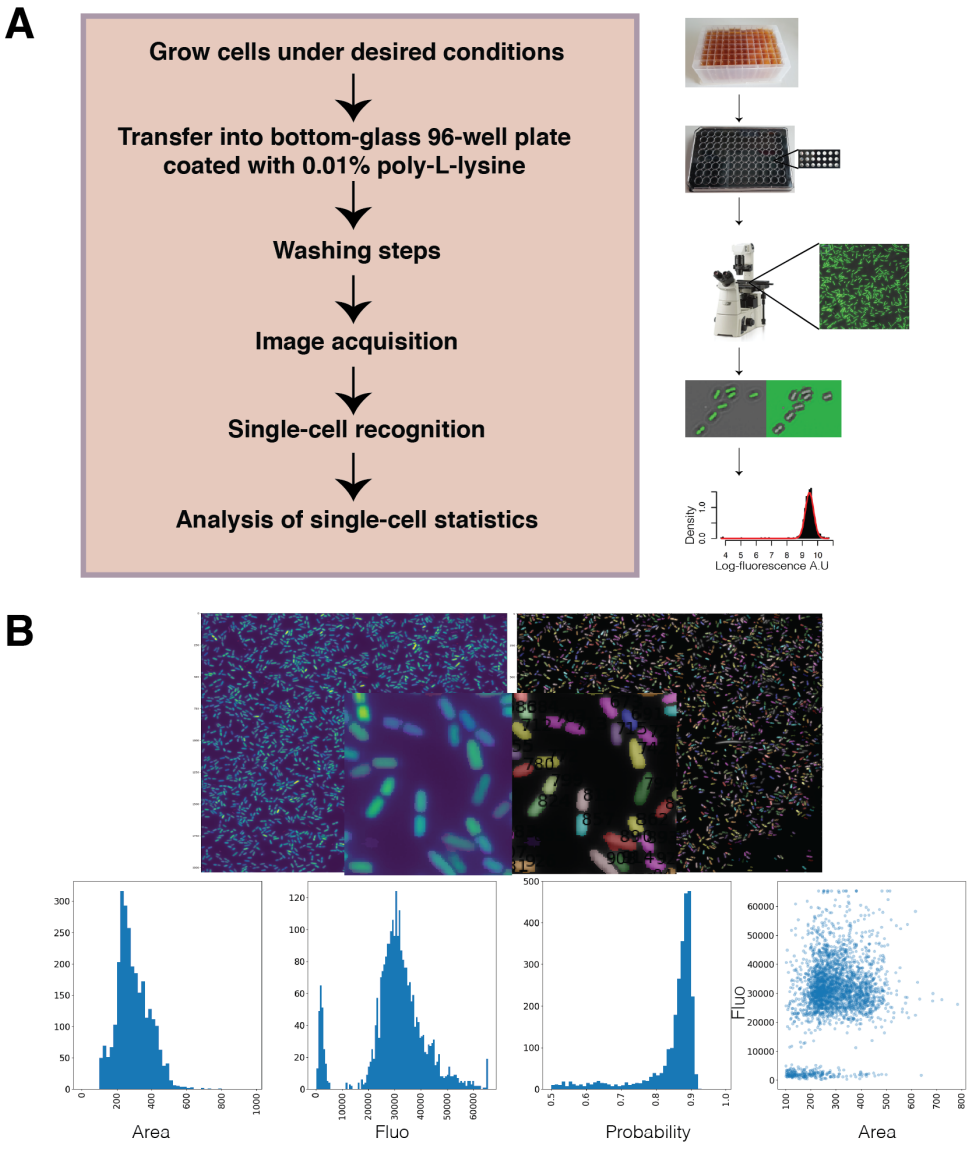


Figure 3.2: High-throughput microscopy method. A: Step-by-step description of the procedure we used to acquire fluorescent snapshots of single-cells attached to the bottom-glass of 96-well plates. See the Supplementary text for a more detailed explanation. **B:** Example of the visual output of the image analysis pipeline we developed to segment the acquired images and extract single-cell statistics. The images are automatically saved as png, allowing a fast visualization of the quality of the segmentation. Each segmented object is assigned a unique ID.

First, a png image saved individually per image allows visual examination of the segmented cells and their masks. It features histograms of the extracted single-cell statistics of the image, such as cell area (in pixels), fluorescence intensity (in pixels), a probability assigned to each mask for it to belong to a cell and a scatter plot of fluorescence intensity against area per segmented object (Figure 3.2B). To plot the histograms, no constraints on probability, area or cellular eccentricity are considered. These can be defined later, as it may happen that the algorithm wrongly segments other objects in the image (e.g. dust particles, Supplementary Figure S3.4 and Supplementary Figure S3.5). These errors, however, will show fluorescence distributions with values close to the background fluorescence of the image (the background in each image changes depending on the fluorescence intensity of the cells, as shown in Supplementary Figure S3.5). To reduce the possible contribution of such errors in downstream analysis, it is important to include thresholds on probability, area and shape, which can be easily set as parameters when running the package. In Supplementary Figure S3.4 we show an image acquired close to the border between wells, and with a few number of cells to illustrate the importance of setting thresholds. In the image, the algorithm recognizes parts of the border as cells, but the masks have small probabilities values, as well as small areas. We typically set a minimal threshold of 0.8 in probability and of 200 pixels in area. Finally, all extracted pixels and statistics are automatically saved in different csv files.

The development of this experimental pipeline gave us the possibility to explore the mean-variance dependency of different promoters in *E.coli* and understand to what extent the dependency, observed when using flow cytometry, was due to shot-noise. In order to explore this question, we used a library of synthetic promoters known to have low noise levels by default [96]. The library is the result of *de novo* evolution in which random pools of DNA sequences (100-150 base pairs in length, ligated into the same plasmid in which the library of native promoters was cloned) were selected to two predefined mean expression levels, medium ($<$ 18.000 gfp molecules per cell) and high ($>$ 18.000 gfp molecules per cell). The selection was accomplished after 5 rounds of FACS sorting and PCR mutagenesis to generate genetic variation. The most interesting features of these promoters is that they show noise levels close to the minimal one observed in native promoters [96] and that they do not contain binding sites for any known transcription factor in their sequences (apart from the housekeeping sigma factor, σ70) [170]. These features made these promoters ideal candidates to study how variance scales with mean.

We first characterized 142 synthetic promoters, each containing a unique sequence from both expression levels (72 medium and 72 high) using flow cytometry. We measured single-cell protein fluorescence levels after growth in four different conditions: minimal media (M9) supplemented with either 0.2% glucose, 0.2% lactose, 0.2% glycerol or 0.2% glucose + 0.4M NaCl (Figure 3.3A). In these conditions, growth rates range from 0.67 to 0.37 h^{-1} in the slowest condition (see Chapter 2). As expected, the promoters showed a clear lower bound on variance as a function of mean in all conditions (Figure 3.3A, blue line, Supplementary Figure S3.6 to see the shot-noise effect on the measurements). We corrected for the dependency, using the simple model described in Chapter 2, in order to be able to compare their noise to the one exhibited by native promoters grown under the same conditions and characterized with flow cytometry, as well. We found that synthetic promoters showed significantly lower noise levels compared to native promoters (Figure 3.3B and Supplementary Figure S3.7). This was not surprising, as it had been reported before [96]. Besides, they also displayed lower plasticity in noise and mean (defined as the variance between all tested conditions, Figure 3.3C and D, respectively), confirming that they are not being subject to regulation (Supplementary Figure S3.8 and Supplementary Figure S3.9 illustrate how each individual promoter moves across conditions in the FCM). It is also noticeable, that highly regulated native promoters exhibit higher noise levels and higher plasticity (due to regulation) than native constitutive ones (Figure S3.10). As elaborated in the previous chapter, this is not surprising in a noise propagation scenario.

After flow cytometry characterization, we randomly selected 18 out of all synthetic promoters (9 medium and 9 high expressers), together with a strain carrying a plasmid without promoter and a native *E.coli* strain without any plasmid (*K12* MG1655). We measured fluorescence levels of the selected strains with our customized microscopy method (Figure 3.2A) after growing them in the exact same conditions as during flow cytometry experiments. In order to remove possible biases due to cell-to-cell inaccuracies in segmentation, we decided to define the fluorescence level of a cell as the average fluorescence of all pixels inside a 3x3 pixel box in the middle of the cell. We then fitted a mixture of two Gaussian's to the log-fluorescence distributions of all cells containing the same construct to further remove the possible contribution of wrongly segmented objects to the statistics. Finally, we inferred the mean and variance of each strain from the Gaussian with the highest probability (Figure 3.4A).

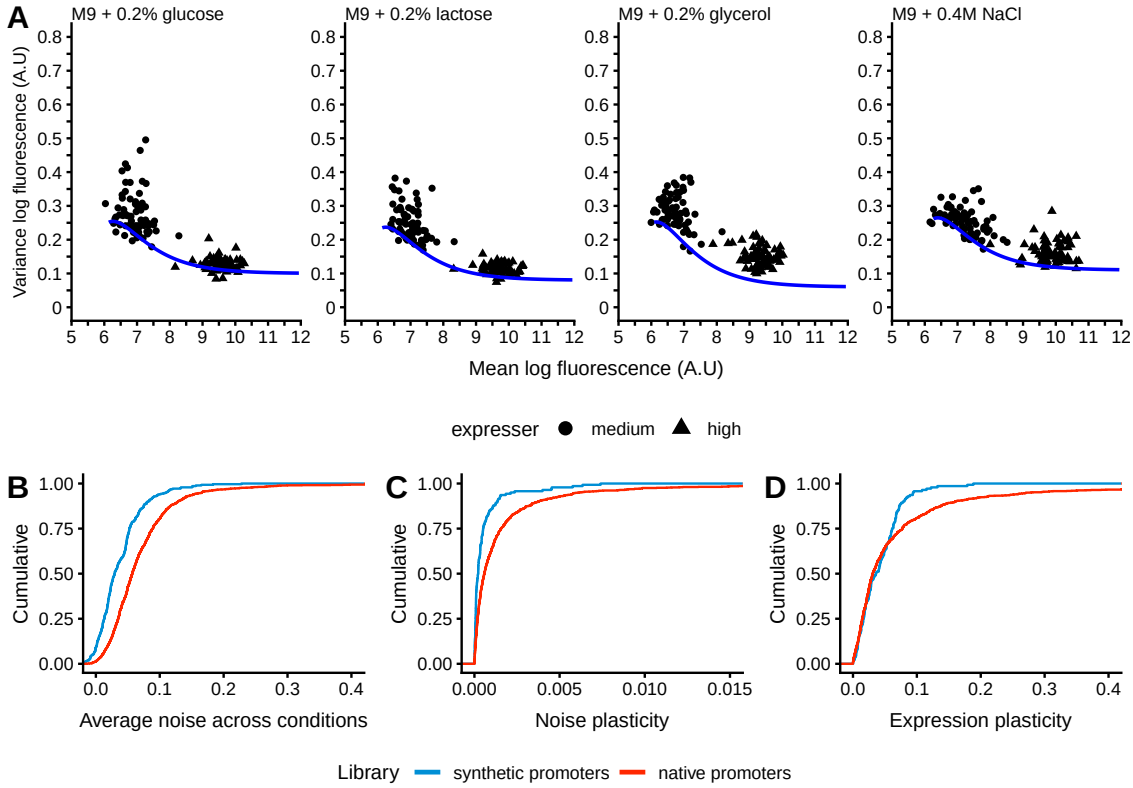


Figure 3.3: Low noise is the default state in unregulated synthetic promoters. A: Variance (y-axis) as a function of mean (x-axis, log-fluorescence) of a set of constitutive synthetic promoters measured using flow cytometry. We measured the library in four conditions (each indicated at the top of the panels) and, as expected, all displayed a very clear lower bound on variance as a function of mean. The blue lines indicate the fitted theoretical minimal variance. The effect of the shot-noise on the CV^2 is illustrated in the Supplementary Figure S3.6. **B:** We calculated the average noise (after correcting for the mean-variance dependency, Figure S3.7) of the synthetic promoters (blue line) and native promoters (red line) over the four measured conditions. Shown are the cumulative of the estimated average noise levels (also referred to as corrected variance in this study). **C:** Cumulative distribution of the noise plasticity (estimated as the variance in noise of each promoter over the four conditions) for the set of synthetic (blue line) and native promoters (red line). **D:** Cumulative distribution of the expression plasticity (estimated as the variance in mean expression of each promoter over the four conditions) for the set of synthetic (blue line) and native promoters (red line).

For our surprise, we found that in all tested conditions, we were not able to see any dependency on variance as a function of mean. On the contrary, all promoters at all expression levels displayed very similar variances. The results are illustrated in Figure 3.4B (Supplementary Figure S3.11 and Supplementary Figure S3.12A). The figure shows the variance as a function of mean in log-fluorescence of the promoters measured using microscopy, together with the predicted theoretical lower bound measured in the flow cytometer. Clearly, the dependency disappears when measuring them under the microscope. Figure 3.4C shows the average variance under the microscope in each condition, calculated over all promoters measured in the same condition (Figure 3.4B, black line). The value stays relatively constant across all, with a slightly higher variance in minimal media supplemented with glycerol (Figure 3.4C). Both medium and high expressers display variances higher than the background fluorescence (Supplementary Figure S3.12B, Supplementary Table S3.1 and Supplementary Table S3.2), suggesting that the differences have a biological origin. Medium expressers show slightly increased variances compared to high expressers (Figure 3.5A, Supplementary Table S3.1 and Supplementary Table S3.2), although their differences are not significant. The results strongly indicate that, in unregulated promoters, the previously observed lower bound on variance as a function of mean of low expressed promoters is an artifact introduced by flow cytometers. However, high expressed promoters measured with flow cytometry show a minimal floor in variance consistent with the minimal average variance observed in the microscope (Figure 3.3A and Figure 3.4C) . This indicates that flow cytometers only give accurate variance estimations of promoters expressing above a certain level (around log(fluorescence) > 8.5) (Supplementary Figure S3.1).

Unfortunately, a direct comparison between the absolute values in mean and variance obtained with flow cytometry and microscopy is not possible. First, the mean fluorescence values are a function of the intensity of the lasers used. Second, in the flow cytometry measurements we define the fluorescence level of a cell as its total fluorescence, whereas in the microscopy measurements we define the fluorescence level as a concentration. Nevertheless, we found that the fold-change in expression between medium and high expressers was very similar between both methods (Figure 3.5B), validating the accuracy of the estimation of mean expression levels in both.

Next, we aimed to further confirm this lack of dependency in native promoters. We decided to choose promoters displaying low noise over a wide range of expression

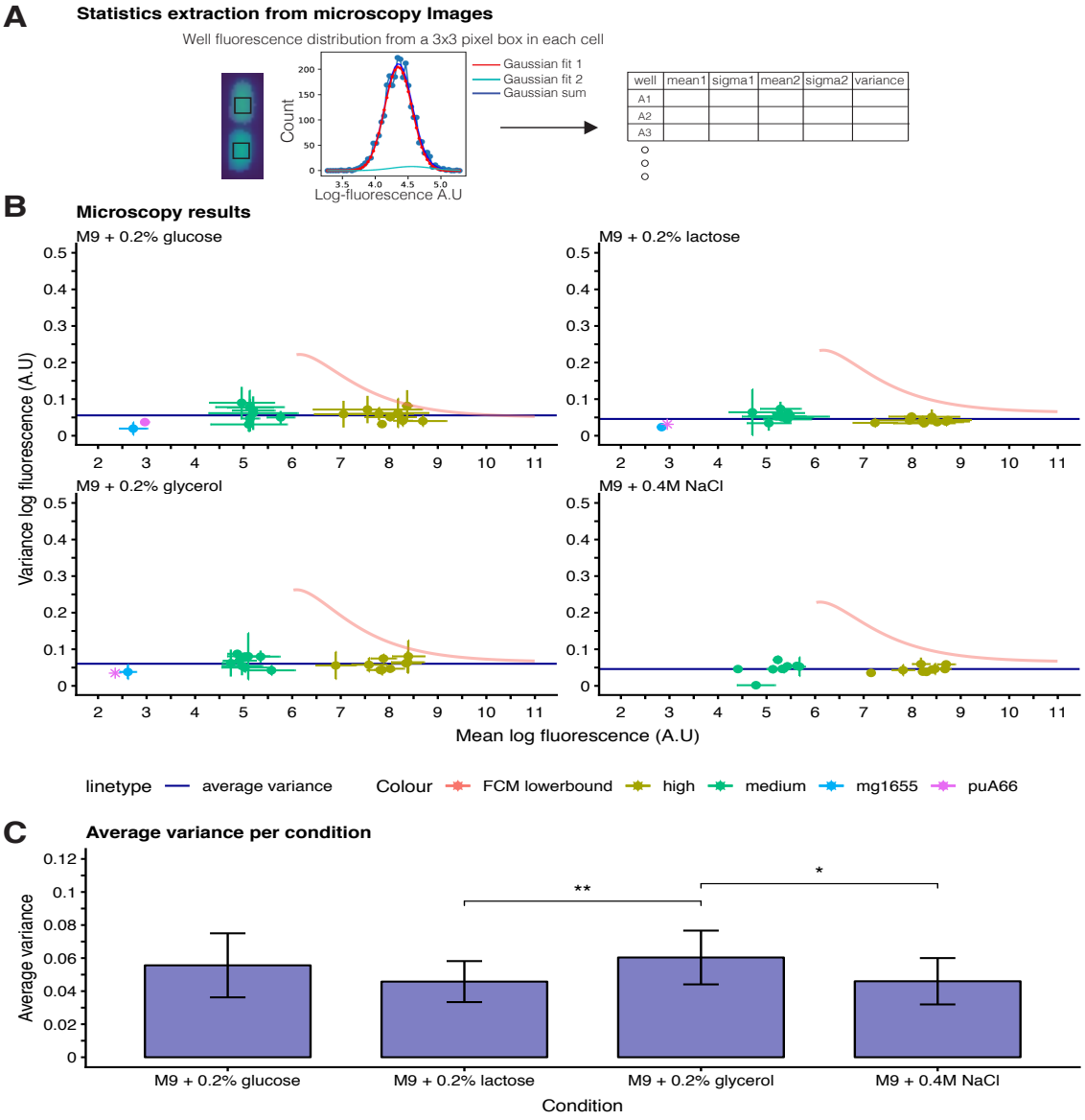


Figure 3.4: The strict lower bound on variance as a function of mean is not present in microscopy measurements. A: In order to remove the contribution of segmentation errors in the final statistics of the images acquired using our microscopy method (Supplementary Figure S3.4 and Supplementary Figure S3.5), we defined the fluorescence level of a cell as the average fluorescence (in pixels) over a 3x3 pixel box in the middle of the cell. We fitted a mixture of two Gaussian's to the distribution of log-fluorescence levels of all cells containing a promoter of interest and extracted the mean and variance from the Gaussian with the highest probability. **B:** Variance (y-axis) as a function of mean (x-axis) of the log-fluorescence of a small subset of synthetic promoters (9 medium and 9 high, together with a strain with an empty plasmid [blue] and the native strain [violet]) characterized with microscopy. The error bars indicate the standard deviation (promoters without error bars are illustrated as stars). The horizontal blue line illustrates the average variance over all promoters measured with microscopy. The pink line shows the fitted lower bound on variance as a function of mean of the promoters measured with flow cytometry. Each panel represents one condition. **C:** Average variances (y-axis, +/- the standard deviation) calculated over all measured promoters in each condition (x-axis). Only the pairs M9 + 0.4M NaCl-M9 + 0.2% glycerol and M9 + 0.2% lactose-M9 + 0.2% glycerol showed significant differences (Wilcoxon rank sum test, p-value = 0.02 and 0.005, respectively).

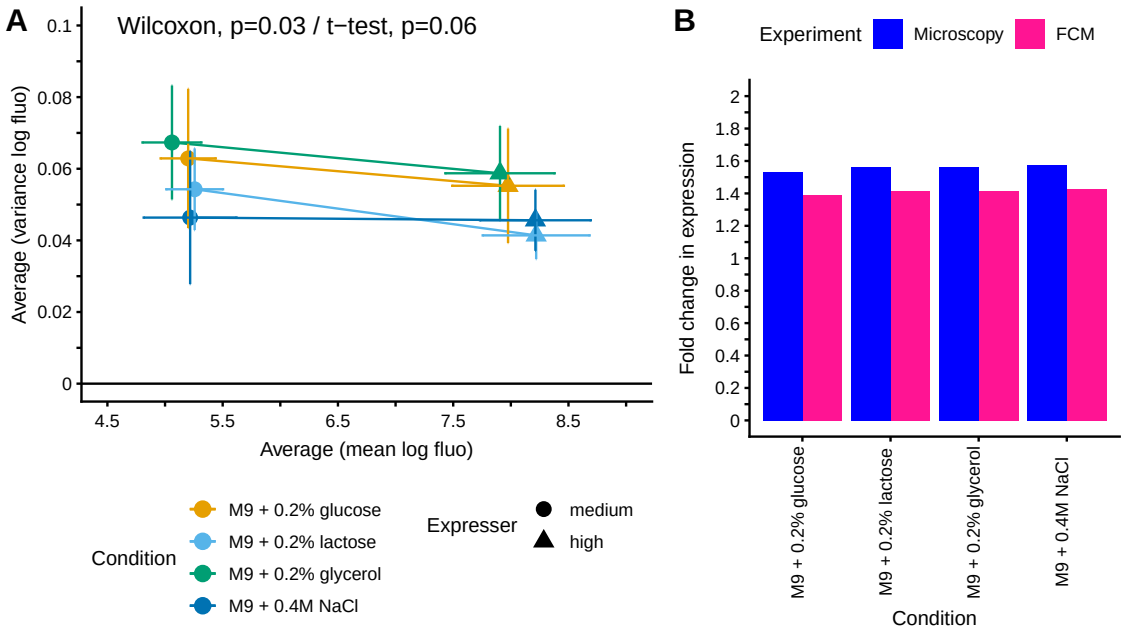


Figure 3.5: Comparing medium and high expressers. **A:** The average mean (+/- sd, x-axis) and average variance (+/- sd, y-axis) over each type of expresser (9 medium and 9 high) in each tested conditions (different colours). The lines link both expressers in each condition. The differences between them are minimal (p-value indicated at the top of the plot estimated after gathering all conditions together). **B:** Fold-change in mean expression between high and medium expressers ($\frac{mean_{high}}{mean_{medium}}$) when measured with microscopy (dark-blue) and flow cytometry (pink).

levels, for which we selected promoters from the library of native transcriptional fusions [128] following the lower bound on variance as a function of mean in different conditions (Figure 3.6 and Supplementary Figure S3.14A).

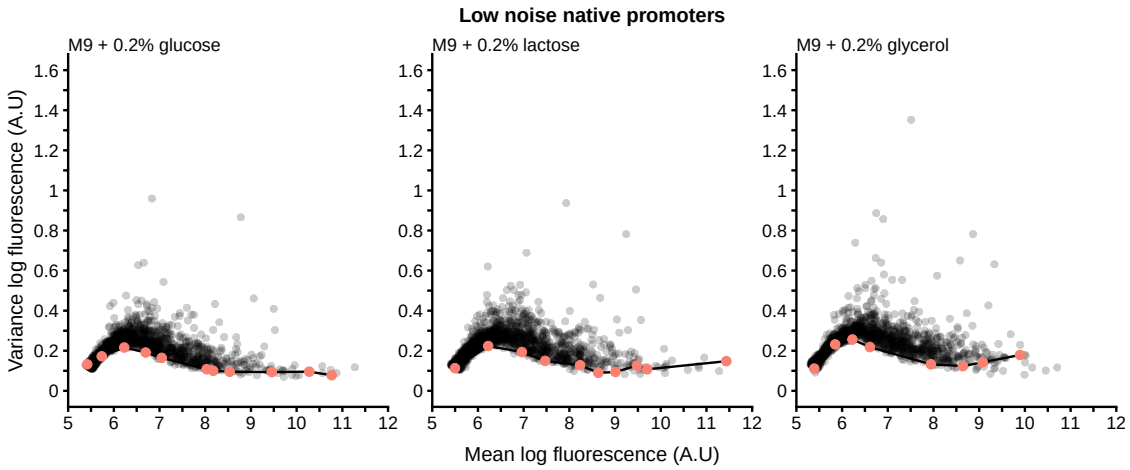


Figure 3.6: Selecting native promoter to study their mean and variance dependency. We selected native promoters (pink points) that followed the minimal variance as a function of mean lower bound across different expression levels in different conditions when measured with FCM. Not the same promoters were selected in each conditions (Supplementary Figure S3.13, Supplementary Figure S3.14B and Supplementary Table S3.3).

We included one extra condition apart from the four in which we grew the set of synthetic promoters: stationary phase after 16 hours of growth in minimal media + 0.2% glucose. Unfortunately, in two conditions we did not obtain replicates (stationary phase 16h and minimal media 0.4M NaCl). Therefore, the results of this two conditions are only shown as Supplementary information (Supplementary Figure S3.14). The name of the genes regulated by the selected promoters is listed in Supplementary Figure S3.13, Supplementary Figure S3.14B and Supplementary Table S3.3. It should be noted that not the same promoters were selected in the different conditions. An important feature of these promoters is that the vast majority are not heavily regulated, according to RegulonDB [20] (Supplementary Table S3.3 and Supplementary Figure S3.15). Overall, the order of expression (from lowest to highest expresser) was almost identical between the measurements done using flow cytometry and microscopy (Supplementary Figure S3.16, with the exception of two promoters, for which we do not have a clear explanation of why they differ). This further validates the suitability of both methods to measure mean expression levels.

However, we found that when characterized using microscopy, the selected native promoters also did not show a mean-variance dependency in any of the measured conditions (Figure 3.7A and Supplementary Figure S3.14C). The average variance was not different between conditions, with the exception of minimal media + glucose 0.2% (although in that particular condition some promoters miss replication, Figure 3.7B, Supplementary Figure S3.14D and Supplementary Table S3.4). The illumination settings changed across replicates and, as a result, no error bar is displayed on their mean estimates. These results further confirm that the variance estimates obtained when using flow cytometry are highly dominated by shot-noise for low expressed promoters. In Figure 3.7B and Supplementary Figure S3.14D we show that, in native promoters, the average variance estimated over all promoters did also not dramatically vary between conditions. Only minimal media supplemented with 0.4M NaCl showed a lower average variance (Supplementary Figure S3.14D), in contrast to the set of synthetic promoters. Although more measurements are necessary in order to draw firm conclusions. To confirm that we were able to identify high noise promoters using our microscopy method, we picked one of the noisiest promoters among all native promoters (gadAX) and measured it in three conditions (Supplementary Figure S3.17). We were able to identify this particular promoter as high noise using our method, as well.

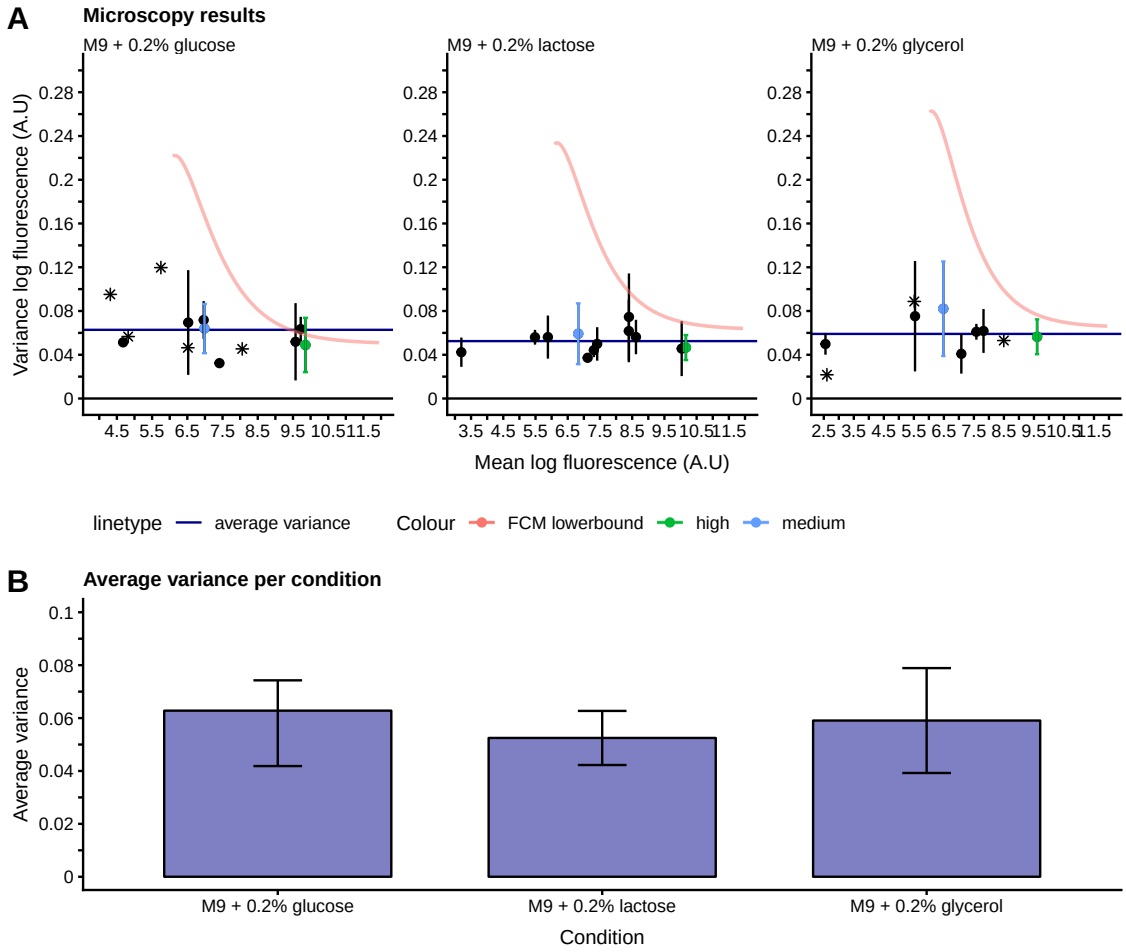


Figure 3.7: Lower bound on variance as a function of mean in the selected native promoters under the microscope. A: Variance (y-axis) as a function of the mean (x-axis) of the selected native promoters characterized with microscopy (black dots, medium expressers: blue dots, high expressers: green dots). The error bars indicate the standard deviation in variance (promoters without error bars are illustrated as stars). The illumination settings changed between days of acquisition and therefore, no error bar on the mean is displayed (see Material and Methods). The horizontal blue line illustrates the average variance over all promoters measured with microscopy. The pink line shows the fitted lower bound on variance as a function of mean of the promoters measured with flow cytometry. Each panel represents a condition. In Supplementary Figure S3.14C two extra conditions (without error bars) are shown. **B:** Average variance (y-axis, +/- the standard deviation) calculated over all measured promoters in each condition (x-axis). None showed large differences.

In Figure 3.8 we compare the average variances obtained with both libraries (native and synthetic promoters) in each condition. The average variances did not show changes across libraries (except in minimal media 0.4M NaCl, which may be due to the lack of replicates in the library of native promoters), confirming that the selected

native promoters in each condition are among the lowest noise promoters in *E.coli* across a wide range of expression levels.

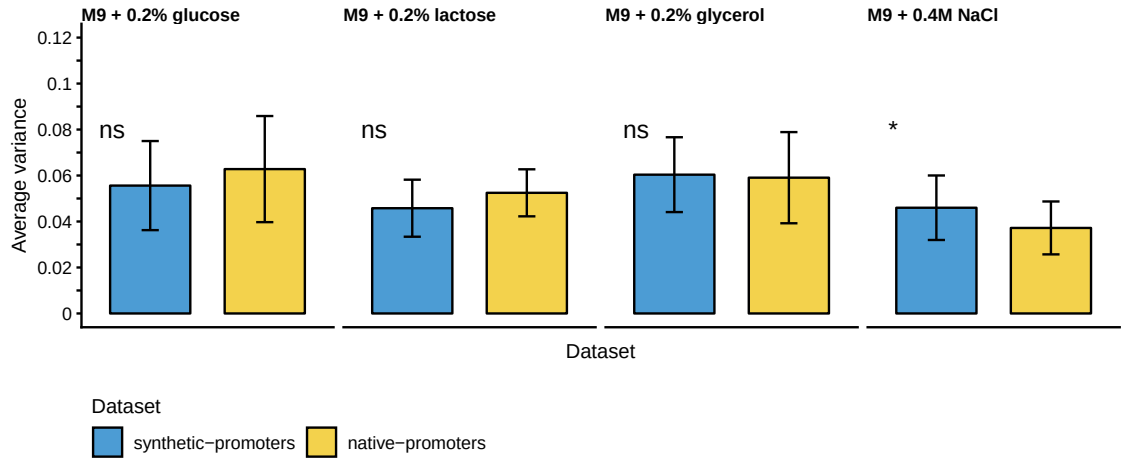


Figure 3.8: The average variance does not change between datasets. Average variance (y-axis, +/- sd) estimated in each condition for each of the measured datasets (x-axis, blue: synthetic promoters, yellow: native promoters). Except in minimal media + 0.4M NaCl (p-value = 0.032, Wilcoxon rank sum test) none showed large changes between them (ns: p-value > 0.05). In the case of M9 + 0.4M NaCl the acquisition of more replicates in the set of native promoters would be desirable in order to draw firm conclusions, as the measurements were only done once.

3.3 Discussion

We determined protein levels from single-cell fluorescence distributions in a set of synthetic constitutive promoters, expressing at two predefined levels (medium and high), and a set of native promoters; low regulated and spanning a wide range of expression levels. Each promoter was measured across a set of different conditions and their mean and variance in log-fluorescence characterized using two independent methods: flow cytometry and microscopy. We confirmed previous reports [122] of flow cytometers introducing a non-negligible shot-noise component, which affects variances estimations. The observed lower bound on variance as a function of mean in the measurements done with flow cytometry (Figure 3.1B) was not reproducible in any of the measurements done with microscopy (Figure 3.4B, Figure 3.7A and Supplementary Figure S3.14C), highlighting the importance of quantifying technical errors. We found that high expressed promoters measured using flow cytometry showed similar variances as the minimal floor in variance observed un-

der the microscope. However, low and medium expressed promoters did not show good agreement (Supplementary Figure S3.1). A careful examination of all possible technical sources of noise that could be introduced by the microscope or the image analysis routine should be performed in order to conclude that the variances measured with microscopy accurately reflect a promoter's real variance. Contrary to the variances estimations, we found good agreement on mean expression levels between both methods. Even though a direct comparison of the numbers is not possible, as discussed in the main text, the fold change in expression between high and medium expressers was very similar in both experimental setups (Figure 3.5B). Native promoters also showed good agreement in mean expression level order when sorted from lowest to highest expressed, although in this dataset a small number of promoters differed between methods, for which we do not have a clear explanation at the moment (Supplementary Figure S3.16).

One of the main observations in this work is that all promoters (native and synthetic), across all conditions and expression levels, showed a relatively constant floor in variance when measured with microscopy (defined as the average variance in log-fluorescence of all promoters per condition, Figure 3.8). Synthetic and native promoters did not show large deviations in the floor between them, except in minimal media supplemented with 0.4M NaCl, although these differences may be due to the lack of replicates in the library of native promoters. The minimal floor in variance is also displayed by high expressed promoters in the flow cytometer and, after correcting for the effects of the shot-noise on the variances estimates, by all promoters at all expression levels (Supplementary Figure S3.7). As introduced at the beginning of the chapter, it has been theoretically and experimentally proved, that the noise level of a promoter can be partly explained by two independent noise sources, each dominating at different expression levels. Intrinsic noise, which is the result of the inherent stochasticity of the molecules involved in gene expression, scales proportionally with the mean level of a promoter and dominates at low expression levels. In [41], it was established that the noise in protein levels is dominated by intrinsic fluctuations only when the number of proteins per cell is lower than 10. Extrinsic noise sources, which scale with the square of the mean, dominate when the number of proteins per cell is bigger than 10. The measurements presented here are only dominated by extrinsic noise fluctuations, as all our promoters are in a regime in which they express more than 10 proteins per cell (Supplementary table S3.5), even those that we consider 'low' expressed. Therefore, we have not been

able to study how intrinsic fluctuations affect the mean and variance dependency. To do so, it would have been necessary to measure promoters that express at a regime of less than 10 proteins per cell. A caveat in doing so, is that the library of promoters that we used is contained in a plasmid and fluctuations in plasmid copy number will contribute additional extrinsic noise. The plasmid used, which contains a pSC101 origin of replication, has been documented to have between 1 to 6 copies per cell [171], which will also increase the number of proteins per cell for genes that are expressed at lower levels. In order to remove the possible effects of plasmid copy number, a solution could be to create a library of chromosomal fusions (*promoter-gfp*). However, even in this scenario the position on the chromosome or the cell-to-cell variability in the number of chromosomes (do to several replication rounds happening during one cell cycle) will introduce additional sources of variability, affecting the estimates, as well. All in all, the presented measurements clearly indicate that constitutive (i.e. unregulated) promoters display low noise by default independently of their expression level, which goes in line with previous reports that the noise of a promoter is determined by the specific features of its sequence and not by its mean expression [90, 95–97]. Moreover, there is a minimal variance level that all promoters will display. Disentangling which noise sources lead to this floor in variance remains an open question, although the data clearly indicates that these sources affect all promoters by the same amount in the different conditions. Studying how the minimal floor in variance varies across conditions gave us some hints of how general condition-dependent physiological changes contribute to it. As presented in Chapter 2, these sources seem to be highly condition-dependent, with slow growing cells showing a higher floor. Here we measured the constitutive promoters in a total of five different conditions, although two (stationary phase and salt) without replicates, so results from these two conditions are not conclusive. We did not observe dramatic differences in the variance floor between conditions for which we had replicates (glucose, lactose and glycerol), although in the flow cytometer these conditions also do not show dramatic differences. The condition with the highest differences in noise floor are stationary phase and rich media (Chapter 2), so it is fundamental to gather more data in these two conditions under the microscope to study whether the relationship between slow growth and high noise still holds. Finally, we also confirmed that deviations from the floor can be achieved via regulation. In Supplementary Figure S3.17, we show microscopy measurements of

one of the most regulated promoters in *E.coli*, gadAX, which displays significant deviations from the minimal floor in variance.

In this study we have shown further evidence of unregulated promoters showing noise levels close to the minimal one *E.coli* promoters display by default and that deviations from this low noise 'state' are a result of regulation (Supplementary Figure S3.17), as extensively presented in Chapter 2. The minimal variance floor constitutive promoters follow, independently of their expression level, arises from general extrinsic noise sources. We hypothesize that differences across conditions in the floor are mostly a result of differences in cell sizes variability. We were able to perform this study thanks to the development of an innovative microscopy method. The method consists of a unique procedure to conduct high-throughput single-cell fluorescence measurements and image analysis of cells attached to the bottom of 96-well plates using microscopy, in a relatively fast and accurate manner. Its greatest advantage is its usability and applicability to different types of biological questions, specially those in which a high number of measurements and single-cell statistics are essential. We believe that other studies will be able to greatly benefit from it.

3.4 Material and Methods

Strains

Library of synthetic promoters

All promoters used were created in [96]: each sequence is the result of the assembly of random nucleotides of around 100-150bp. Sequences were ligated into a version of the low-copy plasmid, puA66, which contains a fast-folding type of GFP (*gfpmut2*) downstream of a strong ribosomal binding site [128]. The plasmid was slightly modified to remove a putative weak σ70 binding site (two point mutations 24bp upstream of the GFP open reading frame, A $\rightarrow$G and T $\rightarrow$G). The sequences were evolved to two predefined expression levels, after 5 rounds of FACS sorting and PCR mutagenesis, and are typically referred to as medium expressers and high expressers. Isolated sequences were transformed into *E.coli* K12 MG1655 and stored at -80°C in LB media supplemented with 7.5% glycerol (and 50 μg/ml Kanamycin). The measured sequences (72 medium and 72 high) were randomly selected from the evolved set.

Library of native promoters

We used the same set of promoters described in the methods section of Chapter 2. The library was created in [128] and comprises around 75% of all intergenic regions (+/- 150 bp of the adjacent regions) in *E.coli* longer than 40bp. The promoter sequences were ligated into the same version of the plasmid as the set of synthetic sequences (except for both point mutations). The strain in which the library was transformed is *E.coli* K12 MG1655 (CGSC#8003).

3.4.1 Growth conditions

We grew both libraries in four different conditions: minimal media (M9) supplemented with 0.2% glucose, 0.2% lactose, 0.2% glycerol and 0.4M NaCl (+ 0.2% glucose). Moreover, we grew the native library of promoters until stationary phase (around 16 hours of growth) in minimal media + 0.2% glucose. As described in the methods section of Chapter 2, we grew all strains in 200 μl in the condition of interest in microtiter plates for two overnights before fluorescence measurements (around 1/2000 dilution between overnights). On the quantification day, cells were diluted 200-fold (500-fold for minimal media with 0.4M NaCl) and grown until mid-exponential phase at 37°C, shaken at 600 rpm. All media were supplemented with 50 μg/ml Kanamycin.

3.4.2 Flow cytometry quantification of fluorescence

In both libraries, we measured the distribution of single-cell log-fluorescence levels with a FACSCanto II (BD Biosciences), fluorescence excitation at 488 nm and 530/30 nm for emission, as described in the methods section of Chapter 2. To extract the mean and variance of each promoter we used a custom R package available in GitHub (`https://github.com/vanNimwegenLab/vngFCM`). A detailed description can be found in Galbusera et al. 2019 [122].

3.4.3 Shot-noise removal

To have a quantitative description of the variance and mean fluorescence levels of the promoters, we had to subtract any non-biological contribution from the measurements, which can arise from autofluorescence or artifacts introduced by the cy-

tometer. In order to subtract the autofluorescence from the statistics, we measured cells carrying a promoter-less plasmid at the same time and in the same conditions as the set of constitutive promoters. Due to the low expression levels in bacteria, a non-negligible shot-noise contribution has to be taken into account. Using a set of artificial beads, we were able to estimate the property of the shot-noise and remove its contribution from the measured fluorescence. A more detailed description is explained in the original study [122]. The CV^2 was computed on the autofluorescence and shot-noise corrected fluorescence intensities. The CV^2 is equivalent to the variance of the log-fluorescence whenever fluctuations are small relative to the mean.

3.4.4 Correcting the minimal variance as a function of mean in the FCM measurements

To correct for the observed dependency between variance and mean in the FCM measurements, we used the model derived in Wolf et al. 2015 [96] and introduced in the Methods section of Chapter 2:

$$var(\log[f_{p,c}]) = a_c(1 - \frac{f_{bg,c}}{\langle f_{p,c} \rangle})^2 + \frac{b_c}{\langle f_{p,c} \rangle}(1 - \frac{f_{bg,c}}{\langle f_{p,c} \rangle})$$

We used the following parameters for the fits:

Native promoters:

Condition	a_c	b_c	$f_{bg,c}$	**% of promoters above background**
M9 glucose	0.05	530	220	54.6
M9 lactose	0.063	550	220	57.2
M9 glycerol	0.065	580	205	59.0
M9 0.4M NaCl	0.065	500	205	51.2

Synthetic promoters:

Condition	a_c	b_c	$f_{bg,c}$	**% of promoters above background**
M9 glucose	0.1	600	230	100
M9 lactose	0.08	570	230	100
M9 glycerol	0.1	610	220	100
M9 0.4M NaCl	0.12	610	230	100

The background fluorescence was estimated from cells carrying a promoter-less plasmid.

As presented in Chapter 2, we defined the corrected variance (in Chapter 2 this term is referred to as noise) of a promoter in each condition as the difference between the measured variance and the fitted minimal variance above a background threshold (2x bg_c in linear scale):

$$corrected\_variance = \sigma^2_{pc} - \sigma^2_{min_c}(\mu)$$

3.4.5 Native promoters selection

To select native promoter along the lower bound on variance as a function of mean, we binned the data in different mean expression levels (in steps of 0.4 A.U. in log-fluorescence) and picked, inside each bin, the promoter displaying the lowest value in variance. Some promoters were removed from the analysis, due to problems in

the image acquisition and are, therefore, not shown in Figure 3.6 or Supplementary Figure S3.13.

3.4.6 Regulatory annotations

All regulatory annotations were done as described in the Methods section of Chapter 2 using RegulonDB v10.5.1 [20].

3.4.7 Microscopy quantification of fluorescence

A detailed description of all critical steps can be found in the Supplementary text section at the end of this chapter. In short, we used Poly-L-lysine 0.01% (Sigma P4832-50ml) to attach single-cells to the glass-bottom of 96-well plates (greiner bio-one, ref. 655090) after growing them as previously described (Chapter 2) in the different conditions. We performed all acquisitions in an inverted Nikon Ti-E microscope equipped with a motorized xy-stage (with a plate holder, Nikon TI-SH-W) and a Piezo stage (Nano-drive 85). We fixed the plates on the stage using custom metal clamps designed to fit to the plates we used. Images were recorded using a CFI Plan Apochromat Lambda DM $\times$100 objective (NA 1.45, WD 0.13 mm) and a CMOS camera (Hamamatsu Orca-Flash 4.0). We maintained the focus over the entire plate using custom runnables (See Supplementary text). The setup was controlled using Micro-manager [168] and its automatizing control possibilities [172]. In concrete, we used their HSC-site generator tool and Multi-Dimensional Acquisition engine to obtain full-plate images. Typically, we obtained 5 positions per well. In each position we acquired z-stacks (-2μM to 2μM from the plane in focus in 0.2μM steps) of the bright-field image and one GFP fluorescence image of the plane in focus. Bright-field images were acquired using 100ms exposure with the transmitted light source at full power (CoolLeD pe-100). For fluorescence acquisitions, the illumination settings varied between experiments. The tables below indicates the settings used in each experiment, together with the number of replicates.

Set of synthetic promoters:

In the set of synthetic promoters the Illumination settings did not change across replicates. Some of the promoter-less strains were measured at 50 ms exposure

time instead of 200ms, although as shown in Figure S3.12 this did not dramatically change the background fluorescence value.

Strain	Condition	N independent measurements	N sequences each measur.	Illumination	Exposure time
MG1655 with puA66	M9 glucose 0.2%	2	1,1	9%	200 ms
	M9 lactose 0.2%	1	1		
	M9 glycerol 0.2%	1	1		50 ms
MG1655	M9 glucose 0.2%	2	1,1	9%	200ms - 50ms
	M9 lactose 0.2%	2	1,1		
	M9 glycerol 0.2%	2	1,1		
medium expressers	M9 glucose 0.2%	3	9,8,8	9%	200ms
	M9 lactose 0.2%	3	9,9,6		
	M9 glycerol 0.2%	3	9,9,8		
	M9 0.4 M NaCl	2	9,9		
high expressers	M9 glucose 0.2%	3	9,9,8	9%	200ms
	M9 lactose 0.2%	3	9,9,8		
	M9 glycerol 0.2%	3	9,9,8		
	M9 0.4 M NaCl	2	9,9		

Set of native promoters:

In the set of native promoters some set of replicates were measured at 9% illumination and 50 ms exposure time, while others at 17% and 200ms exposure. All plots presented in the main text show the mean values of all replicates measured at 17% illumination. However, to estimate the mean and standard deviation of the variance in log-fluorescence, we considered the measurements acquired at 9%, too in order to be able to put error bars in some of the promoters.

Condition	N promoters	N replicates	Illumination	Exposure time
M9 + 0.2% glucose	11	5 promoters ->1 replicate	17%	200ms
		3 promoters ->2 replicates	17%-9%	200ms-50ms
		3 promoters ->3 replicates	17%-9%(1 exp)	200ms-50ms
M9 + 0.2% lactose	10	7 promoters ->2 replicates	17%-9%	200ms-50ms
		3 promoters ->3 replicates	17%-9%(1 exp)	200ms-50ms
M9 + 0.2% glycerol	8	3 promoters ->1 replicate	17%	200ms
		5 promoters ->2 replicates	17%-9%	200ms-50ms
M9 + 0.4M NaCl	12	1 replicate	17%	200ms
Stationary phase 16h	8	1 replicate	17%	200ms

Samples were illuminated using a Lumencor Spectrax (Cyan LED) with excitation (475/35 nm) and emission filters (525/50 nm) used with a dichroic beam-splitter at 495 nm.

3.4.8 Image analysis

The acquired images were analyzed using a stand-alone package in Python developed by Guillaume Witz (*Mathematisches Institue, University of Bern*). The tool is available in GitHub: `https://github.com/urchuegu/DeepPlateSegmenter`. We removed the background offset of the camera, which corresponds to 100 pixels in our setup, in all our data. All relevant statistics (in log-fluorescence) were computed using the package.

The average and standard deviation of the variance in log-fluorescence estimated per promoter was calculated over all collected replicates, independently of the illu-

mination settings used. For the average and standard deviation of the means, only replicates with the same illumination settings were considered.

3.4.9 Statistical analysis

Unless otherwise specified, all comparisons were performed using a non-parametric two-sided Wilcox rank sum test or ANOVA using R, version 3.5.0.

3.5 Acknowledgements

Thank our FACS Core Facility, specially Janine Bögli, for assistance with flow cytometry and Luca Galbusera for assistance with flow cytometry analysis. Thanks to the Werner Siemens Foundation for supporting part of Arantxa's work.

3.6 Author contributions

EvN, TJ and AU designed the study. AU and TJ designed the experiments. AU performed all experiments. GW wrote the image analysis tool and the microscope runnables. AU, EvN, TJ and GW analyzed the data. AU wrote this chapter.

3.7 Conflict of interest

The authors declare no conflict of interest.

3.8 Supplementary information

Table S3.1: Statistical comparison of the variance between background, medium and high expressers gathered across conditions.

Groups	p-value-T.test	p-value-Wilcoxon.rank.sum	significance
background fluorescence - medium expressers	1.10E-09	9.33E-07	****
background fluorescence - high expressers	3.71E-07	1.58E-05	****
medium expressers - high expressers	0.058	0.027	ns/*

Table S3.2: Statistical comparison of the variance between background, medium and high expressers in each condition.

Groups	p-value-Kruskal.Wallis	p-value-Anova	significance
M9 + 0.2% glucose-high	0.072	0.19	ns
M9 + 0.2% lactose-high	0.47	0.52	ns
M9 + 0.2% glycerol-high	0.256	0.49	ns
M9 + 0.4M NaCl-high	0.623	0.6	ns
M9 + 0.2% glucose-medium	0.056	0.09518	ns
M9 + 0.2% lactose-medium	0.594	0.84519	ns
M9 + 0.2% glycerol-medium	0.192	0.56036	ns
M9 + 0.4M NaCl-medium	0.097	0.0021	ns/**

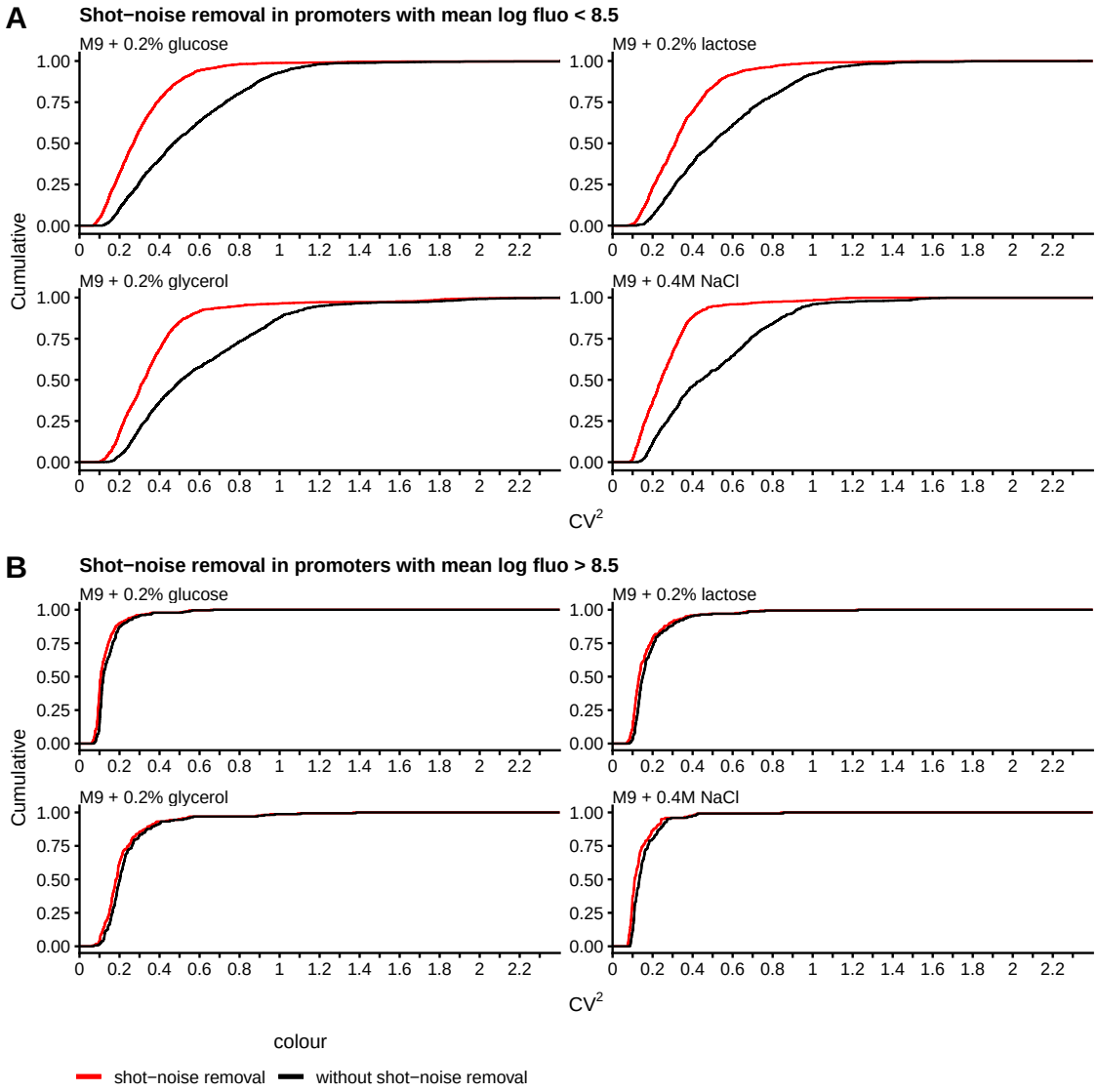


Figure S3.1: Shot-noise removal at different expression levels. A: Cumulative distribution of the CV^2 of promoters with mean expression levels < 8.5 (in log-space) with (red line) and without (black line) shot-noise removal. **B:** Cumulative distribution of the CV^2 of promoters with mean expression levels > 8.5 (in log-space) with (red line) and without (black line) shot-noise removal.

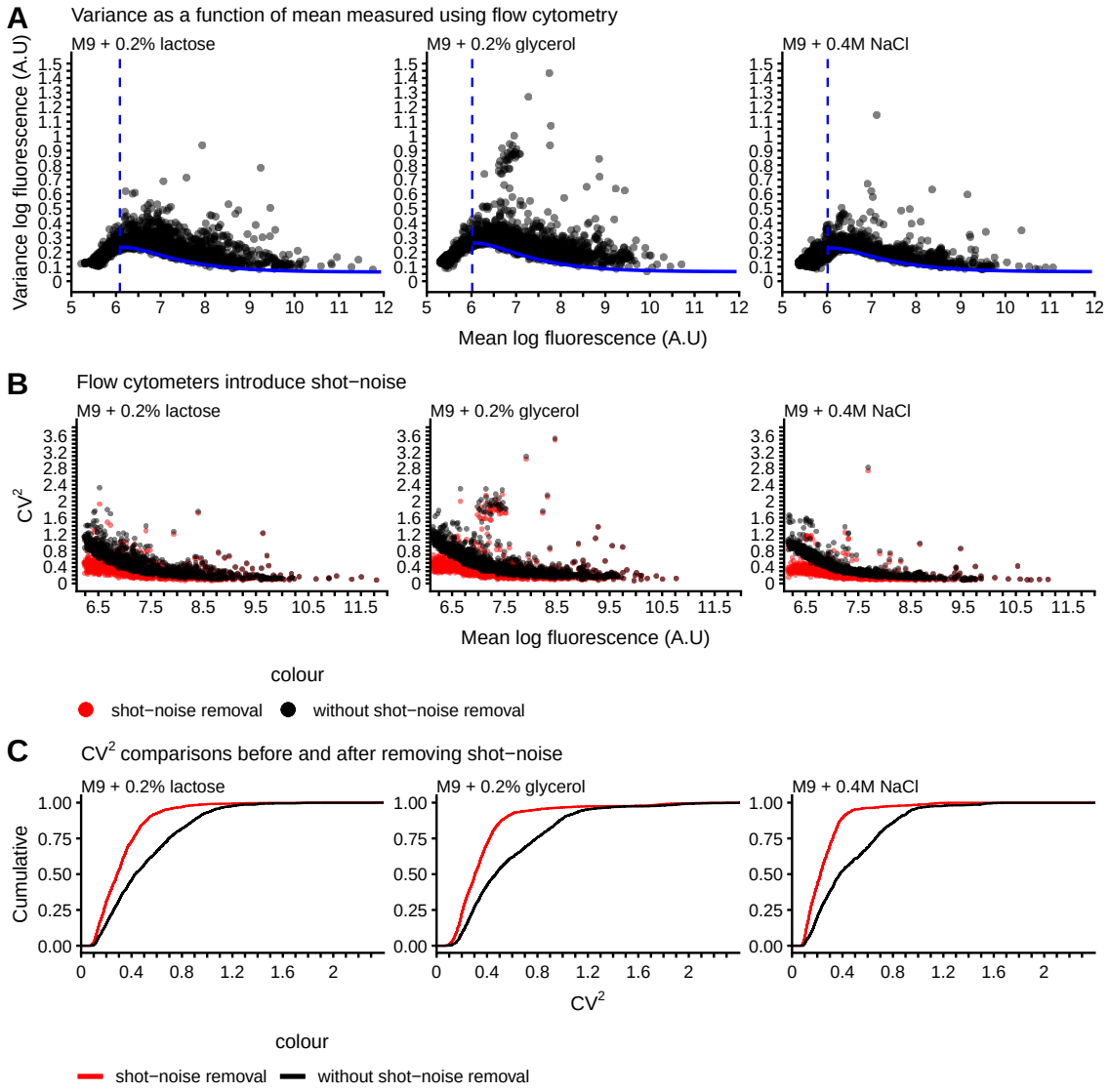


Figure S3.2: Flow cytometers introduce shot-noise. A: Shown are the mean (x-axis) and variance (y-axis) in log-fluorescence of a set of *E.coli* native promoters characterized with flow cytometry and grown in three conditions: minimal media supplemented with either 0.2% lactose, 0.2% glycerol or $0.4M$ NaCl + 0.2% glucose (each point represents one promoter and the condition is indicated at the top of each panel). Clearly, there is a strict lower bound on variance as a function of mean. The blue line indicates the theoretical predicted minimal variance, which breaks for promoters below a background threshold (dashed vertical blue line). **B:** Each point indicates the CV^2 (y-axis) (which is equivalent to variance of the log-fluorescence) as a function of mean (x-axis) of all promoters shown in panel A (above a background threshold, indicated by the vertical blue-dashed line) before (black dots) and after (red dots) shot-noise removal in each condition (indicated at the top each panel). **C:** Cumulative distribution of the CV^2 of the set of promoters represented in panel A before (black line) and after (red line) shot-noise removal in each condition (indicated at the top each panel).
Note: replicate measurements have not been gathered together.

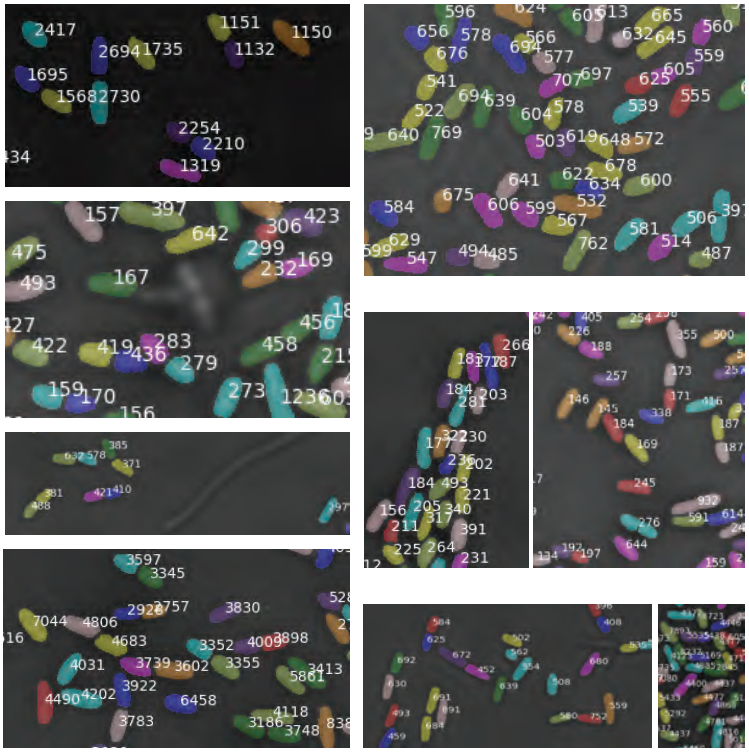


Figure S3.3: Segmentation examples. Shown are the masks of some segmented images interposed with their corresponding phase correlation images. Each cell is identified by a unique identifier.

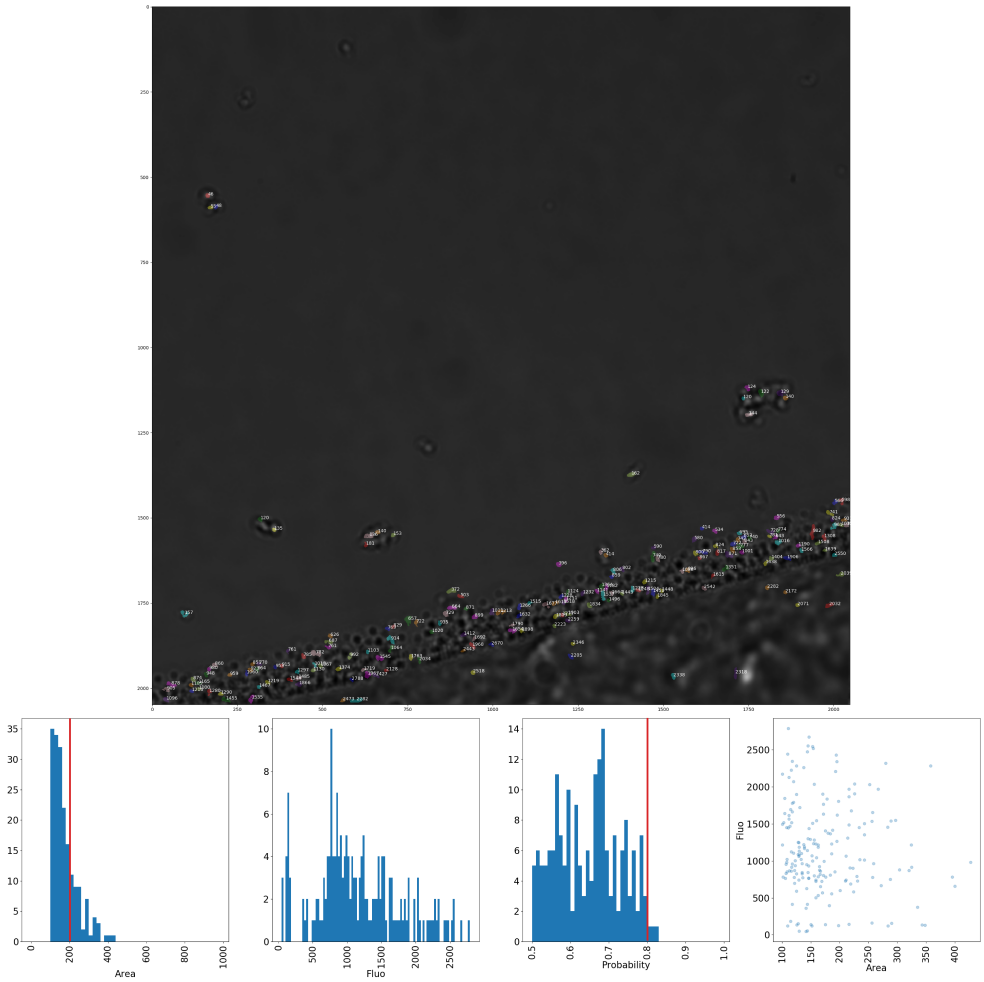


Figure S3.4: Example of a bad segmented image. The image aims to illustrate the output of an image acquired close to the edge of a well containing very few cells. The algorithm wrongly identifies the edges as cells. Nevertheless, as shown in the plots below, their probabilities are very small. We typically set a threshold of 0.8 in probability and 200 pixels in area (red vertical lines).

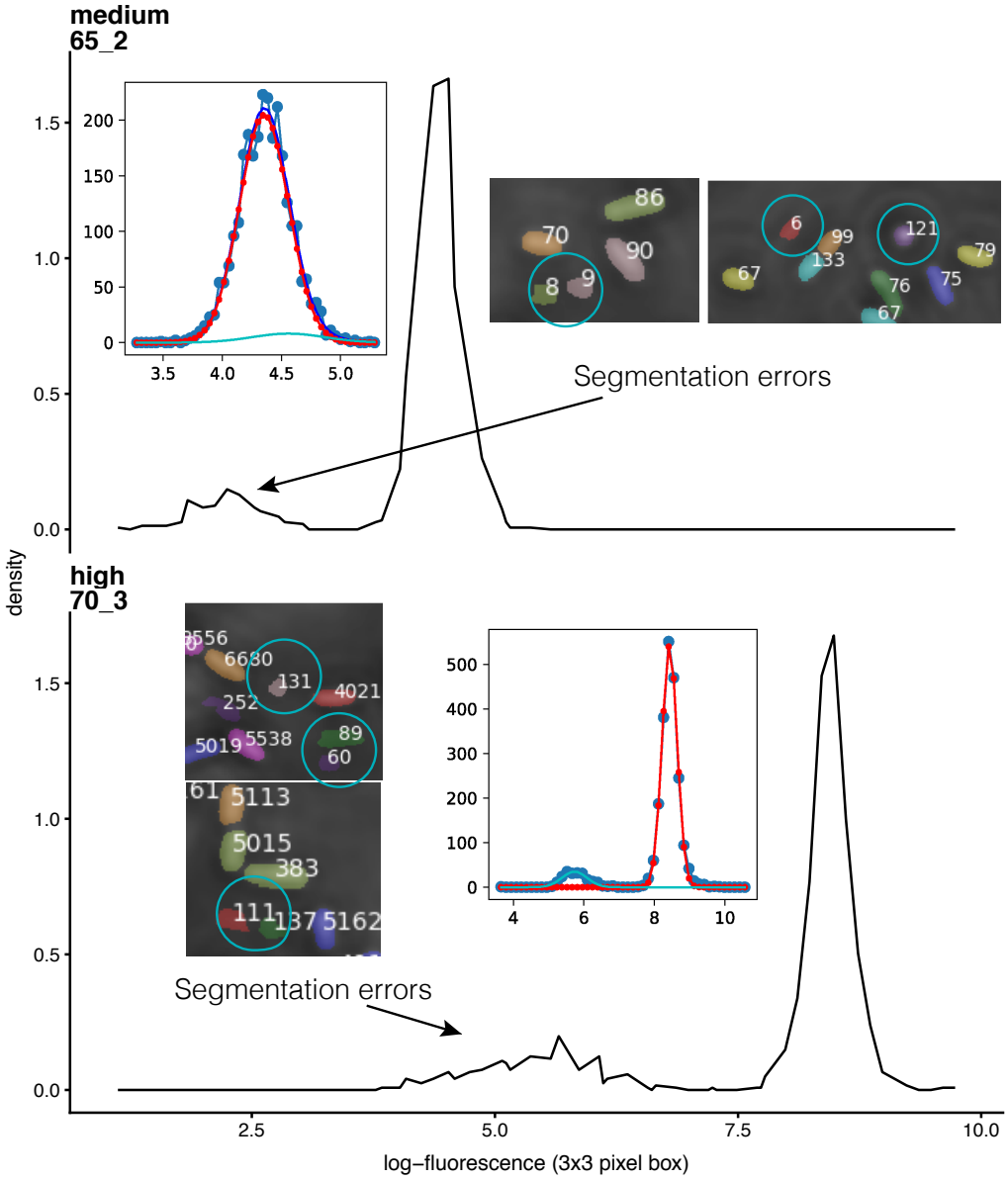


Figure S3.5: Fluorescence distributions of segmentation errors. Example of two promoters and their fluorescence distributions. Wrongly segmented objects (indicated with a circle in the image inlets) will show fluorescence levels close to the background fluorescence of the image. The background fluorescence of the images changes depending on the level of the expression of the fluorophore used. In order to remove possible biases that could arise from wrongly segmented objects, we fit a mixture of two Gaussians to the log-fluorescence data, as shown in the histograms, and extract the mean and variance from the Gaussian with the highest probability.

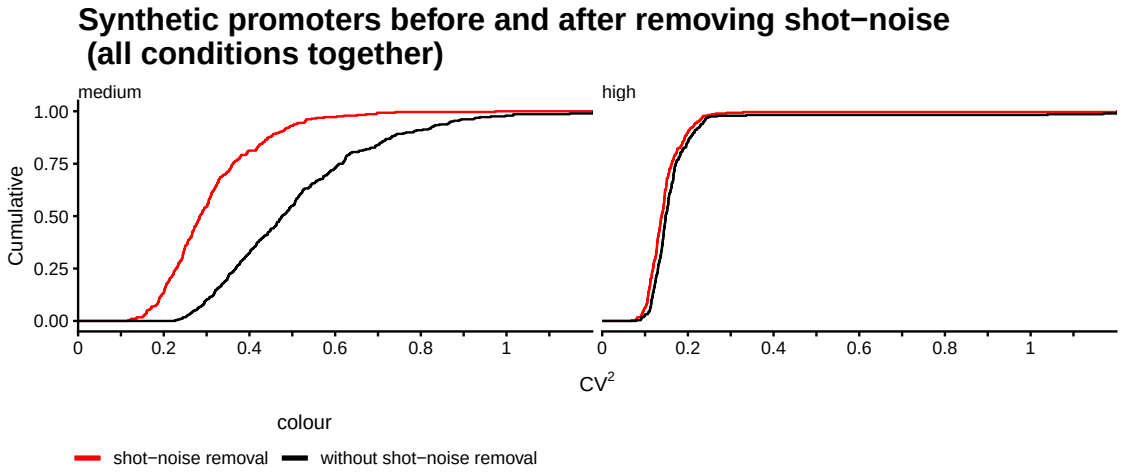


Figure S3.6: Shot-noise removal in the set of synthetic promoters. Contribution of shot-noise to the CV^2 estimations of the library of synthetic promoters. All measured conditions have been gathered together.

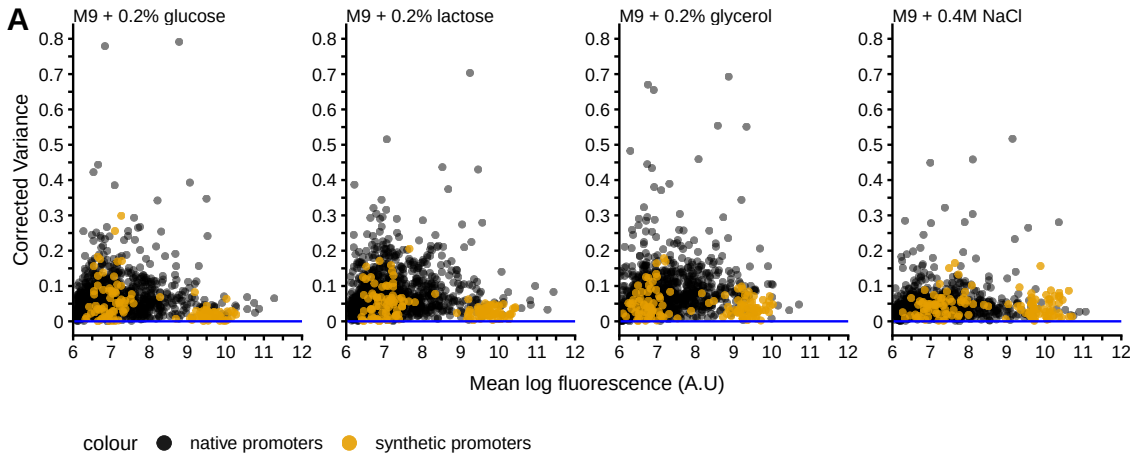


Figure S3.7: Corrected variance (*noise*) in native and synthetic promoters. **A:** Corrected variance (y-axis, also sometimes referred, simply, as noise), defined as the difference between the measured variance and the theoretical predicted minimal variance as a function of mean of all promoters from the library of native (black dots) and synthetic promoters (yellow dots) in each condition (indicated at the top of each panel).
Note: in this plot replicate measurements have been gathered.

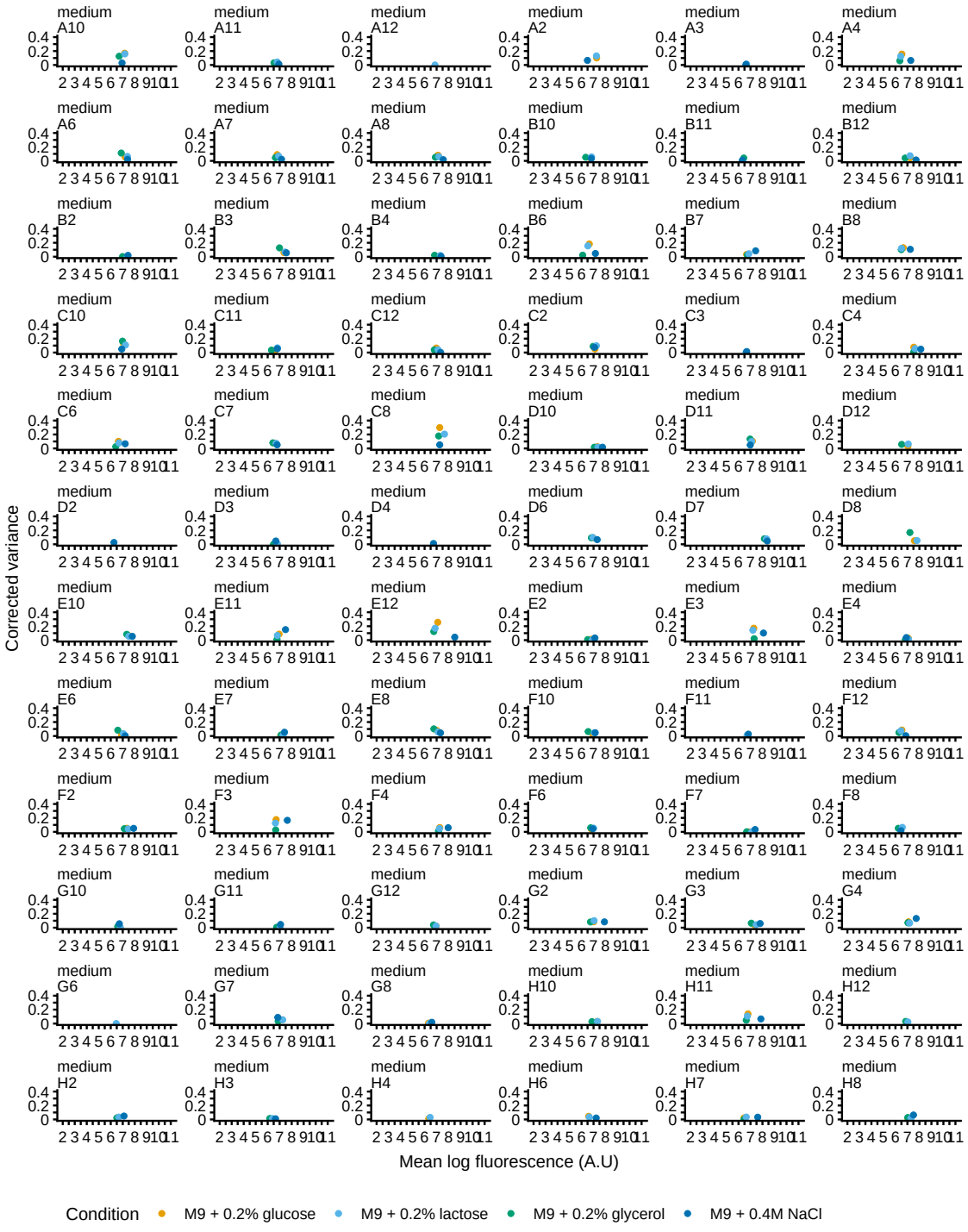


Figure S3.8: Individual medium expressers measured using FCM. Variance (y-axis) as a function of mean (x-axis) of the log-fluorescence of each individual medium expresser (individual panels) measured in four different conditions (indicated by the colour) using flow cytometry.

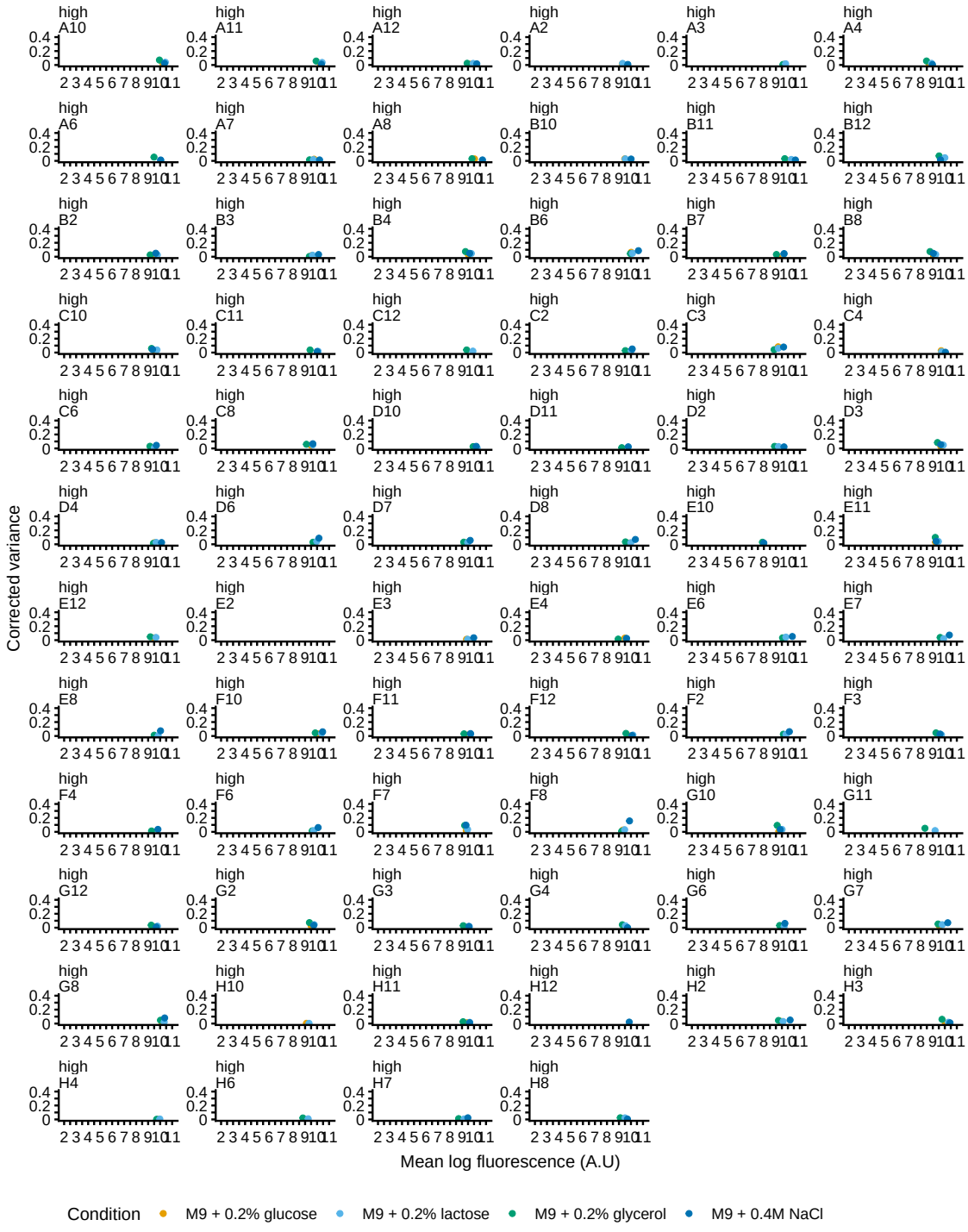


Figure S3.9: Individual high expressers measured using FCM. Variance (y-axis) as a function of mean (x-axis) of the log-fluorescence of each individual high expresser (individual panels) measured in four different conditions (indicated by the colour) using flow cytometry.

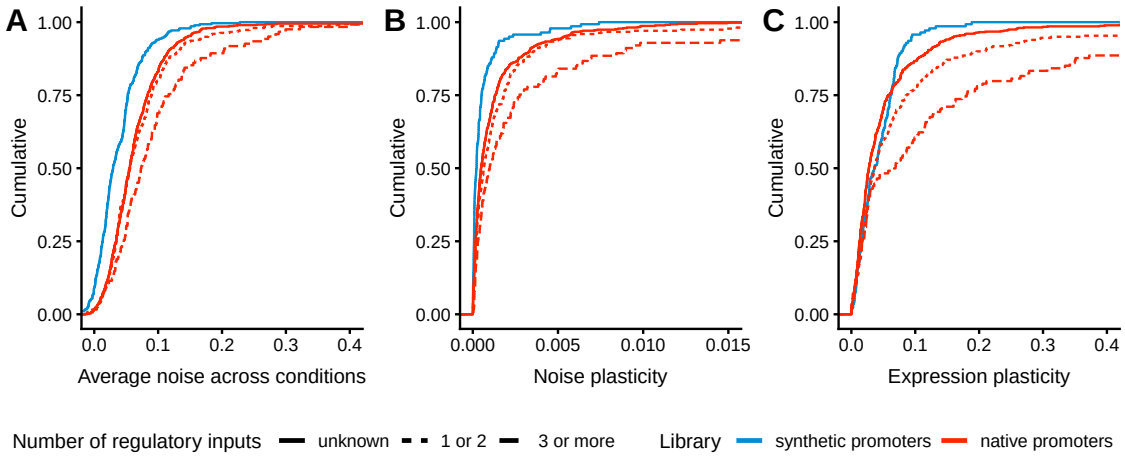


Figure S3.10: Low noise in the set of synthetic promoters. A: We calculated the average noise (after correcting for the mean-variance dependency, Figure S3.7) of the synthetic promoters (blue line) and native promoters (red line) over the four measured conditions. Shown is the cumulative distribution of the estimated average noise levels (also referred to as corrected variance in this study) stratified by annotated unique number of regulatory inputs (various line-types). **B:** Cumulative distribution of noise plasticity (estimated as the variance in noise over the four conditions of each promoter) for the set of synthetic (blue line) and native promoters (red line) stratified by annotated unique number of regulatory inputs (various line-types). **C:** Cumulative distribution of expression plasticity (estimated as the variance in mean expression over the four conditions of each promoter) for the set of synthetic (blue line) and native promoters (red line) stratified by annotated unique number of regulatory inputs (various line-types).
All regulatory interactions have been annotated according to RegulonDB [20].

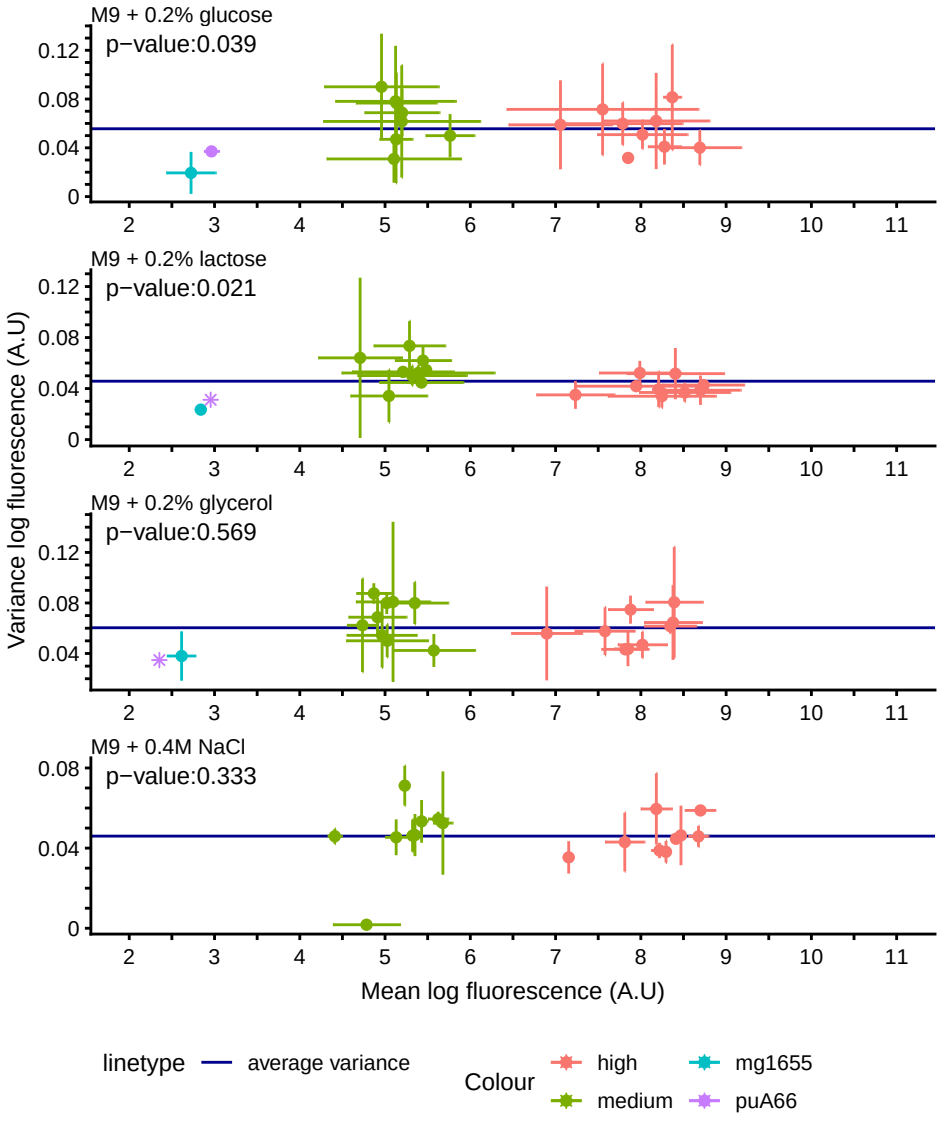


Figure S3.11: Variance as a function of mean as measured using microscopy. Variance (y-axis) as a function of mean (x-axis) of a small subset of synthetic promoters (MG1655: blue, MG1655 + puA66: pink, 9 medium: green and 9 high: red) characterized with microscopy. The error bars indicate the standard deviation among replicates (promoters without replicates are illustrated as stars). Each panel represents one condition. The horizontal blue line illustrates the average variance over all promoters measured in the represented condition.

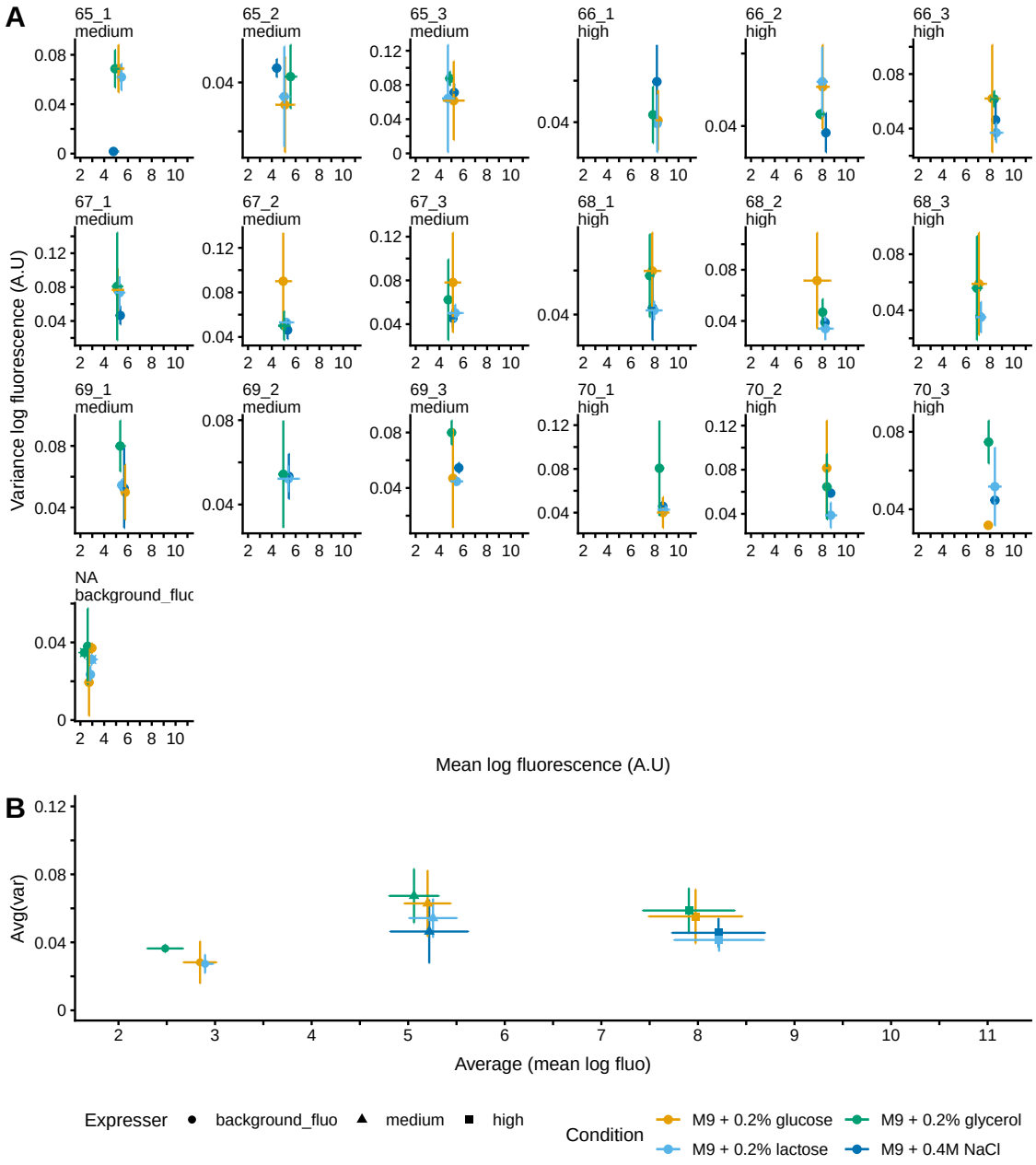


Figure S3.12: Individual synthetic promoters measured using microscopy. A: Variance (y-axis) as a function of mean (x-axis) in log-fluorescence (+/- sd) of each synthetic promoter (individual panels) measured with microscopy in the tested conditions (indicated by the colour). Promoters without replicates are illustrated as stars. **B:** The average variance (y-axis, +/- sd) and average mean (x-axis, +/- sd) over all measured promoters of each expresser in each condition.

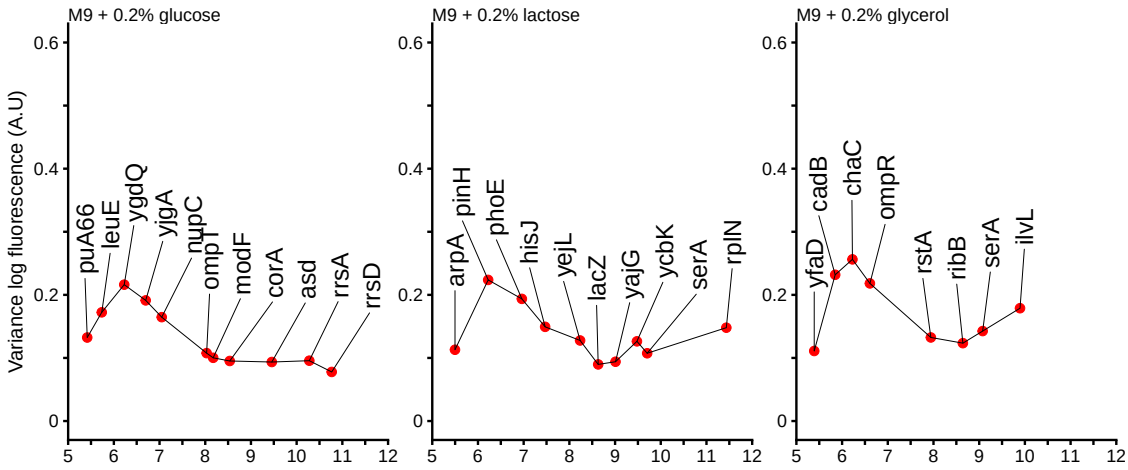


Figure S3.13: Names of the genes immediately downstream (resp. upstream) of the selected native promoters for microscopy characterization.

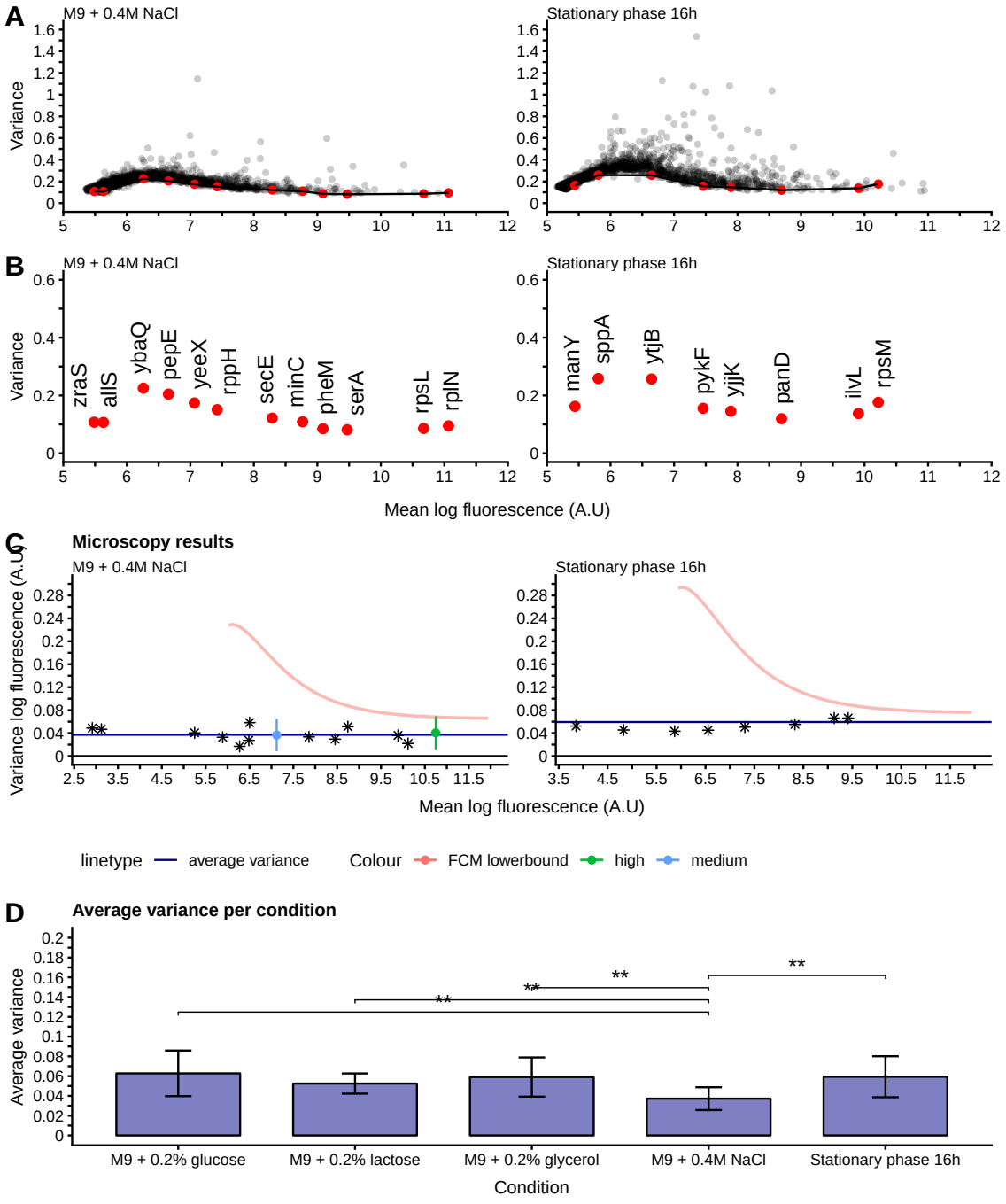


Figure S3.14: Minimal variance as a function of mean in M9 supplemented with 0.4M NaCl, and growth after 16 hours in M9 with 0.2% glucose. A: Selected promoters in M9 + 0.4M NaCl (+ 0.2% glucose) and in 16 hours of growth in M9 0.2% glucose. **B:** Their immediate downstream (resp. upstream) genes. **C:** Variance (y-axis) as a function of mean (x-axis) of the selected native promoters characterized with microscopy. Measurements were acquired only once for this set of conditions, therefore no error bar is displayed. The horizontal blue line illustrates the average variance over all promoters. The pink line shows the fitted lower bound on variance as a function of mean in the flow cytometer. Each facet represents a condition. **D:** Average variance (y-axis, +/- the standard deviation) calculated over all promoters in all conditions (x-axis). M9 supplemented with 0.4M NaCl shows a significantly lower variance compared to each of the other conditions (**: p-value <= 0.01, Wilcoxon rank sum test).

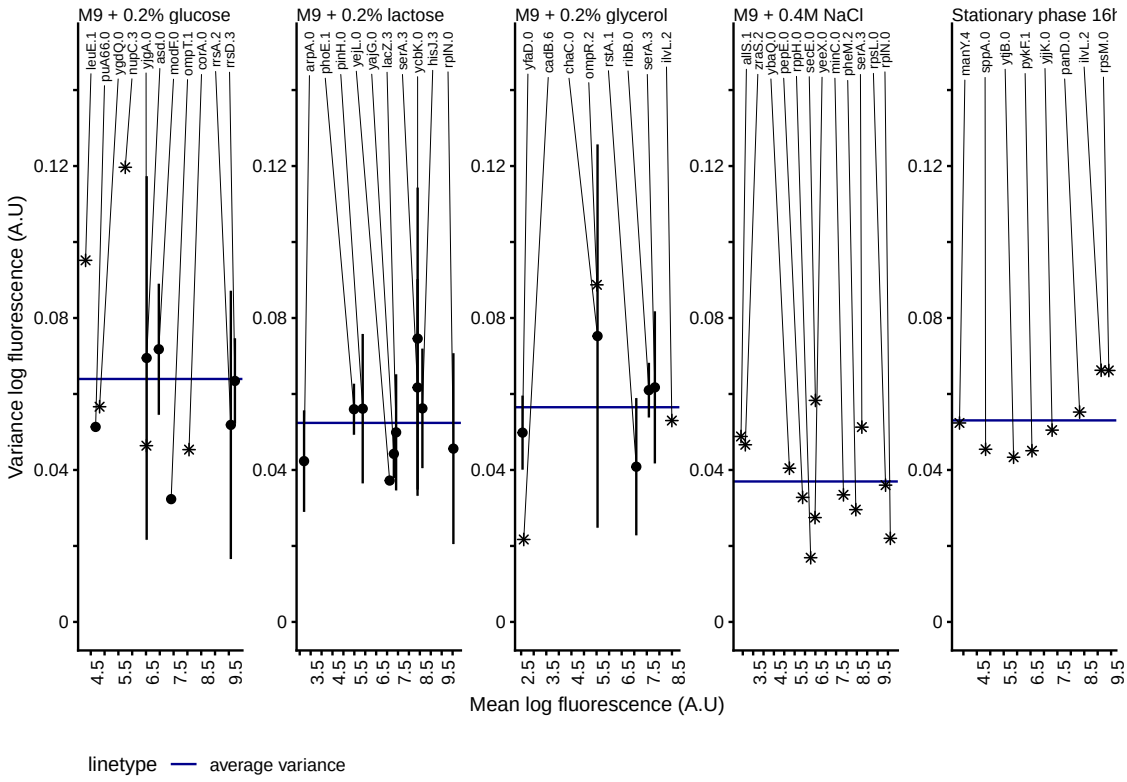


Figure S3.15: Selected native promoters measured with microscopy. Variance (y-axis) as a function of mean (x-axis) in log fluorescence of the selected native promoters, characterized with microscopy, in order to study how their variance scales with mean. Each panel represents one condition, and each promoter has been annotated with the name of the immediate gene downstream (resp. upstream) together with the number of unique regulatory inputs documented in RegulonDB [20]. Interestingly, the majority are not regulated and those that contain the largest number of regulatory inputs (*cadB* and *manY*) are not expressed, suggesting that their regulators may not be active in the measured conditions.

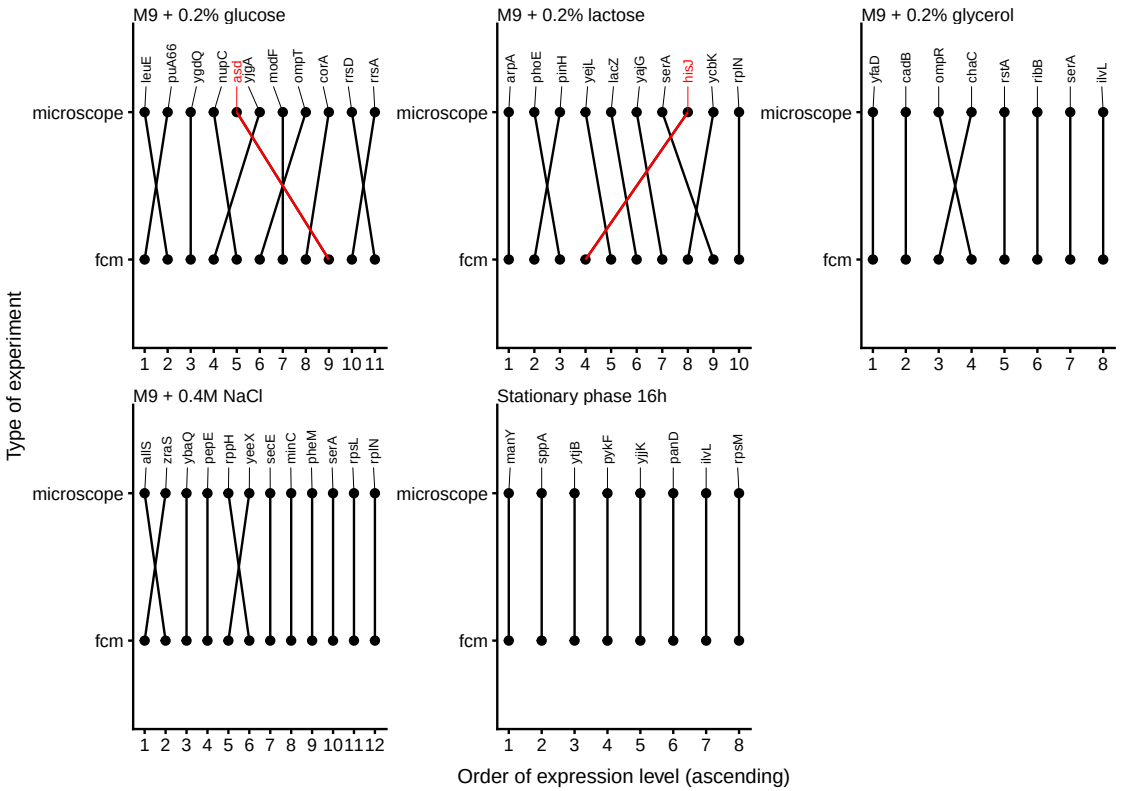


Figure S3.16: Order of appearance (sorted by ascending expression level) in the microscope and flow cytometer (fcm). Each gene was assigned an ID based on expression level (1-n, from lowest to highest expressed, x-axis) in each experimental setup (y-axis). The lines connect the same gene measured in the different setups and the name is indicated at the top. In red are indicated genes that show a significant change in the expression level between both setups.

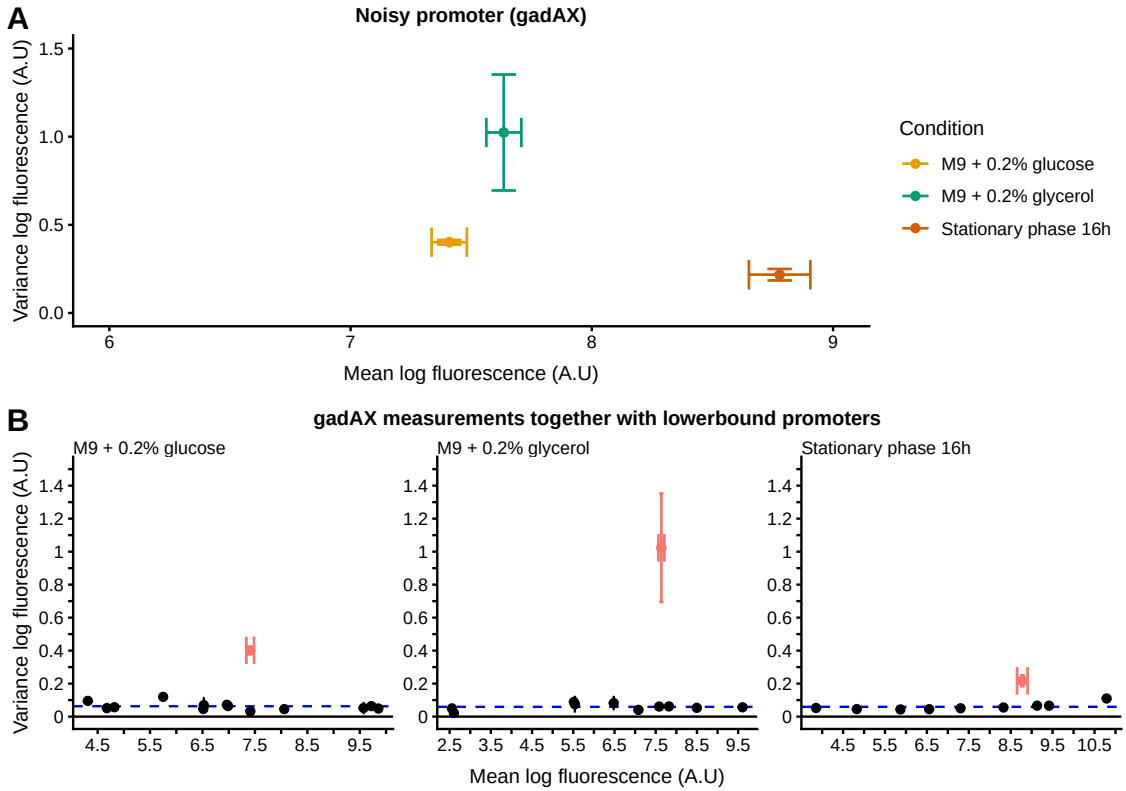


Figure S3.17: Measuring a high noise promoter with the microscopy procedure. A: We selected one of the noisiest promoter in the library of native promoters (gadAX) and measured it under the microscope in three different conditions: minimal media + 0.2% glucose, 0.2% glycerol and 16 hours of growth in 0.2% glucose. **B:** Shown are, in each condition, the variance as a function of mean in log-space of the high noise promoter (red dot) and the native promoters selected to study the minimal variance as a function of mean (black dots).

Table S3.3: Regulatory information of the selected native promoters. Highlighted in black are those promoters with 3 or more annotated unique regulatory inputs as documented in RegulonDB [20].

Condition	Gene	Blattner	Library_plate_id	Library_well_id	Number known regulatory inputs	Sigma factor
M9 + 0.2% glucose	puA66	-	AZ01	F3	0	unknown
M9 + 0.2% glucose	leuE	b1798	AZ11	B5	1	unknown
M9 + 0.2% glucose	ygdQ	b2832	AZ05	D4	0	unknown
M9 + 0.2% glucose	yjgA	b4234	AZ09	F8	0	unknown
M9 + 0.2% glucose	**nupC**	**b2393**	**AZ16**	**G11**	**3**	**Sigma70**
M9 + 0.2% glucose	ompT	b0565	AZ18	H2	1	unknown
M9 + 0.2% glucose	modF	b0760	AZ14	C6	0	unknown
M9 + 0.2% glucose	corA	b3816	AZ16	H5	0	Sigma70
M9 + 0.2% glucose	asd	b3433	AZ18	C9	0	unknown
M9 + 0.2% glucose	rrsA	b3851	AZ19	G2	2	Sigma32, Sigma70
M9 + 0.2% glucose	**rrsD**	**b3278**	**AZ20**	**A4**	**3**	**Sigma32, Sigma70**
M9 + 0.2% lactose	arpA	b4017	AZ05	G2	0	unknown
M9 + 0.2% lactose	pinH	b2648	AZ10	A8	0	unknown
M9 + 0.2% lactose	phoE	b0241	AZ01	A7	1	Sigma70
M9 + 0.2% lactose	**hisJ**	**b2309**	**AZ14**	**H11**	**3**	**Sigma54, Sigma70**
M9 + 0.2% lactose	yejL	b2187	AZ10	A5	0	unknown
M9 + 0.2% lactose	**lacZ**	**b0344**	**AZ14**	**H1**	**3**	**Sigma70**
M9 + 0.2% lactose	yajG	b0434	AZ19	A11	0	Sigma70
M9 + 0.2% lactose	ycbK	b0926	AZ13	C9	0	unknown
M9 + 0.2% lactose	**serA**	**b2913**	**AZ01**	**F4**	**3**	**Sigma70**
M9 + 0.2% lactose	rplN	b3310	AZ19	C9	0	unknown
M9 + 0.2% glycerol	yfaD	b2244	AZ17	B9	0	unknown
M9 + 0.2% glycerol	**cadB**	**b4132**	**AZ17**	**A9**	**6**	**Sigma70**
M9 + 0.2% glycerol	chaC	b1218	AZ16	E7	0	Sigma54
M9 + 0.2% glycerol	ompR	b3405	AZ20	C4	2	Sigma70
M9 + 0.2% glycerol	rstA	b1608	AZ04	H4	1	unknown
M9 + 0.2% glycerol	ribB	b3041	AZ18	C6	0	unknown
M9 + 0.2% glycerol	**serA**	**b2913**	**AZ03**	**C3**	**3**	**Sigma70**
M9 + 0.2% glycerol	ilvL	b3766	AZ01	C1	2	Sigma70
M9 + 0.4M NaCl (+gluc)	zraS	b4003	AZ03	F11	2	Sigma54
M9 + 0.4M NaCl (+gluc)	allS	b0504	AZ13	H5	1	unknown
M9 + 0.4M NaCl (+gluc)	ybaQ	b0483	AZ13	D7	0	unknown
M9 + 0.4M NaCl (+gluc)	pepE	b4021	AZ04	G11	0	unknown
M9 + 0.4M NaCl (+gluc)	yeeX	b2007	AZ17	G10	0	unknown
M9 + 0.4M NaCl (+gluc)	rppH	b2830	AZ13	B7	0	unknown
M9 + 0.4M NaCl (+gluc)	secE	b3981	AZ13	C4	0	unknown
M9 + 0.4M NaCl (+gluc)	minC	b1176	AZ13	A7	0	unknown
M9 + 0.4M NaCl (+gluc)	pheM	b1715	AZ10	B8	2	unknown
M9 + 0.4M NaCl (+gluc)	**serA**	**b2913**	**AZ12**	**C3**	**3**	**Sigma70**
M9 + 0.4M NaCl (+gluc)	rpsL	b3342	AZ19	C1	0	unknown
M9 + 0.4M NaCl (+gluc)	rplN	b3310	AZ19	C9	0	unknown
Stationary phase 16h	**manY**	**b1818**	**AZ17**	**B3**	**4**	**Sigma70**
Stationary phase 16h	sppA	b1766	AZ12	D10	0	unknown
Stationary phase 16h	ytjB	b4387	AZ01	D12	0	unknown
Stationary phase 16h	pykF	b1676	AZ02	A8	1	Sigma38, Sigma70
Stationary phase 16h	yjjK	b4391	AZ14	H3	0	unknown
Stationary phase 16h	panD	b0131	AZ18	B12	0	Sigma70
Stationary phase 16h	ilvL	b3766	AZ01	C1	2	Sigma70
Stationary phase 16h	rpsM	b3298	AZ19	C8	0	Sigma70

Table S3.4: Statistical comparison of the variance between native promoters in each condition.

Condition	p-value-Kruskal.Wallis	p-value-Anova	significance
M9 + 0.2% glucose	0.008	0.22	*/ns
M9 + 0.2% lactose	0.788	0.653	ns
M9 + 0.2% glycerol	0.726	0.662	ns
M9 + 0.4M NaCl	0.875	0.998	ns
Stationary phase 16h	0.439	1	ns

Table S3.5: Estimated number of proteins per promoter. We used the conversion factor determined in [96] (2.88) to transform flow cytometry fluorescence units into number of gfp molecules per cell. The value was established using quantitative Western Blots in eight strains with known FCM intensities. We multiplied the mean fluorescence values in linear scale of each promoter (after autofluorescene removal) by the conversion factor. The values shown for the medium and high expressers are the average number of GFP molecules per cell over all measured promoters, together with the lowest and maximal value measured.

Condition	**Gene**	**Blattner**	**Average number of proteins per cell**
M9 + 0.2% glucose	puA66	-	0
M9 + 0.2% glucose	leuE	b1798	253
M9 + 0.2% glucose	ygdQ	b2832	838
M9 + 0.2% glucose	yjgA	b4234	1,814
M9 + 0.2% glucose	medium	-	2,971 (min: 1,065; max: 11, 911)
M9 + 0.2% glucose	nupC	b2393	3,114
M9 + 0.2% glucose	ompT	b0565	9,589
M9 + 0.2% glucose	modF	b0760	10,422
M9 + 0.2% glucose	corA	b3816	14,201
M9 + 0.2% glucose	asd	b3433	38,069
M9 + 0.2% glucose	rrsA	b3851	85,694
M9 + 0.2% glucose	rrsD	b3278	131,043
M9 + 0.2% glucose	high	-	45,879 (min:17,710; max: 88,321)
M9 + 0.2% lactose	arpA	b4017	0
M9 + 0.2% lactose	pinH	b2648	840
M9 + 0.2% lactose	phoE	b0241	2,996
M9 + 0.2% lactose	medium	-	3,176 (min: 1,282; max: 12, 509)

Table S3.5 continued from previous page

Condition	Gene	Blattner	Average number of proteins per cell
M9 + 0.2% lactose	hisJ	b2309	4,701
M9 + 0.2% lactose	yejL	b2187	10,788
M9 + 0.2% lactose	lacZ	b0344	16,127
M9 + 0.2% lactose	yajG	b0434	23,944
M9 + 0.2% lactose	ycbK	b0926	39,346
M9 + 0.2% lactose	serA	b2913	55,555
M9 + 0.2% lactose	rplN	b3310	285,959
M9 + 0.2% lactose	high	-	53,812 (min: 21,769; max: 103,462)
M9 + 0.2% glycerol	yfaD	b2244	0
M9 + 0.2% glycerol	cadB	b4132	482
M9 + 0.2% glycerol	chaC	b1218	940
M9 + 0.2% glycerol	ompR	b3405	1,673
M9 + 0.2% glycerol	medium	-	2,441 (min: 940; max: 10,622)
M9 + 0.2% glycerol	rstA	b1608	8,625
M9 + 0.2% glycerol	ribB	b3041	16,661
M9 + 0.2% glycerol	serA	b2913	25,720
M9 + 0.2% glycerol	ilvL	b3766	61,714
M9 + 0.2% glycerol	high	-	36,843 (min: 18,185; max: 69,605)
M9 + 0.4M NaCl (+gluc)	zraS	b4003	0
M9 + 0.4M NaCl (+gluc)	allS	b0504	0
M9 + 0.4M NaCl (+gluc)	ybaQ	b0483	979
M9 + 0.4M NaCl (+gluc)	pepE	b4021	1,683
M9 + 0.4M NaCl (+gluc)	yeeX	b2007	2,859
M9 + 0.4M NaCl (+gluc)	medium	-	4,333 (min: 852; max: 15,085)
M9 + 0.4M NaCl (+gluc)	rppH	b2830	4,551
M9 + 0.4M NaCl (+gluc)	secE	b3981	10,548
M9 + 0.4M NaCl (+gluc)	minC	b1176	16,591
M9 + 0.4M NaCl (+gluc)	pheM	b1715	25,910
M9 + 0.4M NaCl (+gluc)	serA	b2913	38,221
M9 + 0.4M NaCl (+gluc)	rpsL	b3342	129,139
M9 + 0.4M NaCl (+gluc)	rplN	b3310	192,297
M9 + 0.4M NaCl (+gluc)	high	-	65,302 (min: 21,682; max: 136,819)

Table S3.5 continued from previous page

Condition	Gene	Blattner	Average number of proteins per cell
Stationary phase 16h	manY	b1818	120
Stationary phase 16h	sppA	b1766	488
Stationary phase 16h	ytjB	b4387	1,983
Stationary phase 16h	pykF	b1676	4,796
Stationary phase 16h	yjjK	b4391	7,699
Stationary phase 16h	panD	b0131	17,631
Stationary phase 16h	ilvL	b3766	64,212
Stationary phase 16h	rpsM	b3298	85,450

3.9 Supplementary text

3.9.1 Extended microscopy protocol

This section contains a detailed description of all steps performed (including critical information) to successfully measure 96-well plates using an inverted fluorescence microscope.

- **Equipment and Materials**
  - Glass-bottom microtiter 96-well plates (Greiner bio-one Ref.655090)
  - Poly-L-lysine 0.01% (Sigma P4832-50ML)
  - Wide-field inverted fluorescence microscope (Nikon Ti)
  - High-numerical-aperture (NA), oil-immersion objective (CFI Plan Apochromat Lambda DM x100 objective [NA 1.45, WD 0.13mm])
  - Immersion oil (Nikon type-A)
  - Motorized x-y stage controller (Nikon TI-SH-W Plate-holder)
  - Plate holder metal clamps
  - Temperature incubator (Life Imaging Services Ice Cube) - OPTIONAL
    - *Depending on the number of positions acquired in each well, the acquisition time for a full plate can range from ∼30 minutes to longer periods of time. Therefore, it can be critical to maintain the cells chilled in order to avoid differences in cellular state between the first and last acquired well. We runned the experiments at around 10 °C.*
  - Piezo stage (Mad City Labs, Nano-Drive 85; Cen-Tech Digital Multimeter) - OPTIONAL

- **Experimental procedure**

  1. **Grow your cells as required**
     - The experiments presented in the main text were performed after growing cells carrying out fluorescent reporters (*GFP*) of interest in microtiter plates in different conditions (e.g. minimal media M9 +

0.2% glucose) for two consecutive overnights after inoculation from a glycerol stock at -80 °C (∼1/2000 dilution between overnights). The day of the experiment a ∼1/200 dilution was performed into 200ml of fresh media and cells were grown between 3 and 16 hours, depending on the condition, before microscopy measurements.

2. **Turn on the microscope and start cooling it down (OPTIONAL)**
3. **Plate preparation**
   (a) Add 25 μl of Poly-L-lysine 0.01% to the bottom of each well. Incubate for 5 min at room temperature.
   Critical! Make sure that the poly-L-lysine covers the whole bottom surface of the well.
   (b) Rinse the wells with dH_2O, first and then with 1x PBS. Repeat twice.
   Critical! Make sure the water and PBS are filtered to avoid any dust particles sticking to the poly-L-lysine. If possible, perform this step in a fume hood.
   (c) Add the growing cultures into each well.
   Critical! Make sure to pipette up and down each culture a few times to separate segregated cells before adding them to the plate. A high fraction of cell clumps will make the analysis of the images very hard, if not impossible.
      – For the experiments presented in the main text, 30 μl of an exponentially growing culture were usually added to each well. For cells already in Stationary Phase ∼25μl of a 1/50 dilution in minimal media were added.
   (d) Centrifuge the plate for 10 min at 500g. This will attach the cells to the poly-L-lysine at the bottom.
   (e) Carefully remove the supernatant.
   (f) Rinse twice with either 1x PBS or the media the cells were growing in.
   Critical! Make sure the media or PBS are clean and filtered to avoid any dust particles sticking to the poly-L-lysine. If possible, perform this step in a fume hood.
   (g) Add immersion oil to the bottom-glass surface of the plate.
   Critical! The oil needs to be distributed carefully to avoid the for-

mation of air bubbles. Air bubbles will prevent the Perfect Focus System (PFS) to find the correct focus across positions. At the same time, it is also important to make sure that the oil is not dripping, in order to avoid damaging pieces of the microscope.

(h) Fill wells with 1x PBS or the media cells were growing in in to avoid them to dry out during image acquisition.

The plate is now ready for image acquisition.

4. **Image acquisition**

(a) Add immersion oil to the objective.

(b) Tight the bottom glass plate with the metal clamps to avoid it from moving while imaging.

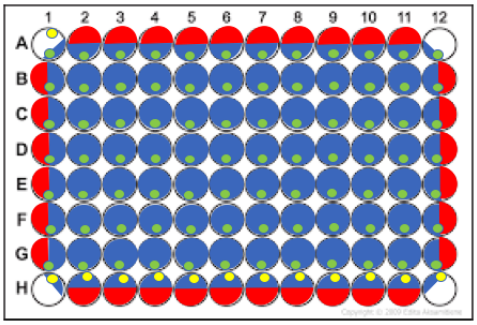
(c) Open Micro-manager (we used version xxx) [168]

(d) Open the HCS Site Generator tool in Micro-manager (plugins → Acquisition Tools → HCS Site Generator, which allows the acquisition of entire 96-well plates) and select the type of plate, number of wells and number of positions to measure in each well. Make sure that the position grid inside each well has a spacing of at least 200μM between positions (to avoid bleaching due to cross-illumination between them). Calibrate the x-y stage position relative to the position of well A1 using the calibration tool.

Critical! Always calibrate the position of the x-y stage before attempting to measure any well using this tool. To do so, position the objective in the middle of well A1 (or a different position in the well, as described below) and use the calibrate option. If failing to do so, the microscope can be very badly damaged!

Important notes on position calibration using HCS:

The wells at the border of the plate (A1-H1, A1-A12, A12-H12, H1-H12) need to be calibrated carefully. Depending on the calibrated position of the x-y stage, the borders of the objective can touch the plate holder in these positions, which can result in equipment damage. The following figure shows the positions that are safe to image without causing any damage:



In the figure, the blue coloured areas are safe positions to image. The red coloured areas are safe too, but if the z-position of the objective is higher than a certain value (this position will change from set-up to set-up) the objective will touch the plate holder, potentially damaging the objective/plate holder. The white coloured areas are Critical! positions, where it is advised not to set the calibration position. The green and yellow spots mark the two ideal places to set the calibration position using the HCS Site Generator tool. If calibrating in the position of marked by the green spot in well A1, all wells except the bottom ones can be safely measured. To safely measure the H1-H12 row, another acquisition needs to be set-up separately after the acquisition of the rest of the wells, where the x-y stage position is calibrated at the position of the yellow spot mark in well A1.

**These notes are based on our set-up. It could very well be, that in a different set-up this problem does not exist*

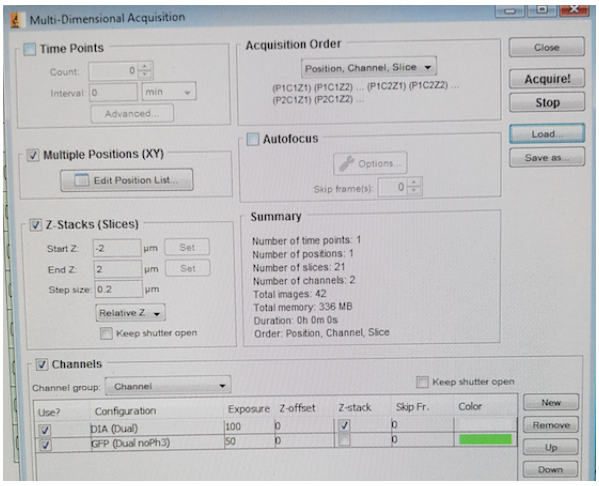
(e) Once the calibration is done, go to a well containing cells (ideally cells highly expressing the used fluorophore) and save the objective position (z-position) in which cells are in focus using the Perfect Focus System.

Critical! Always move between positions (wells) with the objective in a low z-position to avoid damaging it.

Critical! Find the cells using brightfiled (DIA channel), since phase contrast does not work with bottom-glass plates. Once the cells are found, set the correct position using the fluorescence channel, as it is easier to see the exact position at which the cells are in focus using

fluorescence. *Important*!: Do this in a position that you will not acquired afterwards to avoid bleaching.

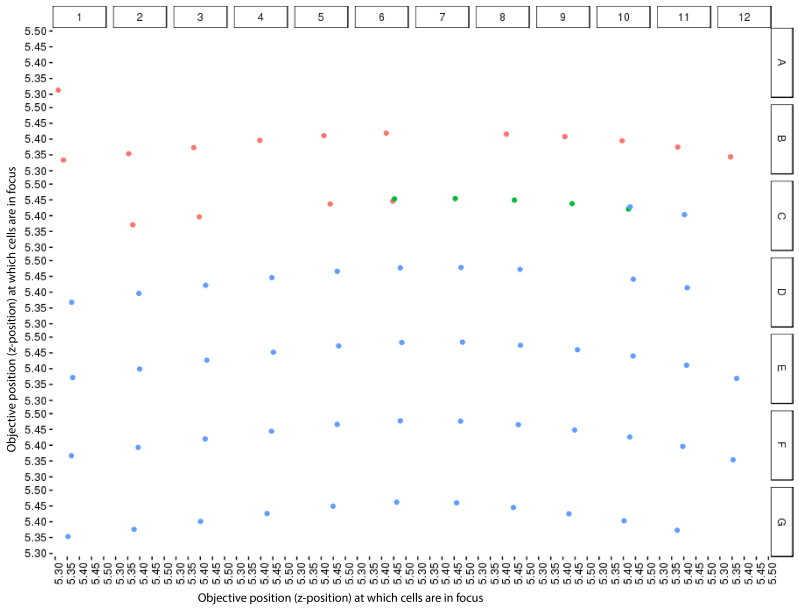
(f) Open the Multi-Dimensional Acquisition engine and select all important parameters to run the experiment (e.g. channels, exposure intensity, exposure time, range of z-stack for brightfield, ...). Below is a snapshot of an example of how the settings of an experiment could look like.



(g) Go to the first well you are interested in acquiring and position the objective close to the focal plane. It is always a good idea to make sure that all settings are correct by acquiring an individual image in a single position, before running the whole plate (again, in a position you will not saved afterwards due to bleaching).

(h) Once all settings are correct. Image the whole plate, by pressing acquire in the MDA engine.

Critical! The z-position at which cells are in focus changes across the plate, as the plate is slightly curved (see picture below).



The picture shows the best z-position (x and y axis) at which cells are in focus in each well (wells A2-A12 and H1-H12 are not shown). The position increases slightly towards the middle of the plate. The maximal difference between z-positions is 171μM (from well A1 to well E7), the difference between neighbouring wells can range from 1μM to 42μM. When using the autofocus hardware integrated in Micro-manager, the experiment is likely to fail and the objective can be very seriously damaged. This is because of the way the hardware tries to find the best z-position between positions. In short, it will always look for the best position between a set lower and upper range starting from the last stored value (initially set by the user). If the Perfect Focus System fails to find the right focus in one position (e.g. due to air-bubbles in the immersion oil, dust in the plate, no cells ...), it will store the z-position at the position where it stopped (which is the highest possible value within the set range). If it fails to find it over consecutive wells, the stored position from image to image will keep increasing, reaching a point where the correct focal plane will be out of the searching range. The objective will end up in a position where it will keep pushing closer and closer against the bottom of the plate from position to position, eventually crashing and breaking. To overcome this issue and successfully be able to find the perfect focus

across the entire plate without breaking the objective, we wrote a small customized runnable to do the following:

- set the speed of the x-y stage controller to slow (we set it a value of 6, as relatively slow speed helps avoiding the formation of air bubbles while moving the objective across positions)
- iteratively search for the perfect focus in small steps of 2μM staring from the optimal objective position (z-position) set by the user. It starts by searching 2μM down from the stored position and then 2μM up, and so on. If after a while the perfect focus is not found, the range is increased, but the runnable prevents the objective to move more than 500μM in any direction to avoid breaking the objective (an already generous range, as the measured highest difference across wells is 171 μM). If the focus is not found, the position at which the Perfect Focus System will start searching in the next image is the last position at which the perfect focus was found (not the position of the previous image).

5. **If successful, save all acquired data, clean and turn-off the microscope**

Most important points in order to succeed from our experience:

- Always calibrate the x-y stage controller within the HCS Site Generator tool before starting any acquisition.
- Always move the objective in a low z-position between positions before starting the experiment.
- Don't set the speed of the x-y stage controller to fast to avoid the creation of air-bubbles in the immersion oil, which will prevent the perfect focus system to find the focus.
- Use custom runnables to ensure that the perfect focus system is able to find the correct focus across all positions in the plate, even if the best z-position changes between them.

This is the code generated to find the correct focus across all positions in the plate. Note that it assumes that there is a Piezo available.

```
//This runnable allows to perform an extended PFS for large samples like plates.
//At each acquisition position, the ZDrive makes a series of attempts to lock
//the PFS by moving the sample above and below the default position.
//The default position is the last position were locking was successfull.

//Before running the MDA, one should be at the first position, turn on the PSF
//and make sure the Z piezo is in the middle of its range.
//PFS is around 3900
//Define names of devices
zStage = "TIZDrive";
PFSname = "TIPFSStatus";
piezoName = "MCL Z Piezo";

//Extended PSF parameters to choose
max_Z_range = 500;//maximum allowed travel distance of Z stage from initial pos.
stepSizeSmall = 0.5;//size in microns of each step
stepSize = 2;//size in microns of each step
stepNumSmall = 4;
stepNum = 73;//number of search steps

//change xyStage speed
mmc.setProperty("TIXYDrive","SpeedX",6);
mmc.setProperty("TIXYDrive","SpeedY",6);

//recover mda settings
settings = mm.acquisitions().getAcquisitionSettings();
num_slices = settings.slices.size();
slice_size = settings.slices.get(num_slices-1);

//function that checks whether PFS found focus
boolean testFocus() {
   try {
      //if anyway out of range dont try to focus
            PSF_status = mmc.getProperty(PFSname, "Status");
            if (PSF_status.equals("Out of focus search range")) {
                  return false;
            }
      mmc.fullFocus();
      mmc.waitForDevice(PFSname);
            }
       catch (Exception ex) {
            //focus failed
            //print("focus failed");
```

```
        return false;
                }
    return true;
   };

//clear any attached runnable
mm.acquisitions().clearRunnables();
//stop PFS
mmc.enableContinuousFocus(false);
//Record the Zdrive position and remember as last working position
ref_Z = mmc.getPosition(zStage);
init_Z = ref_Z;
newZpos = ref_Z;

//
mm.acquisitions().clearRunnables();
settings = mm.acquisitions().getAcquisitionSettings();

runnable = new Runnable() {
        numFrames = settings.numFrames;
        public void run() {
        //print(numFrames);

        boolean success = testFocus();

      for (int i = 0; i < stepNum && !success; i++) {
         if (i < stepNumSmall){
                print("here1");
                newZpos = ref_Z + i*stepSizeSmall;
         }
         else{
                print("here2");
                newZpos = ref_Z + (stepSizeSmall*stepNumSmall) + (i-
stepNumSmall)*stepSize;
         }
         //newZpos = ref_Z + i*stepSize;

         if (newZpos < init_Z + max_Z_range){
                try {
                mmc.setPosition(zStage, newZpos);
                mmc.waitForDevice(zStage);
                } catch (Exception ex) {
                ReportingUtils.showError(ex, "Failed to set Z position");
                }
                success = testFocus();
         }
```

```
        if (i < stepNumSmall){
                print("here3");
                newZpos = ref_Z - i*stepSizeSmall;
        }
        else{
                print("here4");
                newZpos = ref_Z - (stepSizeSmall*stepNumSmall) - (i-
stepNumSmall)*stepSize;
        }
        //newZpos = ref_Z - i*stepSize;

        if (newZpos > init_Z - max_Z_range && !success){
                try {
                mmc.setPosition(zStage, newZpos);
                mmc.waitForDevice(zStage);
                } catch (Exception ex) {
                ReportingUtils.showError(ex, "Failed to set Z position");
                }
                success = testFocus();
        }
      }
      if (success) {
        try {
          ref_Z = mmc.getPosition(zStage);
          print(ref_Z);
        } catch (Exception ex) {
          ReportingUtils.logError(ex);
        }
      } else {
        mmc.setPosition(zStage, ref_Z);
        mmc.waitForDevice(zStage);
      }
      print(ref_Z);
      print(success);

      mmc.setPosition(zStage, ref_Z-slice_size);
        }
};


//attach runnable
// Run the runnable on all frames (at first position, channel, slice)
// Numbers are frames, stage positions, channels, slices (-1 attach to all planes along
given dimension)
mm.acquisitions().attachRunnable(-1, -1, 0, 0, runnable);
```

```
// Start MDA
mmc.enableContinuousFocus(false); // disable PFS before starting MDA
Thread.sleep(500); // wait 0.5s
mm.acquisitions().runAcquisition();
```

4

Summary and future perspectives

The work presented here provides the first systematic investigation of the coupling between gene regulation and gene expression noise across different growth conditions in bacteria, as well as the role of noise propagation in shaping condition-dependent gene expression noise. Previous studies [96] had reported a strong relationship between regulation and noise in *Escherichia coli* growing in minimal media supplemented with glucose, and had suggested that noise propagation from transcription factors to their targets may explain the observed link. Gene expression regulation is mainly controlled by sigma and transcription factors (Chapter 1). In *E. coli* there are roughly 215 transcription factors and 3355 regulatory interactions documented [20], and such complexity entails that noise propagation likely plays a role in shaping genome-wide noise levels. Various studies have shown how noise propagates through regulatory cascades and can affect downstream components experimentally [109, 110, 115] and theoretically [111]. In eukaryotes, for example, it has been established that transcription factor binding and unbinding constitutes an important source of noise [86, 97, 106]. However, the contribution of noise propagation in shaping the variation of expression noise levels across conditions in bacteria had, so far, not yet been systematically quantified until now.

In order to investigate how genome-wide noise properties vary across growth conditions, we first used high-throughput flow cytometry to measure single-cell gene expression distributions of a high fraction of *E.coli* promoters across a diverse range of conditions, and then used computational modeling to understand what drives condition-dependent noise levels (Chapter 2). Moreover, we developed a new technique to measure and analyse single-cell fluorescence measurements using high-

throughput microscopy, as flow cytometry has important limitations when quantifying other pivotal quantities such as cell size and introduces substantial measurement noise (Chapter 3). All in all, the main results of the thesis can be summarized in the following key findings:

1) We show that a promoter's noise can be decomposed into two components. The first component, which we've termed noise floor, is promoter independent and mainly exhibited by constitutive expressed promoters, independently of their expression level. The second component affects regulated genes, which clearly deviate from the noise floor as they display additional noise due to noise propagation from the transcription factors that regulate them. Classically, noise has been decomposed into extrinsic and intrinsic noise [82], and here we've extended the idea by introducing an extra noise propagation component. In both our setups (flow cytometry and microscopy), we were not able to measure the noise contribution of intrinsic sources, as all the constructs express above the threshold at which intrinsic fluctuations dominate [41]. All constructs have been cloned on a low copy plasmid (puA66, puA139, 1-6 copies per cell), and fluctuations in plasmid copy number are likely to contribute to extrinsic noise. Having fluorescent fusions directly integrated in the native locations in the chromosome, could be a possible way to overcome this limitation, although fluctuations in the number of replicated chromosomes may also contribute to add extra sources of noise.

2) In chapter 2 we show that the noise floor is an almost perfect linearly decreasing function of growth-rate, indicating that during adverse conditions, all genes have automatically more heterogeneous expression across cells. This strong association opens the door to questions for which we do not have a clear answer yet. For example, are high noise levels during slow growth a consequence of the changes in gene expression that come with changes in the physiology (i.e dilution rate, mRNA half lives) or has natural selection acted on increasing noise levels in cells that do not grow at their optimum to increase their chances of survival when a more favorable environment comes (bet-hedging)? This observation could, as well, have important consequences for the design, for example, of synthetic circuits in which maintaining low noise levels is fundamental. Although this is a very interesting preliminary observation, flow cytometry has important limitations in allowing us to understand the causes of the observed growth-rate dependent differences in the noise floor. Flow cytometry only allows the comparison of how the relative levels of promoters across

conditions change, as other quantitative estimations of parameters such as cell size variability or growth rate variability across identical single-cells, are not possible [122]. In chapter 3 we overcame some of the flow cytometry limitations by designing a high-throughput microscopy technique. We then used it to investigate whether we still could see a clear noise floor in different conditions by measuring synthetic and low noise native promoters. The set of measured synthetic promoters is an interesting case, as they were evolved from random sequences only on their mean properties, and all showed low noise levels close to the values of the noise floor. This implies that high noise in native promoters must have been positively selected by natural selection [96]. In accordance with the observations in chapter 2, we observed that none of the promoters (native and synthetic) showed deviations from the noise floor. We also did not observe strong differences between conditions in the level of expression noise contributing to the noise floor. However, the conditions measured (glucose, lactose, glycerol), also did not show strong differences in the flow cytometer. It would be informative to measure low noise promoters in rich media using the microscopy setup, as it is the condition that showed the highest differences in the flow cytometry dataset (Figure 2.2C, Chapter 2). A more careful analysis of the dependency of the noise floor to general physiological changes would also be insightful to understand how the noise floor relates to fluctuations in cell size, or condition-dependent changes (e.g growth-rate, number of ribosomes and polymerases, etc). However, this type of questions would require the gathering of data using alternative methods as the ones we used, for example high-throughput versions of a microfluidic device like the mother machine [159], which allows to track expression, cell size and growth rate dynamics in single-cells over long periods of time. Other techniques, such as single-cell mass-spectrometry (sc-MS), could be useful in understanding how the concentration of different molecules (e.g ribosomes, polymerases or transcription factors) fluctuate across cells and environments [173]. However, to date sc-MS is still far from being able to give reliable quantitative and noise-free measurements of single-cell protein levels in bacteria.

As a side note, thanks to the development of the microscopy protocol we confirmed that measurements using flow cytometry were highly dominated by shot-noise, as we had already reported elsewhere [122]. This highlights the importance of carefully examining technical sources of noise when aiming to perform quantitative analyses. Measurements presented in Chapter 2 were corrected for the dependency that arises from the shot-noise before analysis.

3) Due to noise propagation, we show that regulated genes show highly condition-dependent noise properties. Since the amount by which transcription factors fluctuate will most likely change in different conditions, their targets noise levels will be affected accordingly (i.e. in a condition where a transcription factor's activity is highly variable its targets will see their variability increase, in contrast to a condition where the same factor has a more homogeneous activity). We performed systematic computational modeling of each promoter's condition-dependent noise in terms of regulatory binding sites in their sequences to quantify how noise propagates through the regulatory network. This allowed us to identify which transcriptions factors are most responsible for propagating noise in each condition and on average across all. The set of factors that we found consistently appearing through all conditions include, H.NS, Sigma38, CRP, PhoB, GadX and GadW and their appearance is consistent with the chosen conditions and previous knowledge in the literature. A caveat of the model used is, perhaps, its simplicity; it would be interesting to further extend the model by incorporating other possible sources of variability, such as growth rate or cell size and see how the fraction of explained variance may be affected. Or incorporate information on the regulatory function of the factors (i.e. do they activate, repress or both?) to explore how regulatory architectures influence noise propagation. The work presented could be easily extended to study noise propagation in many other different conditions. What kind of factors would have, for example, appeared in a completely different set, such as during anaerobic growth or in biofilms. Models of how noise propagates through gene regulatory networks in a condition-dependent manner have important applications for understanding how gene circuits work and to help in their design.

If the predictions regarding which transcription factors contribute to higher noise levels are correct, it should be possible to experimentally validate them. We started to do some work towards this direction by taking two approaches. First, we chose high noise native promoters for which the model was able to explain a high fraction of their variance in terms of noise propagation from one of the top factors identified in all conditions. We then performed site directed mutagenesis to destroy binding sites for some of these factors and investigate how their noise levels changed. In the second approach, we selected a unique low noise synthetic promoter expressing at a medium level. We created a library of mutants by systematically adding binding sites of some of these factors in different positions along the promoter's sequence. At the moment, either approach has given conclusive results, yet. The difficulty

in analyzing some of the data gathered so far, may come from the fact that we do not have a complete understanding of how many of these factors act *in vivo* or how their activity depends on the interactions with other factors, most likely in very promoter-specific ways. Recent studies have shown that our current understanding of how transcriptional regulation works overall in *E.coli* in terms of transcription factor binding is limited [69].

4) Finally, we identify a new organizing principle of genes based on their regulatory properties. Previously it had been shown that genes are organized along a one-dimensional axis based on their absolute expression level, codon bias, and evolutionary rate [156], and here, thanks to the unique nature of our measurements, we found a second independent axis in which genes are organized by their regulatory complexity, gene expression plasticity and noise. This provides an important new insight into the genome-wide organization of gene regulatory networks and shows that gene expression noise is a trait that is mostly dependent on a promoter's regulatory state.

In this work, we have only presented the quantification of single-cell protein levels. However, since the constructs were designed specifically to measure transcription (all share the same mRNA sequence), we believe that most of the measured fluctuations arise during this step. It would have been, nevertheless, valuable to know how the single-cell mRNAs distributions varied across the environments. Single-cell mRNAs can be measured using a technique called single-molecule fluorescence in situ hybridization (smFISH), which allows to fluorescently 'label' single mRNA in individual cells and count their numbers as fluorescent spots [174]. The fact that all constructs used share the same mRNA sequence, motivated us to try to perform smFISH, as the same set of fluorescently labelled designed probes could be used in all. We performed the technique in a single construct (*placZ-gfp*) and even though we were able to visually see and computationally analyze spots corresponding to single mRNAs, the difference in background fluorescence and the fluorescence coming from a real spot was too small to be able to accurately tell them apart. A possible cause may have been the length of the mRNA sequence (*gfpmut2*) in which only 29 individual probes fit (as an example, other studies, were smFISH has been performed in *E.coli*, used 50 [174] or 72 [90] probes) or simply due to our lack of expertise.

Taken all together, the results of the thesis provide for the first time important insights into the functioning of genome-wide gene regulatory networks, and the way in which noise propagation controls heterogeneity of expression levels across identical cells in a condition-dependent manner. I hope that this work will contribute and help future research projects in understanding why certain regulatory strategies allow homogeneous populations to divide into sub-populations and diversify their adaptation mechanisms, as well as the role of evolution in shaping these strategies. Understanding mechanisms by which bacteria can diversify their response to environmental changes has important implications in many areas of biology: from understanding why individual bacteria may become resistant to antibiotics to rationalizing the design of synthetic bacteria to perform a variety of functions of industrial interest.

References

1. Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental biology* **278,** 274–88 (2005).
2. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).
3. Crick, F. On protein synthesis. *Symposia of the Society for Experimental Biology* **12,** 138–63 (1958).
4. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–563 (1970).
5. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences* **47,** 1588–1602 (1961).
6. Nirenberg, M. W., Jones, O. W., Leder, P., Clark, B. F. C., Sly, W. S. & Pestka, S. On the Coding of Genetic Information. *Cold Spring Harbor Symposia on Quantitative Biology* **28,** 549–557 (1963).
7. Lloyd, G., Landini, P. & Busby, S. Activation and repression of transcription initiation in bacteria. *Essays in biochemistry* **37,** 17–31 (2001).
8. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* **296,** 1285–1290 (2002).
9. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. *Cell* **98,** 811–824 (1999).
10. Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14,** 638–650 (2016).

11. Jovanovic, M., Buck, M., Ye, F., Zhang, X., Kotta-Loizou, I. & Glyde, R. Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Molecular Cell* **70,** 1111–1120 (2018).

12. Murakami, K. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* **5,** 848–864 (2015).

13. Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. & Young, B. *The functional and regulatory roles of sigma factors in transcription* in *Cold Spring Harbor Symposia on Quantitative Biology* **63** (1998), 141–155.

14. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology* **57,** 441–466 (2003).

15. Ishihama, A. Functional Modulation of Escherichia Coli RNA Polymerase. en. *Annual Review of Microbiology* **54,** 499–518 (2000).

16. Brewster, R. C., Jones, D. L. & Phillips, R. Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. *PLoS computational biology* **8,** e1002811 (2012).

17. Kinney, J. B., Murugan, A., Callan, C. G. & Cox, E. C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences* **107,** 9158–9163 (2010).

18. Arnold, P., Erb, I., Pachkov, M., Molina, N. & van Nimwegen, E. MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. *Bioinformatics* **28,** 487–494 (2012).

19. Patrick, M., Dennis, P. P., Ehrenberg, M. & Bremer, H. Free RNA polymerase in Escherichia coli. *Biochimie* **119,** 80–91 (2015).

20. Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J. S., Alquicira-Hernández, K., Muñiz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D. A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C.-F., Galagan, J. & Collado-Vides, J. RegulonDB v 10.5: tackling challenges

to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Research* **47,** D212–D220 (2019).

21. Maeda, H. Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research* **28,** 3497–3503 (2002).

22. Lee, D. J., Minchin, S. D. & Busby, S. J. Activating Transcription in Bacteria. *Annual Review of Microbiology* **66,** 125–152 (2012).

23. Hook-Barnard, I. G. & Hinton, D. M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene regulation and systems biology* **1,** 275–93 (2007).

24. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The Transition between Transcriptional Initiation and Elongation in E. coli Is Highly Variable and Often Rate Limiting. *Molecular Cell* **24,** 747–757 (2006).

25. Bidnenko, E., Bidnenko, V., Grylak-Mielnicka, A. & Bardowski, J. Transcription termination factor Rho: a hub linking diverse physiological processes in bacteria. *Microbiology* **162,** 433–447 (2016).

26. Mitra, A., Kesarwani, A. K., Pal, D. & Nagaraja, V. WebGeSTer DB-A transcription terminator database. *Nucleic Acids Research* **39,** D129–D135 (2011).

27. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3'-end chronicles. *Journal of molecular biology* **412,** 793–813 (2011).

28. Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. An Allosteric Path to Transcription Termination. *Molecular Cell* **28,** 991–1001 (2007).

29. Andreeva, I., Belardinelli, R. & Rodnina, M. V. Translation initiation in bacterial polysomes through ribosome loading on a standby site on a highly translated mRNA. *Proceedings of the National Academy of Sciences* **115,** 4411–4416 (2018).

30. Laursen, B. S., Sorensen, H. P., Mortensen, K. K. & Sperling-Petersen, H. U. Initiation of Protein Synthesis in Bacteria. *Microbiology and Molecular Biology Reviews* **69,** 101–123 (2005).

31. Mitarai, N., Sneppen, K. & Pedersen, S. Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. *Journal of Molecular Biology* **382,** 236–245 (2008).

32. Salis, H. M., Mirsky, E. a. & Voigt, C. a. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* **27,** 946–950 (2009).

33. Molina, N. & Van Nimwegen, E. Universal patterns of purifying selection at noncoding positions in bacteria. *Genome Research* **18,** 148–160 (2008).

34. Li, G.-W. How do bacteria tune translation efficiency? *Current Opinion in Microbiology* **24,** 66–71 (2015).

35. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics* **12,** 32–42 (2011).

36. Karlin, S. & Mrazek, J. Predicted highly expressed genes of diverse prokaryotic genomes. *Journal of Bacteriology* **182,** 5238–5250 (2000).

37. Petry, S., Weixlbaumer, A. & Ramakrishnan, V. The termination of translation. *Current Opinion in Structural Biology* **18,** 70–77 (2008).

38. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic acids research* **38,** 750–3 (2010).

39. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *Journal of bacteriology* **119,** 736–47 (1974).

40. Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences* **99,** 9697–9702 (2002).

41. Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A. & Sunney Xie, X. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329,** 533–538 (2010).

42. So, L.-H. H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R. & Golding, I. General properties of transcriptional time series in Escherichia coli. *Nature Genetics* **43,** 554–560 (2011).

43. Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R. & Heinemann, M. The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology* **34,** 104–110 (2016).

44. Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T. & Williamson, J. R. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology* **11,** e784–e784 (2015).

45. Sepúlveda, L. A., Xu, H., Zhang, J., Wang, M. & Golding, I. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science* **351,** 1218–1222 (2016).

46. Garcia, H. G., Sanchez, A., Kuhlman, T., Kondev, J. & Phillips, R. Transcription by the numbers redux: experiments and calculations that surprise. *Trends in Cell Biology* **20,** 723–733 (2010).

47. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J. & Phillips, R. Transcriptional regulation by the numbers: Models. *Current Opinion in Genetics and Development* **15,** 116–124 (2005).

48. Jacob, F. F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. eng. *Journal of Molecular Biology* **3,** 318–356 (1961).

49. Camilli, A. & Bassler, B. Bacterial Small-Molecule Signaling Pathways. *Science* **311,** 1113–1116 (2006).

50. Feklístov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual review of microbiology* **68,** 357–76 (2014).

51. Jishage, M. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38. *Journal of Bacteriology* **177,** 6832–6835 (1995).

52. Paul, B. J., Berkmen, M. B. & Gourse, R. L. DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences* (2005).

53. Cavanagh, A. T. & Wassarman, K. M. 6S RNA, a Global Regulator of Transcription in Escherichia coli , Bacillus subtilis , and Beyond. *Annual Review of Microbiology* (2014).

54. Shah, I. M. & Wolf, R. E. Novel protein-protein interaction between Escherichia coli SoxS and the DNA binding determinant of the RNA polymerase α subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promot. *Journal of Molecular Biology* **343,** 513–532 (2004).

55. Turnbough, C. L. & Switzer, R. L. Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors. *Microbiology and Molecular Biology Reviews* (2008).

56. Ishihama, A. Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiology Reviews* **34,** 628–645 (2010).

57. Balderas-Martínez, Y. I., Savageau, M., Salgado, H., Pérez-Rueda, E., Morett, E. & Collado-Vides, J. Transcription Factors in Escherichia coli Prefer the Holo Conformation. *PLoS ONE* **8,** e65723 (2013).

58. Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology* **6,** 482–489 (2003).

59. Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8,** 450–461 (2007).

60. Cho, B. K., Knight, E. M. & Palsson, B. Genomewide identification of protein binding locations using chromatin immunoprecipitation coupled with microarray. *Methods in Molecular Biology* **439,** 131–145 (2008).

61. Cho, B. K., Palsson, B. & Zengler, K. Deciphering the regulatory codes in bacterial genomes. *Biotechnology Journal* **6,** 1052–1063 (2011).

62. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316,** 1497–1502 (2007).

63. Van Nimwegen, E. Finding regulatory elements and regulatory motifs: a general probabilistic framework. *BMC bioinformatics* **8 Suppl 6,** S4 (2007).

64. Bulyk, M. L. Computational prediction of transcription-factor binding site locations. *Genome biology* **5,** 201 (2003).

65. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* **9,** 770–780 (2008).

66. Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I. & Karp, P. D. The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research* **45,** D543–D550 (2017).

67. Suzuki, H. *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature genetics* **41,** 553–562 (2009).

68. Balwierz, P. J., Pachkov, M., Arnold, P., Gruber, A. J., Zavolan, M. & Van Nimwegen, E. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Research* **24,** 869–884 (2014).

69. Larsen, S. J., Röttger, R., Schmidt, H. H. & Baumbach, J. E. coli gene regulatory networks are inconsistent with gene expression data. *Nucleic acids research* **47,** 85–92 (2019).

70. Yamamoto, K., Watanabe, H. & Ishihama, A. Expression levels of transcription factors in Escherichia coli: Growth phase- and growth condition-dependent variation of 90 regulators from six families. *Microbiology* **160,** 1903–1913 (2014).

71. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences* **43,** 553–566 (1957).

72. Spudich, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262,** 467–471 (1976).

73. Lehner, B. & Kaneko, K. Fluctuation and response in biology. *Cellular and Molecular Life Sciences* **68,** 1005–1010 (2011).

74. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology* **13,** 497–508 (2015).

75. Kussell, E. & Leibler, S. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science* **309,** 2075–2078 (2005).

76. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467,** 167–173 (2010).

77. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297,** 1183–6 (2002).

78. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31,** 69–73 (2002).

79. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135,** 216–226 (2008).

80. Sanchez, A., Choubey, S. & Kondev, J. Regulation of Noise in Gene Expression. *Annual Review of Biophysics* **42,** 469–491 (2013).

81. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature reviews. Genetics* **6,** 451–64 (2005).

82. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 12795–800 (2002).

83. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology* **4,** e309 (2006).

84. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304,** 1811–1814 (2004).

85. Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y. & Barkai, N. Noise in protein expression scales with natural protein abundance. *Nature genetics* **38,** 636–43 (2006).

86. Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. & Weissman, J. S. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840–846 (2006).

87. Shahrezaei, V. & Swain, P. S. The stochastic nature of biochemical networks. *Current Opinion in Biotechnology* **19,** 369–374 (2008).

88. Munsky, B., Neuert, G. & van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **336,** 183–7 (2012).

89. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123,** 1025–1036 (2005).

90. Jones, D. L., Brewster, R. C. & Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* **346,** 1533–1536 (2014).

91. Yang, S., Kim, S., Rim Lim, Y., Kim, C., An, H. J., Kim, J. H., Sung, J. & Lee, N. K. Contribution of RNA polymerase concentration variation to protein expression noise. *Nature Communications* **5** (2014).

92. Zopf, C. J., Quinn, K., Zeidman, J. & Maheshri, N. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Computational Biology* **9** (ed Kondev, J.) e1003161 (2013).

93. Schwabe, A. & Bruggeman, F. J. Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophysical Journal* **107,** 301–313 (2014).

94. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics* **43,** 95–100 (2011).

95. Silander, O. K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U. & Ackermann, M. A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. *PLoS Genetics* **8,** e1002443 (2012).

96. Wolf, L., Silander, O. K. & van Nimwegen, E. Expression noise facilitates the evolution of gene regulation. *eLife* **4,** e05856 (2015).

97. Carey, L. B., van Dijk, D., Sloot, P. M. A., Kaandorp, J. A. & Segal, E. Promoter Sequence Determines the Relationship between Expression Level and Noise. *PLoS Biology* **11,** e1001528 (2013).

98. Blake, W. J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R. & Collins, J. J. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Molecular Cell* **24,** 853–865 (2006).

99. Lehner, B. Conflict between noise and plasticity in yeast. *PLoS Genetics* **6,** e1001185 (2010).

100. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular systems biology* **4,** 170 (2008).

101. Barkai, N. & Shilo, B.-Z. Variability and Robustness in Biomolecular Systems. *Molecular Cell* **28,** 755–760 (2007).

102. Carey, J. N. & Goulian, M. A bacterial signaling system regulates noise to enable bet hedging. *Current Genetics* **65,** 65–70 (2019).

103. Patange, O., Schwall, C., Jones, M., Villava, C., Griffith, D. A., Phillips, A. & Locke, J. C. Escherichia coli can survive stress by noisy growth modulation. *Nature Communications* **9,** 5333 (2018).

104. Mitosch, K., Rieckh, G. & Bollenbach, T. Noisy Response to Antibiotic Stress Predicts Subsequent Single-Cell Survival in an Acidic Environment. *Cell Systems* **4,** 393–403 (2017).

105. Becskei, A., Kaufmann, B. B. & Van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics* **37,** 937–944 (2005).

106. Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422,** 633–7 (2003).

107. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* **98,** 8614–8619 (2001).

108. Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. *Molecular cell* **45,** 483–93 (2012).

109. Pedraza, J. M. Noise Propagation in Gene Networks. *Science* **307,** 1965–1969 (2005).

110. Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V. & Tans, S. J. Stochasticity of metabolism and growth at the single-cell level. *Nature* **514,** 376–379 (2014).

111. Tkačik, G. & Walczak, A. M. Information transmission in genetic regulatory networks: A review. *Journal of Physics Condensed Matter* **23,** 153102 (2011).

112. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467,** 174–178 (2010).

113. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 3581–6 (2005).

114. Hornung, G. & Barkai, N. Noise propagation and signaling sensitivity in biological networks: A role for positive feedback. *PLoS Computational Biology* **4,** 0055–0061 (2008).

115. Hooshangi, S. & Weiss, R. The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos* **16** (2006).

116. Bruggeman, F. J. & Teusink, B. Living with noise: On the propagation of noise from molecules to phenotype and fitness. *Current Opinion in Systems Biology* **8,** 144–150 (2018).

117. Munsky, B., Trinh, B. & Khammash, M. Listening to the noise: Random fluctuations reveal gene network parameters. *Molecular Systems Biology* **5,** 318 (2009).

118. Kittisopikul, M. & Suel, G. M. Biological role of noise encoded in a genetic network motif. *Proceedings of the National Academy of Sciences* **107,** 13300–13305 (2010).

119. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Stochastic Switching of Cell Fate in Microbes. *Annual Review of Microbiology* **69,** 381–403 (2015).

120. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330,** 1099–102 (2010).

121. Klumpp, S. & Hwa, T. Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Current opinion in biotechnology* **28,** 96–102 (2014).

122. Galbusera, L., Bellement-Theroue, G., Urchueguia, A., Julou, T. & van Nimwegen, E. Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. *bioRxiv* (2019).

123. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Research* **18,** 1084–1091 (2008).

124. Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461,** 193–198 (2009).

125. Shahrezaei, V. & Marguerat, S. Connecting growth with gene expression: of noise and numbers. *Current Opinion in Microbiology* **25,** 127–135 (2015).

126. Keren, L., van Dijk, D., Weingarten-Gabbay, S., Davidi, D., Jona, G., Weinberger, A., Milo, R. & Segal, E. Noise in gene expression is coupled to growth rate. *Genome Research* **25,** 1893–1902 (2015).

127. Ackermann, M., Stecher, B., Freed, N. E., Songhet, P., Hardt, W.-D. & Doebeli, M. Self-destructive cooperation mediated by phenotypic noise. *Nature* **454,** 987–990 (2008).

128. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G. & Alon, U. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods* **3,** 623–628 (2006).

129. Freed, N. E., Silander, O. K., Stecher, B., Böhm, A., Hardt, W.-D. & Ackermann, M. A simple screen to identify promoters conferring high levels of phenotypic noise. *PLoS genetics* **4,** e1000307 (2008).

130. Schaechter, M., MaalOe, O. & Kjeldgaard, N. O. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *Journal of General Microbiology* **19,** 592–606 (1958).

131. Sánchez, Á. & Kondev, J. Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences* **105,** 5081–5086 (2008).

132. Giese, K. C., Michalowski, C. B. & Little, J. W. RecA-Dependent Cleavage of LexA Dimers. *Journal of Molecular Biology* **377,** 148–161 (2008).

133. Phillips, I., Culebras, E., Moreno, F. & Baquero, F. Induction of the SOS response by new 4-quinolones. *Journal of Antimicrobial Chemotherapy* **20,** 631–638 (1987).

134. Salmon, K. A., Hung, S. P., Steffen, N. R., Krupp, R., Baldi, P., Hatfield, G. W. & Gunsalus, R. P. Global gene expression profiling in Escherichia coli K12: Effects of oxygen availability and ArcA. *Journal of Biological Chemistry* **280,** 15084–15096 (2005).

135. Arense, P., Bernal, V., Iborra, J. L. & Cánovas, M. Metabolic adaptation of Escherichia coli to long-term exposure to salt stress. *Process Biochemistry* **45,** 1459–1467 (2010).

136. Guttenplan, S. B. & Kearns, D. B. Regulation of flagellar motility during biofilm formation. *FEMS Microbiology Reviews* **37,** 849–871 (2013).

137. Amsler, C. D., Cho, M. & Matsumura, P. Multiple factors underlying the maximum motility of Escherichia coli as cultures enter post-exponential growth. *Journal of Bacteriology* **175,** 6238–6244 (1993).

138. Sernova, N. V. & Gelfand, M. S. Comparative Genomics of CytR, an Unusual Member of the LacI Family of Transcription Factors. *PLoS ONE* **7,** e44194 (2012).

139. Kram, K. E., Geiger, C., Ismail, W. M., Lee, H., Tang, H., Foster, P. L. & Finkel, S. E. Adaptation of Escherichia coli to Long-Term Serial Passage in Complex Medium: Evidence of Parallel Evolution. *mSystems* **2,** 00192–16 (2017).

140. Vassinova, N. & Kozyrev, D. A method for direct cloning of Fur-regulated genes: Identification of seven new Fur-regulated loci in Escherichia coli. *Microbiology* **146,** 3171–3182 (2000).

141. Dorman, C. J. H-NS: a universal regulator for a dynamic genome. *Nature Reviews Microbiology* **2,** 391–400 (2004).

142. Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A. & Takahashi, H. Heterogeneity of the principal sigma factor in Escherichia coli: the rpoS gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase Escherichia coli. *Proceedings of the National Academy of Sciences* **90,** 3511–3515 (1993).

143. Landini, P., Egli, T., Wolf, J. & Lacour, S. sigmaS, a major player in the response to environmental stresses in Escherichia coli: Role, regulation and mechanisms of promoter recognition. *Environmental Microbiology Reports* **6,** 1–13 (2014).

144. Dong, T. & Schellhorn, H. E. Control of RpoS in global gene expression of Escherichia coli in minimal media. *Molecular Genetics and Genomics* **281,** 19–33 (2009).

145. Guido, N. J., Lee, P., Wang, X., Elston, T. C. & Collins, J. A Pathway and Genetic Factors Contributing to Elevated Gene Expression Noise in Stationary Phase. *Biophysical Journal* **93,** L55–L57 (2007).

146. Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T. & Gerdes, K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Reviews Microbiology* **13,** 298–309 (2015).

147. Kvint, K., Farewell, A. & Nyström, T. RpoS-dependent promoters require guanosine tetraphosphate for induction even in the presence of high levels of σ(s). *Journal of Biological Chemistry* **275,** 14795–14798 (2000).

148. You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y. P., Lenz, P., Yan, D. & Hwa, T. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500,** 301–306 (2013).

149. Santos-Beneit, F. The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology* **6** (2015).

150. Tucker, D. L., Tucker, N., Ma, Z., Foster, J. W., Miranda, R. L., Cohen, P. S. & Conway, T. Genes of the GadX-GadW regulon in Escherichia coli. *Journal of Bacteriology* **185,** 3190–3201 (2003).

151. Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R. & Hwa, T. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature* **528,** 99–104 (2015).

152. Kleman, G. L. & Strohl, W. R. Acetate metabolism by Escherichia coli in high-cell-density fermentation. *Applied and Environmental Microbiology* **60,** 3952–3958 (1994).

153. Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences* **102,** 14338–14343 (2005).

154. Drummond, D. A., Raval, A. & Wilke, C. O. A single determinant dominates the rate of yeast protein evolution. *Molecular Biology and Evolution* **23,** 327–337 (2006).

155. Drummond, D. A. & Wilke, C. O. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* **134,** 341–352 (2008).

156. Koonin, E. V. Are there laws of genome evolution? *PLoS Computational Biology* **7,** e1002173 (2011).

157. Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S. P., Hengartner, M. O. & von Mering, C. PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life. *Molecular & Cellular Proteomics* **11,** 492–500 (2012).

158. Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A. & Jun, S. Robust growth of escherichia coli. *Current Biology* **20,** 1099–1103 (2010).

159. Kaiser, M., Jug, F., Julou, T., Deshpande, S., Pfohl, T., Silander, O. K., Myers, G. & van Nimwegen, E. Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software. *Nature communications* **9,** 212 (2018).

160. Sánchez, Á. & Kondev, J. Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences* **105,** 5081–5086 (2008).

161. Zong, C., So, L. H., Sepúlveda, L. A., Skinner, S. O. & Golding, I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Molecular Systems Biology* **6,** 440 (2010).

162. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional Pulsing of a Developmental Gene. *Current Biology* **16,** 1018–1025 (2006).

163. Chong, S., Chen, C., Ge, H. & Xie, X. S. S. Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158,** 314–326 (2014).

164. Ha, T., Cole, J. A., Luthey-Schulten, Z. A., Fei, J. & Peterson, J. R. Effects of DNA replication on mRNA noise. *Proceedings of the National Academy of Sciences* **112,** 15886–15891 (2015).

165. Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V. & Tans, S. J. Stochasticity of metabolism and growth at the single-cell level. eng. *Nature* **514,** 376–379 (2014).

166. Jones, D. & Elf, J. Bursting onto the scene? Exploring stochastic mRNA production in bacteria. *Current Opinion in Microbiology* **45,** 124–130 (2018).

167. Adan, A., Alizada, G., Kiraz, Y., Baran, Y. & Nalbant, A. Flow cytometry: basic principles and applications. *Critical reviews in biotechnology* **37,** 163–176 (2017).

168. Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D. & Stuurman, N. Advanced methods of microscope control using μManager software. *Journal of biological methods* **1,** 10 (2014).

169. Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., Macdonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W. & Swedlow, J. R. Metadata matters: access to image data in the real world. *The Journal of cell biology* **189,** 777–82 (2010).

170. Wolf, L. *Evolution of transcriptional regulation in Escherichia coli* PhD thesis (University of Basel, 2014), 145.

171. Peterson, J. & Phillips, G. J. New pSC101-derivative cloning vectors with elevated copy numbers. *Plasmid* **59,** 193–201 (2008).

172. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer Control of Microscopes Using µManager. *Current Protocols in Molecular Biology,* 1–14 (2010).

173. Doerr, A. Single-cell proteomics. *Nature Methods* **16,** 20 (2019).

174. Skinner, S. O., Sepúlveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nature Protocols* **8,** 1100–1113 (2013).

Acknowledgments

'There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.'

– J.R.R. Tolkien, *The Hobbit*

I like Tolkien's quote, because it symbolizes well the beauty of embarking in new projects, and of always keeping an eye open. Many times we start a new journey having certain expectations of the output we will obtain from it, however in the majority of cases we end up discovering things and taking unexpected turns we had never even contemplated or imagine we would. This is of course not different during a PhD project. When I started this journey, I never thought that I would end up being the scientist and person I have become today, and part of it is due to all the people that have accompany me through the process.

I would like to start by thanking my both co-supervisors Erik van Nimwegen and Thomas Julou for their constant support and mentorship throughout these years. It has not always been an easy journey for me, and both managed to always find ways to encourage and motivate me to keep progressing on my work. I will always be grateful to Erik, for giving me the chance to join his group and the opportunity to learn advanced quantitative techniques. He has always been extremely generous in allowing me to go to conferences inside and outside Switzerland, which helped me broaden my horizons as a scientist. Thomas's daily guidance and discussions helped me keep in track. I will always be grateful to him for teaching me the importance of research integrity and good data analysis practices, specially for introducing me to the 'tidyverse-world' and R, in general, without which I would have had a much harder time analyzing my data.

I would also like to thank Olin Silander, for supporting me during the start of my PhD, Urs Jenal and Gasper Tkačik for accepting being part of my PhD Committee Thesis and for their useful comments. Thanks to Janine Bögli for her never-ending assistance with the flow cytometer, which, in its high-throughput mode, did not always behaved as smoothly as we would have liked to. Thanks to Guillaume Witz for being such an amazing collaborator and colleague. For always finding time to discuss with me and implementing such great computational tool for image analysis, a vast majority of my work would not have been possible without you.

Special thanks to Luca Galbusera, for your patience and for always being prone to answer all 'stupid' questions I may have had about data analysis, flow cytometry, and everything in general. Not only for your support as a work-colleague, but also as a friend. Thanks to Gwendoline Bellement for being my 'PhD sister' during all these years, for all your support, for all the good, and, sometimes, crazy times together. I am extremely happy that I could share these years with you. Also special thanks to Athos, Anne and Mathias for being such good friends and colleagues and for all the good times together inside and outside the lab. And, of course, thanks to all current and former members of the van Nimwegen lab, working with all of you has been one of the best parts of doing a PhD. Although not always easy, I have all in all immensely enjoyed being able to work with people with such different backgrounds as mine. I strongly believe that it is in the interplay between disciplines where great science happens and I could not have been in a better group in this sense.

Of course, any of this would have been possible without the support of my family and friends:

Gracias a mis padres, por vuestro apoyo y constante confianza. Gracias por estar ahí cuando lo he necesitado y por darme libertad para hacer y ser quien he querido ser, pero a la vez por darme raíces firmes que me recuerdan de dónde vengo y las cosas que realmente importan. Gracias, mamá, por tu generosidad, por todos tus sacrificios para darnos las mejores oportunidades, pero, sobre todo, por estar siempre ahí, en lo bueno y en lo malo. Para mí eres mi mayor ejemplo a seguir, por tu fortaleza, alegría, por ser un ejemplo de superación y por ser capaz de sacar y resolver todo lo que se te ponga por delante. Gracias, papá, por haberme transmitido desde pequeña la pasión por la ciencia, el valor del conocimiento y el pensamiento crítico. Gracias por tu paciencia, por estar siempre ahí para escucharme y guiarme, por enseñarme que el trabajo tiene recompensa y por ser otro gran ejemplo de

superación. Gracias a ti soy científica y, ahora, una doctora más en la familia. Gracias a mis hermanas por los buenos momentos juntas, las risas y por saber que pase lo que pase siempre nos tendremos las unas a las otras. Por supuesto, gracias a todos los demás miembros de mi gran familia, me siento tremendamente afortunada y feliz de poder contar con todos vosotros y vuestro apoyo. Por último, me gustaría agradecer a todas las personas que, de una manera u otra, han estado junto a mí durante este viaje. Gracias sobre todo a Ismael, por ser mi mejor compañero y por ser capaz de hacerme ver siempre lo bueno durante los momentos más duros de la tesis.

Gracias a todos por haber estado conmigo, aunque la distancia no siempre haya sido fácil, vuestro apoyo ha logrado traspasar fronteras y ha sido fundamental durante estos años de aprendizaje tan importantes e intensos para mí.