Dynamics of gene expression in the genotype-phenotype map

Doctoral Thesis presented by

Djordje Bajić

Facultad de Ciencias
February 2016

Dynamics of gene expression in the genotype-phenotype map

Abstract

Genetic and environmental components can combine in quite complex ways to determine the phenotype of living organisms. Broadly, the goal of this thesis is to understand some of the design principles and constraints driving this assembly. We first study how genetic interaction networks – composed of phenotypically relevant interactions between genes – change in response to perturbation in their elements. Such networks at the genome-scale are progressively contributing to map the molecular circuitry that determines cellular behaviour. To what extent this mapping changes in response to different environmental or genetic conditions is however largely unknown. In Chapter 1 we assembled a genetic network using an *in silico* model of yeast metabolism to explicitly ask how separate genetic backgrounds alter the overall structure of the network. Backgrounds defined by single deletions induce particularly strong rewiring when the deletion corresponds to a catabolic or central metabolic gene, evidencing compensatory versatility. We found as well that weak interactions and those linking functionally separate genes tend to be more unstable. Overall, these patterns reflect the distributed robustness of core metabolic pathways. We examined as well a second class of evolutionary-motivated background, defined as a neutral mutation accumulation. The observed genetic network instability (predominantly in negative interactions) together with an increase in essential genes reflects a global reduction in buffering. Notably, rewiring of the genetic network is associated as well to a diminished environmental plasticity, what emphasizes a mechanistic integration of genetic and environmental buffering. More generally, this work demonstrates how the specific mechanistic causes of robustness influence the architecture of

multiconditional genetic interaction maps.

In Chapters 2, 3 and 4, we shift to systems that regulate the expression of genes. The plastic expression of different phenotypes enables organisms to respond to a wide variety of environmental changes, adapting their homeostasis. The dynamics of this plasticity can gen particularly interesting when operating mechanisms involve feedback, for instance when a gene encodes its own activator or repressor. The integration of positive and negative feedbacks can establish intricate patterns such as multistability, pulsing or oscillations. This depends on the specific characteristics of each interlinked feedback. In Chapter 2, we investigate a circuit associated with a dual, positive and negative transcriptional autoregulatory motif derived from the multiple antibiotic resistance system (*mar*) of *Escherichia coli*. Our results show that this motif enhances response speedup when it incorporates a linear positive feedback. Linearity also anticipates a homogeneous population phenotype anda higher input sensitivity, which we corroborate experimentally. As the motif is embedded in a broader regulatory network, we also studied how the system integrates additional cross-talks. Notably, the presence of an accessory positive regulation scales the response so that the circuit becomes unresponsive to other (metabolic) stress signals. Overall, we found that an antagonistic autoregulatory motif genetically encoded as a bicistron represents a versatile stimulus-response mode of control through the action of the positive-feedback regulation.

Beyond precise and specific regulatory systems such as *mar*, in Chapter 3 we explore the possibility that more broad and "stereotypic" expression programs also exist. We firstly analyzed a genome-scale expression dataset comprising single gene deletions in 25% of *Saccharomyces cerevisiae* genes. Our analyses suggest that tens of broad expression programs exist that explain more variation in this dataset than expected at random. We further find that

these programs seem to be activated also in conditions different to gene deletion, such as environmental perturbation or upon experimental compensatory evolution. These results suggest the possibility that broad, unspecific, "educated guess" gene expression responses have evolved as an adaptation to uncertain environments.

Finally, in Chapter 4, we focus on a phenomenon by which the ability of expression change (plasticity) appears coupled to uncontrolled, stochastic expression variation (noise). This coupling can constrain gene function and limit adaptation. We examine the factors that contribute at the molecular level to modulate this coupling. Both transcription re-initiation and strong chromatin regulation are generally associated to coupling. Alternatively we show that strong regulation can lead to plasticity without noise. The nature of this regulation is also relevant, with plastic but noiseless genes subjected to broad expression activation whereas plastic and noisy genes experience targeted repression. This differential action is particularly illustrated in how histones influence these genes. The cost of coupling plasticity to noise seems to be then compensated by a wider regulatory versatility. Contrarily, in genes with low plasticity, translational efficiency is the main determinant of noise, a pattern we found linked to gene length. Genome architecture (particularly, neighboring genes) appear then as a modifier only effective in highly plastic genes. In this class, we confirm bipromoters as a architecture capable to reduce coupling (by reducing noise) but also highlight its limitation (as they could also decrease plasticity). This presents ultimately a paradox between intergenic distances and modulation, with short intergenic distances both associated and disassociated to noise at different plasticity levels. In summary, balancing the coupling among different types of expression variability appears as a potential shaping force of genome architecture and regulation.

Dynamics of gene expression in the genotype-phenotype map

Resumen

El fenotipo de los organismos vivos es el resultado de una compleja combinación de componentes genéticos y ambientales. Desde un punto de vista general, esta tesis tiene como objetivo tratar de entender algunos de los principios de diseño y limitaciones de tiene este ensamblaje. En el primero de los trabajos presentados se estudia cómo las redes de interacción genética (compuestas de interacciones fenotípicamente relevantes entre genes) cambian en respuesta a perturbaciones en algunos de sus elementos. Este tipo de redes a escala genómica están contribuyendo de manera creciente a mapear los circuitos moleculares que determinan el comportamiento celular. Hasta qué punto este "mapa" cambia en respuesta a diferentes perturbaciones genéticas o ambientales? Tratando de responder a esta pregunta, en el Capítulo 1 hemos ensamblado este tipo de redes en de manera sistemática diferentes *fondos genéticos* usando un modelo *in silico* del metabolismo de la levadura. Los fondos genéticos correspondientes a enzimas del catabolismo o metabolismo central indujeron una reorganización de la red particularmente fuerte, indicando una versatilidad en los mecanismos de compensación. Asímismo, las interacciones más débiles y aquellas entre genes funcionalmente distantes aparecen como las más inestables. Estos patrones reflejan la robustez distribuída de las rutas catabólicas y del metabolismo central. Por otro lado, también hemos examinado un tipo de fondo genético evolutivamente motivado, definido por la acumulación sucesiva de deleciones neutrales. La inestabilidad observada (predominantemente en interacciones negativas), junto con un incremento en el número de genes esenciales, refleja una reducción global en los mecanismos de compensación. De manera particularmente interesante, hemos obser-

vado que la reorganización de la red genética está asociada a una reducción en la plasticidad ambiental. Esto pone de manifiesto que los mecanismos que subyacen a la robustez genética y a la ambiental son esencialmente los mismos. De manera más general, este trabajo muestra cómo los mecanismos específicos de robustez afectan la arquitectura multi-condicional de los mapas de interacción genética.

En los capítulos 2, 3, y 4, estudiamos diferentes aspectos de los sistemas que regulan la expresión de los genes. La expresión plástica de diferentes fenotipos hace posible que los organismos puedan responder a un amplio rango de cambios ambientales, adaptando su homeostasis a éstos. Las dinámicas específicas de esta plasticidad son particularmente interesantes cuando el mecanismo implica retroalimentación; por ejemplo, cuando un gen codifica su propio activador o represor. La integración de auto-regulaciones positivas y negativas puede establecer complejos patrones fenotípicos, como multiestabilidad, pulsos de actividad o oscilaciones. Esto depende de las características específicas de cada uno de los sistemas de retroalimentación implicados. En el Capítulo 2, estudiamos un motivo que contiene tanto una autoregulación positiva como una negativa, usando como modelo el operón de resistencia múltiple a antibióticos (*mar*) de *Escherichia coli*. Nuestros resultados demuestran que eeste sistema acelera la respuesta al incorporar una retroalimentación positiva lineal. Se demuestra experimentalmente que esta linealidad también produce una respuesta homogénea en la población y una alta sensibilidad. Por otro lado, también estudiamos cómo se integra este "motivo" en la red de regulación mayor. En este sentido, observamos que la presencia de una autoregulación positiva adicional es capaz de desacoplar el sistema de señales metabólicas. Finalmente, examinamos la influencia de posibles arquitecturas alternativas, mostrando cómo codificar la autoregulación dual antagonística en forma de bi-cistrón representa un versátil

sistema estímulo-respuesta.

Además de sistemas regulatorios específicos y precisos como *mar*, en el Capítulo 3 exploramos la posible existencia adicional de sistemas regulatorios "estereotípicos", más generales e inespecíficos. Para ello, analizamos en primer lugar un conjunto de datos experimentales en los que la expresión génica a escala genómica fue medida para deleciones en un único gen, que engloba un 25% de los genes de *Saccharomyces cerevisiae*. Nuestros análisis sugieren que existen decenas de programas globales e inespecíficos. Además, encontramos evidencia de que estos mismos programas también pueden encontrarse en otros tipos de perturbaciones, como las ambientales y tras evolución experimental compensatoria. Estos resultados indican la posibilidad de una respuesta global e inespecifica como potencial estrategia adaptativa en un ambiente incierto.

Finalmente, en el Capítulo 4, transladamos nuestra atención al fenómeno por el que la capacidad de un gen de cambiar su expresión génica en respuesta a cambios ambientales (plasticidad) se correlaciona con una variabilidad incontrolada y estocástica (ruido). Este acoplamiento puede limitar la función génica y la adaptación. Examinamos por tanto los factores a nivel molecular que pueden contribuír a su modulación. Tanto la re-iniciación transcripcional como la regulación a nivel de cromatina se presentan asociados a este acoplamiento. Alternativamente, demostramos cómo una regulación fuerte también puede ser ejercida sin incrementar el ruido. La naturaleza de esta regulación también es relevante; la plasticidad desacoplada del ruido se obtiene mediante mecanismos de activación generales. Mientras tanto, la regulación por represión específica está asociada a ruido, como pone también de manifiesto la influencia de las histonas. Nuestros resultados indican que el coste del ruido se ve compensado por una mayor versatilidad regulatoria. Por el contrario, en genes poco plásticos el

ruido viene determinado fundamentalmente por la eficiencia traduccional, un patrón que encontramos asociado a la longitud de los genes. En consecuencia, la arquitectura genómica (particularmente la influencia de genes vecinos) constituye un modificadorsólo en genes plásticos. En estos últimos, confirmamos que los promotores bi-direccionales pueden reducir el ruido, pero también reducen la plasticidad. Constituyen por tanto un mecanismo limitado para desacoplar plasticidad y ruido. En resumen, nuestros resultados sugieren que equilibrar diferentes tipos de variabilidad constituye potencialmente una fuerza modeladora de la arquitectura y regulación de los genomas.

Contents

Author List

The following authors contributed to Chapter 1: Clara Moreno Fenoll, Juan F. Poyatos
The following authors contributed to Chapter 2: Guillermo Rodrigo, Ignacio Elola, Juan F. Poyatos
The following authors contributed to Chapter 3: Balázs Papp, Juan F. Poyatos
The following authors contributed to Chapter 4: Juan F. Poyatos

Listing of figures

Listing of tables

0

Introduction

As we know it, life is the phenomenon that fixes the rules of the inert into complex sets of instructions, called genomes. Since it appeared in this form (under unclear circumstances), life has endured for a few billion years on earth. This remarkable stability is rooted in two basic forms of information flow. The first one is related to the information stored in a genome – the *genotype* – becomes explicit as a physico-chemical context defining a set of reactions and interactions. This ultimately results in the functions, traits and behaviors that conform the *phenotype*.

The second flow of information occurs in the opposite direction. It is related to how phe-

notypes influence the content of genomes in the evolutionary process. Evolution operates on populations, modifying the frequencies of the genotypes according to their phenotypic stability in a changing environmental and ecological setting. It is clear that these opposite directions of information flow are sides of the same coin, and neither can be fully understood without the other.

In this dissertation, we ask *how* the genetic and functional architecture of different biological systems (such as metabolism or gene expression) ultimately impacts the phenotype. When talking about architecture, we imply both the components of a system and the interactions (again, information fluxes) between them. Moreover, in all studied systems, we focus on two broad questions. Firstly, we try to understand the mechanism, i.e. the physical properties leading to the observed phenotype. Second, we try as well to understand the general principles and strategies that drive the evolution of such systems.

As defined by Wilhelm Johannsen in 1903, the genotype is "the sum total of all the *genes* contained in a gamete or in a zygote", where *gene* is "nothing but a very applicable little word useful as an expression of the unit-factors, elements or allelomorphs in the gametes"[112]. Note that the physical nature of "genes" was not known at the time, so the term "gene" was intended to be "useful" and "applicable" ("as to the nature of the genes, it is as yet of no value to propose any hypothesis, but that the notion *gene* covers a reality is evident from Mendelism"[112]). However, the "unit factor" notion falls short as a conceptual tool to understand the mapping of phenotypes to genotypes. For instance, Mendel had already observed by 1866 cases where multiple traits appeared under the control of a single "factor", being inherited in a correlated manner[149]. In 1910, Ludwig Plate coined the term "pleiotropy" to describe this phenomenon (from Greek πλείων/pleion, meaning "more", and τρόπος/tropos,

meaning "way")[166,204].

The conception of genes as independent "unit factors" – the so-called "Panglossian paradigm"[85] – is further challenged by the observation that they often combine in complex ways to determine a phenotype. This was first described by William Bateson in 1909[17], when he found deviations from the expected segregation ratios (based on the action of individual genes) in the outcome of a dihybrid cross. Bateson named this phenomenon *epistasis* (Greek επί/epi, meaning "upon", and στάσις/stasis, "stand"), as the effect of some mutations "stood above", or masked, the effect of others. In 1918, in one of his most influential articles, Ronald Fisher extended Mendelian inheritance to continuous, quantitative traits[66]. More importantly, Fisher's work placed inheritance in a population framework, laying the groundwork for rigorous study of evolutionary phenomena. The notion of epistasis used in Fisher's models differs from Bateson's: it means any statistical departure from the expected additive combination of two different loci in their effects on a phenotype. This can be written as $W_{xy} = w_x + w_y + \epsilon$, where W is the measured phenotype, w represents the individual effects of loci x and y, and ϵ represents the deviation due to epistasis.

The bottomline is that the genotype-phenotype relation is in no way linear or easily decomposable in univocal "unit factors". This fact not only makes many aspects of biological systems hardly understandable from a reductionist approach; it has also broad evolutionary consequences. For instance, pleiotropy leads to adaptive compromises between different traits, with broad evolutionary implications[228]. Pleiotropy also represents a central topic in the discussion about evolution of complexity, since Fisher proposed his geometric model[67]. It has also broad implications in evolution of diverse fundamental phenomena such as senescence[235], disease[48] or evolution of cooperation[69].

Evolutionary implications of epistasis are as well crucial – paraphrasing Richard Lewontin, "it is the context of genetic background effects that is so essential in the determination of the subsequent evolution of these populations"[132]. Epistasis between two genes (with respect to fitness) implies that these can not be optimized independently by adaptive forces. As first emphasized by Sewall Wright, this leads to multiple-peaked fitness landscapes, with potentially dramatic consequences on plausible evolutionary paths[238]. To what extent these consequences actually apply has been a long-standing question. Unlike the nice two or three-dimensional textbook simplifications, real fitness landscapes are highly multi-dimensional – and thus difficult to comprehend and analyze[165]. Notwithstanding these issues, epistasis has been proposed as an explanation for key evolutionary phenomena. For instance, speciation can be caused by reproductive isolation of sub-populations evolving towards different fitness peaks[3]. Recently, several studies have found strong empirical evidence that epistasis – not between genes, but residues inside a protein – has been a central factor in the evolution of proteins[159,168,27].

Moreover, in the first half of the XX century, geneticists principally focused on the genotype-phenotype relations, without understanding of the underlying mechanisms. It was not until the late thirties that Beadle and Tatum pioneered the use of genetics for functional inference. This genetic approach established as one of the most successful in the "molecular revolution" of the forthcoming decades. Somewhat ironically – because the dominant reductionist approach commonly neglected the context dependence of gene function. Much later, in the eighties, interest in epistasis grew again, but most often as a tool to understand quite intertwined regulatory hierarchies[103,8], such as those directing vulva formation in *Caenorhabditis elegans*[206].

The main limitation for this approach was scalability, as the number of genotypes to be generated grows quadratically with the number of genes. The problem was partially solved by the development of high throughput technologies, at least in some species. The first such method capitalized on the availability of comprehensive deletion libraries and easy pairwise combination of mutations by mating in *Saccharomyces cerevisiae*[220]. As a result, systematically generated and growingly exhaustive genetic interaction networks started being assembled. This was accomplished mainly in yeast[220,221,203,46] with some notable exceptions, such as a RNA interference-based *C. elegans* genetic interaction network[126]. These genetic interaction networks recapitulated quite well current knowledge (e.g. clustering of functionally related genes), allowing for functional discovery by placing unknown genes within a context.

As genetic networks represent a sort of "functional landscape" of cells[46], interest also grew in understanding the molecular details of genetic interactions[125]. For instance, negative (aggravating, or synergistic) epistasis is associated to redundancy (e.g. duplicated genes or functionally equivalent pathways)[106,227]. Thus, it can be highly informative about the structure of compensatory mechanisms in the cell. Positive epistasis (alleviating, or masking, in the Bateson's sense) can arise between genes that belong to the same functional unit, such as a protein complex or a linear metabolic pathway[192,46,12]. But still, the molecular details are unknown for the most part of interactions identified in high throughput studies, and our ability to predict them is almost nonexistent.

The availability of genetic networks also fueled system-level inquires into the functional architecture of cells. For instance, genetic networks exhibit hubs – genes with a high number of interactions. There are some well understood examples, such as chaperons which assist folding of many other proteins[185,243], or chromatin regulators affecting the expression of many

genes[126,163]. But in most cases, it is unclear why some genes exhibit so many interactions. Hubs are nevertheless important, because they represent phenotypic "capacitors" – for instance, hubs have been observed to buffer environmental variation in yeast[130].

In addition, it has been observed that epistasis is very frequent between distinct functional modules[221,46]. Furthermore, genetic networks exhibit "monochromaticity", a property by which genes can be organized in modules in such a way that all interactions between any two modules are of the same sign[193]. In other words, genetic interactions reflect also higher order functional links, above the individual gene level. These studies also nourished from complementary *in silico* approaches providing more or less realistic models, which can greatly help to conceptualize and understand empirically observed patterns[210]. For example, although most of the studies focus on epistasis relative to growth rate (for both methodological and evolutionary reasons), this can be easily extended using in silico modeling to a bigger, systematically generated phenotypic space.

To what extent are genetic interactions themselves context-dependent? For both pragmatism and rigour, modern analyses of genetic interactions usually consider them against a fixed reference background. But in real populations, genetic background differs from individual to individual. How stable are genetic networks to these variations?[87] This question has potential implications in both basic and applied genetics. If small variations in the genetic context lead to different epistatic – and thus also fitness – landscapes, this could mean that different evolutionary pathways are available to different individuals in a population. From a more applied point of view, the issue becomes especially important considering the prevalence of epistasis in genetic determinants of complex disease[47,152,45,30].

However, this issue has not been addressed systematically to date. One small-scale study

in Drosophila showed that indeed, genetic network rewiring could appear as consequence of different genetic backgrounds[88]. Also, several high-throughput experiments have been carried out that suggest significant rewiring, although in response to environmental or evolutionary change, rather than genetic[98,11,186,94]. The potential for rewiring has been also shown in yeast in response DNA damage in *S. cerevisiae*[11]. Rewiring was specifically observed in functionally relevant modules, but other interactions (e.g. among housekeeping genes) appeared conserved.

In the first chapter of this thesis, we systematically analyze the rewiring of genetic networks in response to modification of genetic background. For this, we used an in silico modeling framework known as *constraint based modeling*[170,68,158]. These models optimize the fluxes through a reconstructed metabolic network to maximize an objective output (in our case, this was biomass production, as a proxy for fitness). The key advantage of this model is a fast and automated generation of genotypes, and computation of flux and growth rate predictions for each of these. This allows evaluation of genetic and metabolic effects simultaneously, enabling functional interpretation. We systematically examined two types of genetic backgrounds: inactivation of genes associated to active reactions, and neutral deletion accumulation lines. We ask: i) how ubiquitous is rewiring, and how it correlates to functional changes (e.g. metabolic fluxes or functional modules); ii) how rewiring upon perturbation is associated to the structure of the *wt* network; iii) how rewiring in response to neutral deletion accumulation (in a constant environment) predicts reduction in environmental plasticity.

The biochemical or structural functions specified by genes (and interactions among them) are in no way the sole determinant of the phenotype of an individual. Genomes also respond to the internal and external environment by means of expressing different sets of genes, ade-

quate for different conditions or situations. Biological functions able to sense environmental changes adjust the genomic output to maintain homeostasis are greatly valuable in an unpredictable environment.

The presence of such mechanisms becomes first apparent as non-random associations between environmental conditions and observed phenotype, in absence of genotypic differences. For example, Richard Woltereck used the term "reaktionsnorm" (norm of reaction) to describe differences in development in the *Daphnia* water flea, caused by environmental differences[236]. Later, during his thesis, Jacques Monod observed that when grown in both glucose and lactose, bacteria consumed the latter only after the former was exhausted[151]. Instead of "adaptation" (the term used by Karlström some years before to describe similar phenomena[96]), Monod and his colleagues preferred the term "induction". Crucially, they realized that certain genetic structures were able to interact with the environment, *inducing* the expression of the adequate enzymes. These investigations led him along with François Jacob to characterize in 1961 the *lac* operon, the first known *genetic inducible system*[109]. This was the first direct proof that molecular mechanisms modify the output of genomes depending on the external conditions in which it develops.

Furthermore, until the decade of 1960, genetics and biochemistry were fairly separated disciplines. Genetics was a theory-dominated, mostly statistical field, aiming to build models to explain phenotypic or evolutionary observations. On the other hand, biochemistry was mainly focused on understanding metabolic reactions and pathways experimentally, making use of thermodynamic and kinetic formalization. Although Beadle and Tatum had already anticipated that enzymes should arise from genes, no mechanistic link for this association was known. Therefore, Monod's work opened the path for the unification of genetics and

biochemistry in a single discipline, sometimes called molecular genetics. The expression of genes and their regulation (their "reaktionsnorm") could be for the first time characterized quantitatively in molecular detail[190,214].

Other forms of regulation, apart from the repressor-mediated model were also subsequently described, such as those based on activator (e.g. the CAP protein in the *lac* operon[52] or bi-regulator molecules (e.g.[197,64]). One of the key architectures that were recognized were *circuits*, in which one effect of a regulatory gene is to regulate its own expression[82]. Regulatory circuits imply feedback control, which bears profound consequences on the dynamical properties of the system[214]. For instance, positive feedbacks can lead to a delayed and heterogeneous output[138], and even to multistability and hysteresis[19,117]. Negative feedback, on the contrary, reduces cell to cell variation and enhances response speed[29,20].

The natural tendency was, however, to consider all these different regulatory architectures as special cases, result of the particular needs, evolutionary forces and constraints acting on the gene under study[213]. When the availability of enough data allowed so, questions emerged as well about the design principles of regulatory networks in a more global sense[3]. One of the most surprising results in this respect was the presence of specific "motifs" in regulatory networks[196]. As an analogy to well-known sequence motifs, network motifs represent patterns (of connectivity, in this case) that appear in different parts of the network at much higher frequencies than expected by random[196,5]. For instance, in the *Escherichia coli* network, three motifs appear over-represented: the "feed-forward loop", the "single-input" motif, and the "dense overlapping regulons"[5].

These findings prompted two basic questions. First, which are the functional properties added by these motifs? We already mentioned some of the features of simple feedback cir-

cuits; more complex motifs will arguably lead also to more elaborated information processing tasks. For example, feed-forward loops can function (depending on the structural type) as persistence detectors[139], pulse generators[15] or response accelerators[140]. Secondly, network motifs are not isolated, but embedded into the bigger regulatory network, which is manifest as additional links to other sub-systems[107]. How independent are the dynamic properties of a motif, and how is it modulated by cross-talks?

In the second chapter of the thesis, we address these questions in a regulatory circuit consisting of a dual autoregulation, involving both a positive and a negative feedback. What are the consequences of integrating these two feedbacks – with *a priori* contrasting properties – in a single architecture? To investigate this, we use the *Escherichia coli mar* operon[2] as an experimental model. We first characterize the dynamic behavior using standard kinetic modeling, and confirm the predictions experimentally (with both bulk and single-cell measures). We study the effect of some specific implementation features in the *mar* operon, such as the weakness of the positive feedback and the copper-mediated sensing. Using mathematically controlled comparison we analyze the effect of hypothetical alternative implementations, stressing the effects of the particular characteristics of our model system. Finally, we demonstrate how cross-talks (interference of metabolic signals) can be buffered by means of additional regulatory inputs.

Systems such as the above described, with (sets of) specific effectors (or "transcription factors") operating on genes to produce a fine-tuned phenotype, represent one of the main paradigms of gene expression regulation[171]. But additional regulation "layers" also exist. In bacteria, one example is the use of alternative σ factors[160]. Although a σ factor is an essential component of the prokaryotic transcriptional machinery, alternative σ factors can bias it

towards promoters with specific sequence motifs[18]. In this way, they enable "promoter selection" by polymerases, switching between frequently used global expression programs (e.g. between growth and stress/stationary phase[100]). In eukaryotes, similar promoter selection mechanisms have greatly diversified. The eukaryotic the pre-initiation complex (the polymerase plus the "basal transcription factors") counts tens of subunits, some of which have prominent roles in determining global expression programs in response to a variety of situations[86]. For example, while the basal factor `TFIID` is linked to housekeeping gene expression in *Saccharomyces cerevisiæ*, the alternative `SAGA` is associated to stress-response functions[104,38].

In addition, eukaryotic DNA *in vivo* appears wrapped around octameric units called nucleosomes, with a crucial regulatory role[215]. Firstly, nucleosomes present a physical barrier to transcription initiation[3]. Thus, they must be removed, or *remodeled* to allow transcription to proceed. As well, nucleosomes can be covalently modified in a wide variety of forms (acetylation, methylation, ubiquitination, etc.)[32]. The specific combinatorics of these modifications greatly impact the expression of genes, to the extent that a "histone code" has been proposed to govern the activity of eukaryotic genomes[3]. Moreover, regulatory effects of both gene-specific or general transcription factors are often elicited via chromatin-mediated mechanisms – to continue with the previous example, both `TFIID` and `SAGA` harbor histone acetylation activities[86], and many transcription factors activate expression by recruiting chromatin modifiers or remodelers[28,172].

In turn, this regulatory complexity has been proposed as the main source of evolution of *organismic complexity*[129]. This very term is somewhat vague and ambiguous. Without entering the debate here, we can attempt to operatively define it as the diversity of phenotypic responses a genome is potentially able to articulate. The most prominent example is per-

haps the sophisticated regulatory mechanisms at the basis of major developmental events in multicellular organisms, allowing highly contrasting phenotypes to be differentiated in cells sharing the same genome. As a result, *Drosophila melanogaster* develops much more elaborate cell and tissue types than *Caenorhabditis elegans*, despite having about six thousand less genes[129].

What are then the *design principles*[202] governing the evolution of such multiplicity of intertwined regulatory layers? One strategy is clearly to evolve well-defined regulatory mechanisms that lead to specific phenotypes, such as in the *lac* or in the *mar* systems. Evolution of this kind of mechanisms can take place only when the challenge is well-specified and frequently encountered during the evolutionary history of the organism, such as switching between common carbon sources, or facing "expected" stresses such as heat shock in *Escherichia coli*. But organisms face a very wide array of perturbations, many of which are unforeseen and/or unpredictable[133]. We could then hypothesize that regulatory evolution also enables some sort of unspecific, *stereotypical* responses, available to deal with "unknown" challenges. There is increasing evidence of such behavior. For example, one would expect that *Saccharomyces cerevisiæ* cells will respond to a given environmental insult by expressing the specific genes aimed to neutralize that particular stress. Instead, yeasts elicit a "environmental stress response" (ESR), containing responses to multiple heterogeneous stress conditions and encompassing a sizable fraction of the genome. In another striking example, when yeasts are submitted to an unforeseen challenge, unspecific and global gene expression reprogramming can as well be observed (although convincing explanation for this is still lacking).

In the third chapter of this thesis, we explore the existence of stereotypical gene expression responses by analyzing a genome-wide expression dataset consisting of single gene deletions

in about one quarter of *Saccharomyces cerevisiæ* genes. We first estimate that on the order of tens of global expression "signatures" explain significantly more variation than expected by chance. We then ask how general are these signatures, by trying to detect them in response to other, different perturbations, such as environmental or evolutionary. Our results suggest that yeast cells have indeed evolved "generic" modes of gene expression response. Although these responses seem functionally heterogeneous and rather unspecific, analyses suggest that at least some might be related to fundamental cross-talks between metabolism and gene expression.

Although all the above discussed regulatory mechanisms determine the expression level of genes, they do it in a probabilistic, rather than deterministic manner[189]. Gene expression is a random involving a small number of diffusing molecules, and is therefore inherently random. Ultimately, "noise" in gene expression translates in phenotypic heterogeneity across individuals of a population, even in absence of genetic or environmental differences.

It is being increasingly recognized that cells and populations can capitalize on this heterogeneity to produce meaningful outcomes[71,62]. For instance, heterogeneity in a trait potentially conferring an advantage can represent an effective bet hedging strategy[131]. In the context of circuits that display multiple metastable states, noise is crucial for switching between one or another. This is observed, for instance, in developmental contexts where cells of a tissue are required to randomly differentiate different phenotypes ("division of labor")[134,62]. In unicellular organism populations, this stochastic switching can even be more effective than sensing, when frequency of switching is similar to frequency of environmental change[118]. However, these potentially beneficial effects of noise seem to be linked to quite specific functional contexts. From a more general perspective, noise is considered as a rather detrimental feature that

limits the precision in genetic control of the phenotype[72].

Moreover, the process of gene expression is a complex one, involving several steps and many different effectors. The final noise level in protein amount – ultimately influencing the phenotype – does not depend equally on all these steps. Transcription, which involves the lowest number of molecules, is the primary responsible for most of the observed noise[23]. Particularly relevant are bursts in promoter activity[161]. These can be in turn "amplified" by means of an efficient mRNA translation, leading to high levels of noise at the protein level. Conversely, the combination of increased transcriptional but lower translational efficiency can reduce noise, while maintaining the average level of expression[175,176]. Both transcriptional and translational efficiency can be modified at the genetic level (respectively, by modifying promoter strength or use of codons with more or less abundant tRNAs, for example[24]). As well, we mentioned the property of regulatory feedback architectures as to modulate noise[5]. Therefore, noise can be controlled by genetically encoded variation in parameters or architectures involved in the expression of genes, and is in consequence subject to evolutionary tuning.

In eukaryotes, the transcriptional process incorporates several additional features that affect noisy expression[24,223]. First, eukaryotic promoters must undergo chromatin remodelling to become transcription-permissive. Second, the eukaryotic transcriptional machinery (the "pre-initiation complex", counting tens of subunits) must be sequentially assembled on the promoter. Finally, the pre-initiation complex must not be assembled from scratch for each transcribed mRNA molecule. Instead, it can "fire" several initiation events once assembled, a process known as "reinitiation"[55]. All these distinctive features can be as well subject of evolutionary modulation. For example, the presence of a TATA-box motif in a subset of yeast

promoters increases re-initiation and transcriptional bursting, and is therefore associated to higher noise[14]. Some genes exhibit as well a stable nucleosome at their promoter[215], what is also linked to noisy expression.

Moreover, these mechanisms not only contribute to noise, but are as well key for gene expression regulation. At least in some genes, this leads to an evolutionary trade-off between beneficial (controlled) and detrimental (uncontrolled) expression variability[124]. In the last chapter of this thesis, we exhaustively examine different molecular strategies and features that could uncouple noise from plasticity (i.e. regulated responsiveness to environmental perturbation). We firstly show how the elements involved in eukaryotic transcription can combine in different ways, leading to different degrees of coupling. We also examine how this integrates in the context of the general expression strategies observed in yeast (e.g. growth versus stress). Finally, we study also the influence of genomic neighborhood.

1

Rewiring of genetic networks in response to the modification of genetic background

Gene action is commonly determined by its interactions with other genes. This includes genes known to be associated to the action under study, but also those whose association is less expected or their biochemical properties still unknown. Both classes of interactions can now be effectively mapped at a large scale by following two complementary strategies.

The first one relies on the progress of experimental tools to produce genetic perturbations in large numbers and to automatically quantify their effects[13]. The effect most commonly

quantified is growth rate (but see, for instance,[130,201]). These tools are providing the first global genetic landscapes of cells, e.g.,[181,46]. A second approach benefits from the advance of computational methods capable to predict phenotypes. Metabolic constraint-based models are particularly useful in this regard, as they incorporate genomic information (of metabolism) into an *in silico* framework that can estimate cell growth under specific conditions[177]. Notably, flux balance predictions have been confirmed experimentally, e.g., [200]. Single mutant fitness and their corresponding genetic interactions can also be produced in this framework[210].

These strategies are currently being combined to better interpret the molecular underpinnings of genetic interactions (i.e., epistasis), both negative and positive. Negative epistasis is observed when the fitness defect of a double mutant is lower than that expected from single mutant values, and reflects functional redundancy. For instance, the presence of negative epistasis between the urmylation pathway and the elongator complex in yeast suggested that both jointly modify certain transfer RNAs[46]). In a metabolic context, negative epistasis typically indicates buffering between metabolic routes that are able to compensate for the inactivity of each other, e.g. synthesizing the same component[164]. Positive genetic interactions, in contrast, are commonly observed between genes that constitute a multi–protein complex or metabolic pathway, i.e., genes being part of the same functional unit[203]: a mutation in one of its constituents can inactivate this unit, reducing the effect of other perturbations in additional components.

Large scale approaches lead as well to the identification of system-level patterns, when the interactions are represented as genetic networks. For instance, the network presentation of high-throughput data of *Caenorhabditis elegans* and *Saccharomyces cerevisiae* clearly identified the presence of genetic hubs, that are mostly associated to chromatin regulation[3,46].

Another feature, revealed for the first time with flux balance modeling, is monochromaticity; the specific distribution of epistasis types in the interactions within/between functional modules[193]. This characteristic was later confirmed by metabolic experiments[210] and high-throughput data[46], in which a specific distribution of epistasis strengths was additionally identified[169].

All previous properties implicitly suggest genetic networks as a stable configuration, a view partially influenced by the constant conditions in which interactions were examined. However, recent studies are emphasizing that this stability should not be necessarily the case. Genetic interactions and, more broadly, genetic networks were shown to change depending on the particular context where fitness is evaluated[241,203,98,11,94]. Rewiring is further confirmed by means of comparative analysis across organisms[181,56,73,186]. Moreover, the "instability" of these networks should not come as a surprise; earlier works already discussed the influence of context (environmental or genetic) on the phenotypic effect of mutations and their interactions[35,36], a phenomenon that can directly influence evolutionary dynamics[???]. To what extent genetic networks are context-dependent is nevertheless mostly unknown.

Here we ask how the structure of a genetic network reorganizes in response to changes in genetic background. To this aim, we mapped genetic interactions between metabolic genes by using a computational model of metabolism in *Saccharomyces cerevisiae*. We consider two broad (genetic) background classes. The first class corresponds to single gene deletions of each of the enzymes that are active (i.e., showed nonzero flux) in the wild-type (WT). The rewiring patterns found stress the different organization of biosynthetic and catabolic routes and how this impacts their capacity to compensate change. A second class presents neutral backgrounds, i.e. trajectories of accumulated neutral deletions, resembling evolutionary

genome reduction processes. This highlights how cryptic variability modifies buffering in genetic networks, and how the new network structures also reflect in different levels of environmental plasticity. We additionally corroborate some of these patterns with inspection of experimental data.

1.1 The model genetic network and types of interactions

We considered the `iND750`[59] genome-scale model as the *wt* genotype throughout the analyses in this chapter. This model incorporates all the necessary complexity of *Saccharomyces cerevisiae*'s metabolism (e.g., it is fully compartmentalized), and the validity of its predictions have been experimentally corroborated. Using this model, we computed the fitness (i.e. growth rate) for all the single and double deletion mutants. Then, a genetic interaction between two genes is defined as the difference between the observed double mutant effect and the expected one according to the single mutant effects. According to the multiplicative model[226], $\epsilon = W_{AB} - W_A W_B$.

We applied a scaling method that fits all epistasis values between -1 and 1, revealing a characteristic tri-modal distribution[193] (fig 1.1). In addition to the classical sub-division of genetic interactions into positive and negative, this scaling reveals an additional sub-division of interactions into weak and strong ones. Strong negative (or synthetic lethal, SL) interactions, $\widetilde{\epsilon} = -1$, involve two unique alternatives for an essential function. Strong positive (SP) ones, $\widetilde{\epsilon} = 1$, imply an absolute functional dependence, like the one found in genes that act sequentially in a parallel pathway, or sub-units of the same protein complex. Fig 1.1B represents schematically the kind of architecture underlying both SN and SP interactions.

Moreover, weak interactions reveal qualitatively different phenomena. Weak negative in-

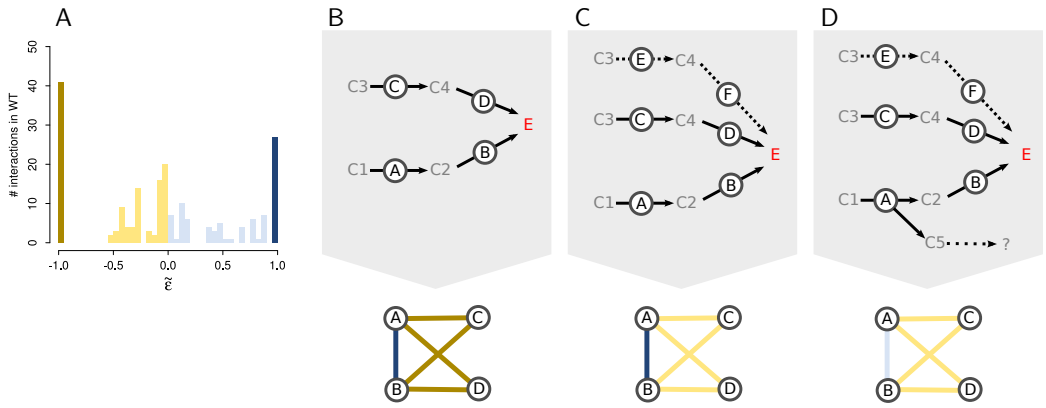


Figure 1.1: Types of epistasis and underlying metabolic architectures A) Definition of scaled epistasis [193] and distribution in the *S. cerevisiae* iND750 model network. Yellow and blue are used respectively for strong-negative and strong-positive interactions; for weak interactions, we use lighter hues. B) Classical example of strong negative epistasis (synthetic lethality) arising from two parallel pathways leading to the same essential compound E. If one of the pathways is more efficient than the other, genes within it will display also strong positive epistasis. C) The presence of an additional alternative transforms strong into weak negative interactions. D) If one of the genes in the most efficient pathway has additional functions (represented by the question mark), positive epistasis will be as well weak.

teractions (WN, $-1 < \widetilde{\epsilon} < -0.01$) appear when there exists an additional (less efficient) solution to the two main functional alternatives represented by the WN-interacting genes (fig 1.1C). The multiplicity of alternatives with different efficiency normally results from the presence of biochemically different ways to perform a given function (like synthesizing a compound). Weak positive interactions (WP, $0.01 < \widetilde{\epsilon} < 1$) emerge among two genes when their contribution is only partially joint (whereas for another part of their contribution they are independent). This could be interpreted as a form of "multi-functionality" (fig. 1.1D). Weak interactions arise then from rather complex metabolic architectures. As we will show, this has key implications for understanding how genetic networks rewire and change.

We first analyzed a dataset comprising the single deletion genotypes of all metabolically active nonessential genes from the *wt* model. These deletions normally result in the blockade of some reactions and the corresponding reconfiguration fluxes across the metabolic network[164,25]. How is readjustment of fluxes reflected in rewiring of the genetic network?

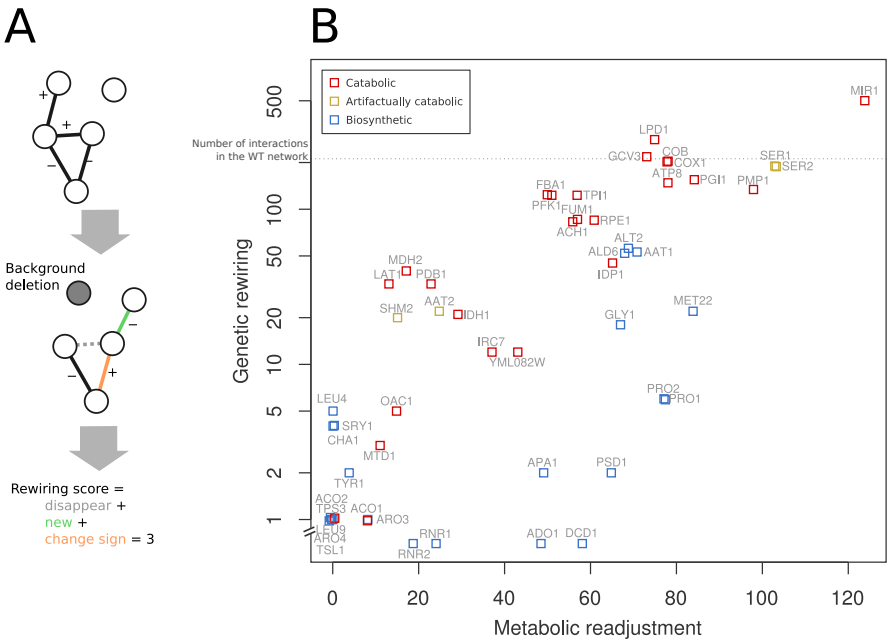


Figure 1.2: Genetic rewiring correlates with metabolic readjustment. A) Cartoon that illustrates the kind of alterations in the genetic network that contribute to the rewiring score: new interactions + interactions that changed sign + disappeared links (excluding those of the gene acting as background). (B) Association between genetic rewiring and metabolic readjustment (as number of reactions that modified their relative flux) in response to deletions in active genes, i.e., backgrounds; there exists 207 interactions in the WT network (dashed line), and 277 active fluxes in the corresponding metabolism (coordinates incorporate some noise to help visualization, and the y-axis logarithmic scale is broken to locate backgrounds with no genetic rewiring). Notably, catabolic genes exhibit much stronger rewiring than biosynthetic ones, regardless of the strength of metabolic readjustment. Note that some genes are artifactually catabolic in the model

To understand this correspondence, we quantified the number of modified interactions

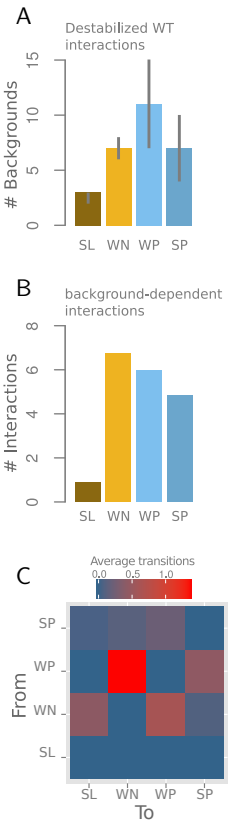
with respect to the WT genetic network as a simple rewiring score (Fig. ??A), and also counted the number of altered fluxes as measure of the underlying metabolic readjustment. Figure ??B shows the strong correlation between these two scores (Spearman $\rho = 0.72$, $p < 10^{-8}$). We subsequently partitioned the metabolic measure into a qualitative and quantitative component (i.e., active fluxes that become inactive or vice versa, and changes in relative flux level in reactions that remain active, respectively). Qualitative changes predict the strength of genetic rewiring (multiple linear regression $p < 10^{-8}$) while quantitative ones do not ($p = 0.97$). Thus, genetic rewiring denotes a redistribution of metabolic fluxes to alternative reactions and pathways.

Is rewiring associated to some particular pathways or functions? We noted that backgrounds corresponding to catabolic (or central-metabolic) functions display stronger rewiring than biosynthetic ones (Fig. ??B). (Note that, although some apparently biosynthetic genes such as `SER1`, `SER2` or `SHM2` also produce strong rewiring, the model artifactually predicts them a role in ATP generation). To analyze this more quantitatively, we focused on metabolites instead of genes or reactions. Depending on the available fluxes, a given metabolite can be produced in different amounts (of course, it is consumed at the same rate to satisfy mass balance). We devised a Ξ_m^{bg} score quantifying the change in the production of metabolite m observed in background bg respect to the wt (Methods). The change of most metabolites significantly correlates with genetic rewiring (table 1.1), consistently with the above shown association between rewiring and metabolic readjustment. However, it is best predicted by the variation metabolites associated to catabolic/central metabolic pathways (e.g., currency metabolites, glycolytic or TCA cycle intermediaries). This confirms that genetic rewiring is strongly associated to changes in central metabolism.

Metabolite	ρ	Metabolite	Spearman ρ	Metabolite	Spearman ρ
h[c]	0.89†	13dpg[c]	0.76‡	paps[c]	0.15
h[m]	0.88†	3pg[c]	0.76‡	so3[c]	0.15
nad[m]	0.86†	ficytc[m]	0.76‡	so4[c]	0.15
nadh[m]	0.86†	focytc[m]	0.76‡	so4[e]	0.15
nad[c]	0.85†	o2[m]	0.76‡	trdox[c]	0.15
nadh[c]	0.85†	o2[c]	0.76‡	trdrd[c]	0.15
nadp[c]	0.83†	o2[e]	0.76‡	ppi[c]	0.15
nadph[c]	0.83†	dhap[c]	0.76‡	2cpr5p[c]	0.13
pi[c]	0.82†	fdp[c]	0.76‡	3ig3p[c]	0.13
thf[c]	0.81†	glu_L[c]	0.76‡	anth[c]	0.13
co2[m]	0.81†	h2o[e]	0.76‡	pran[c]	0.13
mlthf[c]	0.8†	g6p[c]	0.76‡	utp[c]	-0.1
gly[m]	0.79†	glc_D[e]	0.76‡	25aics[c]	0.1
thf[m]	0.79†	f6p[c]	0.76‡	34hpp[c]	0.1
mlthf[m]	0.79†	pyr[m]	0.64‡	5aizc[c]	0.1
h2o[c]	0.79†	icit[m]	0.62‡	aicar[c]	0.1
pi[m]	0.79‡	1agly3p_SC[c]	0.62*	air[c]	0.1
for[c]	0.79‡	oaa[c]	0.6*	dcamp[c]	0.1
3php[c]	0.79‡	akg[m]	0.6*	fgam[c]	0.1
pser_L[c]	0.79‡	acald[c]	0.57*	fpram[c]	0.1
ser_L[c]	0.79‡	2pg[c]	0.57*	fprica[c]	0.1
10fthf[c]	0.78‡	pep[c]	0.57*	fum[c]	0.1
methf[c]	0.78‡	cit[m]	0.56*	gar[c]	0.1
nh4[m]	0.78‡	nadp[m]	0.56*	gln_L[c]	0.1
co2[c]	0.78‡	ac[c]	0.55*	imp[c]	0.1
nh4[c]	0.78‡	accoa[c]	0.55*	pi[e]	0.1
atp[m]	0.77‡	coa[c]	0.55*	pram[c]	0.1
akg[c]	0.77‡	pyr[c]	0.54*	ctp[c]	-0.09
gly[c]	0.77‡	coa[m]	0.54*	dudp[c]	-0.09
h2o[m]	0.77‡	accoa[m]	0.54*	dump[c]	-0.09
glc_D[c]	0.76‡	asp_L[c]	0.54*	prpp[c]	0.08
atp[c]	0.76‡	oaa[m]	0.53*	gdp[c]	-0.04
adp[c]	0.76‡	nadph[m]	0.51*	gtp[c]	-0.04
adp[m]	0.76‡	mal_L[c]	0.46*	g1p[c]	0.04
q6[m]	0.76‡	hco3[c]	0.43*	ps_SC[m]	0.04
q6h2[m]	0.76‡	asp_L[m]	0.43*	tre6p[c]	0.04
co2[e]	0.76‡	aps[c]	0.15	udp[c]	0.04
g3p[c]	0.76‡	pap[c]	0.15	udpg[c]	0.04

Table 1.1: Metabolites and genetic network rewiring Correlation between Ξ_m^{bg} and genetic rewiring for each metabolite. Metabolites are colored according to broad functional context where they appear. Red - catabolism; blue - biosynthesis; violet - currency metabolites and related molecules (e.g. H^+ or phosphate); yellow - metabolites belonging to artifactually catabolic pathways (see below), gray - other types of metabolites (such as water, inorganic nutrients, etc.). $^{\dagger}P < 10^{-20}$, $^{\ddagger}P < 10^{-10}$, $^{\star}P < 0.01$

We next asked how does the type of genetic interaction affect their propensity to rewire. Instability of each *wt* interaction was obtained by simply counting the number of backgrounds in which it changes. Although in general *wt* interactions tend to be conserved, weak and positive ones are significantly more unstable than strong and negative, respectively (fig. 1.3A). The same is observed for background-induced interactions, which tend to be weak and positive (fig. 1.3B). Weak interactions also exhibit a tendency to change their sign (fig. 1.3C) indicating frequent functional re-purposing.



D

Subset	Test group	Mean instability	*p*-value
Strong	Intra-group	2.7	2.3×10^{-5}
	Inter-group	5.1	
	Positive	7.1	2.1×10^{-9}
	Negative	2.4	
Weak	Intra-group	8.3	0.97
	Inter-group	8.7	
	Positive	10.9	1.8×10^{-7}
	Negative	7.1	
Intra-group	Positive	7.1	0.013
	Negative	4.2	
	Strong	2.7	9.7×10^{-8}
	Weak	8.3	
Inter-group	Positive	10.3	1.78×10^{-10}
	Negative	6.0	
	Strong	5.1	$1.9 \times 10{-}7$
	Weak	8.7	
Positive	Intra-group	7.1	0.014
	Inter-group	10.3	
	Strong	7.1	1.8×10^{-4}
	Weak	10.9	
Negative	Intra-group	4.2	0.002
	Inter-group	6.0	
	Strong	2.4	5.2×10^{-18}
	Weak	7.1	

Figure 1.3: Different types of genetic interaction exhibit different stability A) Instability of interactions as a function of sign and strength (median with upper and lower quartiles). Weak interactions are more unstable than strong ones of the same sign, and positive interactions are more unstable than negative. B) Average number of new interactions of each type by background. C) Average number of transitions between interaction types per background. D) Using different subsets of the data (first column) we computed the differential stability of different types of interactions (positive vs. negative, strong vs. weak and intra vs. inter-module. Mean and Wilcoxon test p-values are shown.

A look at the *wt* network shows that these unstable interactions (i.e. positive and/or weak)

form a densely interconnected "central" component of the network (fig 1.4A-E). In contrast, the most stable ones (synthetic lethal) are located at the periphery of this component, or even in completely separated clusters. We observed additionally that weak interactions typically connect genes from different metabolic modules (see Methods for definition), in comparison to strong ones (Fisher's test, $p < 0.0002$). As expected, inter-module interactions tend to be more unstable than intra-module, but this effect is independent of sign or strength (fig. 1.3D).

In fig. 1.4F, we explicitly show how these unstable interaction concentrate at catabolic / central metabolic modules. Most notably, they involve glycolysis, TCA cycle and oxidative phosphorylation, and also the glycine/serine module (remember its artifactual role in ATP synthesis). In rewired networks, new interactions also arise preferentially between modules (90% of cases, compared to 78% in the wt network, Fisher's test $p < 10^{-7}$). They also attach most frequently to catabolic modules; in contrast, they appear less frequently between biosynthetic modules, and in a more background-specific manner (fig. 1.4G).

Moreover, these observations also suggest that the response to background change is very much linked to the rewiring of genetic hubs. For instance, nodes with most of their connections weak (mean $|\widetilde{\epsilon}| < 0.5$) have a higher degree ($\sim 8$ connections on average), than nodes with mainly strong interactions (mean $|\widetilde{\epsilon}| > 0.5$), which present ~ 3 (Wilcoxon-test $p = 0.003$). In fig. 1.5A, we show in detail an example of the rewiring of a hub (PFK1) in different backgrounds. Despite the relative instability of their interactions, hubs exhibit a high connectivity conservation (fig. 1.5B). In other words, there is a tendency to substitute lost interactions by new ones (fig. 1.5C), which reflects a remarkable functional versatility. We further display how hubs principally rewire their interactions in response to the deletion of other hubs (fig. 1.5D); an example of this is shown for the ΔPFK1 background (fig. 1.5E). Fi-

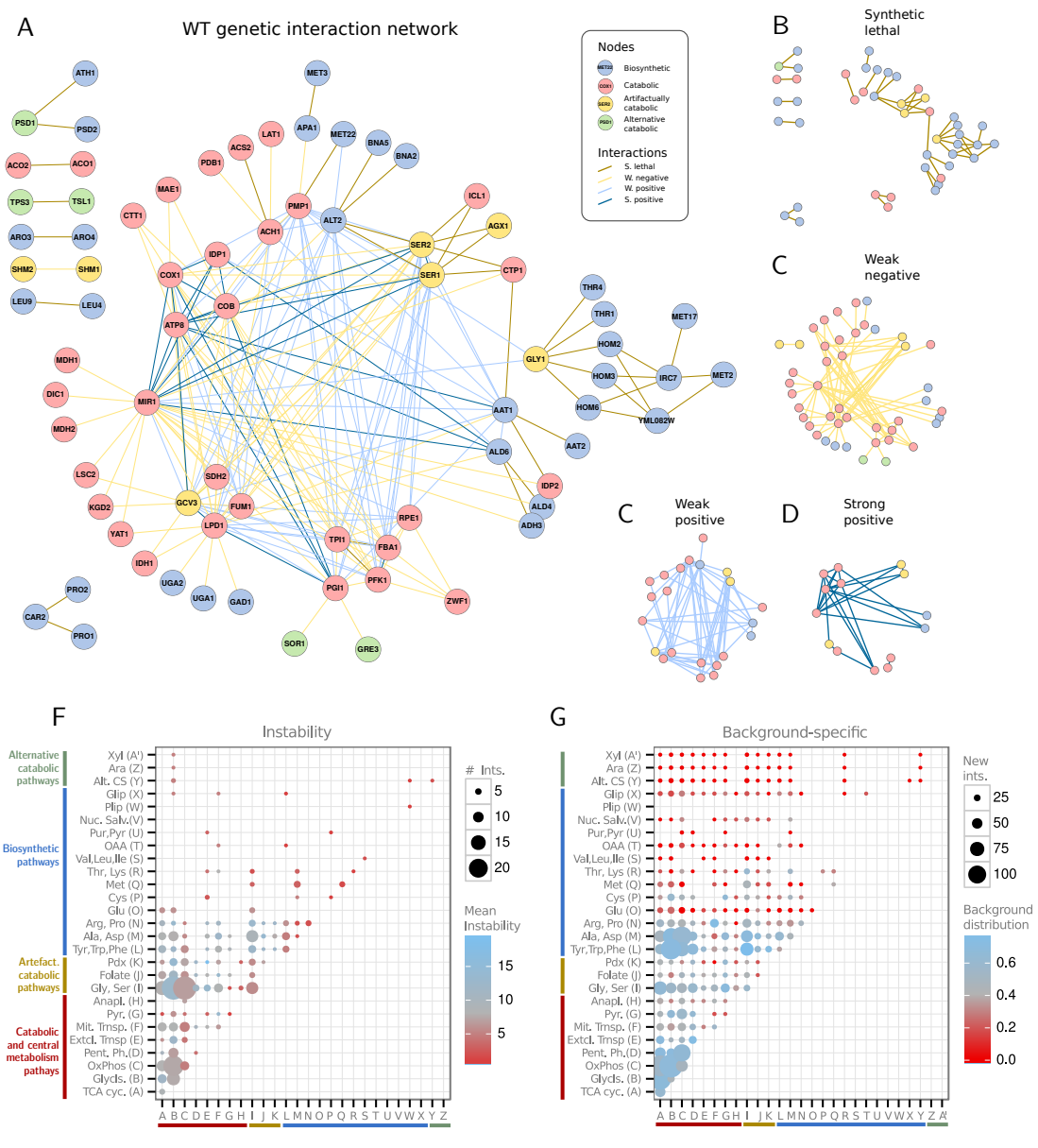


Figure 1.4: Function and genetic network instability A) The wt genetic network with nodes colored by function and interactions by class. B - E) Decomposition of the network into the four interaction classes. Note that whereas synthetic lethal links broadly appear at biosynthetic genes and at the periphery, the unstable types (weak and/or positive) connect catabolic in a densely connected central component. F) Number of interactions of the wt network between the corresponding metabolic annotation modules (dot size), and their average instability (dot color; measured as in fig. 1.2). Catabolic modules associate with more instability than biosynthetic ones. G) New interactions (i.e., absent in WT network) between modules (dot size), and their distribution among backgrounds (dot color; distribution quantified as normalized Shannon entropy, Methods). Catabolic modules are characterized by the emergence of many new interactions in different backgrounds. Among biosynthetic modules, new interactions are more rare and background-specific.

nally, we also detected a number of genes that acquire the role of genetic hubs in some specific backgrounds, and that frequently corresponds again to catabolic functions (fig. 1.5F).

In sum, the network exhibits a central, highly connected component, with abundant weak and positive interactions connecting different pathways, that is highly unstable to genetic perturbation. This organization can be explained by the action of a number of versatile genes capable to contribute to fitness in different ways (by altering their function) and to partially buffer each others' action (see also fig 1.1). The abundance of WP interactions emphasizes the different means to contribute to fitness. For example, although glycolysis (e.g., genes such as `TPI1`, `FBA1`) and the TCA cycle (`LPD1`, `FUM1`) work in coordination to supply reduced equivalents to oxidative phosphorylation, they can also readjust their metabolic role when one of the subsystems is compromised (what causes WP links between `TPI1` and `LPD1` or `FBA1` and `FUM1`; Fig. 1.4). Moreover, WN interactions indicate distributed buffering[3], i.e. different sets of genes that can implement a particular metabolic task in biochemically different ways, and in consequence, with different efficiency. An example corresponds to the mitochondrial phosphate importer `MIR1`, that negatively interacts with genes are able to jointly compensate for this deficiency. These include genes that contribute to the phosphate import strictly speaking (by antiport with malate), but also many that reduce the redox imbalances generated by extruding malate from mitochondria. The resulting broad redistribution of fluxes across central metabolic pathways (Fig. 1.6) has far reaching consequences for the genetic network: around 150 *wt* genetic interactions disappear (more than 70%) and almost 400 new ones appear. We simplified this huge amount of rewiring in Δ`MIR1` in fig. 1.7 to emphasize underlying principles.

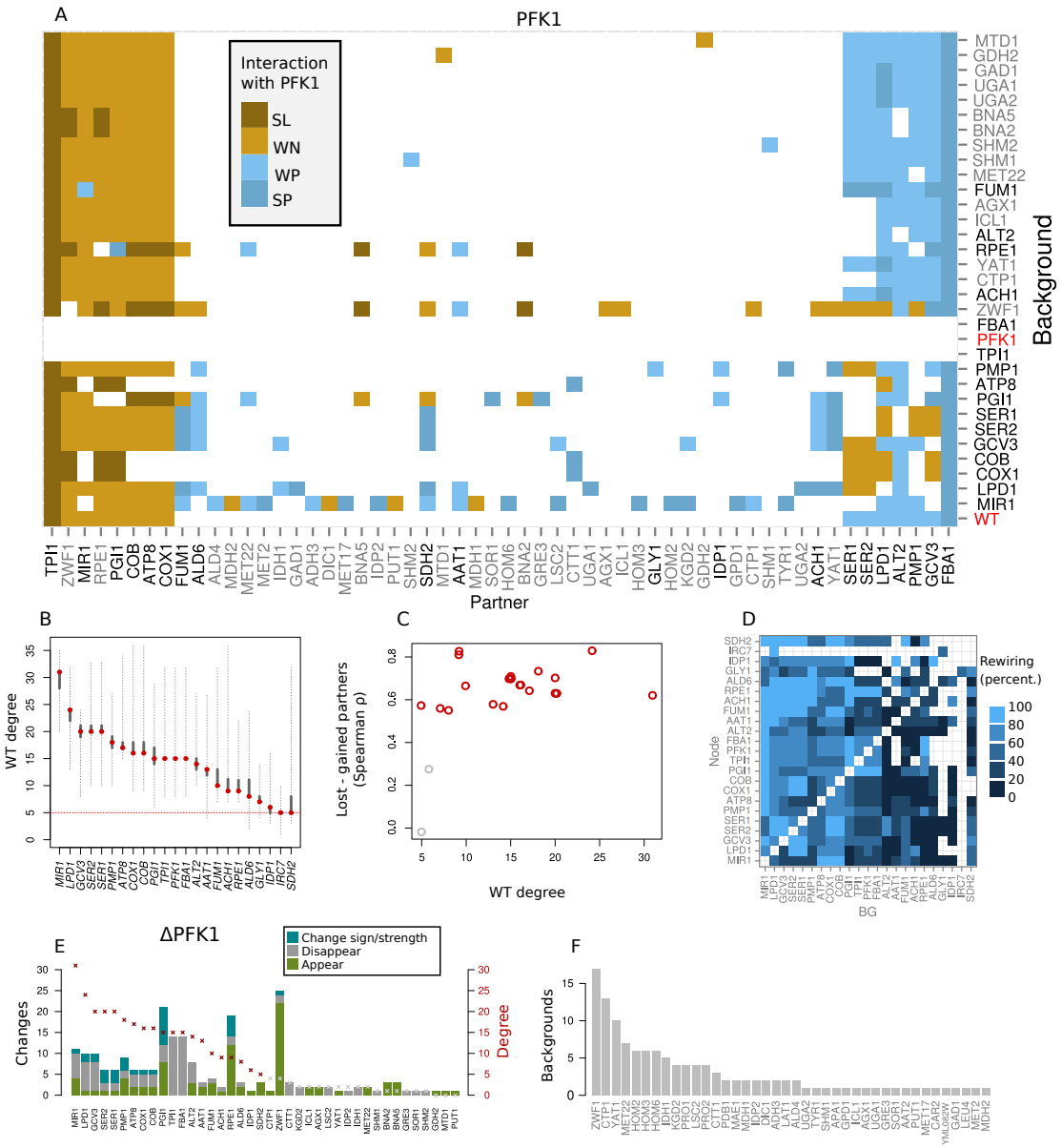


Figure 1.5: Rewiring of genetic hubs. A) Exhaustive map of genetic rewiring of PFK1 and its interaction partners (horizontal axis) in all backgrounds (vertical). wt hubs are highlighted in black, emphasizing hub-hub connectivity and rewiring. B) Degree distribution of wt hubs (degree >5) among backgrounds. We show wt degree (red dots), degree percentiles 10 to 90 (thick gray lines), and distribution range (dashed lines). C) Negative correlation between gained and lost genetic interactions in hubs (red color indicates p <0.001). D) Rewiring of hubs in response to hub deletions. We depict percentiles of a normalized rewiring score (number of changed interactions / wt degree). Empty spaces correspond to absence of rewiring. E) Rewiring experienced by genes in the ΔPFK1 background. Dots show the wt degree of each gene. Note that most rewiring is observed in genes that are functionally related, (i.e. other glycolytic genes, and pentose phosphate pathway, which assumes glucose breakdown ΔPFK1. F) List of genes acting as condition-dependent hubs, and number of backgrounds in which they exhibit such large connectivity; many of these genes are related to catabolism.

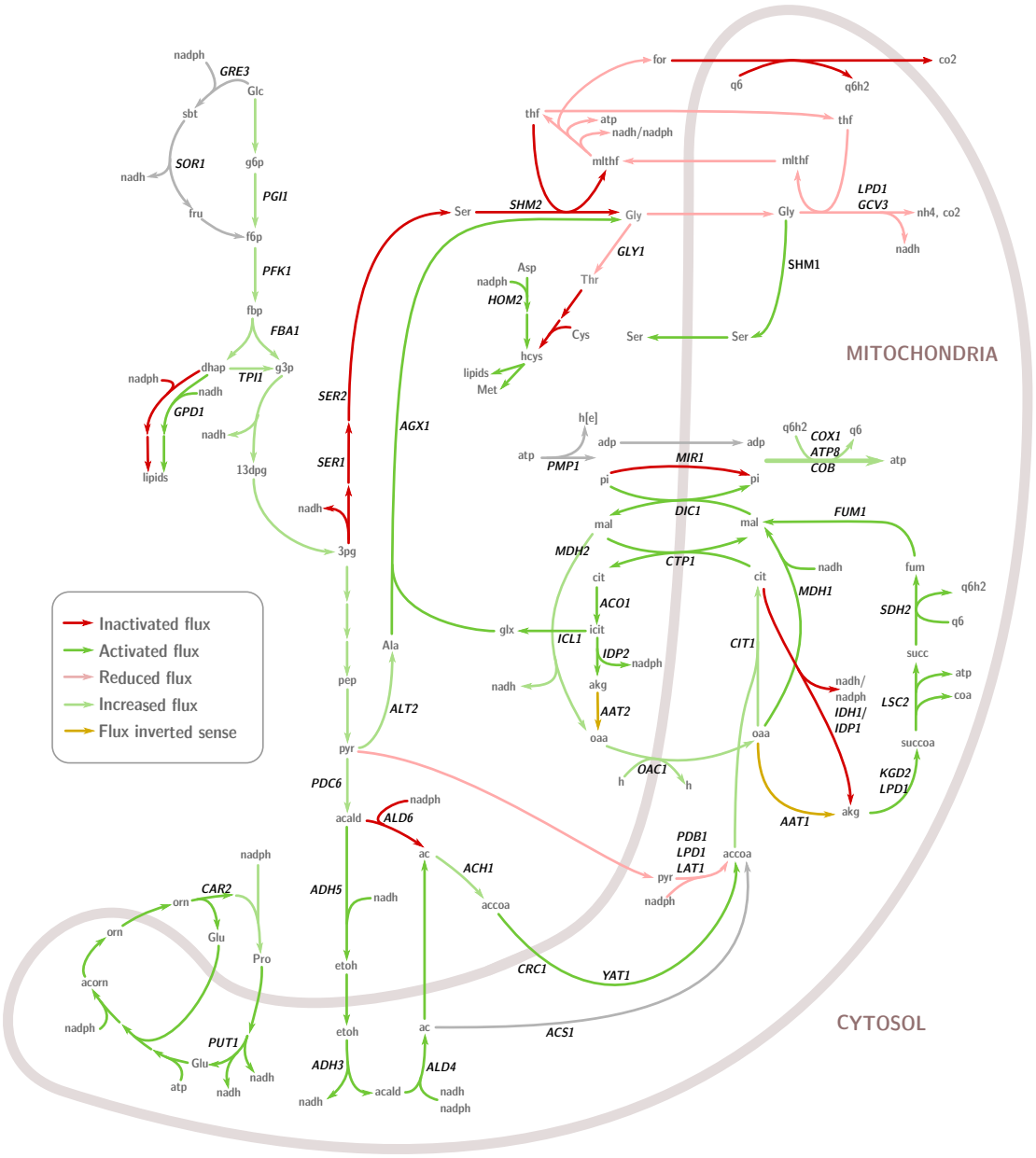


Figure 1.6: Distributed metabolic buffering of `MIR1` deletion Schematic representation of the metabolic changes that occur upon `MIR1` deletion. Colors indicate whether a given flux has increased, decreased, shut off or on, or changed direction, as shown in the legend.

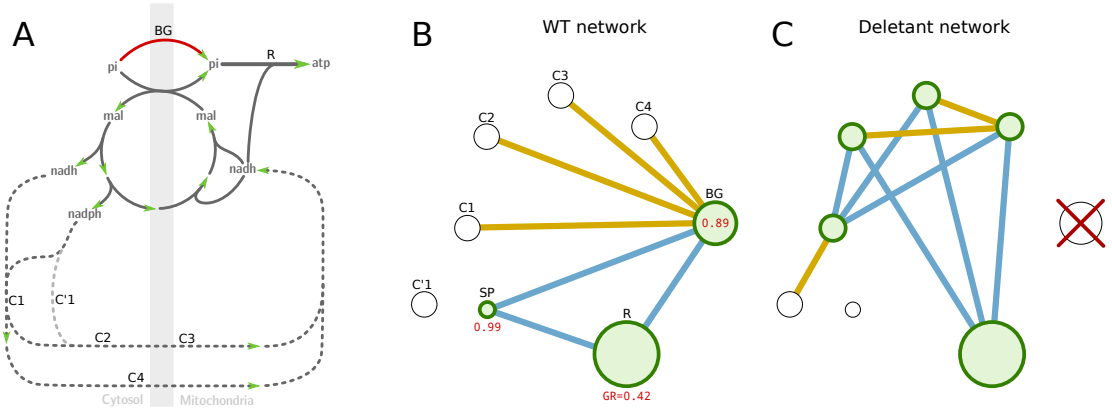


Figure 1.7: Principles of genetic rewiring illustrated in the ΔMIR1 background. A) Upon deletion of MIR1, a compensatory mechanism is activated (C1). This compensation can produce side effects (in the case of MIR1, extrusion of malate from mitochondria) unless it is fully equivalent (e.g., duplicated genes). Additional mechanisms correct these secondary imbalances (C1, C2, C3, and C4, dashed lines). B) WT genetic network, with nodes associated to flux-carrying reactions colored green, and single fitness contribution as node size. MIR1 interacts negatively (yellow) with compensatory mechanisms (C1, C2, C3, C4). Positive interactions (blue) occur either with the targets of its function (e.g., with respiration, R, as it is the ultimate target of phosphate import to mitochondria) or with genes that stop functioning when MIR1 is compromised (generally displaying lower fitness effect; SP for strong positive partner). C) In the rewired network, the target mechanism (respiration) interacts positively with those genes that function compensating the MIR1 deletion (C1, C2). These compensatory mechanisms can interact among them positively or negatively, depending on the underlying functional relationship. Some of them can have in turn their own compensating mechanisms in the new background (e.g., C′1 buffers C1) also producing new negative interactions.

1.4 Genetic rewiring in neutral backgrounds indicates reduction in buffering

We now introduce a second class of genetic backgrounds in which a set of neutral gene deletions accumulate (Methods, fig. 1.8A for a case trajectory). How does the network rewire in response to these backgrounds? We answer this question by discussing first which genes appear in neutral backgrounds and how their deletion modifies the network.

Note that a subset of these deletions correspond to nodes of the *wt* network. These nodes can only have negative interactions (their deletion could not be neutral otherwise) which reflects that they constitute the "first order" compensatory mechanism for their interaction

partners. Their deletion can reveal "second order" buffering mechanisms as new negative interactions. This is observed for instance in the deletion of SDH2 (fig. 1.8B), which is able to compensate for several genes in the *wt* (e.g. GCV3, LPD1, etc). SDH2 deletion uncovers secondary buffering mechanisms (e.g., by FBP1 or GLY1). Changes of sign in *wt* links are observed as well; for example, interactions with MIR1 or ZWF1. Typically, these changes occur in central metabolism, where different pathways are able to contribute to fitness either feeding one another linearly (e.g., to ultimately supply redox power to oxidative phosphorylation) or serving as substitutes to each other (e.g., as direct mechanisms of ATP synthesis by substrate-level phosphorylation). Also, neutral deletions can force metabolism to switch fluxes to pathways of equivalent efficiency; this can in turn reveal remove or add positive interactions to the network (e.g., ALT2–GCV3 link in Fig. 1.8B). Note also how genes that become nodes in one step of the trajectory can produce relatively strong rewiring when they undergo subsequent deletion (e.g., FBP1, that turns into a node after SDH2 mutation to be deleted in a later step, fig. 1.8A). Last (but not least), deletions of genes participate in synthetic lethal interactions will of course produce new essential genes, which is often observed (fig. 1.8A).

Deletions of genes not present in the *wt* network – such as those participating in "second-order" compensation mechanisms – can also modify the network. Examples of this are the deletion of MIS1 or HXK1/HXK2 in fig. 1.8A. The latter is especially illustrative, as three different genes are capable of performing the glucose phosphorylation reaction: HXK1, HXK2 and GLK1. The deletion of HXK2 first unveils a synthetic lethal interaction between the other two (fig. 1.8C). Later, when HXK1 is also deleted, GLK1 becomes essential. In addition, HXK2 and HXK1 (but not GLK1) were additionally capable of fructose or mannose activation (besides glucose). Elimination of these functions induces many new constraints to catabolic function,

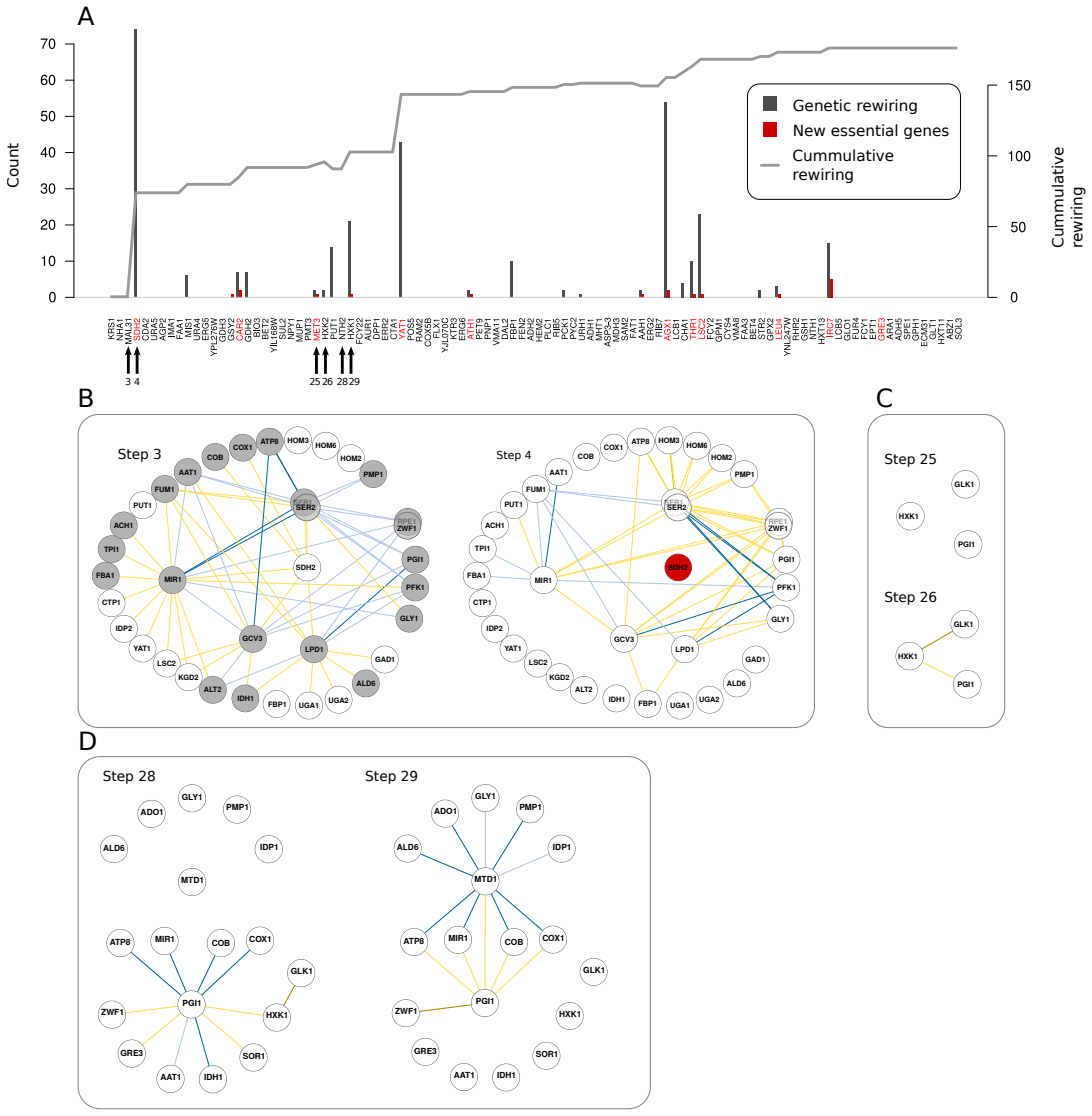


Figure 1.8: Detailed view of a neutral deletion trajectory. A) Genes were consecutively removed (from left to right), accumulating network rewiring (curve). Removed nodes of the wt network are highlighted in red. Bars depict rewiring score (gray) and increase in essential genes (red), both with respect to the previous step. Rewiring occurs upon some few critical deletions that produce strong rewiring. The appearance of essential genes is concomitant with genetic rewiring. B-D) Examples of rewiring sub-networks at three critical steps of the trajectory. B) SDH2 is able to compensate for the loss of several genes (weak negative links). Its removal reveals second-order alternatives as new negative interactions. Also, several genes undergo changes in their functional role, that are reflected in interaction sign changes and emergence of new positive interactions. C) The deletion of one of the three alternatives for hexokinase (HXK2), the other two constitute a synthetic lethal link (HXK1 and GLK1). HXK1 can also phosphorylate fructose, which is the reason of the weak negative link with PGI1. D) Subsequent deletion of HXK1 further rewires PGI1 interactions (e.g., the weak negative link with pentose phosphate pathway gene ZWF1 becomes synthetic lethal, as the only alternative for initial glucose processing) and induces several positive links in central metabolism, e.g., MTD1, involved in the (artifactual) glycine fermentative pathway (color code of genetic interactions as previous figures).

which uncover new genetic interactions: negative ones, between glycolysis and respiration, and positive ones, between respiration and folate pathway genes (fig. 1.8D).

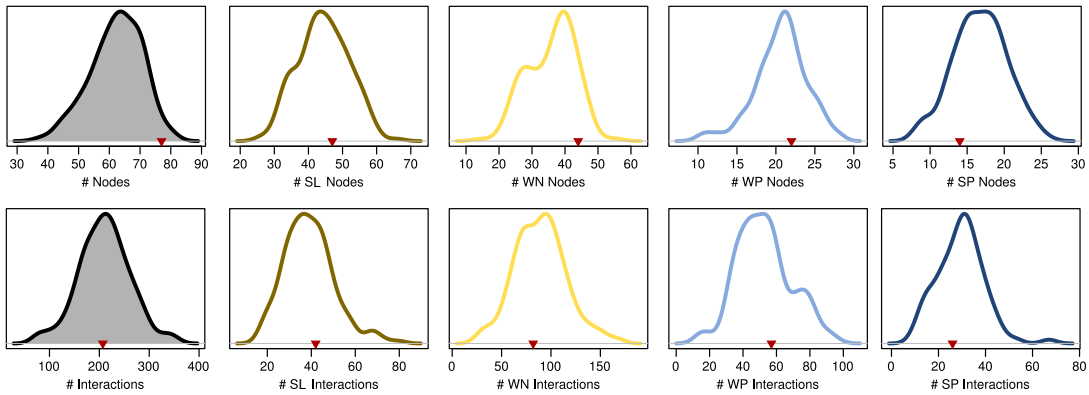


Figure 1.9: Neutral deletion trajectories modify the size of the network. Distribution of number of nodes (above) and edges (below) in genetic networks resulting from neutral deletion trajectories. It was measured considering the entire network (left, filled in gray) or each of the four types of interactions separately (right).

In order to identify common patterns of network rewiring, we analyzed a group of 200 alternative neutral backgrounds similar to the example in fig. 1.8. We found first that the corresponding networks exhibit a smaller number of nodes (98% contain less nodes than WT). Second, networks tend to exhibit more epistatic interactions per node (166 out of 200 present larger average epistasis). Both patterns are related mostly to negative interactions (fig. 1.9). Also, the negative component of the network undergoes a significantly stronger rewiring than the positive one (fig. 1.10). This suggests overall that previously hidden phenotypic effects unveiled as a result of the global reduction in buffering mechanisms. This is further evidenced by an increase in the number of essential genes, which is observed in 93.5% (187/200) of genotypes and usually occurs concomitantly with rewiring of the network (fig 1.8).

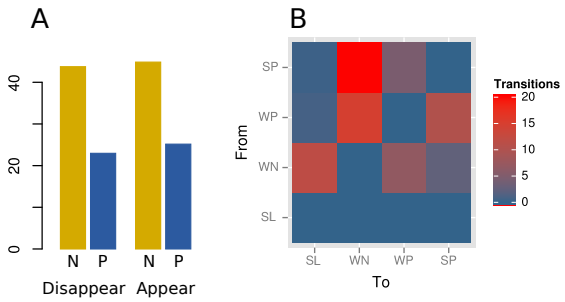


Figure 1.10: Rewiring of negative and positive interactions. A) Average number of negative (N) and positive (P) interactions observed to disappear and appear in genetic networks in response to neutral deletion trajectories. B) Average number of transitions between interaction types observed when comparing these networks with the wt one. Overall, there is a stronger tendency to rewiring in negative interactions.

1.5 Genetic rewiring is associated to diminished environmental plasticity

One could expect that not all accumulated deletions in the neutral backgrounds impair metabolism in the same way, and thus different backgrounds could cause contrasting metabolic plasticity. This can be detected by generating diverse random environments[231] (Methods) in which fitness is computed. The resulting growth measures did reveal the cryptic variability linked to the neutral trajectories (Fig. ??A).

To relate this variability to specific rewiring patterns (in wt conditions), we assembled the genetic networks of the 100 genotypes with the highest and lowest median growth among the environments considered (HG and LG, respectively; note that these are the 200 networks considered in the previous section). LG genotypes present a significantly stronger rewiring (Fig. ??B), which especially affects the types of unstable interactions previously identified as associated to catabolic / central metabolic pathways. As a result, LG networks were smaller both in number of nodes and edges (fig 1.11), which can be attributed again to negative interactions. Consistently, also the number of essential genes was higher in LG networks, including

several crucial catabolic components (e.g., ATP8, FBA1, PGI1, fig. ??C).

Although the number of new interactions did not differ between LG and HG (mean = 60.6 and 61.0 respectively, $p = 0.96$), a subset of them appeared more frequently in LG genotypes –notably, negative ones between pentose phosphate pathway (ZWF1 and RPE1) and other catabolic genes (e.g., PGI1, MIR1 or LPD1, Fig. S21B-C). Sign and/or strength change was also considerably stronger in LG genotypes (mean = 24.3 interactions/genotype) as compared to HG (mean = 13.4 interactions/genotype, $p = 6.7 \times 10^{-14}$, Fig. S21D-E). As a result of the stronger rewiring, LG networks exhibited higher negative-to-positive interaction ratios than HG ones (mean = 1.65 *vs.* 1.53, Wilcoxon test $p = 0.0005$).

These results evidence that the loss of compensatory mechanisms, mostly associated to catabolism and central metabolism, links the amount of genetic network rewiring to the reduction of environmental plasticity. This can be explained by the fact that carbon sources other than glucose usually require only few transformation steps before being incorporated into the core pathways, e.g., at different steps of glycolysis or TCA cycle. Some sources are alternatively transformed into glucose by means of gluconeogenesis. Although "central" pathways are used then with relative independence of the external carbon source, they can be used differently: some branches that are optimal in one environment can be sub-optimal in another, where they can nevertheless compensate for the loss of the optimal one. This is further corroborated by the differential distribution of deletions between LG and HG genotypes (Fig. ??D). LG genotypes are enriched in 26 specific deletions that can be grouped in i/ genes important for the initial processing of different carbon sources (e.g., PNP1, XYL2, XKS1, GAL1, etc), ii/ gluconeogenesis (e.g., PCK1), and iii/ key central metabolic genes, such as SDH2, KGD1, or LSC1. Although neutral in glucose minimal medium, these genes are able

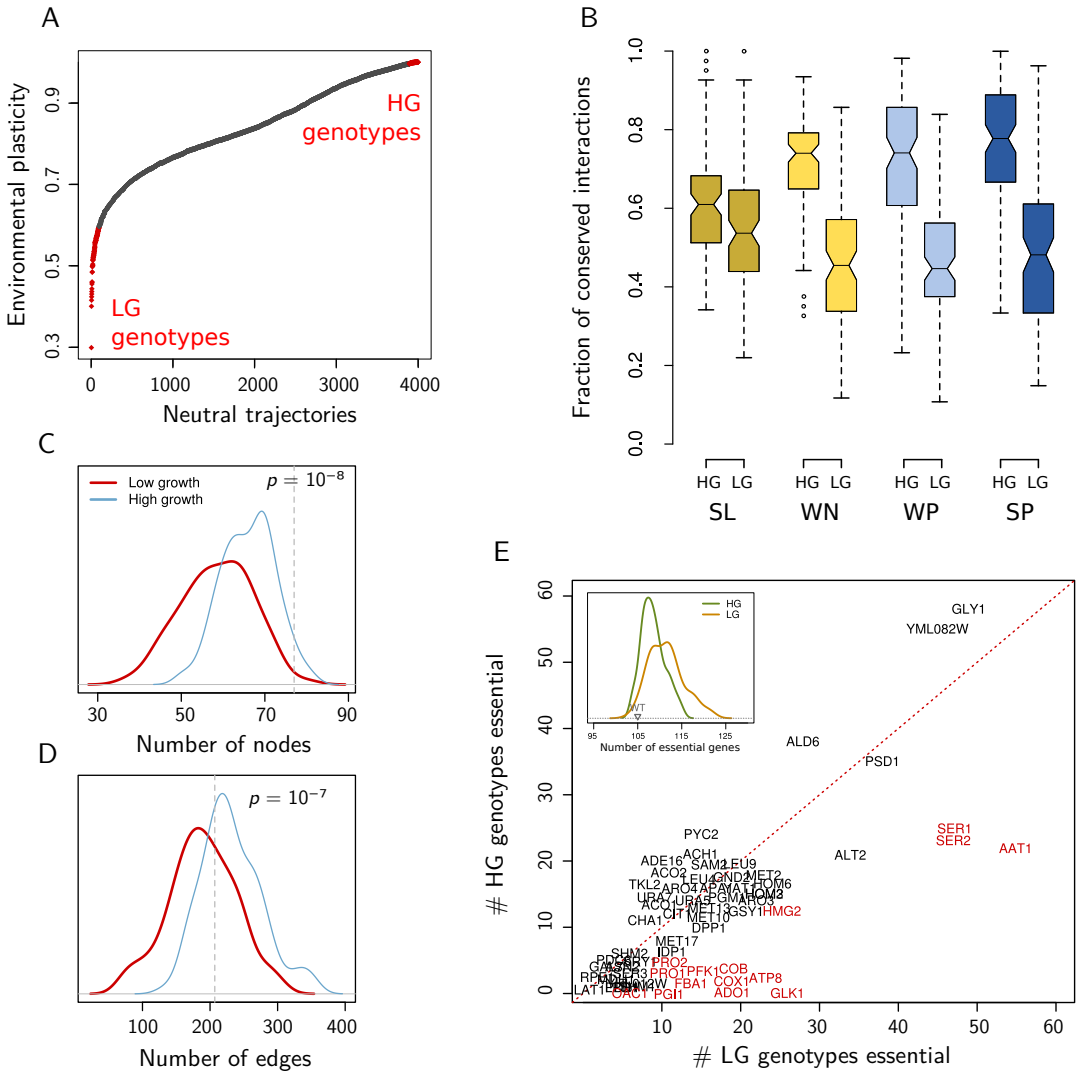


Figure 1.11: Genetic rewiring predicts loss of environmental plasticity. A) Environmental plasticity exhibited by 4000 metabolisms, each originated from a random neutral deletion trajectory. Plasticity of a metabolism is scored as the median growth rate in 1000 randomly generated environments, normalized to the growth rate of the wt in the same environments. We highlight in red the groups with highest and lowest growth (HG and LG, respectively). B) Fraction of conserved interactions by type in LG or HG genotypes. C) Distribution of the number of nodes, and D) edges, in LG and HG genetic networks (the number observed in the wt network is shown as vertical dashed lines). E) Frequency at which different genes are essential among LG (horizontal axis) and HG genotypes (vertical axis). Genes never becoming essential were omitted. We highlight in red those genes that become essential in LG genotypes significantly more than in HG ones (Fisher tests p<0.001)

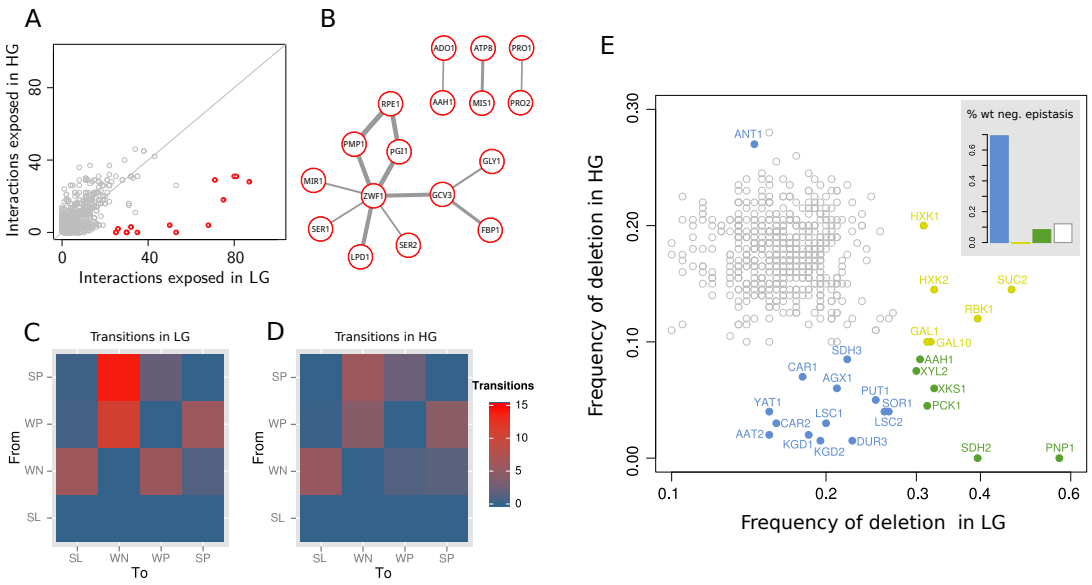


Figure 1.12: Specific rewiring patterns at low-growth genotypes. E) We performed a bootstrapping analysis to check for enrichment/depletion of specific deletions in LG or HG trajectories. Genes significantly enriched (or depleted) in the HG genotype, LG genotype or both were correspondingly colored in blue, yellow and green ($p < 0.01$, after multiple testing). In the inset, we show the percentage of genes with genetic interactions in the WT network in these significant groups.

to compensate for defects in the optimal pathways (evidenced by their multiple negative interactions), but importantly can take over their role under different carbon sources.

1.6 Discussion

Different genetic backgrounds can modify gene interactions and consequently rewire genetic networks. Here, we systematically examined how backgrounds impact networks by using an *in silico* model of yeast metabolism[177] (Methods).

We first asked to what extent genetic rewiring is necessarily coupled to strong metabolic readjustments. To this aim, we analyzed a class of backgrounds defined by single deletions of active enzymes (i.e., that exhibit nonzero flux in WT conditions). Using a simple score to

measure metabolic reorganization, we confirmed this association (fig. 1.2. More specifically, genetic network rewiring was linked to qualitative changes in metabolism (i.e. switching to alternative metabolic pathways) but not quantitative ones (e.g. change in relative flux through active pathways).

We observed as well that a group of backgrounds corresponding to deletions in catabolic and central metabolic nodes was associated to a particularly strong genetic network rewiring. These backgrounds constitute hubs in the genetic interaction network, forming a densely connected sub-network characterized by the abundance of positive and weak interactions between different functional modules. Most of the background-specific (i.e., not observed in the WT) interactions and hubs are as well related to catabolism, and enriched by inter-module weak negative epistasis (figs. 1.3–1.5). Therefore, the structure of the wild-type network is already representing the functional associations most sensitive to background change, and in consequence anticipating its own instability.

Likewise, our analysis clarifies the linkage among fitness contribution, pleiotropy, and network connectivity (node degree)[210]. That a particular gene is nonessential but contributes to fitness implies the existence of a number of inefficient distributed buffering mechanisms[164] of the type observed in catabolism. Pleiotropy (Methods) is strongly related to catabolic genes, due to the ubiquity of currency metabolites in different pathways (e.g. biosynthetic routes of biomass constituents). Pleiotropy is thus present in catabolic genes and absent in biosynthetic ones, and its correlation with fitness contribution and node degree could in the end denote the distributed robustness of the catabolic subsystem. As expected, both pleiotropy and fitness contribution anticipate rewiring (fig. 1.13).

The strong correlation between rewiring and perturbation of currency metabolite balances

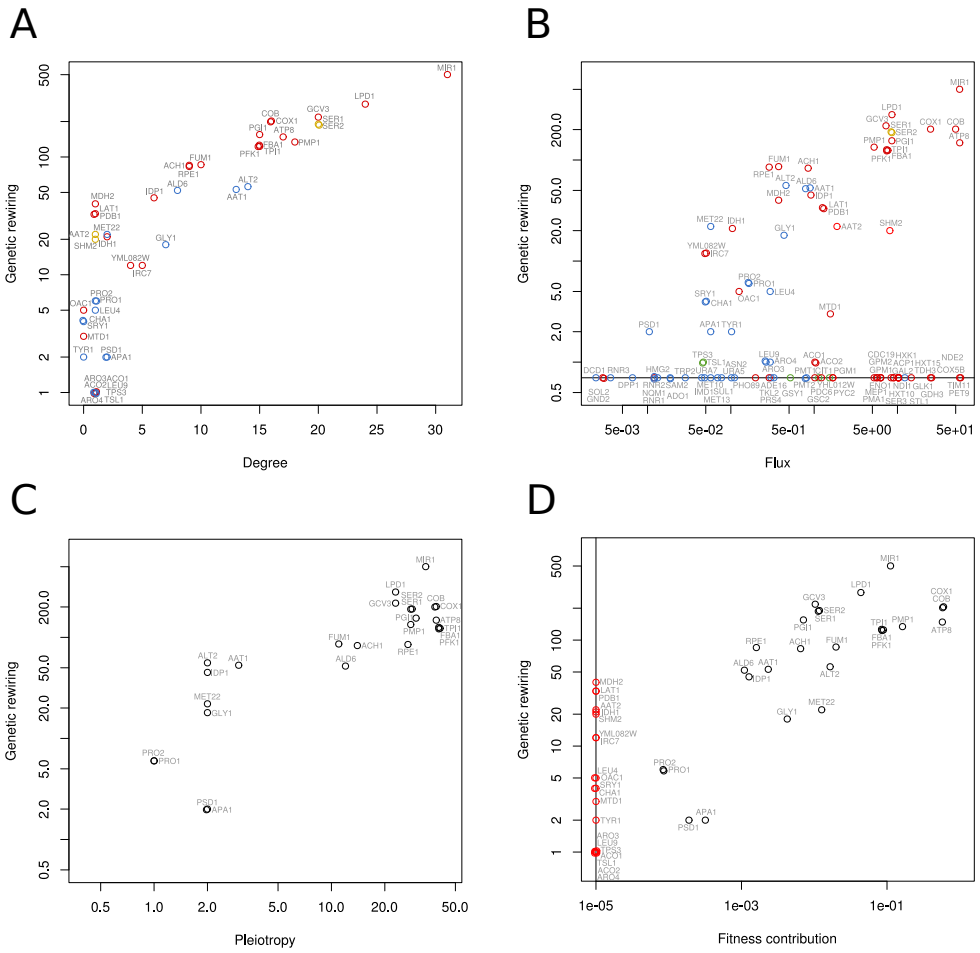


Figure 1.13: Rewiring, degree, fitness, pleiotropy and flux. Correlation between genetic rewiring in different backgrounds and their A) degree, B) total flux through the associated reactions, C) pleiotropy, and D) single-gene fitness contribution (those without effect shown in red).

(table 1.1) is helpful in understanding these patterns. The existence of multiple NAD(P)H and ATP producing enzymes in catabolic pathways enables them as potential substitutes to each other, i.e., as regulators of currency metabolite homeostasis[74]. But these mechanisms are not equivalent biochemically and consequently not equally optimal (figs. 1.6 and 1.7). Catabolism exhibits then a functional degeneracy in which qualitatively different catabolic

configurations lead to similar but not identically efficient solutions[3]. This explains many of the phenotypic features corresponding to catabolic genes. In particular, it causes catabolic genes to be typically nonessential, but often fitness contributing. Degeneracy explains too the pervasiveness of weak and unstable interactions which reflects either fitness contributions that are partially shared (weak positive), or deficient buffering (weak negative, see also fig. 1.1). Notably, background-dependent interactions tend to be weak as well. In sum, the distributed nature of catabolic processing determines the transient and context-dependent functional associations that define its genetic network.

Biosynthetic pathways exhibit in contrast a much different architecture. They usually display relatively isolated and linear configurations, each of them usually containing biochemically unique metabolites that simply act as intermediates for the synthesis of a particular compound. This limits the buffering possibilities compared to catabolism, what is manifested in the enrichment of essential genes, and also in a genetic landscape dominated by redundancy-based synthetic lethal interactions (Fig. 1.4). These SL interactions form smaller (peripheral) clusters and exhibit a marked stability that only becomes disrupted when one of the partners is deleted or becomes essential (fig. 1.4-1.5).

These findings are consistent with several evidences from previous experimental studies on the rewiring of genetic interactions across species (studies not always linked to metabolism). For instance, SL pairs and interactions within functional modules were found considerably conserved, while interactions between modules remodeled[56,181] –both signals confirming what we observed–, and the change of epistasis sign that we detected in catabolic nodes could indicate a sort of functional re-purposing[73]. Indeed, the analysis of this comparative data[73] corroborated that weak interactions are less conserved (4.9% of the weak are conserved *vs.* 8.7%

conservation of the strong, χ^2-test $p < 2.2 \times 10^{-16}$), and that positive ones are more unstable (0.8% conserved *vs.* 5.4% conservation of negative ones, χ^2-test $p < 2.2 \times 10^{-16}$). We additionally used this dataset to compare interactions involving catabolic and biosynthetic genes only. We found that interactions involving central metabolic genes are significantly more unstable (39.4% of pairs are not conserved) compared to interactions involving the rest of metabolic genes (29.0% not conserved, χ^2-test $p < 2.2 \times 10^{-16}$). Moreover, we also recognized an association between interaction type and metabolic function (as we found *in silico*) in a set of genetic interactions recently measured experimentally between metabolic genes (Fig. S24)[210].

We examined a second class of backgrounds that are rather defined by (the accumulation of) neutral deletions (fig. 1.8)[162,54]. These trajectories generally originated metabolisms with a higher incidence of essential genes and smaller but more densely connected genetic networks (fig. 1.9). This denotes overall a global reduction in buffering. Neutral backgrounds also modify environmental plasticity (i.e., capacity for robust growth in a range of environments) to a different degree (figs. 1.11– 1.12). Notably, genetic networks associated to more limited plasticity present the strongest genetic rewiring, again observed in interactions associated to catabolic function. The mechanistic explanation is that, after usually few initial specific processing steps, all contrasting carbon sources enter the common catabolic core (glycolysis, TCA cycle, respiration). Mutations that are neutral in glucose minimal medium (affecting less efficient catabolic routes) can nevertheless represent the most efficient catabolic processing alternatives in other carbon sources.

The connection between environmental and genetic robustness[148] would further predict that the patterns identified in response to the alteration of genetic background could be sim-

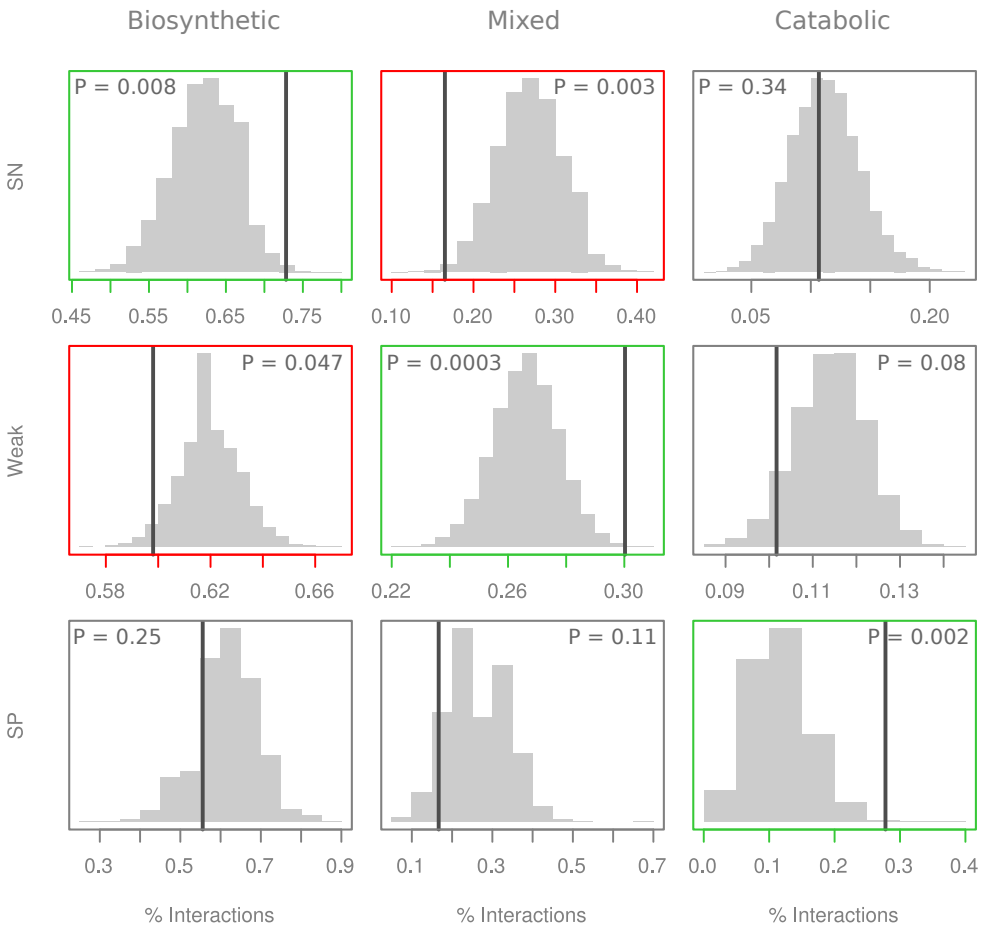


Figure 1.14: Biosynthesis and catabolism associate respectively to SL and SP interactions We selected the high-confidence genetic interaction dataset in Szappanos *et al.*[210] the interactions between genes present in the `iND750` model of *S. cerevisiae* used here. Interactions were classified as strong positive (upper quartile among positive), strong negative (idem among negative) and weak (the rest). For functional classification, we conservatively considered a gene as ``catabolic'' if it belongs to either glycolysis, TCA cycle or oxidative phosphorylation; the rest were considered ``biosynthetic''. According to this classification, an interaction can be catabolic, biosynthetic, or mixed (one catabolic and one biosynthetic partner). In each strength category, we computed the percentage belonging to each functional class (vertical lines). To assess if any class is particularly enriched or depleted, we randomized 10000 times the assignment between strength and functional category (distributions shown as histograms, with the corresponding p-values). Plot boxes were colored red if significant depletion, and green if significant enrichment.

ilarly recognized in reaction to environmental change. To test this hypothesis, we characterized rewiring of a recently assembled (yeast) genetic network after several DNA-damaging

treatments[94]. Our predictions are qualitatively confirmed, with weak interactions being more unstable than strong, and positive more unstable than negative (fig. 1.15B–C, compare with Fig. 1.3). Interactions among genes that are functionally related were also more stable (fig. 1.15C). Furthermore, treatment-specific links occur between functionally different genes (92% as compared to 87% in the untreated network, χ^2-test, $p = 3 \times 10^{-15}$, and are more often weak (fig. 1.16).

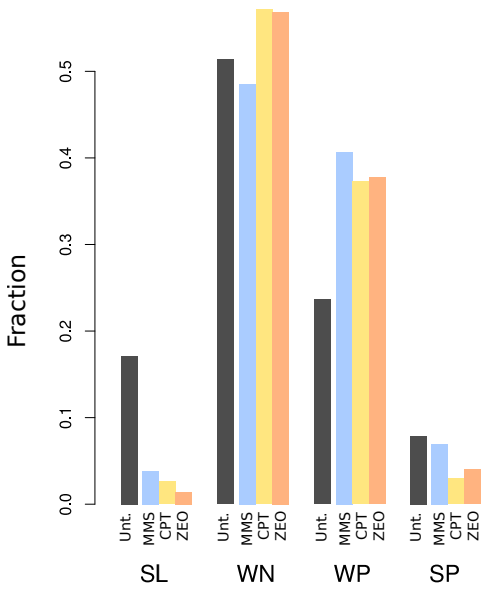


Figure 1.16: Rewiring distribution under DNA damage We computed, for all interactions appearing in at least one treatment, but not in the untreated network (i.e., treatment-specific interactions), the proportion of each strength/sign category. Treatment specific interactions are enriched in weak (especially weak positive) and depleted in strong (especially, in strong negative).

In summary, we showed how distinct functional structures within the metabolic system, i.e., biosynthesis and catabolism, determine both the architecture of the network and its rewiring (fig. 1.15E), an interpretation that is naturally coupled to the two main sources of robustness, i.e, redundancies and distributed compensation[3]. These predictions are based on global features of metabolism, what overcomes the limitations associated to FBA modeling (that sometimes generates artifacts due, for instance, to its latent optimality assumptions[210], but that nevertheless can provide useful conceptual guidelines to the associated biology[3]). Differential network mapping should thus consider the specific mechanistic causes of robustness in the system under study to accurately interpret the dynamic rewiring of genetic networks in health and disease[105,75].

1.7 Methods

Genotypes and Simulations.

We considered the iND750 genome-scale model as the WT genotype[59]. This model incorporates all the necessary complexity of *Saccharomyces cerevisiae*'s metabolism (e.g., it is fully compartmentalized), has been empirically corroborated, and also reduces the computational load associated to the background analysis. We studied two types of genotype derived from this model. Single deletion genotypes were obtained by deleting each gene present in the model individually. Neutral deletion trajectories were obtained by successively deleting genes that have no effect on phenotype[162] (i.e., optimal growth does not change) until reaching 100 deleted genes. All optimizations were performed using Flux Balance Analysis (FBA)[170]. WT growth conditions correspond to glucose minimal medium and aerobiosis (glucose: 18.5 mmol gr^{-1} h^{-1}, unlimited O_2). Fluxes through all reactions in the solution of a given genotype were normalized by the amount of biomass produced. This enables the comparison of different solutions with distinct growth rates.

Generation and processing of genetic networks.

Optimal growth of all single and double deletion mutants, encompassing all nonessential genes in a given genotype, was computed using FBA[170]. The mutant/WT growth ratios obtained were used to compute the epistasis (ϵ) that incorporated a multiplicative model and posterior scaling[193]; interactions with $|\epsilon| < 0.01$ were not considered. An additional processing was applied to the networks to simplify functional redundancies that are non informative and do not contribute to the system-level analysis discussed in the manuscript. Namely, we identified all sets of genes coding for exactly the same reactions, and excluded all but one from

each set for further analysis. This simplifies positive interactions associated to subunits of the same complex (e.g., mitochondrial ATP synthase has 15 essential subunits in the model, which results in $\binom{15}{2} = 105$ positive epistatic interactions), and also negative interactions that exist between equivalent gene duplicates coding for fitness contributing reactions. This dataset reduction was applied in all our analyses, unless otherwise specified.

Advantages of Flux Balance Analysis.

FBA was considered a suitable tool for this study due to several reasons – beyond the obvious advantage of avoiding the complications of producing the very large number of required genotypes experimentally. First, it simplifies several layers of biological complexity (e.g., gene expression or enzymatic activity regulation) by means of optimality assumptions in a relatively realistic way[170]. While this could lead to some artifacts, they do not modify in any case the general conclusions of our analysis (see also Supplement). Second, the model enables a straightforward interpretation of the phenotype as an univocal consequence of the structure of the underlying metabolic reaction network. We can thus imagine the *in silico* model as a biological "organism" *per-se*, that can provide broad conceptual guidelines for a comprehensive interpretation of the rewiring of genetic networks associated to real biological systems (not necessarily restricted to metabolism).

Pleiotropy.

We computed the pleiotropy of each nonessential and fitness contributing gene in the WT genotype following[210]. The method basically consists in optimizing for the production of a given biomass constituent individually (instead of using the entire biomass reaction) in presence and absence of a given gene. The number of affected constituents represents a rigorous

measure of pleiotropy for that specific metabolic gene.

Rewiring and metabolic modules.

Background dispersion was quantified in Fig. ??B as normalized Shannon entropy S_{MP} (MP denotes "module pair"). This is defined as $S_{\mathrm{MP}} = -\sum_{i=1}^{n} k_i \log k_i / \log N$, with n being the number of backgrounds with new interactions between the two modules, and k_i the number of interactions appearing in the background i (divided by the total number of interactions considering all backgrounds). This was normalized by $\log N$, where N is the total number of analyzed backgrounds where any new interaction appears between any two modules (N = 37). The figure illustrates how catabolic modules are characterized by appearance of many different new interactions in different backgrounds. Conversely, much fewer interactions appear among biosynthetic modules, these being generally much more background-specific. The notation of the modules in the figure can be found in Table S6.

Random environments.

1000 random environments were generated in which each of 107 organic nutrients was assigned a probability of being present from an exponential distribution (with mean = 0.1[231]). After defining the particular set of nutrients, their dosage was randomly obtained by applying an uniform distribution between 0 and 20 mmol gr^{-1} h^{-1}. All environments considered were aerobic (i.e., unconstrained O_2 availability).

Experimental datasets used Data on instability of interactions in response to environmental change (yeast cells growing in rich media and in the presence of three distinct DNA damaging agents: Methyl methanesulfonate, Camptothecin, and Zeocin) was obtained from[94]. We considered as *not* significant epistasis those values below 2 and above -2.5 (following the orig-

inal reference). We defined strong positive interactions as those in the upper quartile among positive ones, and similarly for negative ones. The instability of each category in Fig. ??A was quantified as the average number of treatments where the interaction changes or disappears (out of three). We quantified the functional similarity of the genes constituting an interaction as the ratio between the number of shared functional classes (i.e., biological process annotations as in[94]) and the minimal number of classes that one of the genes of the pair presents. If this score was bigger than 0.1 then genes were considered functionally "close", and "distant" otherwise. In addition, we considered an interaction "stable" if it remained within the same category (sign, strength) in all conditions, and unstable otherwise (Fig. ??B).

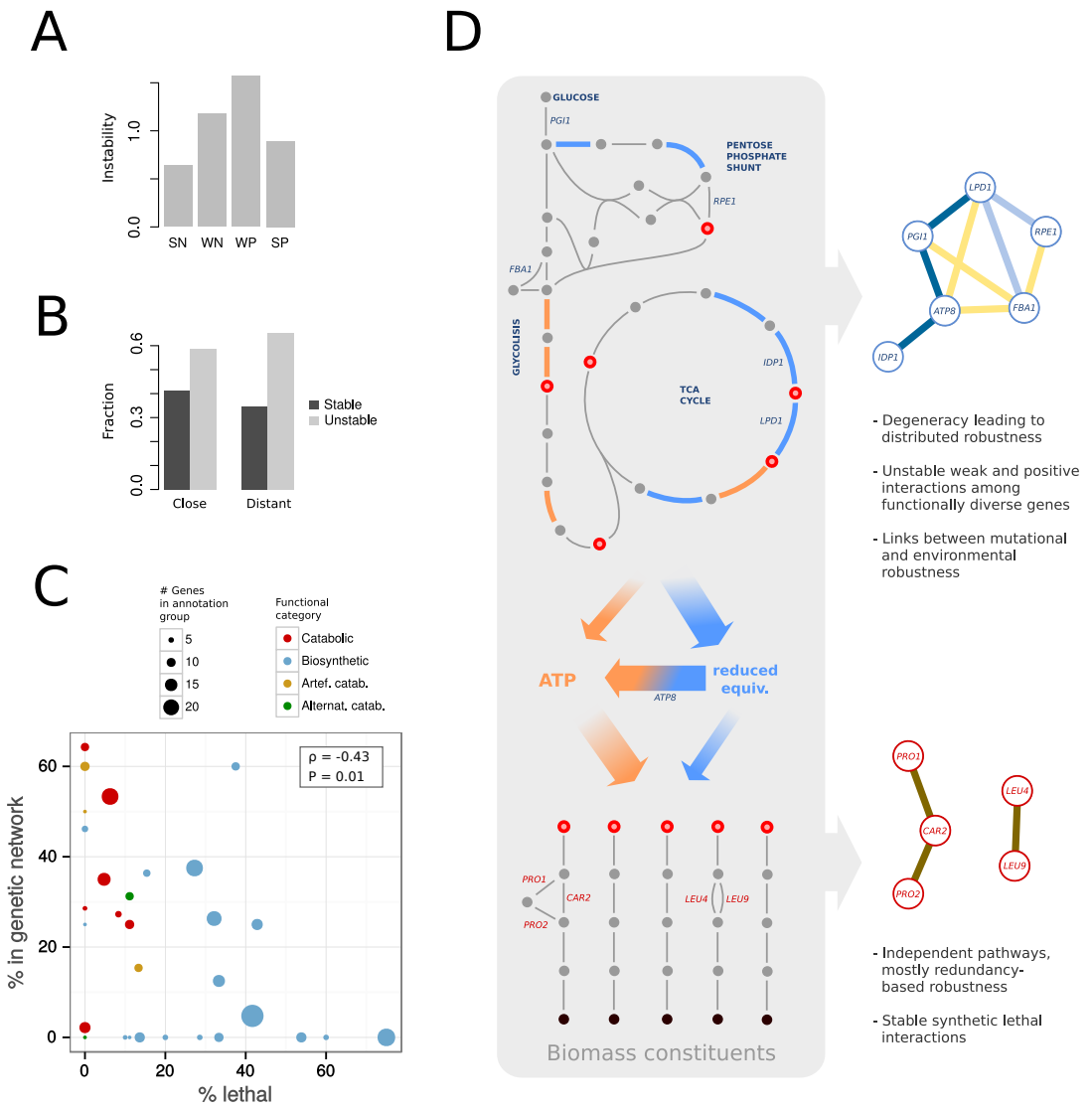


Figure 1.15: Robustness influences rewiring patterns. A) Relative instability of each interaction type after environmental change (interactions were experimentally obtained in rich media and in presence of three different DNA damaging agents [94], Methods). This pattern is qualitatively similar to the one predicted in response to modifications of genetic background (fig 1.3). B) Higher proportion of unstable interactions among weak, positive, and C) functionally distant genes (χ^2-test, $p = 0.003$, for the latter case, Methods). D) Each dot represents a metabolic module, with size and color indicating respectively number of genes and functional category. We show percentage of essential genes in each module (as a proxy of robustness; horizontal axis), and number of (nonessential) genes that are nodes in the WT genetic network (as a proxy of genetic landscape; vertical axis). D) The architecture of catabolism and biosynthesis (left) determines the resulting genetic network and its stability (right). We show in blue the reactions producing NAD(P)H, and in orange those producing ATP; metabolites (dots) that constitute biosynthetic precursors are highlighted in red. Some representative genes and their corresponding genetic interactions are included (color code of genetic interactions as previous figures).

2

Dual autoregulation enhances the dynamic response of a resistance phenotype

Understanding gene-expression dynamics depends in part on our ability to *decode* the function of transcriptional networks[173]. In recent years, this goal has been partially achieved with the identification of recurrent simple transcriptional architectures within these networks[4], e.g., autoregulation (also termed autogenous control[82,190]), feed-forward loop (FFL) regulation, and others[5]. These motifs were interpreted as fundamental regulatory units what triggered the interest in analyzing motif operation. For instance, we now know that neg-

ative autoregulation speeds up responses[182,29], while positive autoregulation induces delays and can cause bistability[138]. FFLs can in turn function as signal-persistence detectors[139], with more elaborated motifs implementing arguably more sophisticated functions –like hierarchical modules that can coordinate the expression of a group of genes with a single input[5].

How do these motifs aggregate and eventually work together as a network? This question raises a number of issues. One could ask, for example, to what extent the properties demonstrated with the characterization of network motifs in isolation remain valid when they are part of more intricate regulatory modules. Several studies seem to suggest so[61], but certainly more work is needed. In a broader context, these studies can also help us to discern how cross-talks are minimized in complex networks[198]. A somehow related matter is how motifs that enable contrasting behaviors could nevertheless appear combined. This problem of how antagonistic traits become integrated is not unusual in other biological areas, and widely corresponds to the understanding of how biological systems –and regulatory systems in particular[92,211]– ultimately face tradeoffs.

Here, we focus on this latter issue. We consider a situation in which both positive and negative autoregulation appear together. On one side, positive autoregulation can slow down the response and diversify cellular behavior[138,111]. On the other, negative autoregulation can speed up dynamics and increase the sensitivity of the system[182,155]. We examine how these two motifs work in combination by identifying a natural scenario in which this situation is observed. This is the case of the regulation of the multiple antibiotic resistance (*mar*) phenotype in *Escherichia coli*. The *mar* phenotype is a physiological response that capacitates bacteria to tolerate several stresses, including antibiotics like tetracycline or chloramphenicol[79]. Interestingly, the phenotype is coupled to an operon architecture harboring a repressor (MarR) and

an activator (MarA), implementing an antagonistic autogenous control of gene expression, and whose genetic architecture is intriguingly unique within *E. coli*'s transcriptional regulation (figure 2.1).

In this specific model system, we examine *i*) response dynamics, *ii*) integration of additional regulatory links, and *iii*) single-cell behavior. We also explore two additional questions, one is specific (which limits present the *mar* system on its speedup), the other more general (how genetic design influences the action of the dual module). Inspection of all these topics helps us to illustrate how tradeoffs between two antagonistic motifs can be resolved, and also shows the particular way in which antagonistic autogenous control determines the dynamical properties of the *mar* response.

2.1 The *mar* operon as a model of dual autogenous regulation

In bacteria, adverse circumpstances usually trigger a phenotypic response, usually involving a number of physiological changes that help defending against the effects of stress[207]. One of these physiological programs available to *Escherichia coli* corresponds to the multiple antibiotic resistance (Mar) phenotype. This response capacitates bacteria to tolerate several toxins, including antibiotics like tetracycline or chloramphenicol[79], but also phenolic compounds such as salycilate or other repellents[183,42]. That this response connected for the first time antibiotic resistance to the bacterial chromosome, rather than being caused by a plasmid-borne gene, prompted the search for a better understanding of its genetic architecture. In this way, we currently recognize that the Mar phenotype is coupled to a unique operon architecture harboring a repressor (MarR) and an activator (MarA), and that it is additionally modulated by other transcriptional factors (e.g., SoxS or Rob)[41,42,144].

The expression of the *marRAB* operon is then consequence of the inactivation of the repressor MarR – which represents the sensor of the stress – and a later boost in the level of MarA, which works as the actuator of the system. Increase of MarA abundance acts subsequently on a relatively large regulon that includes genes contributing to efflux pumps, e.g., *acrAB-tolC*[135,70], membrane permeability systems, e.g., *micF-ompF*[43], etc.

The genetic circuit that orchestrates expression of *marRAB* incorporates several feedbacks (figure 2.1) involving a crucial combination of both negative and positive autogenous control[82,190]. Notably, autogenous control was shown to provide very suitable features for stress response[29,199], such as speedup of dynamics[190,182] or diversification of cellular behavior[138,111]. How both types of control act together and the consequences of this combined regulation for the mounting of the antibiotic resistance remains, however, an open question.

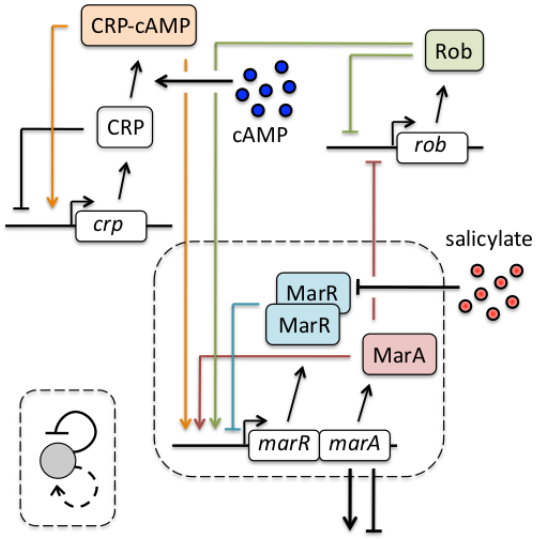


Figure 2.1: Regulatory architecture of the *marRAB* operon
The mar network regulates a response to a number of toxic compounds, including antibiotics. It includes a dual autogenous control motif constituted by an activator (MarA; acting as a monomer), and a repressor (MarR; acting as a dimer). The motif can sense as well metabolic signals (through CRP), and it reads additional positive inputs (Rob).

2.2 MarR gives rise to a fast response of the *mar* circuit

To examine the activation dynamics of the *mar* phenotype, we used a chromosomally integrated yellow fluorescent protein (YFP) under

the control of the *mar* promoter (P_{mar}-*yfp*). We monitored YFP dynamics at high resolution in reaction to salicylate (fig. ??A), and quantified the time to reach half its steady state as response time (t_{50}). With 0.5 mM salicylate, induction of the phenotype is carried out in about 15 min. This is almost six times faster than the (experimentally measured) cell-cycle time. Note that this reference time corresponds to a null model in which we can conceive the phenotype to be under the control of a simple regulation (i.e., non-autogenous control). That leads to present a regular exponential increase [$(1 - e^{-\mu t})$, with $\log(2)/\mu$ denoting the cell-cycle time] to reach the induced steady state.

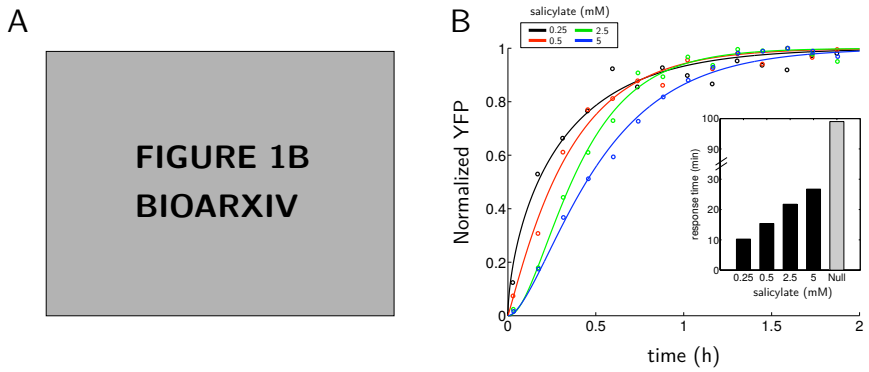


Figure 2.2: Fast dynamic response of the *mar* circuit. A) Dynamic response upon induction with 0.5 mM salicylate (error bars correspond to the mean and standard deviations of three independent replicates; fluorescence values normalized by the maximum). The red line corresponds to the fit to $(1 - e^{-\lambda t})^m$ with $\lambda = 2.92 \pm 0.31\,\mathrm{h}^{-1}$ and $m = 1.08 \pm 0.15$. This gives a response time (time to reach half of the steady state concentration) of t_{50} = 15.35 $\pm$ 3.87 min. This time is faster than the cell-cycle time ($t_{\mathrm{cell-cycle}}$ = 86.64 $\pm$ 5.42 min, obtained from the experimental growth curves, $\mu = 0.48 \pm 0.03\,\mathrm{h}^{-1}$), which represents t_{50} of a null model that would assume constant transcription rate. B) Normalized response of the system to four different concentrations of salicylate. Experimental values (small circles represent averages of three replicates) and theoretical models [lines are the fit to $(1 - e^{-\lambda t})^m$] are shown. The inset shows the corresponding t_{50} values (gray bar denotes response time of the null dynamics, $t_{\mathrm{cell-cycle}}$ = 99.02 $\pm$ 9.43 min, which was calculated with the growth rate at 5 mM salicylate, $\mu = 0.42 \pm 0.04\,\mathrm{h}^{-1}$)

We expected the negative autoregulation elicited by MarR to be the primary responsible for the observed speedup[190,182,29]. If this is the case, we would expect that t_{50} would change with

the amount of salicylate, as dosage eventually determines the number of free MarR molecules and therefore the extent of the feedback repression (note that MarR is a strong repressor[194]). Indeed, the higher the dosage the slower the observed response (fig. 2.2B).

2.3 The effect of copper as mediator of MarR inactivation

Interestingly, we observed that the behavior of the system was still far (in terms of response time) from that predicted without negative autoregulation at saturating levels of inducer, in which free MarR molecules should be mostly absent. We hypothesized that this could be linked to the particular action of salicylate, which inactivates MarR by shifting the intracellular redox equilibrium of copper from Cu^{2+} to Cu^{+} (the system is not induced in anaerobic conditions, when this balance is potentially modified, fig. 2.3A). More in detail, Cu^{2+} oxidizes a residue of MarR that causes tetramerization and repressor dissociation from the *mar* promoter[97]. If this mechanism could not ultimately titrate all repressor molecules, it would impose a mandatory reduction of response time.

We included this effect explicitly in a mathematical model of the circuit that also incorporates the specific regulatory architecture (see appendix A for detailed description of the model, including nominal parameter values). Simulation of the dynamics confirms how the fraction of non-oxidized MarR at saturating concentrations of inducer (α) eventually modulates response time. This is illustrated in fig. 2.3B, where we represented how response time is affected by this parameter (accounting for different intracellular concentrations of Cu^{2+}). By using recent data on the action of Cu^{2+} on MarR[97], we estimated that the specific value of α should be around 5% (appendix A and fig. 2.3C). This value predicts well the t_{50} times observed experimentally (inset in Figure ??D).

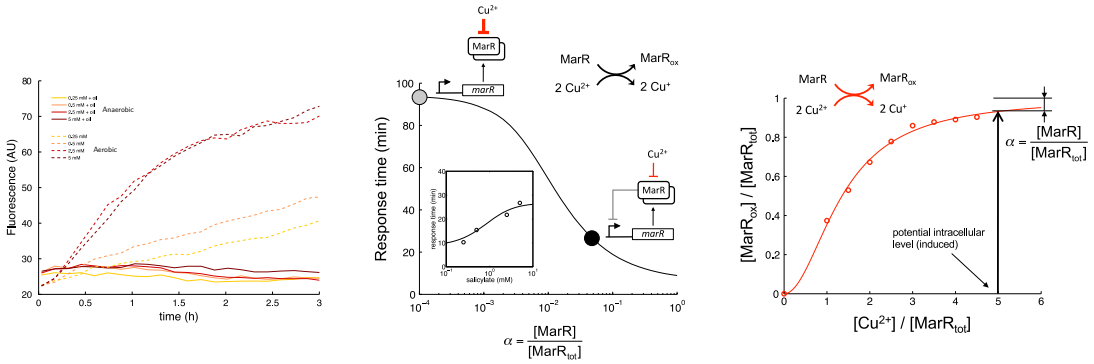


Figure 2.3: The effect of copper in MarR inactivation A) Dynamic response of the wild-type system (YFP fluorescence) upon induction with different concentrations of salicylate in aerobic (dashed lines) and anaerobic (solid lines) conditions. Fluorescence values are ?-replicate averages. Anaerobic conditions were achieved by using mineral oil to cover the cultures. At least partially, the lack of responsiveness of the *mar* circuit to salycylate under anaerobic conditions may be explained by the fact that copper appears to shift intracellularly from Cu^{2+} to Cu^{+} in absence of oxygen[?]. B) Model predictions of how t_{50} would change as a function of the remaining fraction of non-oxidized MarR (α) in saturating salicylate conditions (10 mM). The black point corresponds to the predicted value with $\alpha \simeq 5\%$, which we estimated from experimental data (Supplemental Information), while the gray point describes a limiting regime in which most MarR molecules are titrated (i.e., no repression). The inset shows, with that value of α, a good correspondence between model predictions and experimental values of the response to different dosages of salicylate. C) Transfer function of copper regulation relating the fraction of oxidized MarR ($MarR_{ox}$) with the relative amount of Cu^{2+}. Experimental values (small circles, data taken from Hao et al., 2014[?]) represent averages of replicates of the relative to the maximum oxidation. The line corresponds to $\frac{\left(\frac{[Cu^{2+}]}{[MarR_{tot}]}\right)^2}{K_{ox}^2+\left(\frac{[Cu^{2+}]}{[MarR_{tot}]}\right)^2}$, fitted with $K_{ox} = 1.35$. At first order, $\alpha \approx \left(K_{ox}\frac{[MarR_{tot}]}{[Cu^{2+}]}\right)^2$. The intracellular copper levels upon indiction with salycilate may be about $10\mu M$ (Rensing-Grass, 2003) and the concentration of MarR about $2\mu M$ (MartinRosner2004). This gives $\alpha \approx 5\%$.

2.4 Positive autoregulation by MarA speeds up the response

If the dual circuit accelerates the response due to the action of the negative autoregulation, what is the effect of the positive one? In general, both the strength and nonlinearity of the feedback could alter expression dynamics of a gene under dual autogenous control (fig. 2.4). Note first that the mere presence of a positive autoregulation could decrease response time by reinforcing promoter strength (understood as the binding affinity of the activator to the promoter, fig. 2.4C). In contrast, either high expression fold-change or feedback nonlinearity

generate delays (fig. 2.4D).

These latter aspects are minimized in the *mar* system due to the weak and monomeric action of MarA[141]. Thus, the influence of MarA represents effectively an increase in promoter strength *on the fly* –i.e., as long as the system is expressed– that combined with the negative autoregulation (through MarR) additionally speeds up the response (as compared to the negatively autoregulated circuit). This is demonstrated by both simulations and experiments (fig. 2.5).

Moreover, the combination of autoregulations originates as well pulses in promoter activity (2.6A; we calculated promoter activity by combining independent data of YFP and cell growth dynamics, see appendix A). This pulse is a consequence of an inflexion point in activity (fig. 2.6B), which also reflects in the dynamics of the activator due to its short half-life (fig. 2.6D; we considered the concentration of MarA proportional to promoter activity). The specific geometry of the (MarA, MarR) two-dimensional phase space (fig. 2.6C, mathematical model in appendix A) further highlights i) the different time scales of the system (fast activation since MarA is actively degraded by the Lon protease[91], ii) slow repression since MarR presents a low translation rate[145]), and ii) how the activator and repressor present similar protein expression levels –quick degradation compensates inefficient translation (note that results in Figure 2 correspond to a Δ*rob* strain to control for the additional effect of Rob, see below).

2.5 Effect of additional feedbacks by Rob and marB

Other molecular components related to the *mar* circuit could also affect its response time. In particular, we first investigated the influence of Rob, the main transcriptional cross-talk

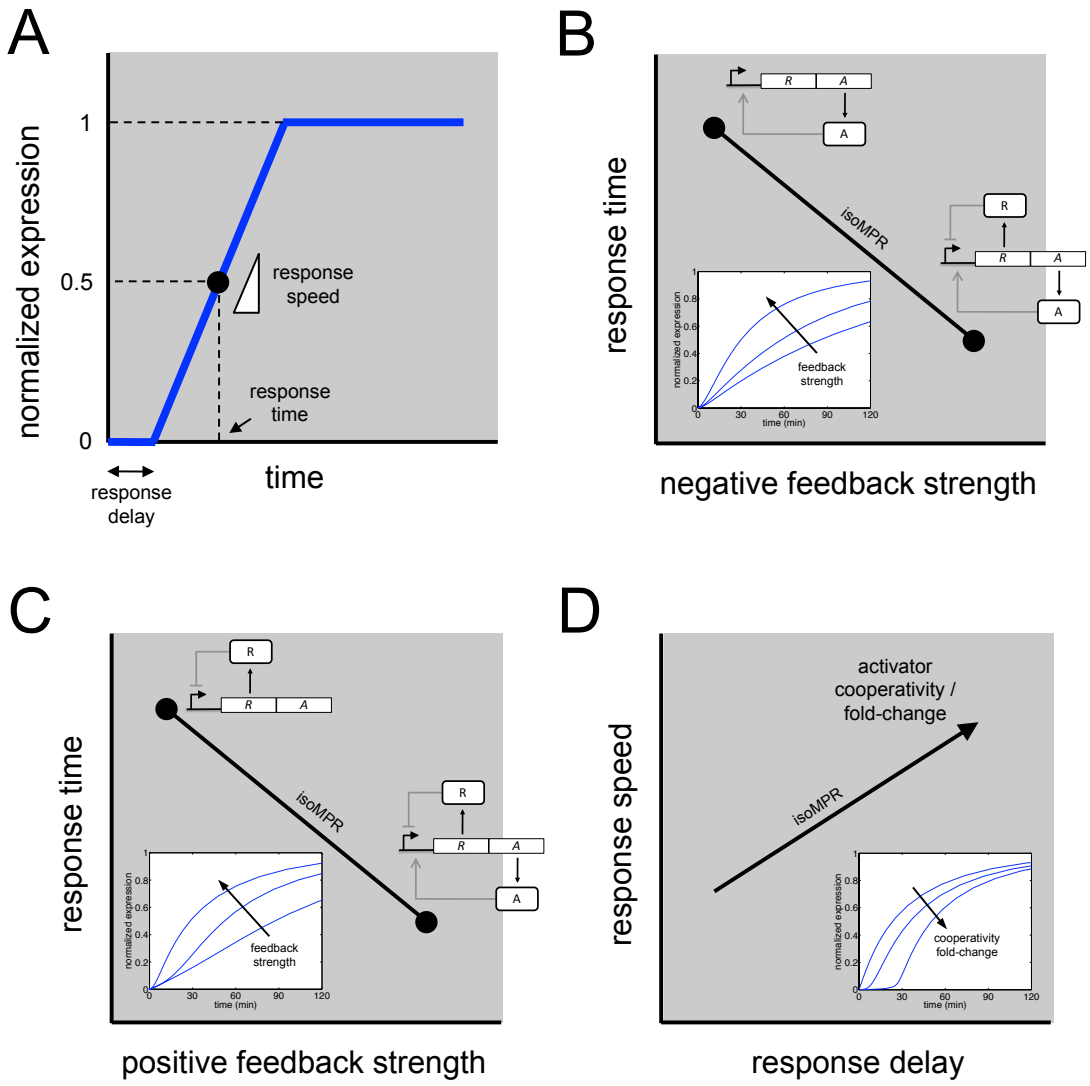


Figure 2.4: Analysis of the dynamic response of a circuit with antagonistic autoregulation A) Scheme of the dynamic response. B) The response time depends with a decreasing trend on the negative feedback strength (understood as the binding affinity of the repressor to the promoter), being constant the maximal production rate of the operon (isoMPR). The inset shows model simulations. Note that here the repressor is a dimer. C) The response time depends with a decreasing trend on the positive feedback strength (understood as the binding affinity of the activator to the promoter). The inset shows model simulations. Note that here the activator is a monomer. D) The response speed and response delay are positively correlated and they depend with an increasing trend on the degree of multimerization of the activator (cooperativity) and its ability to stimulate of RNA polymerase (fold-change). The inset shows model simulations.

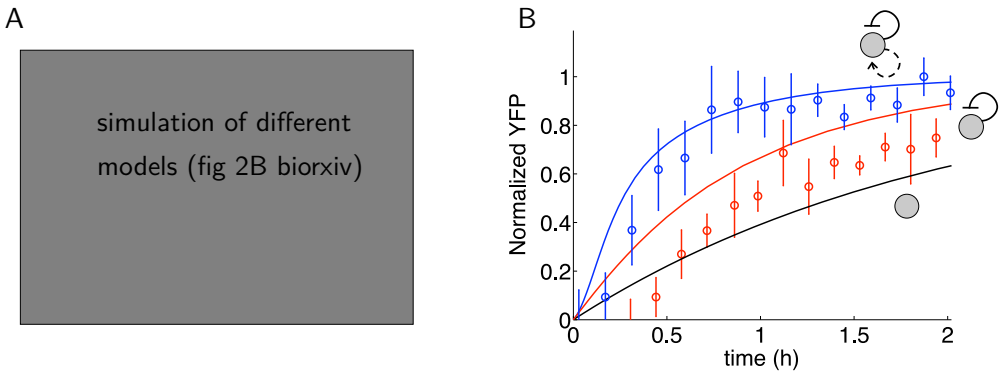


Figure 2.5: Positive autoregulation influence in the mar response. A) Model simulations of MarR dynamics [normalized by the steady state (ss) value] upon induction with 5 mM salicylate for alternative regulatory implementations to control the mar phenotype. We compared dual autogenous control with (wild-type, solid blue line) or without Rob activation (δ*rob*, dashed blue line), single autogenous control (positive or negative, dotted blue lines), and no regulation (black line). All simulations performed by using nominal parameter values (Methods). B) Dynamic response upon induction with 0.5 mM salicylate of Δ*rob* and Δ*rob*Δ*marA* systems (blue and red, respectively; error bars correspond to the mean and s.d. of three independent replicates; fluorescence values normalized by the maximum). Solid blue/red lines are simulations of each system with a fully predictive mathematical model. For Δ*rob* (dual autoregulation), we obtained a response time (time to reach half of the steady state concentration) of $t_{50} = 24.77 \pm 10.04$ min, by fitting to $(1e^{\,t})^m$, which is faster than for Δ*rob*Δ*marA* (single autoregulation), $t_{50} = 65.67 \pm 15.66$ min (U-test $p < 0.01$; fitted curves not shown). The cell cycle time is $t_{cc} = 83.18 \pm 2.09$ min (obtained from the experimental growth curves, $\mu = 0.50 \pm 0.01 h^{-1}$), which represents t_{50} of a null model that would assume constant transcription rate (black curve).

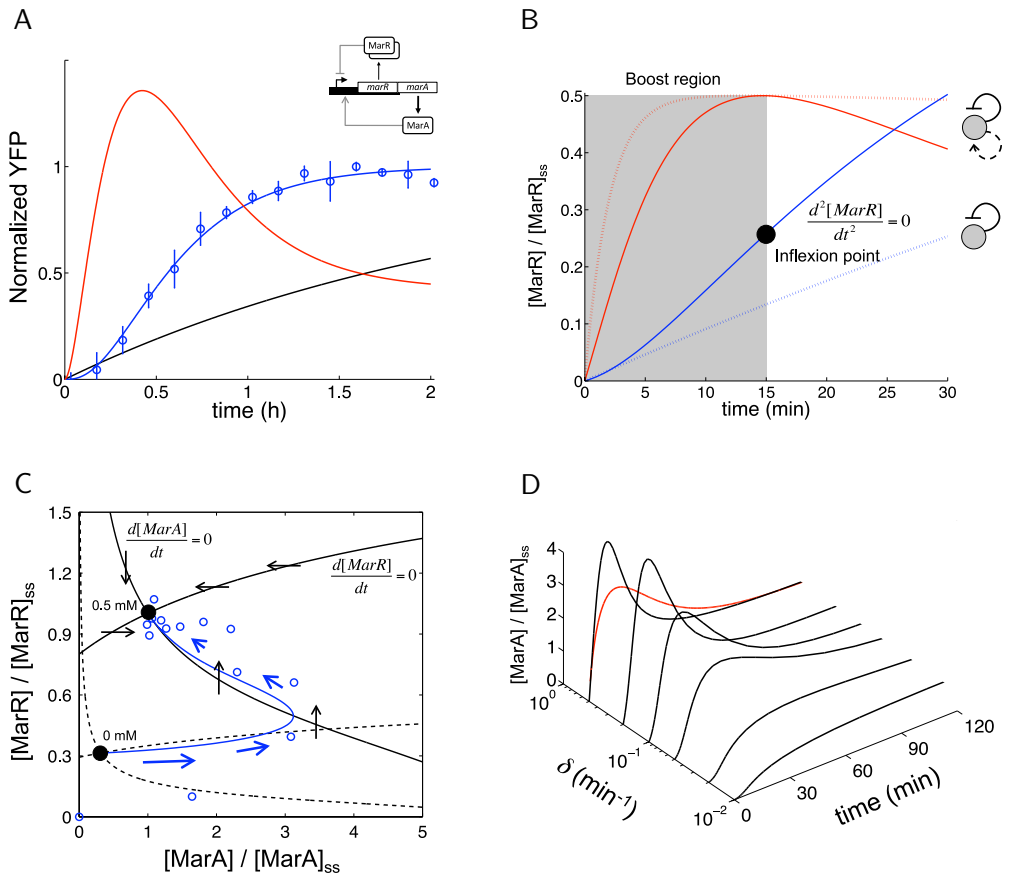


Figure 2.6: Positive autoregulation influence in the mar response. A) Normalized response upon induction (5 mM salicylate) for the antagonistic autogenous system (blue circles; mean and standard deviations of three independent replicates; fluorescence normalized by the maximum; Δ*rob* system). We fitted these data to $\frac{[YFP](t)}{[YFP]_\infty} = \left(1 - e^{\lambda t}\right)^m$ (λ and m being the two parameters that control the dynamics of the response, and $[YFP]_\infty$ the steady-state concentration) with $\lambda = 2.81 \pm 0.68h^{-1}$ and $m = 2.99 \pm 1.39$ (blue line). We then obtained the corresponding promoter activity (red line). This was compared to the simple transcriptional unit (black line, $\mu = 0.43 \pm 0.01h^{-1}$). The response time of the system was $t_{50} = 33.64$ min, and the boost time (i.e. when promoter activity is maximal), was $\tau = 25.81$ min. See Appendix ?? for calculation details. B) Controlled comparison between the dynamics (model-based simulations) of circuits with dual (solid line) or negative autogenous regulation (dotted line) at short times. Note the inflexion point (black point) where the dynamics changes its curvature and promoter activity is maximal (red line, normalized to help visualization). The highlighted (boost) region associates to increasing promoter activity. C) Two-dimensional phase space associated to ($[MarA]$, $[MarR]$) dynamics. Nullclines (black curves; solid for induced and dashed for uninduced situations) represent the trajectories where only the concentration of MarR or MarA changes. Steady states (black points) are given by the intersection nullclines. We represent a trajectory upon induction (0.5 mM); solid line is the simulation and circles the experimental data. Arrows represent direction and strength of change. F) Model simulations of the MarA concentration (relative to steady state value) upon induction with 0.5mM salicylate for different degradation rates of this protein (δ). When the protein is unstable, it presents a pulse-like dynamics, ``following'' promoter activity (red line denotes experimental promoter activity).

experienced by the system under physiological conditions[40]. Our model predicted a slight delay when no feedback through Rob was included (fig 2.5A). To test this, we constructed a Δ*rob* strain in the previous P_{mar}:*yfp* background. The response of the Δ*rob* strain is still fast compared to the cell-cycle time, but slower than the wild-type system [$t_{50} = 33.64 \pm 4.25$ min (Δ*rob*) vs. $t_{50} = 26.65 \pm 3.99$ min (wild-type), *U*-test $p < 0.01$]. This difference in dynamics is linked to the effect of Rob on the transcription rate of the *mar* promoter. Since the presence of Rob can be understood as an enhancing effect on promoter activity (fig. 2.7A-C), its absence necessarily causes a slower response (simulations confirm how response time increases by weakening promoter strength).

In addition to MarR and MarA, the marRAB operon contains another open reading frame coding for marB[41], a small periplasmic protein. Recently, MarB has been identified as a repressor of the marRAB promoter[157], establishing a new putative negative feedback into the system. To study this novel regulation on the dynamic response, we knocked out the *marB* gene. A quantification of transcripts revealed that MarB is not able to repress the expression of the marRAB operon in presence of salicylate, although MarB does repress the operon in absence of it (Fig. S2N), which is tune with previous results[224]. We confirmed by fluorescence assays the observation that MarB does not repress the marRAB promoter upon induction with salicylate, even for moderate concentrations of this inducer (fig. 2.7D). While MarB does not impact the dynamic range, we observed that the Δ$marB$ system exhibits a slower response to salicylate. Our results suggest that the bacterial cell might exploit MarB as an enhancer of MarR to tightly repress the operon. They also revealed that Rob and MarB are regulatory elements that contribute to a fast mounting of the mar phenotype.

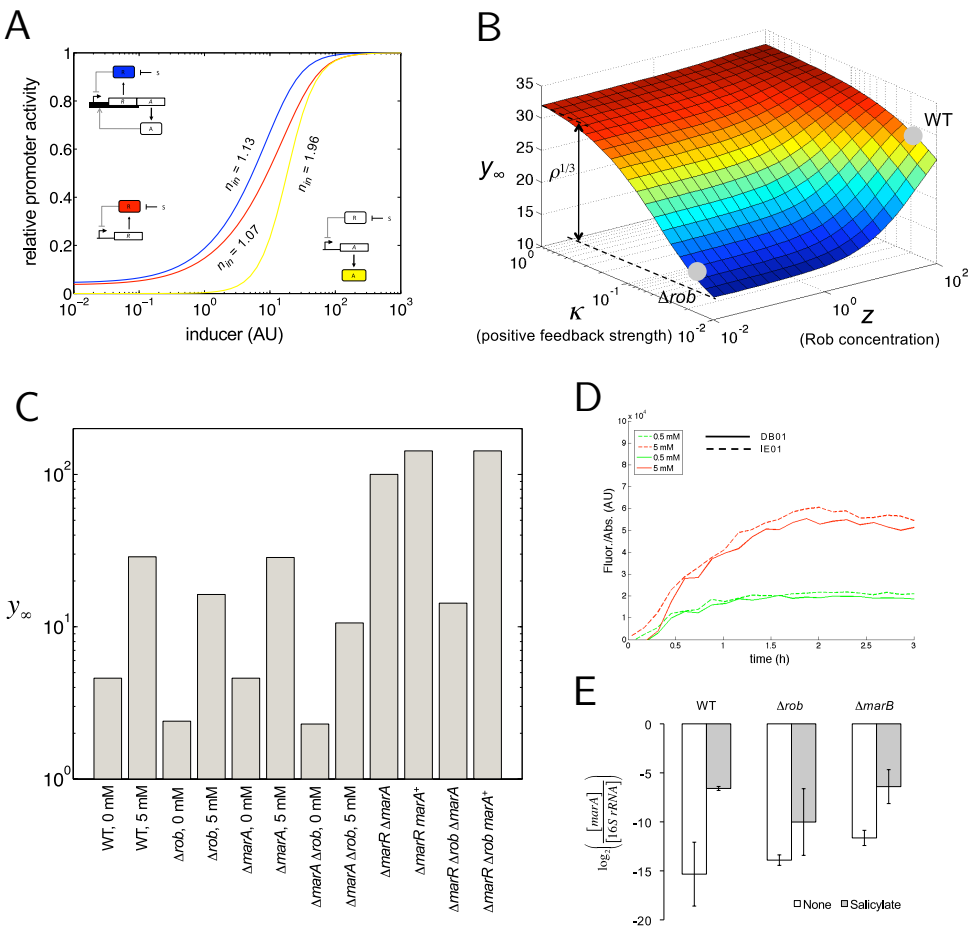


Figure 2.7: Influence of Rob and MarB. A) Model predictions of relative promoter activity in different systems (in steady state) as a function of the inducer. We show the wt circuit (blue), a negatively auto-regulated sensor gene (red) and a consititutively expressed sensor gene (yellow). For simplicity, here we took $v_s = 1$ and $\alpha = 0$. The rest of parameter values as in table A.1. B) The positive feedback strength (parameter κ) and Rob concentration (parameter z) modulate the steady state of the system (simulations for 5mM salicylate). Despite, the output dynamic range is almost constant. C) Model predictions for different genotypes and salicylate levels. The predictions show $\pi_{mar}(5mM)/\pi_{mar}(0mM)|_{WT} \approx \pi_{mar}(5mM)/\pi_{mar}(0mM)|_{rob} \approx \alpha^{2/3} = 7.3$, and $\pi_{mar}(5mM)/\pi_{mar}(0mM)|_{WT} \approx \rho = 10$. In addition, they show $\pi_{mar}(WT)/\pi mar(rob)|_{0mM} \approx \pi_{mar}(WT)/\pi_{mar}(rob)|_{5mM} \approx \rho^{1/3} \approx 2$. Finally, they show $\pi_{mar}(marA^+)/\pi_{mar}(marA)|_{marR} \approx 1.4$, and $\pi_{mar}(marA^+)/\pi_{mar}(marA)|_{marR,rob} \approx \rho = 10$. These predictions are in tune with previous experimental results (MartinRosner, 1997). D) Dynamic response of the Δ*marB* system to salicylate. Solid lines correspond to the Δ*marB* system, whereas dashed lines correspond to the wt. In both cases, we represent the normalized fluorescence (YFP) with time. Fluorescence values are averages of replicates. E) Characterization by qRT-PCR of the intracellular RNA concentration of gene marA for different strains (wild-type, Δ*rob*, Δ*marB*), with 5mM and without salicylate. We report the ratio between the expression of gene *marA* and a constitutive housekeeping gene (16S ribosomal RNA, rRNA). Error bars correspond to standard deviations.

2.6 Rob reduces the cross-talk between antibiotic and metabolic stresses

The regulation of the *mar* promoter by additional transcription factors makes it responsive to other signals[184] like oxidative stress (through SoxS), DNA supercoiling (through Fis), or metabolic/catabolic stress (through CRP, fig. 2.1). Since cAMP-CRP signaling is instrumental to coordinate gene expression in multiple situations[240], we decided to analyze how the promoter integrates both the antibiotic and metabolic signals as a two-dimensional transfer function[115]. Because the *mar* and *crp* operators overlap, we extended our model by considering competitive binding between the activators MarA, Rob and CRP (fig. 2.8A). In fig. 2.8B, we show the promoter activity obtained experimentally (in steady state) for several combinations of salicylate and cAMP. The response appeared almost independent of cAMP. We observed however a significant effect of cAMP on the Δ*rob* strain (fig. 2.8C). This effect was predicted by our simulations (fig. 2.8D-E). Thus, Rob appears as a regulatory element that *E. coli* employs, in addition to fine tune response time, to isolate the *mar* phenotype from metabolic signals such as cAMP.

2.7 The *mar* circuit presents wide input and moderate output dynamic ranges

We extended the characterization of the *mar* phenotype with the quantification of the input (R_{in}) and output (R_{out}) dynamic ranges of the regulatory module[83]. To this aim, promoter activity was experimentally measured in steady state for different concentrations of salicylate. We then fitted the curve to a sigmoidal to obtain $R_{in} = 31.11 \pm 2.46$ and $R_{out} = 8.85 \pm 0.04$ for the wild-type system (fig. 2.9A). This regime of values, which are captured by our model (fig. 2.9B), demonstrates that dual regulation increases the sensitivity of the response with respect to a circuit without feedback, in a qualitatively similar manner to negative autoregu-

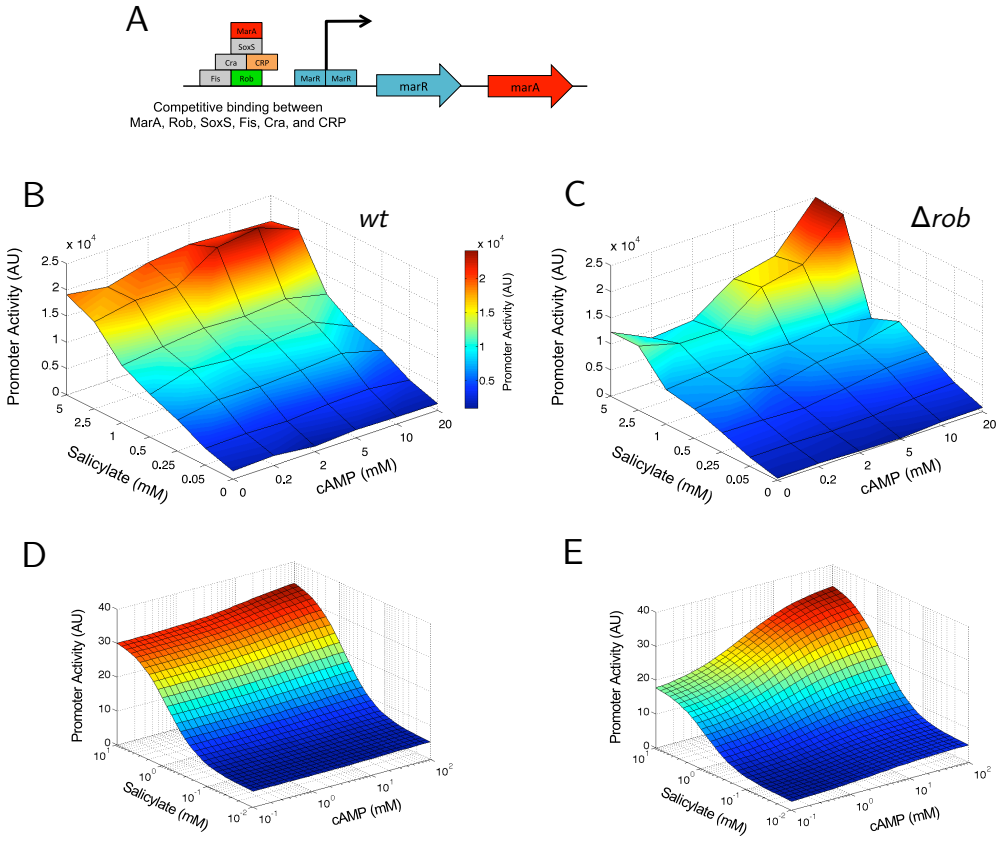


Figure 2.8: Two-dimensional transfer function of the *mar* circuit response. Promoter activity in steady state [arbitrary units (AU)] is represented for each of the 42 combinations of the two input signals, salicylate and cAMP. A) Wild-type system; B) Δrob system. In C) and D) we show model-based predictions. We used the parameter values shown in table A.1.

lation (fig. 2.7A)[155,137].

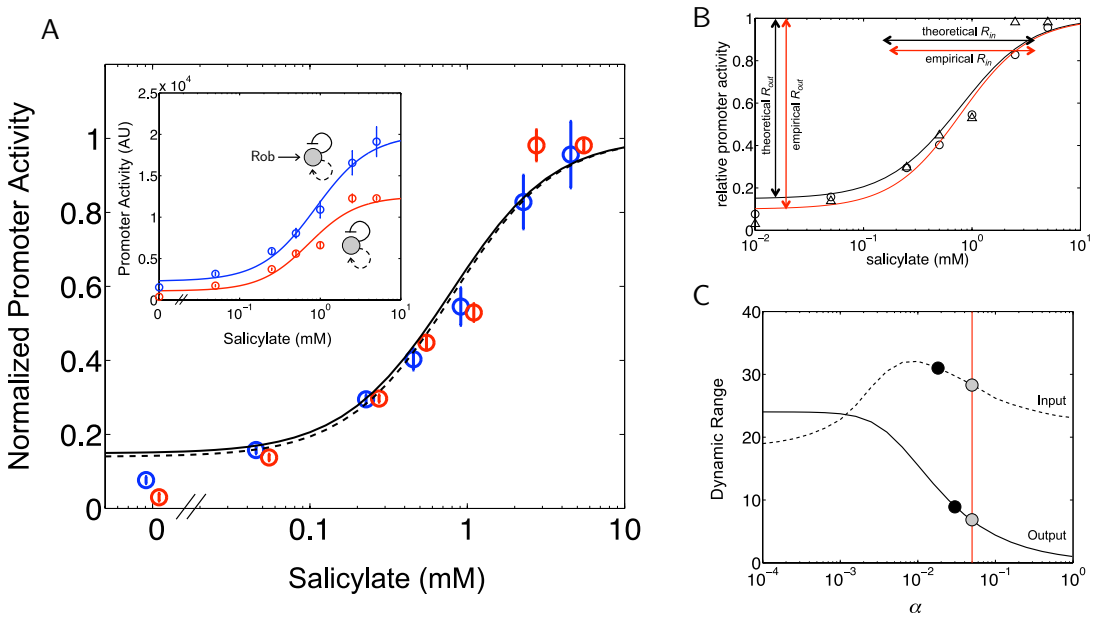


Figure 2.9: Dynamic range in the *mar* circuit A) Normalized promoter activity in steady state (relative to the maximum) is represented for each concentration of salicylate. Open circles and error bars correspond to experimental data (blue for the wt and red for the δ*rob* systems), while solid and dashed lines correspond to model predictions (wild-type and δ*rob*, respectively). The inset shows promoter activity in steady state [arbitrary units (AU)]. A Hill-like model was fitted independently to obtain the input and output dynamic ranges [$R_{in} = 31.11 \pm 2.46$ and $R_{out} = 8.85 \pm 0.04$ (wt), $R_{in} = 20.88 \pm 1.17$ and $R_{out} = 11.48 \pm 1.03$ (δ*rob*)]. B) Model prediction of the transfer function of the system. Relative promoter activity (in steady state) of the system as a function of salicylate. Solid black line represents model simulations, whereas circles (wild-type) and triangles (δ*rob*) are experimental data. We took the parameter values shown in table A.1. In addition, the Hill-like model $\Pi_{mar} = \frac{\Pi_{mar}^{min} + ([Sal]/K_s)^{n_s}}{1 + ([Sal]/K_s)^{n_s}}$ adjusted to all experimental data (from wt and δ*rob* systems; red line), gave $\Pi_{min} 0.10 \pm 0.04$, $K_S = 0.79 \pm 0.10$mM, and $n_S = 1.36 \pm 0.21$. The plot shows comparatively theoretical and empirical values of input and output dynamic ranges. C) Model simulation of the change in input (dashed) and output (solid) dynamic ranges of the system with the remaining fraction of non-oxidized MarR upon induction with salicylate (α). Black points correspond to the experimental values reported for the wild-type system, whereas gray points are model predictions for the chosen parameterization. The value used was $\alpha \approx 5\%$, inferred from the oxidation curve with copper (vertical red line).

Rob appears not to change much the previous input/output ranges but rather to scale the response according to simulations of our model (fig. 2.7B). The scaling factor can be approximated by $\rho^{1/3}$, with ρ denoting the activation fold by MarA and Rob (see Supplemental In-

formation). To evaluate these predictions, we produced a experimental dose-response curve of the Δ*rob* strain and fitted it again to a sigmoidal. The data confirmed that absence of Rob does not alter much the range values (fig. 2.9A), but scaled the expression of the system (the wild-type dose response is about 1.6-fold that of the mutant). Indeed, the fold change observed ($1.6^3 \simeq 4$) agrees with experimental estimations of the activation fold by MarA[143,40] (see Figure S3D for other experimentally obtained fold-change values that are captured by our model).

We asked also if the remaining non-oxidized MarR (upon induction with salicylate) influenced dynamic ranges as it did previously with response time. Figure 2.9C shows the (model) predictions of the change in R_{in} and R_{out} with α, together with the experimental values reported in this work. Note that the output dynamic range follows a similar dependence with α as the response time (i.e., the higher is the derepression of the system, the higher and slower is the induction). A fast response then exhibits moderate R_{out}. This trade-off is nicely captured by solving analytically the model, giving $R_{out} = \alpha^{-2/3}$ and $t_{50} \propto \alpha^{-1}$ (Supplemental Information). We could then calculate in an alternative manner α from the experimental value of R_{out}. We obtained a value of $\alpha \simeq 4\%$, which corroborates the value inferred from the oxidation curve with copper ($\alpha \simeq 5\%$). In contrast to the monotonous trend of R_{out} with α, R_{in} presents a maximum, although not very pronounced.

The polycistronic implementation of dual autogenous control shows higher sensitivity and faster response

An alternative genetic implementation of dual autogenous control could involve dual regulators. For instance, in *E. coli* there exist three systems regulated by this type of architecture:

CRP, ChbR, and LldR. These transcription factors work as repressors that turn into activators in response to cAMP[108], chitobiose[167], and lactate[1], respectively (table ??. Based on these cases, we imagined an *hypothetical mar* circuit in which the oxidized MarR functions as an activator (fig. 2.10). We modified accordingly our model to solve the new dynamics (appendix A). The natural circuit showed higher R_{in} (approximately 5-fold to the hypothetical one), whereas the hypothetical circuit exhibited much higher R_{out} (about 9-fold to the natural one, fig. ??A).

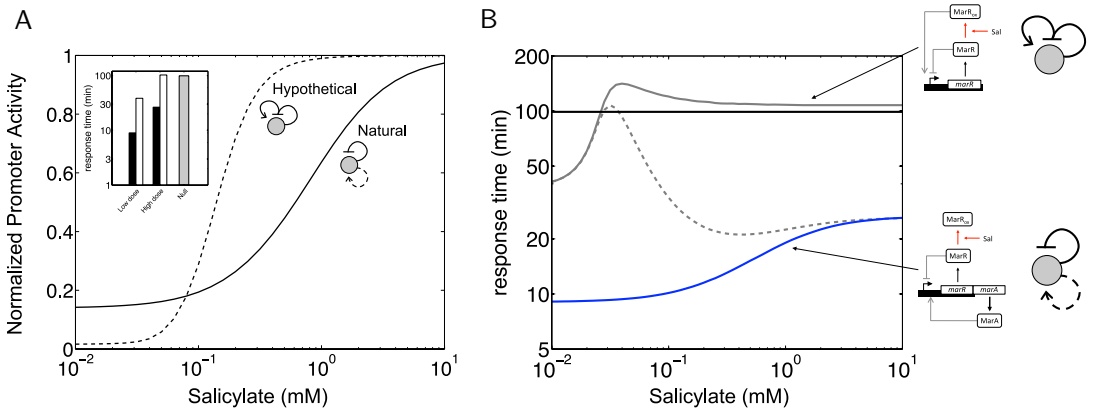


Figure 2.10: Comparison with a hypothetical alternative architecture A) Model predictions of normalized promoter activity with salicylate. We show the Δ*rob* circuit (solid line; $R_{in} = 27.51$ and $R_{out} = 7.17$), whereas the dashed line corresponds to a hypothetical circuit where the oxidized MarR acts as an activator ($R_{in} = 5.16$ and $R_{out} = 62.17$). The inset shows the corresponding response times of these two implementations of dual autogenous control. Black bars correspond to the natural circuit, whereas hatched bars to the hypothetical circuit (gray bar is for the null model). Low dose corresponds to 0.01mM and high dose to 10 mM. B) Simulations to analyze the response time of two implementations of the dual autogenous control system. The blue curve corresponds to the wild-type mar circuit. The gray curves correspond to a hypothetical situation where the oxidized MarR acts as an activator. Solid gray curve for competitive binding between the repressor and activator, and dashed gray curve for independent binding. Parameter values from table A.1.

Table 2.1: *E. coli* operons with dual autogenous regulation from RegulonDB [187].

Regulator	Operon	Note	Molecular implementation
ChbR	*chbBCARFG*	Repressor that becomes activator in response to chitobiose. AraC/XylS family of transcriptional regulators.	Multifunctional regulator
CRP	*crp*	Repressor that becomes activator in response to cAMP. CRP-FNR superfamily of transcription factors.	Multifunctional regulator
LldR	*lldPRD*	Repressor that becomes activator in response to lactate. GntR family of transcriptional regulators.	Multifunctional regulator
marR, marA	*marRAB*	Control of multiple antibiotic resistance. AraC/XylS family of transcriptional regulators.	Distributed regulation

This suggests that the specific architecture of the *mar* circuit, which is unique in the *E. coli* genome (table 2.1), preferentially evolved to exhibit higher sensitivity to gradients of pollutant concentrations, while maintaining moderate fold change. Alternatively, responses governed by dual regulators could require a wider output range to either fine tune very large regulons (this applies for instance to CRP; note that the *crp* and *mar* modules are connected) or induce fairly digital responses [e.g., to activate the response under a very narrow signal range, as it is the case of ChbR, which activates the expression of the *chb* operon only when sufficient flux through their associated pathway is sensed[167]]. In addition, the response dynamics of the natural circuit would be faster than the hypothetical one (inset in fig. ??A and ??B). More in detail, we found that the hypothetical system would respond similarly to one under simple regulation at high dosages (activation would dominate repression) but exhibit speedup at very low dosages. In the latter situation, there would be a combined action of positive and negative regulation over the promoter, similar to the one governing the induction of the *mar* circuit.

The regulatory architecture of the marRAB operon is suitable to present excitable behavior (and even to produce sustained oscillations) in the non-induced state, due to the interlinked, delay-based positive and negative feedbacks[99,93]. For this architecture, oscillations appear when a clear separation of time scales between activator and repressor dynamics exists. This implies that i) the activator degradation should be stronger than the repressor one (in our case, MarA is quickly degraded by a protease, whereas MarR is diluted due to cell growth rate), and ii) the activator translation rate should be also stronger than the repressor one (here, the RBS of MarA is stronger than the one of MarR). Nevertheless, simulation of the stochastic dynamics of the system (appendix A, fig 2.11A) displays fluctuations out of the (deterministic) equilibrium caused by the mixture of gene expression noise[153] and the action of dual regulation (fig. 2.11D).

The dynamics of the circuit can be further contemplated by modifying some of its basic attributes. Firstly, in the mar system, the activator (MarA) works as a monomer. A variant model that introduced cooperativity in both regulators – leading to much faster dynamics – did display oscillations (fig. 2.11C and F). Moreover, when competitive binding between MarA and MarR is considered (in particular, that MarA prevents MarR binding[77], stochastic pulses can appear, even with a monomeric activator (fig. 2.11B and E). Previous experiments support however relatively independent action of MarA and MarR[146], aside from their operators do not overlap.

To confirm the predictions of our original model, we followed experimentally the dynamics of a single cell in absence of salicylate. Figure 2.11G illustrates a representative trajectory that corroborates the noisy dynamics around steady state without the emergence of signifi-

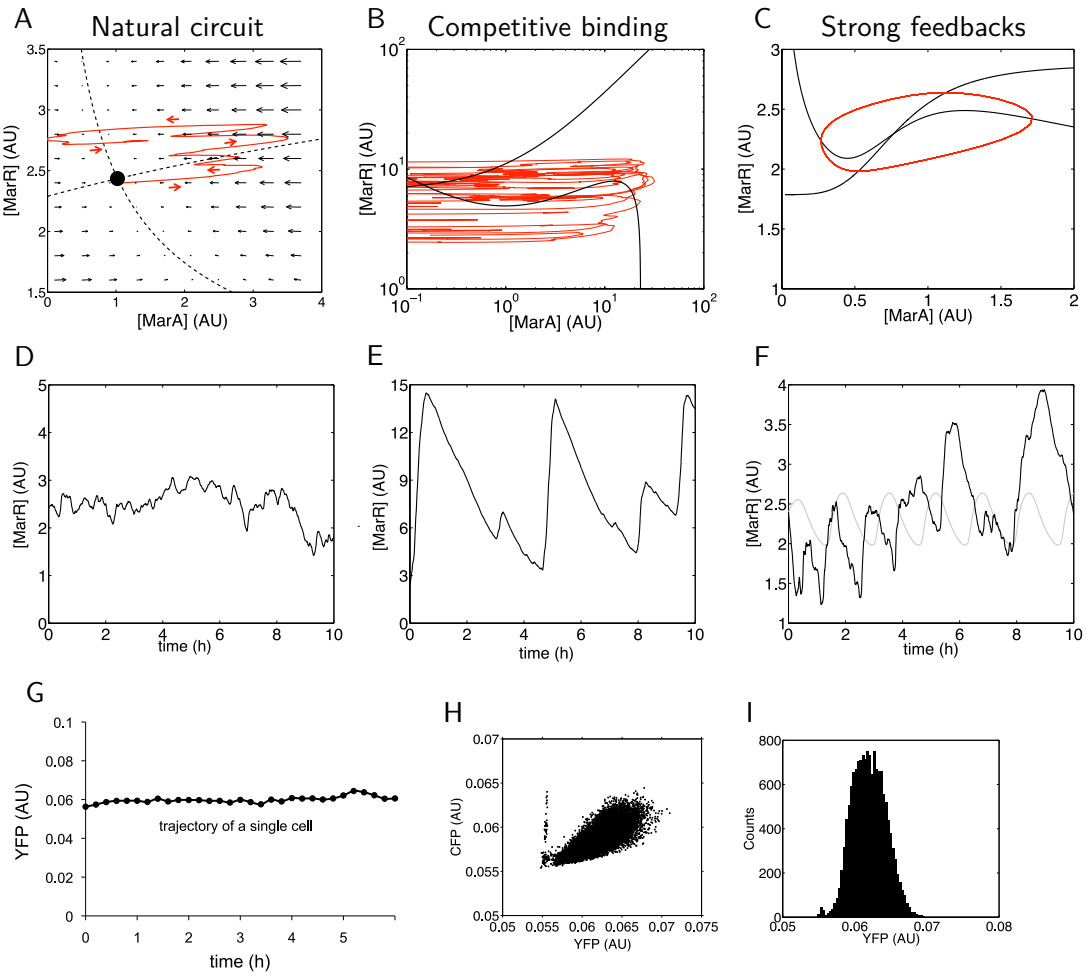


Figure 2.11: Hypothetical effect of strong and competitive autoregulation A) We plot part of a trajectory in the phase space of the system (red arrows indicate time evolution). This representation highlights how gene expression levels fluctuate around the deterministic steady state (black point) by the combined action of intrinsic/extrinsic noise, and dual autogenous control. Dashed lines correspond to the nullclines of the system, and field arrows denote the strength of change towards the steady state. We used the following parameter values: $\delta = 0.15\,\mathrm{min}^{-1}$, $\mu = 0.01\,\mathrm{min}^{-1}$, $\pi_0 = 0.2\,\mathrm{min}^{-1}$, $\rho = 10, \kappa = 1, \beta = 5$, and $z = c = 0$. B) Phase space of the system in absence of salicylate showing stochastic pulses when MarA competitively prevents MarR binding. To model this, we replaced y_0 in Eqs. (S11) by $y_0/(1+x)$. The black lines correspond to the nullclines, with a different shape than for the wt system. The red curve represents a stochastic trajectory. We For these plots, we used the parameter values shown in table A.1, with $z = 0$. C) Phase space of the system in absence of salicylate with cooperativity in marA. We assumed high cooperativity (Hill coefficient of 4) for both MarA and MarR. Here, the red trajectory is deterministic. D-F) Sample trajectories in time for the models in A-C, respectively. G) Representative trajectory (YFP) of a single cell (wild-type system). In the inset, trajectory for the Δrob system. H) Scatter plot of single-cell YFP and CFP corresponding to an experiment that follows colony growth and gene expression dynamics without salicylate (wt). Note the correlation in the fluctuations in both reporter proteins (expressed from different promoters). I) Distribution of YFP for all single cells at all time points (wild-type) confirms a unimodal distribution corresponding to a continuous production of MarR.

cant oscillations (much of the cell-to-cell variability in YFP and CFP is correlated, fig. 2.11H). The histogram of variability on YFP suggests a constant production of the proteins of the *mar* operon and the absence of any oscillatory/pulsing dynamics[153] (fig. 2.11H).

2.9 The *mar* response is relatively homogeneous across a population

We additionally investigated the single-cell dynamics in the presence of stress by tracking cells upon induction with 5 mM salicylate at time 0 (fig 2.12A; CFP signal also shown as control). We asked to what extent the *mar* phenotype is expressed differentially, e.g., only triggered in a subset of the population[44]. This is motivated again by the presence of a positive feedback, an architecture typically related to heterogeneous gene expression even when there is no bistable behavior[219]. By following the dynamics of a growing population of cells, we observed however that activation is relatively homogeneous, with unimodal distributions being clearly identified at different times (fig 2.12B), or at various salicylate dosages at steady state (fig 2.12C). We also noticed that noise decreases with salicylate dosage (fig 2.12AD-F). Because our model predicts higher variability by considering dimerization of the activator and absence of external transcriptional signals (fig 2.12G), we attribute the observed coherent behavior across the population to the linear positive feedback. This is in tune with previous work showing that the presence of coupled positive and negative feedback loops contributes to reduce noise in gene expression[111].

To quantify nevertheless the variability that we did observe in the single-cell response, we fitted the mean YFP expression of each cell lineage to an exponential model ($(1 - e^{-\lambda t})^m$, fig. 2.13A-B). This fit allowed us to describe the expression of the phenotype as a composition of two features: response speed (a measure of how much the expression changes per unit of

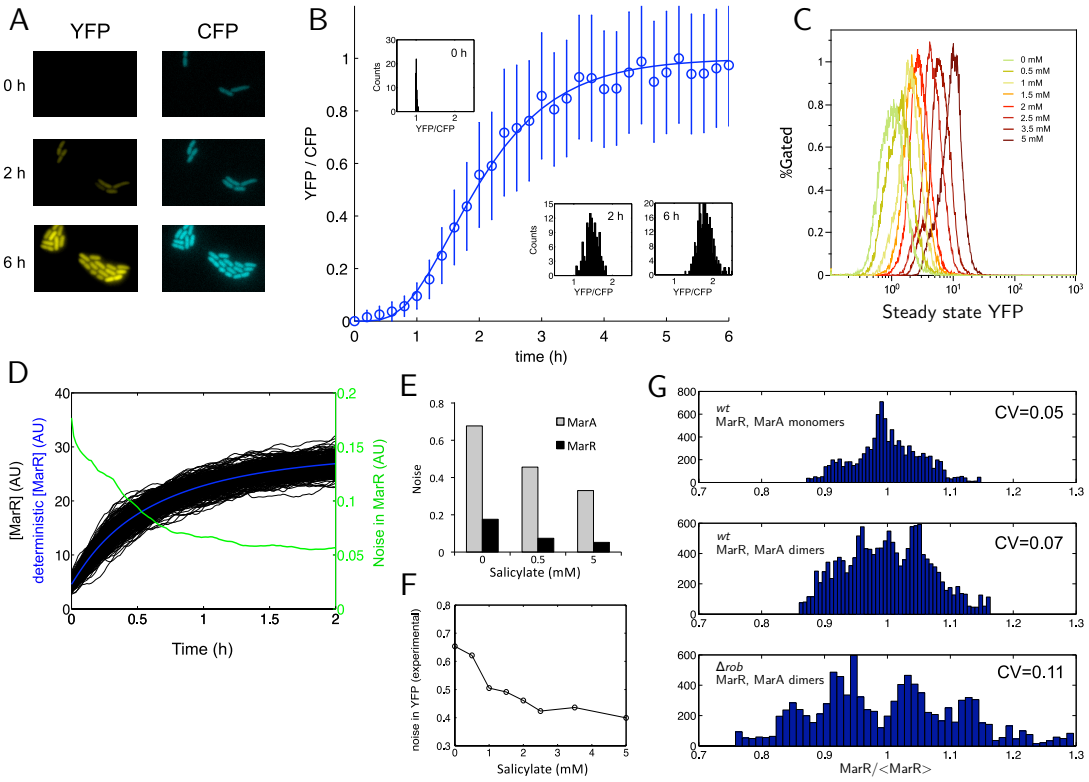


Figure 2.12: Homogeneous response of *mar*.A) Response of the mar promoter (YFP signal) and a control promoter (CFP signal) upon induction with 5 mM salicylate at time 0 (δrob system). B) Graded transcriptional activation of the mar phenotype, measured as the ratio YFP/CFP, in a growing population of single cells upon induction with 5 mM salicylate (insets are histograms at different times, illustrating the relatively homogeneous response). C) Flow cytometry analysis (using YFP intensity) of the δrob system, for several conditions of salicylate. D) Model simulations to study the stochastic response. Different trajectories (simulating different single cells) are shown in black. In green, we represent noise in gene expression (coefficient of variation). In blue, we show the dynamics in the deterministic regime (which approximates very well to the average). E) Model predictions of noise in expression of MarA (η_x, gray bars) and MarR (η_y, black bars) as a function of salicylate.F) Experimental noise calculated with flow cytometry data, for the wild-type system, to validate the results reported in (E). G) Stochastic simulations of distributions of gene expression at 0.5 mM salicylate. On top, wild-type system; in the middle, wild-type system assuming that MarA and Rob work as dimers; on bottom, Δrob system assuming that MarA and Rob work as dimers.

time) and delay (time before increasing expression) (recall figure2.4). These properties are determined by λ and m, respectively, the two parameters characterizing our fit (fig. 2.13D). Since delays in the lineages could be partly associated to peculiar growth conditions rather

than the dynamics linked to the regulatory motif, we used response speed as score to compare to a system without autoregulation. Response speed in the latter case equals the rate of cell division (μ, see Methods). However, fig. 2.13C. shows how the presence of feedbacks always implies a faster speed (note as well the underlying variability).

2.10 Discussion

The mechanism of autogenous control of gene expression involves a genetic program in which the protein encoded by the structural gene is working as its own regulatory element. This entails a number of functional advantages with respect to the *classical* model of regulation[110,82]. Notably, these advantages are specific on whether the autoregulation is positive (activator-controlled) or negative (repressor-controlled)[190]. We focused here on an inducible system that presents both types of autogenous regulation (positive and negative) within the architecture that associates to the *mar* phenotype.

Dual control necessarily integrates contrasting properties of the dynamics. For instance, response time could be expected to be either fast or slow according to earlier studies on negative[182,29] and positive[138] autogenous systems. We observed experimentally a rapid induction that is modulated by the action of salicylate on the strong repression of MarR (fig 2.2). Thus, the faster the response, the stronger the repression which also results in lower expression levels at equilibrium (fig 2.4B). In addition, the speedup appeared influenced by copper signaling, since the accumulation of intracellular Cu^{2+} in response to salicylate oxidizes MarR preventing its binding to DNA[97]. Because the intracellular concentration of this cation is bounded (to avoid toxicity), copper balance imposes a maximal repressor abundance for the promoter to be derepressed (appendix A and fig 2.3C); a constraint that provides a rationale for the

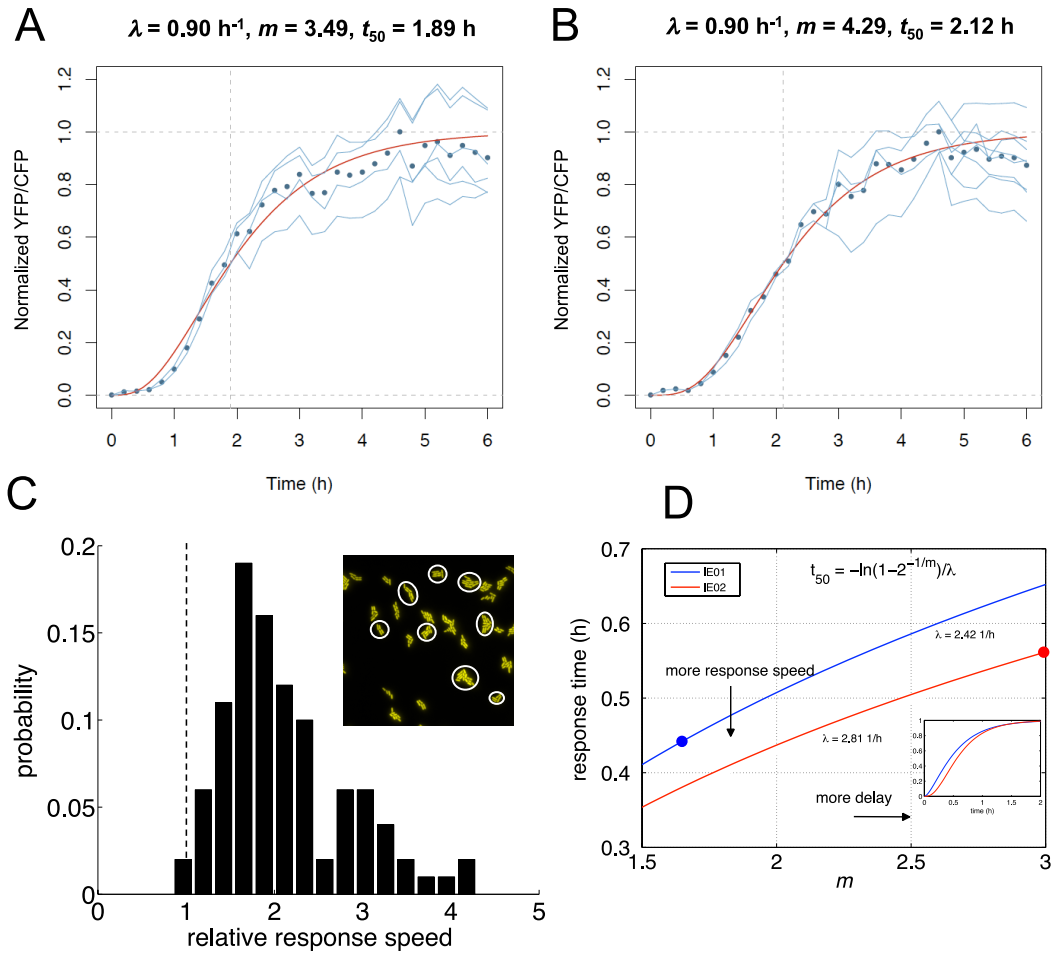


Figure 2.13: Variability in response time of individual lineages A-B) Two representative lineages (δrob strain, 5 mM salicylate added at $t = 0$) showing different response times. A lineage starts from one cell (or few) and grows and divides with time to form a colony. For each lineage we fitted (red curve) the exponential model $(1 - e^{-\lambda t})^m$. We then obtained the t_{50} values. C) Variation in relative response speed (as the ratio between response speed and growth rate) among 100 different lineages within the population. The image shows a subset of lineages. D) Map of the response time, which is determined by i) the time lag to react (empirical parameter m), and ii) the speed at which the expression increases (empirical parameter λ). Open circles correspond the biological systems. The inset shows the experimentally determined dynamics.

limited translation rate observed for MarR[145].

Moreover, the positive autoregulation rather than delaying the response contributes to its

speedup due to the increase *on the fly* of the transcription rate of the promoter (fig 2.4C); dynamics of wild-type and $\Delta marA$ systems shown in fig 2.5B). This coordination between repressor and activator generates pulses in promoter activity (fig. 2.6A), which can be "followed" by the actuator of the regulon (MarA) due to its short half-life[91]. The features that effectively minimized the delay of the response are the weak activation and monomeric action of MarA (fig 2.4D). These attributes keep the phenotype buffered in absence of signal (fig 2.11)[29], limit large-amplitude (and heterogeneous) transient responses (fig 2.12)[34], and ultimately set apart the observed dynamics from that characterized in other biological systems also regulated by interlocked positive and negative feedback loops, e.g.,[208,111].

Interestingly, the essential weakening of the positive control (through MarA) becomes somehow neutralized by the additional regulation of Rob, an interaction that can be effectively interpreted as a mean to increase promoter strength (fig. 2.9A). The absence of Rob causes not only a slightly slower response, but also a reduction in signal specificity (fig. 2.8). While the wild-type circuit decreases the potential cross-talk between antibiotic and metabolic stress signals (the latter connected to the *mar* system through CRP, Figure ??A), the absence of Rob amplifies the transfer between them.

The dual autogenous control of the *mar* phenotype is also unique in its genetic organization. Indeed, this system is the only one in *E. coli* presenting a polycistron carrying both repressor and activator. However, we did identify alternative designs of dual autogenous control by means of dual regulators in this bacterium. These regulators switch between repression or activation in the presence of its cognate inducer (CRP, for instance, works in this way[108]). What are the differences in dynamics between these two implementations? To analyze this, we considered an hypothetical *mar* system in which oxidized MarR turns into an

activator (fig. 2.10). This more compact genetic design exhibits a smaller input range (i.e., the reaction to a signal gradient is more sigmoidal), a wider output range (what enables, for example, to regulate differentially large regulons as it could be the case of CRP), and a slower response (with no speedup at high dosages). This exemplifies that the dynamics corresponding to a given regulatory logic is certainly influenced by its genetic implementation [95,93,34], and that the *mar* control circuit is quite distintive in its organization and function, triggering a response that is fast, graded, sensitive, robust, and coherent across the population.

2.11 Methods

Strains, culture media and reagents

. Two-color fluorescent reporter E. coli strains (IE01, IE02, TC01 and TC02) were engineered to measure the activity of the mar promoter. The strain IE01 contains a chromosomal copy of the yfp gene under the control of the mar promoter, and the cfp gene expressed with a constitutive promoter. The strains IE02 and TC01 were constructed by deletion of the rob and marA genes, respectively, on IE01 with the application of a knockout protocol 34 . The strain TC02 was constructed as the double knockout of rob and marA genes. Medium LB was always used for overnight cultures. Minimal medium M9 (M9 salts 1x, MgSO 4 2 mM, CaCl 2 0.1 mM, glucose 0.4%, casamino acids 0.05%, vitamine B1 0.05%) was used to grow cells during characterization experiments. To induce the mar circuit, different concentrations of salicylate and cAMP (Sigma Aldrich) were used. Note that glucose inhibits the production of internal cAMP. When appropriate, kanamycin was used at 50 μg/mL. See also Supplemental Information for extended experimental procedures.

Quantification of fluorescence in a cell population

. Cultures (2 mL) inoculated from single colonies (three replicates) were grown overnight in LB medium supplemented with glucose 0.4% at 37 0 C and 170 rpm. Cultures were then diluted 1:200 in M9 minimal medium and were grown for 2 h at the same conditions. Cultures were then used to load the wells (200 μL) of the microplate (Thermo Scientific) with salicylate and cAMP when appropriate. The microplate was assayed in a Victor X2 (Perkin Elmer) measuring absorbance (600 nm), YFP (497/16 nm, 535/40 nm), and CFP (434/17 nm, 479/40 nm) for 4 h at 37 0 C with shaking. Analysis of fluorescence data described in Supplemental Information.

Quantification of fluorescence in single cells

. Culture (2 mL) inoculated from a single colony was grown overnight in LB medium supplemented with glucose 0.4% at 37 0 C and 170 rpm. Culture was then diluted 1:200 in M9 minimal medium and was grown for 4 h at the same conditions. Culture diluted 1:10 (2 μL) was then used to load the agarose pad. Before characterization, 5 mM salicylate was added to induce cells. Agarose pads were monitored in an inverted microscope Axiovert200 (Zeiss) with objective 100X/1.45 oil Plan-Fluar at 30 0 C. Cell images were acquired from the bright-field and fluorescence channels, YFP (490-510 nm, 510-560 nm) and CFP (426-446 nm, 460-500 nm), using software MetaMorph (Universal Imaging). Analysis of single cell images described in Supplemental Information.

Empirical fitting of the dynamic behavior

. A generalized exponential model $\frac{[YFP](t)}{[YFP]_\infty} = \left(1 - e^{\lambda t}\right)^m$ was used to describe the dynamics of the system upon induction with salicylate, at both population and single cell levels (see e.g. Figs. S5A, S9B). From this, we can compute the time to reach half of the steady state value of expression as $t_{50} = \frac{-1}{\lambda} ln\left(1 - \frac{1}{2^{1/m}}\right)$. While λ is proportional to the response speed evaluated at t_{50} ($\frac{\mathrm{d}ln[YFP]}{dt} \approx ln(2)\lambda$ for $m \gg 1$), m describes the response delay (Fig. S3A). Note that both parameters are correlated according to our experiments, and ultimately determine response time (Fig. S8D). We applied these expressions to fit the response of the population (and the quantification of different t 50 , e.g., Fig. S1A), and of the lineages (Fig. S11). To obtain the relative response speed of each lineage, we computed the ratio between the response speed and the corresponding cell growth rate (μ). The latter defines the response speed associated to the expression dynamics of a non-autogenous regulated gene (this can be fitted as $1e^{\mu t}$). In addition, a sigmoidal model was used to describe the dose-response curve (see e.g. Fig. S14A). Parameters for the empirical models were obtained through nonlinear regression with our own experimental data. Bootstrapping was applied to calculate the errors associated to the measurements of response time and input/output dynamic range. U-tests were performed to compare distributions of inferred parameters.

Modeling the mar circuit

. A system of differential equations was constructed to model the dynamic response of the system. The model considered as variables the concentrations of MarA and MarR. The concentration of Rob was considered constant. Model parameters were mainly obtained from previous experimental data. The model was numerically and analytically solved. The model

was perturbed to account for the dynamics of other regulatory architectures. Langevin approach was followed to simulate stochastic dynamics. See also Supplemental Information for extended mathematical models.

3

Genome-wide stereotypical gene expression responses in yeast

Genomes store information that can be used to adjust physiology to an array of conditions. As these conditions fluctuate, not all of this information is needed all the time. Moreover, the spurious expression of genes can be detrimental in many ways: it is energetically costly [227], it can be toxic or interfere with other cellular processes. Thus, regulating genomic expression emerges as a major evolutionary need. What are the general strategies that evolution has found for an adequate control of gene expression?

One of the major paradigms in regulation – the dominant one in prokaryotes – responds

to a quite straightforward logic. A gene (or operon) needed in a given condition is operated by a molecular device that senses this condition, so that expression is only active when relevant. Because of their well specified nature, most of the known, well-characterized regulatory mechanisms respond to such logic[189] (see, for example, Chapter 2). However, this "dedicated" control strategy presents several potential limitations. Firstly, such mechanisms are likely to fix in the population only if they provide an effective advantage frequently enough during the evolutionary history of the organism[10]. However, there is potentially a great diversity of infrequent or unseen conditions which taken as a whole can appear quite often. How do cells respond then to conditions for which they are not specifically adapted?

There is an additional limitation of these "dedicated" regulatory systems. Namely, adding new genes implies that they must also be accomodated in the broader regulatory network. Each nonregulatory gene that is incorporated in the genome imposes then a bigger need for new regulatory ones. This somewhat abstract concept (well-known in complex systems theory[76]) is explicited in bacteria as a nearly-quadratic scaling of the number of transcriptional regulators to the overall number of genes in the genome[49,33]. Ultimately, it imposes an upper limit (a regulatory "overhead") to further increase in genome complexity[174]. At some point, adding new genes is not worth given the massive increase in regulatory infrastructure required.

In the case of eukaryotes, this exponential scaling is as well observed, although the exponent is significantly lower[222]. In consequence, the number of genes quite often exceeds by several fold the theoretical maximum imposed by regulatory overhead in prokaryotes. The most likely explanation is the emergence of new regulatory "technologies"[147] with a fundamentally different strategy, enabling parallelization or multitasking[50] on genomes. Examples

could be RNA[147] or chromatin-based regulation – indeed, both mechanisms play a central role in the most complicated regulatory programs that qualify cells with identical genomes to differentiate in highly contrasting phenotypes[58,37].

We still know very little about the general principles orchestrating the coherent integration of these different regulatory layers[133]. One hypothesis is the existence of "stereotypical" repsonses[78]. Adaptively, such programs could provide several advatages. Firstly, they could represent ways to deal with not previously encountered challenges. Some "unknown" type of stress will likely affect homeostasis in aspects the cell does in fact "know" how to deal with. Then, "trying" a generic response could, if not assure, at least elevate the possibilities of survival. Moreover, generic responses could also enable cross-protection between different types of stress. This was suggested as a plausible explanation for the large and unspecific "environmental stress response" in yeast, and similar observations in *Arabidopsis thaliana*[133].

Beyond adaptive speculations, there is also extensive evidence for the existence of such responses at the mechanistic level. For example, chromatin effectors (such as remodellers or modifiers) typically affect much larger gene sets than canonical transcription factors[127]. Moreover, although eukaryotic genes are monocistronic, they cluster in so-called "chromatin domains" – regions of the genome spanning multiple genes that exhibit a specific, functionally relevant chromatin structure[89,16]. Also, general transcription factors (essential components of the transcriptional machinery) often harbor chromatin modifying activities[32,113]; for example, both `TFIID` and `SAGA` catalyze histone acetylation[121]. Finally, chromatin regulators also display great functional redundancy, manifest as high correlations between gene expression responses to the deletion of different chromatin effectors[86].

In this chapter, we systematically explore the existence of stereotypical expression pro-

grams in yeast. We start by analyzing a whole-genome expression dataset encompassing 1484 single gene deletions in yeast[2] . Using principal component analysis, we first extract gene expression signatures that appear recurrently in different genotypes. We attempt to statistically estimate the number of these explaining more variation than expected, given the structure of the data. We then show that these patterns are not specific only to deletion, but are also present in response to diverse environmental perturbations, as well as in experimentally evolved yeast lines. Finally, functional characterization of one of these responses suggests that it could be related to cross-talk between metabolism and gene expression.

3.1 Limited number of expression modes in response to genetic perturbation

We firstly used principal component analysis to find recurrent expression patterns in the deletion dataset, This method finds the expression "modes" that explain most variation in the data; individual genotypes represent then linear combinations of these modes with different ponderation factors for each one(fig. 3.1). For instance, the first mode explains almost 25% of the total variation present in the dataset; this mode has been characterized elsewhere and hence is not the focus here. But next modes explain as well a quite high percentage of the variation.

By definition, this method finds as many modes as columns in the dataset. Therefore, we need a way to statistically estimate the number of modes that explain significantly more variation than expected. This can be achieved by comparing the percentage of variation explained by each mode in the original M dataset with the percentage explaiend in randomized matrices. The main question is then which null model to use. For example, using either column

or row shuffle to randomize the matrix preserves in either case one of the marginal distribution. Optimally, we need however to preserve both the genotype and gene-wise variation. In order to do so, we attempted a regression-based method that considers both margins to find the random structure of the absolute expression matrix (rather than the one expressing variation).

More specifically, the differential expression dataset M has the form $M_{ij} = log(R_{ij}/G_{ij})$, with R and G representing respectively the absolute gene expression values from the mutant and the wt. We can then consider that the expression level R_{ij} is the realization of a random Poisson variable, with parameter $\lambda_{i,j}$[154]. The value of λ for each cell will depend on both the horizontal and vertical margin as $log\,\lambda_{ij} = \eta + \theta_i + \delta_j$, where θ_i accounts for gene (row) effects, δ_j for genotype (column) effects, and η is a constant. The parameter values for each row and column can be obtained using the "iterative proportional fitting" algorithm, which gives the maximum likelihood estimates of λ_{ij}. The distributions defined by these λ values can be then sampled, resulting in random matrices that keep the overall structure of the original dataset, but break nonrandom associations between genes or genotypes. Indeed, this method is able to predict differentially expressed genes in a satisfactory way ??.

Overall, our analyses indicate that a limited number of modes, that we can grossly estimate in the order of tens and below 100, can be considered as significantly recurrent in the deletion dataset (fig. 3.1).

3.2 Structure of recurrent expression signatures

It is known that yeast genes have different intrinsic tendency to change their expression[38]. This leads to the question whether SVD modes describe different patterns of variation in the

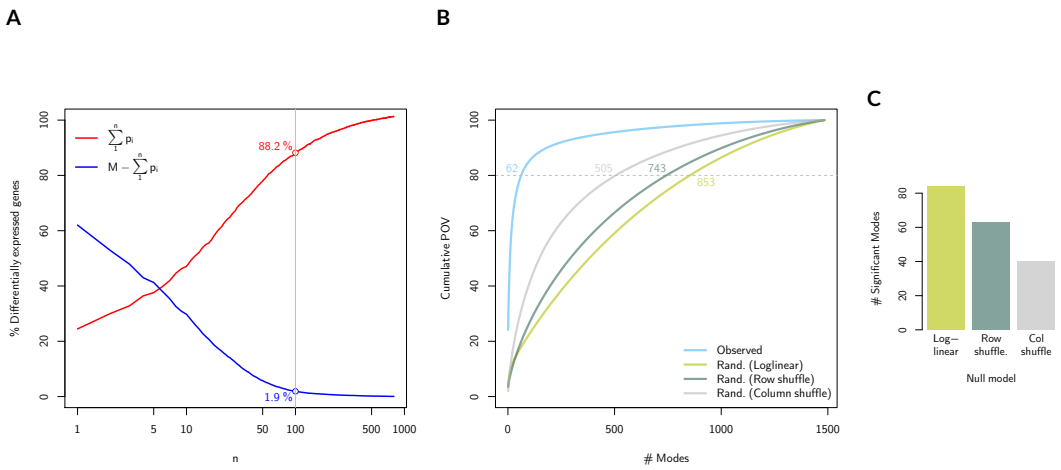


Figure 3.1: Significantly recurrent expression modes in yeast deletions. A) The M matrix can be decomposed in two matrices: the sum of projections of modes 1 to n (red), and the sum of projections from $n+1$ to k (blue), where k is the total number of modes. For each of these matrices, one can compute the number of differentially expressed genes ($FC > 1.7$, and $p < 0.05$ in the original p-value matrix). As an example, we highlight the value for $n = 100$. B) Cumulative fraction of variation explained by the first n modes, in the observed dataset and under different null models. For the null models, lines represent the mean for 100 random matrices, but the entire distribution is not wider than the line thickness. C) The number of modes that explain a bigger fraction of variance in the observed data than in any random matrix (1000 randomizations), for three different null models.

same subset of (highly variable) genes, or rather changes in different subsets. Pairwise comparison reveals that any two modes share on average 20% of affected genes (figure 3.2A). This number is much higher than the expectation, probably reflecting an over-representation of highly variable genes (e.g. TATA-box containing, figure 3.2B). Still, most of the genes affected by two modes are different (figure 3.2A). When observing groups of genes changing in several modes, in general the observed patterns were different (example in figure 3.2C). Therefore, although modes partially reflect different change patterns at the same gene sets, they describe mostly variation patterns at different subsets of genes.

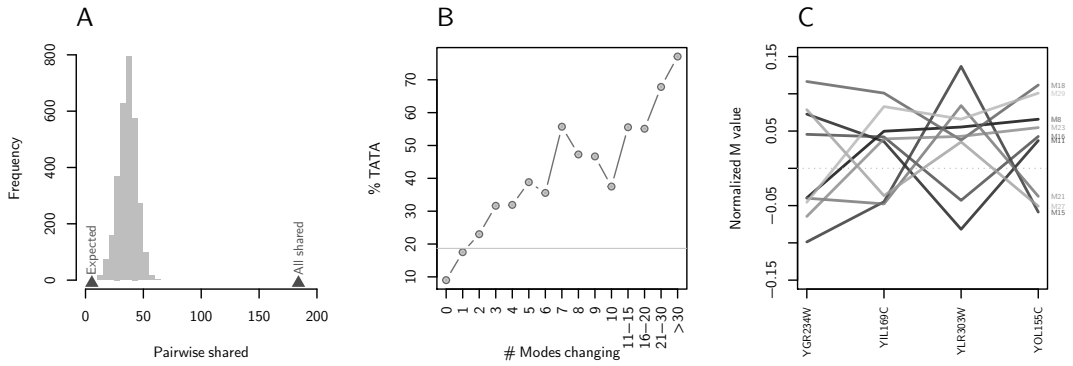


Figure 3.2: Comparison between different recurrent profiles. A) We considered the upper and lower 1.5% of genes affected by each mode (total 3% genes affected in each mode). If all genes have equal probability of being affected in any mode, the expectation would be for any given pair to share $6109(0.03^2) = 5.5$ differentially expressed genes. On the other hand, if two modes contain the same differentially expressed genes, they would share $6109(0.03) = 183$. Histogram shows the distribution of observed values, considering modes 2 to 80. B) Frequency of TATA-box containing genes considering the number of modes in which these genes are affected. C) Four genes whose expression is affected in 9 different environments.

3.3 Response to deletion mimics environmental response programs

We hypothesized that recurrent expression responses to deletion could mimic responses to environmental changes. To test this, we measured how many modes in a given dataset are correlated to at least one environment from the compendium (Methods).

It is not straightforward to compare the profile of a mode with that of an environment. Firstly, profiles in the environment compendium come from different experimetal settings. Second, second, we do not know the number of significantly up and downregulated genes in each environment. Thus, we took the following approach. In all profiles, we classified genes as either upregulated (upper 1.5%), downregulated (lower 1.5%) or not differentially expressed (rest). To compare two profiles, we can build then a 3×3 contingency table. We further reduced it to a 2×2 table by considering only the corners. Then, we compute the

Cramér V score (ϕ_c) and its p-value (derived from the χ^2 test). Using the upper and lower 1.5%, as we do, the expectance if profiles are independent is to find less than 6 genes in these four cells, with no possibility of a significant χ^2 p-value.

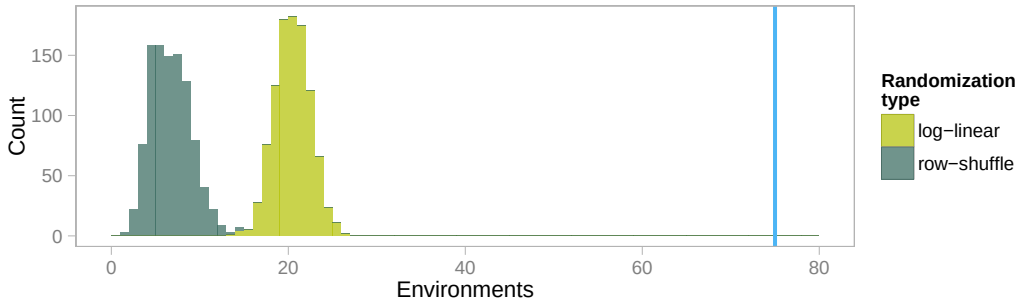


Figure 3.3: Presence of recurrent modes in environmental profiles. Vertical blue line represents the number of modes in the original dataset that show similarity to at least one environment (criteria for this similarity are described in the main text). Histograms show the distribution of this same number in randomized datasets using two different randomization types (Methods).

We found indeed that the number of modes significantly correlated to at least one environmental profile is significantly higher in the real dataset than what one could expect from the randomized ones (fig. 3.3). This result is robust to alterations in the criteria (we used $\phi_c > 0$ and $p > 0.0001$). We conclude that modes from the deletion dataset tend to mimic environmentally induced profiles.

3.4 Compensatory evolution tends to modify gene expression along deleteome-established modes

If the modes represents general, "stereotypical" patterns of gene expression, we would also expect to find them in other types of perturbation, beyond the deletions and the environmental profiles used so far. To check this, we used 8 yeast lines that were evolved in the lab to compensate for single gene deletions until practically recovering the original fitness[3]. Im-

portantly, expression changes in these lines did not generally restore *wt* expression values, but showed alternative ones. Note as well that these profiles were measured in the same technological platform and laboratory as the "deleteome" dataset; therefore, we could compute the exact percentage of variation explained by modes in each of these profiles. Qualitatively, the presence of modes in these profiles is comparable to the deletion genotypes, and is always higher than 50% (figure 3.4).

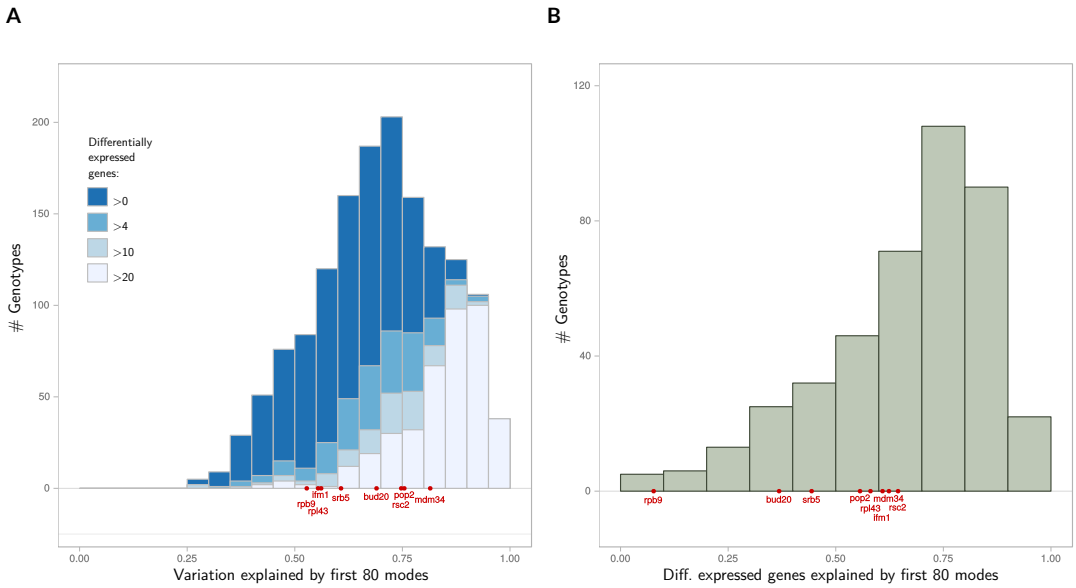


Figure 3.4: Presence of modes in compensatory evolution lines A) Histograms show the distribution of the percentage of explained variation by the first 80 modes in the deletion dataset. Deletions subsets with a minimal number of differentially expressed genes are as well shown. Red dots represent the same measure for compensatory evolution lines. B) The histogram shows the distribution of the number of genes with a differential expression value solely explained by the first 80 modes. Red dots represent again the same measure in compensatory evolution lines.

3.5 Signals of homeostasis in recurrent expression modes

Finally, we asked whether we can detect any sign of functional enrichment in any of these modes. Mode 2, for example, explains 6% of the variation in the dataset. Firstly, we observed

that deletion genotypes where this mode explains most variation quite often correspond to genes in nutrient signalling pathways (cAMP, TOR). Second, we also detected presence of several subunits of the mediator complex. Among genes that appear as downregulated in this mode, we observed several metabolic categories – most notably, mitochondrial membrane and oxidoreduction processes and vacuolar protein catabolism. Finally, we observed also that environments most correlated to this mode were greatly enriched in changes affecting aerobic/anaerobic metabolism (either natural, via changes in oxygenation, or chemical, via respiratory inhibitors, fig. 3.5). But other perturbations were also found, such as osmotic or chemical ones with unknown mechanism of action.

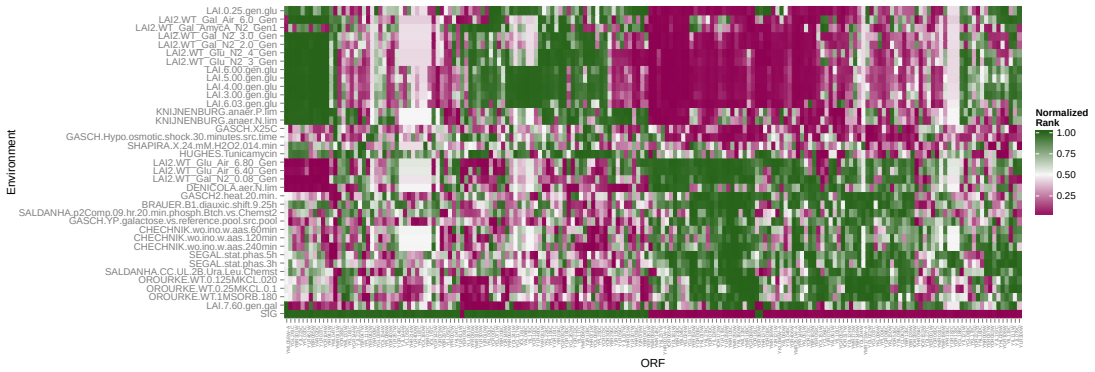


**Figure 3.5: Environmental profiles correlated to ``Mode 2''.** Heatmaps depicting the correspondence of the candidate environments (rows) to ``Mode 2''. Columns represent genes (only those significantly altered in M2). Both columns and rows are ordered using hierarchical clustering to help visualization. The first row (labeled as SIG) corresponds to the significantly altered genes in M2. Shown are rank-transformed differential expression values.

3.6 Discussion

We explored in this chapter the existence of stereotypical gene expression responses in yeast. Stereotypical responses are expected to appear recurrently under different perturbations. We first detected candidate "modes" using principal component analysis. Then, we tried to esti-

mate how many of these modes explain more variation than expected at random. The precise estimation depends however on the null model. Ideally, we would like to preserve the exact distribution of variation generated by each deletion, and the exact distribution of variation propensity of each gene. Doing this exactly is a quite difficut task; therefore, we attempted a method that tries to preserve these distributions statistically. In this way, we estimated that in the order of tens of "significantly recurrent" gene expression modes are observed in yeast.

If these responses are truely stereotypical, we would expect to find them not only under one type of perturbation (e.g. genetic). Indeed, we found signal of their presence in i) gene expression profiles arising from environmental perturbation, and ii) those arising from compensatory (exoperimental) evolution in response to deletions. Note that although these latter did recover the *wt* fitness value, they did not revert to the *wt* gene expression regime; instead, they showed other, alternative gene expression changes.

What is the functional meaning of stereotypical responses? A possible speculation is that, when they face a perturbation without disposing of an adapted response, they try to "learn" how to deal with it by some sort of "regulatory noise". Some of the cells might then find a positive outcome, enabling the survival of the population.

The existence of such a strategy is not demonstrated and represents pure speculation, but there is a number of evidences pointing to it. Firstly, a striking set of experiments has shown that, when cells are faced to an unforeseen challenge, they elicit a global and highly unspecific gene expression response, enabling the populaton to recover the *wt* fitness in a short period of time[26]. Second, one would expect that at least some stereotypic responses could reflect modification in fundamental cellular functions, likely to be affected in any sort of known or unknown stress. In general, our enrichment analyses did not show clear functional links of

specific modes to any particular function. However, one of the identified modes seems to be linked to nutrient signalling and energy metabolism. Finally, it is known that a sizeable fraction of yeast genes encodes a high level of gene expression variability in their promoters (by means of TATA box elements or nucleosomes). Although functionally, heterogeneous, this group is depleted in housekeeping genes, but enriched in diverse stress-related functions. This is exactly the class of genes that we would expect to be altered as part of stereotypical responses. Indeed, we observed a strong tendency of TATA box-cntaining genes to be part of stereotypical expression responses.

In sum, we found evidence of global stereotypical gene expression programs in yeast, that appear to be elicited under a diverse range of perturbations. Still, more work is needed to fully demonstrate the true reach of such programs, and better characterize the way they operate. Nevertheless, the hypothesis of such a scenario challenges the view of a biological system as one that is perfectly optimized by evolution. It suggests that, when facing uncommon perturbations, cells do have in their adaptive toolbox some "generic" instruments to deal with the uncommon and uncertain.

3.7 Methods

The deletion data compendium

The complete dataset is a matrix with 6170 rows (genes) and 1484 columns (deletion genotypes). This is the dataset in the supplementary material in[2]. Additional pprocessing involved removing the wt-variable genes as in the original publcation (untrusted experimental data)[2].

Environmental gene expression compendium

We downloaded the expression profiles corresponding to environmental conditions from Saccharomyces Genome Database (SPELL)[101]. Biological replicates were averaged together in each case. In the case of two-channel arrays, original log_2 ratios were kept. In the case of one-channel arrays, log_2 ratios were calculated, taking as reference the array from the same dataset that was most simmilar to a wt control in standard culture conditions. After compiling the dataset, profiles where more than 75% of the genes were missing values were removed. In the remmaining, KNN-imputation of missing values was performed using the function provided in R package "impute".

4

Balancing noise and plasticity in eukaryotic gene expression

Variations in gene expression are observed between closely related species, even when the specific gene coding sequence is largely conserved, e.g.,[63]. Within a species, expression can fluctuate following a perturbation, environmental or genetic. Gene expression can even vary in absence of perturbations and among identical individuals – this being often interpreted as detrimental noise[72]. What molecular factors determine these fluctuations? Are these factors subjected to selection pressures? And which general trends on expression vari-

ability can be identified at the genomic level?

Partial answers to these questions were recently reached by using high-throughput experiments on the budding yeast *Saccharomyces cerevisiae*. Noise was quantified for $> 2,500$ proteins using GFP-tagged yeast strains[156], validating the contribution of mRNA dynamics to protein noise. For instance, chromatin dynamics – transitions between active/inactive states – was shown to correlate with noise. But also protein function, with stress-related genes displaying much higher noise than housekeeping, which suggests differences in selective pressure[156]. Moreover, the use of a compendium of genome-wide expression profiles in four yeasts[218] allowed to quantify both plasticity (responsiveness to change in external conditions) and divergence of *S. cerevisiae* genes' expression.

Additionally, several studies enabled the identification of key molecular determinants of expression variability. Genes presenting a TATA box in their promoter showed higher interspecies variability (controlling for function) which suggested the influence of transcription re-initiation mechanisms[24]. Similarly, response to mutations (using mutation-accumulation experiments[119]) identified TATA boxes and *trans*-mutational target sizes (number of proteins influencing the expression of a focal gene) as determinants of neutral variability. Finally, the production of a complete nucleosome occupancy map covering $\sim 81\%$ of the genome[122] helped to characterize how different (absolute and relative) promoter occupancy levels further control variability.

These initial findings lead to new questions. For instance, are all these aspects of variation (short-term –noise/plasticity– *vs.* long-term –divergence) linked to a unifying promoter structure? This is clearly suggested in some studies, with an emphasis on the role of chromatin regulation[215,38,39]. Although this strategy could be advantageous (e.g. in terms of regulatory

economy[232]), it clearly leads to functional trade-offs – need of bipolarity in genome-wide transcription[14,116], presence of gene classes requiring precise but plastic expression[124], etc.

Here, we attempted to identify the molecular strategies available to decouple noise and plasticity. We first revisited the influence of chromatin regulation. While strong chromatin regulation is always reflected on higher plasticity, whether it is coupled to noise depends also on other factors, such as promoter nucleosomes or TATA boxes. The type of chromatin control also plays an important role, with a contrast between global and specific regulation. We alternatively find that noise uncouples from plasticity in low-plastic genes due to changes in translational efficiency. These distinct modes are confirmed by the modulating effect of genomic neighborhood on coupling, with short intergenic distances and bipromoter architecture can both be related to high and low noise.

4.1 Chromatin regulation not always couples noise to plasticity

TATA boxes and high nucleosomal occupancy at the proximal regions of transcriptional starting sites (TSSs) have been identified as fundamental promoter features leading to gene expression variability[14,218,119]). Both features were further argued to couple two specific forms of variability, i.e., expression noise and plasticity. Linkage between noise and plasticity was also associated to highly dynamic chromatin, as quantified by histone exchange rates[124]. However, histone exchange rates do not fully describe the many *trans* factors influencing nucleosome dynamics.

To better understand how such factors determine the noise-plasticity coupling, we used a score that assesses chromatin regulation effects (CRE), i.e., how much the expression of a given gene varies when deleting its *trans*-acting chromatin regulators[205,215,38,39] (Methods).

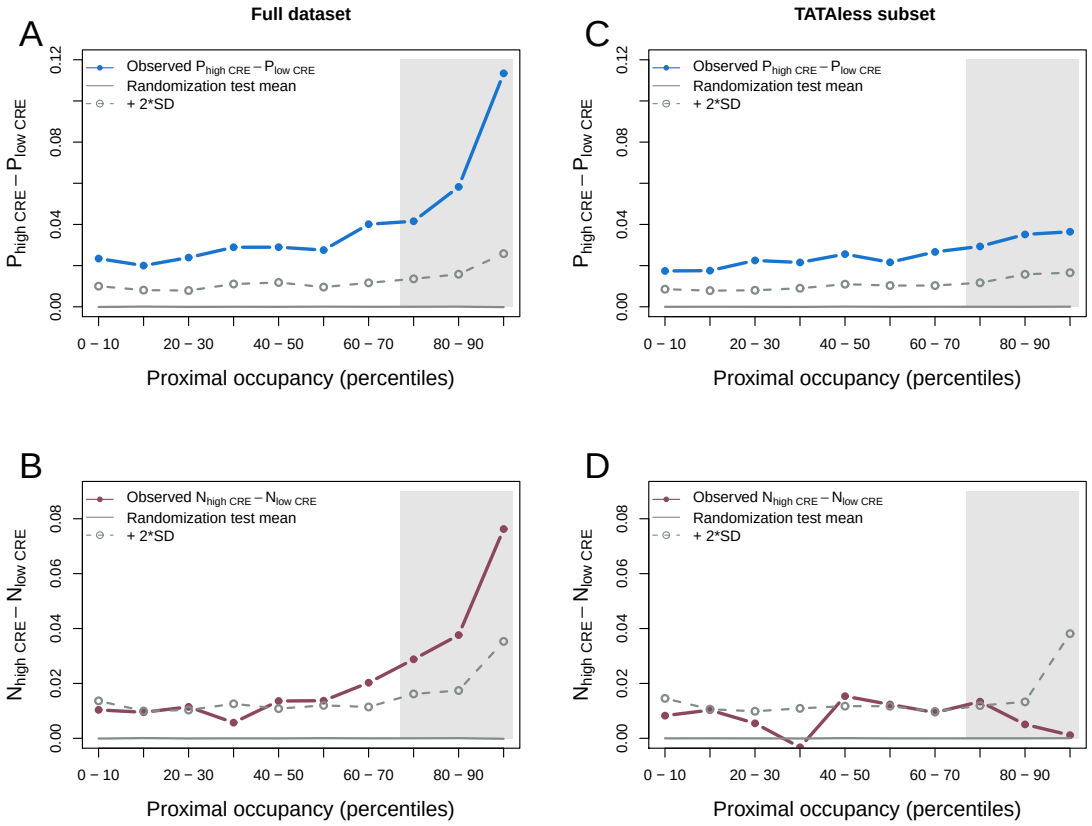


Figure 4.1: Proximal nucleosomal occupancy, chromatin regulation, and the noise-plasticity coupling. We divided the dataset ($n = 2045$) in ten equally sized bins of increasing proximal nucleosome occupancy. In each bin, we computed the median chromatin regulation effect (CRE). We plotted the difference in plasticity (A, blue curve) or noise (B, red curve) of genes above/below this median and contrasted the observed values with those expected randomly (permutation test in each bin to depict significance, shown as the mean --gray curve-- and mean plus two standard deviations --dashed gray curves-- obtained with 10000 randomizations). Plasticity is always enhanced by strong chromatin regulation; however, regulation enhances noise only in promoters with high proximal nucleosome occupancy and TATA box (shaded area in A,B). An identical analysis is shown in (C) and (D), but excluding genes with TATA-containing promoters. High occupancy does not lead to increased plasticity/noise in this case (shaded area in C,D).

CRE correlated with high plasticity as expected (Spearman's correlation coefficient $\rho = 0.57$, $p < 10^{-20}$, $n = 2045$). We then grouped genes in terms of proximal nucleosome occupancy and computed mean plasticity score for genes exhibiting either high or low CRE within each group. We found that chromatin regulation can induce a relatively high level of plasticity

independent of nucleosomal occupancy and presence of TATA promoters (fig. 4.1). A well-placed proximal promoter nucleosome and the presence of a TATA box can further enhance plasticity, but at the cost of increasing noise. Promoter nucleosomes likely increase sensitivity to chromatin regulation[39], but also lead to an equilibrium between open and close promoter states that can increase noise. TATA boxes increase transcriptional initiation efficiency, but also cause bursting[24]).

The crucial effect of the high (proximal) nucleosomal occupancy to enhance coupling is emphasized by the low noise exhibited by a subset of TATA-containing genes with a pronounced nucleosome depleted region (NDR) similar to TATAless genes. Thus, TATA boxes only produce noise in promoters with high proximal nucleosome occupancy (fig. 4.2). Moreover, if the noise-plasticity coupling has its origin in the stability of the transcriptional apparatus at the promoter, we should also observe coupling in TATAless genes with a SAGA-dominated initiation (that also produces bursting transcription[242]). Even this group being small, this is what we observe (noise-plasticity correlation in TATAless and SAGA dominated genes, $\rho = 0.51$, $p = 1.2 \times 10^{-5}$, $n = 66$, fig. 4.2D)

4.2 Noise-plasticity coupling is influenced by *trans*-regulatory strategy

To further understand what determines the coupling (or uncoupling) of noise and plasticity, we inspected potential qualitative differences in the type of chromatin regulation. We computed the mean effect on expression of a compendium of mutations in regulators[205] (CRE score before represents a subset, see Methods) on plastic genes. This analysis highlighted a strong anti-correlation between the effect of perturbations in genes with low noise but high plasticity (LNHP, see Methods for definition of these classes) and high-noise high-plasticity

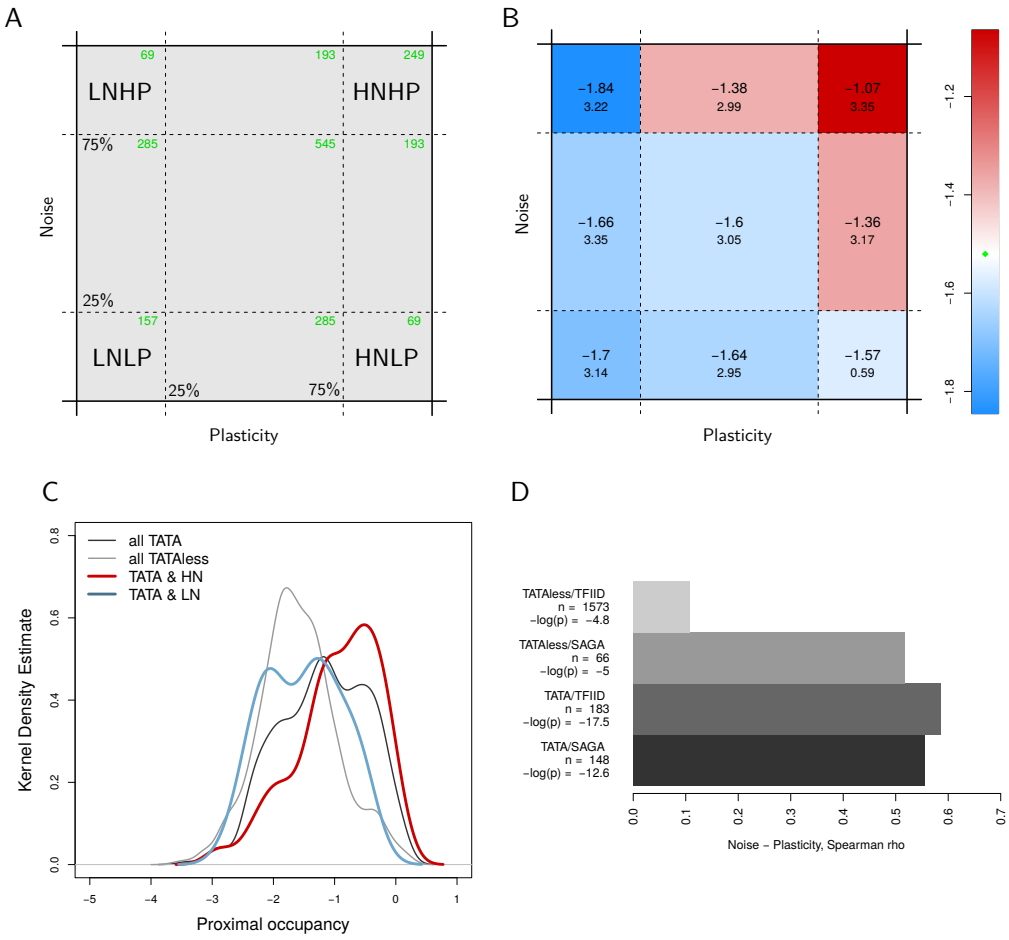


Figure 4.2: Noise-plasticity coupling modulation A) Sketch representing the categories of noise and plasticity used througout the chapter. Bounds represent percentiles 25 and 75 in both cases. Corners correspond to classes ``low noise - low plasticity'' (LNLP), ``high noise high plasticity'' (HNLP), etc. The number of genes in each box is shown in green. B) Each region is colored according to the mean proximal nucleosome occupancy (as $\log 2$ ratio, see Methods); the value is shown in the center of each box. . The whole-dataset mean is shown in green in the colorbar. Below, we show the $-\log_{10}(p)$, obtained by randomization (10000 times) and corrected for multiple testing. Promoters with the highest occupancy levels concentrate in the HNHP group. C) Distribution of proximal nucleosome occupancy for TATA-containing subset of HN and LN groups, and for the whole TATA and TATAless subgroups. In contrast to low noise ones, noisy TATA-containing genes show ``covered'' promoters. B) Spearman ρ correlation between noise and plasticity for TATA-containing or TATAless, and SAGA-dominated or TFIID-dominated genes. SAGA-dominated transcription drives noise-plasticity coupling even in TATAless genes, evidencing the role of transcription initiation.

genes (HNHP, $\rho = -0.83, p < 10^{-16}, n = 170$, fig.4.3). This confirms mechanistically a complementary program of regulation between these two groups of genes (enriched by growth –ribosomal– and stress genes, respectively[14,133]), to be added to the previously observed differences in promoter nucleosome occupancy (fig. S5) and histone modification enrichment[116].

More specifically, LNHP genes are most affected by perturbations that cause a decrease in expression (fig. 4.3A, red dots), implying a dominantly activating function. Most of them correspond to general transcription factors (20 out of 41). Particularly significant is the presence of many perturbations involving TBP-associated factor 1 (`TAF1`), an essential component of the general transcription factor TFIID. This factor is part of the eukaryotic transcription preinitiation complex (PIC), and is involved in the transcription of $\sim 90\%$ of yeast's genes[104] (with the rest $\sim 10\%$ dominated by SAGA).

Additionally, perturbations at `TAF1` affect LNHP genes significantly more than low plasticity (LP) ones (fig. 4.3B), despite the transcription of these being also TFIID-dominated. A possible explanation could be related to the fact that `TAF1` is also associated to chromatin regulation activity (histone acetyltransferase)[104,60], which has already been linked to expression variability[38]. Indeed, LNHP genes are known to be hyperacetylated[178]. This feeds the speculation that while at constitutively expressed LP genes `TAF1` could function as a merely a structural component of the transcription initiation machinery, in LNHP genes it gains a regulatory function, modulating the level of expression.

On the other hand, HNHP genes are regulated by SAGA (instead of TFIID), which also displays a HAT activity. Although the HAT activity of SAGA and that of TFIID could be equivalent and capable of compensate one the absence of another[104], they are differentially

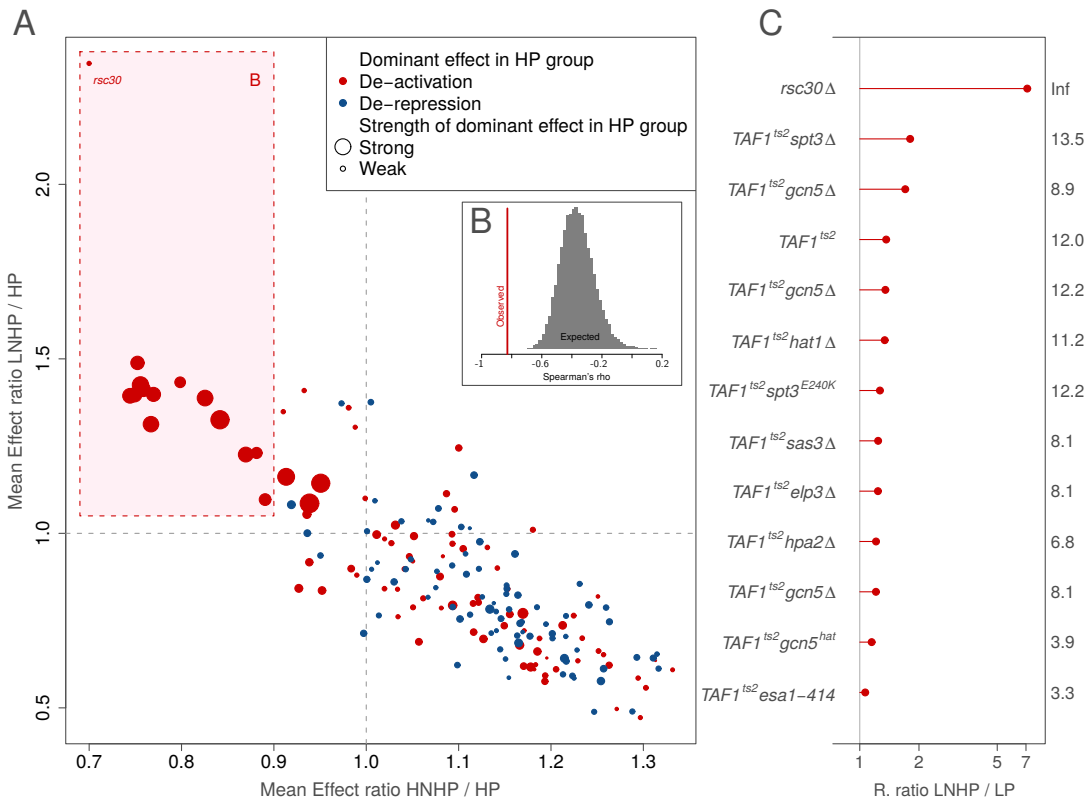


Figure 4.3: Distinct chromatin regulation strategies to achieve noisy or quiet plasticity. A) Each dot represents the mean effect in the expression of a set of genes in a subclass (HNHP, x coordinate; LNHP, y coordinate; normalized by effect in HP class) when a particular regulator is mutated [205]. A ratio > 1 thus implies that the corresponding subclass is more strongly influenced by certain regulator than the full HP group. A strong negative correlation is found indicating that many regulators are highly specific to either HNHP or LNHP genes. (The correlation is stronger than expected by randomizing the HN and LN labels, as x and y axes are not completely independent, see inset). This confirms that these groups are enriched by complementary functional classes (stress and growth related genes, respectively) which are generally regulated in opposite sense [14,133,116]. Dot colors denote the dominant effect of the regulator on the HP class (blue: regulator is dominantly repressing expression, red: regulator is dominantly activating) while sizes describe the strength of the dominant effect; e.g., LNHP genes are frequently affected by strong chromatin activators. C) We examined in detail the effects on LNHP genes (box in A). Except *rsc30* (a regulator of ribosomal proteins [6]) all these mutations involved TAF1, a TFIID subunit [104,60]. This essential factor regulates $\sim 90\%$ of the genes in the genome, not including most of HNHP (which are regulated by SAGA) but including almost all LP genes (see main text). Nevertheless, we observed that all these mutations affected significantly more strongly LNHP than LP genes (Kolmogorov-Smirnov tests with FDR-corrected -log(p-value)'s shown at the right).

involved in regulating the growth/stress genomic expression programs. In addition, HNHP genes are affected by perturbations – mostly at chromatin regulators, 92 out of 135 – that cause an expression increase, implying a repressing function (fig. 4.3A, blue dots). Altogether, these results indicate that fundamentally different regulatory strategies that function in the broad "growth" and "stress" categories could have a decisive influence on noise/plasticity coupling.

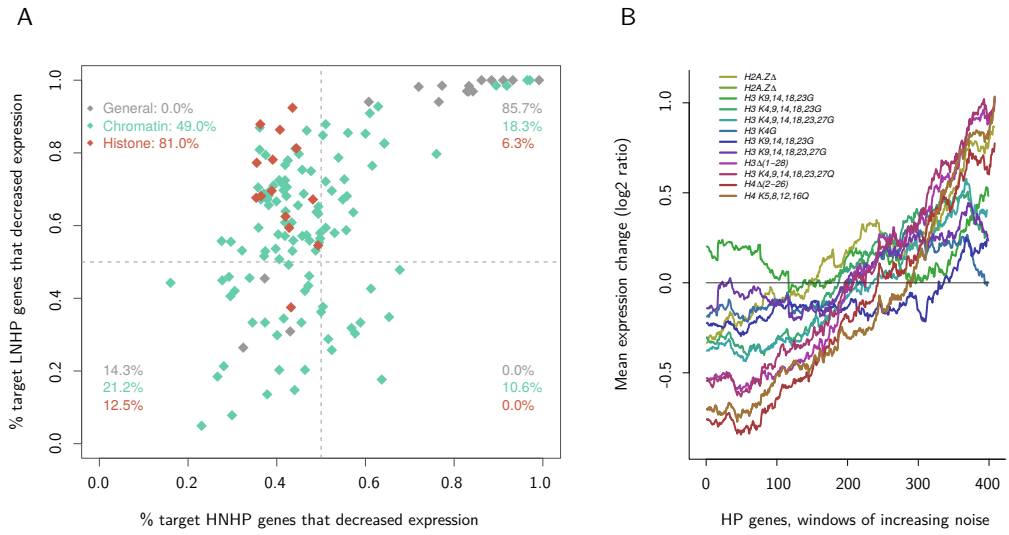


Figure 4.4: Dual action of histones. Both *trans*-acting chromatin regulators and histones tend to have opposite effects in HNHP and LNHP genes. For each mutation from [205], we plot the fraction of genes in the HNHP group and in the LNHP group that decreased expression. Consistently with that observed in fig. 4.3, we find that mutating as much as $\sim 50\%$ of chromatin regulators results in the de-activation of the majority of the LNHP genes, but de-repression of most of the HNHP genes. In addition, and perhaps more importantly, we observe that 81% of mutations in histones also exhibit this behavior. This indicates that histones by themselves are needed for repression of HNHP genes and, at the same time, activation of LNHP. B) We ordered the HP group ($n = 511$) by ascending noise, and performed a sliding window analysis (window size $= 100$ genes) of the expression values after different histone perturbations (which correspond to the orange dots shown in fig. 3 in the main text). Genes whose plasticity is not coupled to noise tend to reduce their expression level; therefore, histones act favoring their expression. On the contrary, the same perturbations tend to increase the expression of genes with high noise, indicating a repressive action.

We also found that regulators whose deletion results in decreasing expression level in the

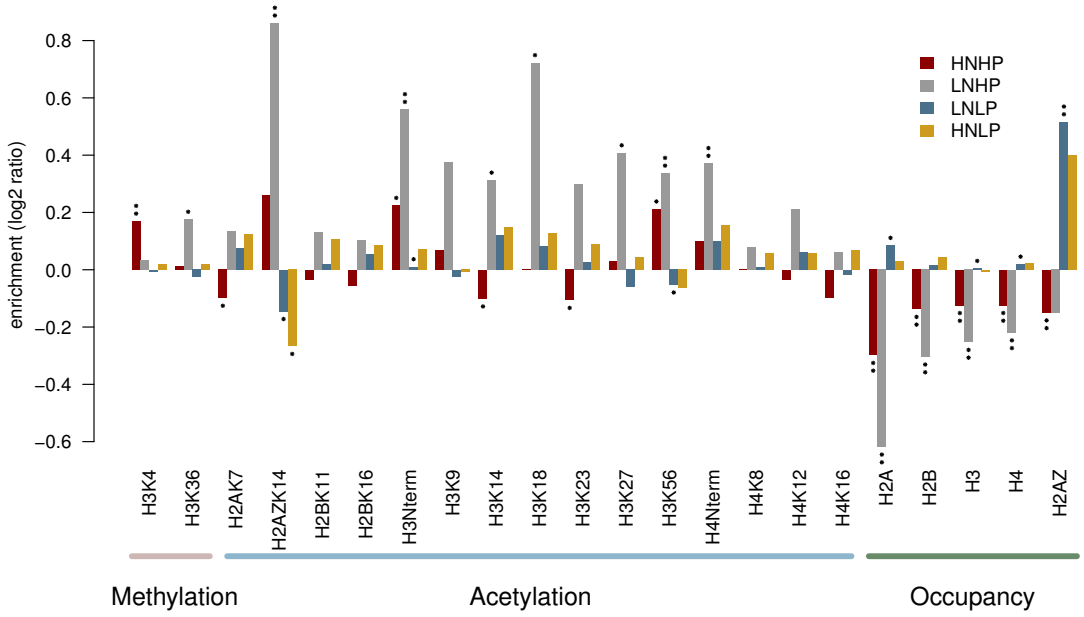


Figure 4.5: Histone modifications in noise-plasticity classes. We show the enrichment in different histone modifications at promoters, using nucleosome-normalized data. We used ChromatinDB (*http://www.bioinformatics2.wsu.edu/cgi-bin/ChromatinDB*)) as data source. At each bat, one asterisk is shown if enrichment is significant at $< 10^{-2}$, or two asterisks if it is significant at $< 10^{-4}$ (corrected for multiple testing [21]).

majority of LNHP genes tend in comparison to increase it in the majority of HNHP (fig. 4.4). As expected, this is not observed for general transcription factors (that generally reduce expression level), but it is for many deletions of chromatin regulators and, notably, histones. As much as 81% of histone deletions caused a generalized increase in the expression of most HNHP genes, while they decreased expression in LNHP. How can histones repress transcription in a set of genes and, at the same time, activate it in another? A recent result can help us understand this [116]. Namely, LNHP genes are greatly enriched in activating marks (mostly acetylations, fig. 4.5), and a strong change in acetylation level is observed when these genes are repressed [116]. Therefore, acetylated histones are probably essential for the expression of these genes. On the contrary, HNHP genes do not show such changes in histone acetylation sta-

tus, but they reduce their occupancy level when activated. Therefore, histone deletion is in this case more likely to obstruct the formation of repressive nucleosomes, resulting in a more frequently open promoter and increasing expression level.

4.3 Noise in not plastic genes arises from enhanced translational efficiency

We noted that LP genes also present differential coupling to noise. In contrast to HP genes, this difference does not seem to respond to transcriptional-based determinants. LP genes hardly present TATA boxes ($26/513$, 5.1% in LP; $343/1529$, 22.4%, in the rest, χ^2 test $p < 10^{-15}$), display pronounced NDRs (mean proximal nucleosomal occupancy LP: -1.70, $n = 513$; rest: -1.46, $n = 1532$, KS-test $p = 7.6 \times 10^{-7}$, see also Methods) and are poorly regulated by chromatin (mean CRE LP: 0.58, $n = 513$; rest: 0.70, $n = 1532$, KS-test $p < 10^{-15}$). A notable feature of these promoters is their enrichment in histone variant `H2A.Z` at promoters, which has been already noted[116] and is thought to help stabilizing the NDR (with our dataset, mean LP: 0.41, $n = 365$, rest: 0.12, $n = 1144$, KS test $p = 1.2 \times 10^{-9}$). Indeed, we did not find differences in all these factors when considering low and high noise subgroups within the LP set.

We thus inspected if uncoupling could be associated in these genes to translation as this is known to control noise[161,176]. Our analysis shows that noise in LP genes is correlated with translational efficiency[136] and ribosomal density[7] ($\rho = 0.22, p = 7.9 \times 10^{-5}$, and $\rho = 0.21, p = 1.6 \times 10^{-4}$, respectively; $n = 327$, and fig.4.6) while we did not observe this in highly plastic genes ($\rho = 0.06, p = 0.32$ and $\rho = 0.08, p = 0.14$, respectively; $n = 312$). If translation controls noise in LP genes, then noise should also covariate with factors influencing translation efficiency. In LP genes, translational efficiency correlated more strongly

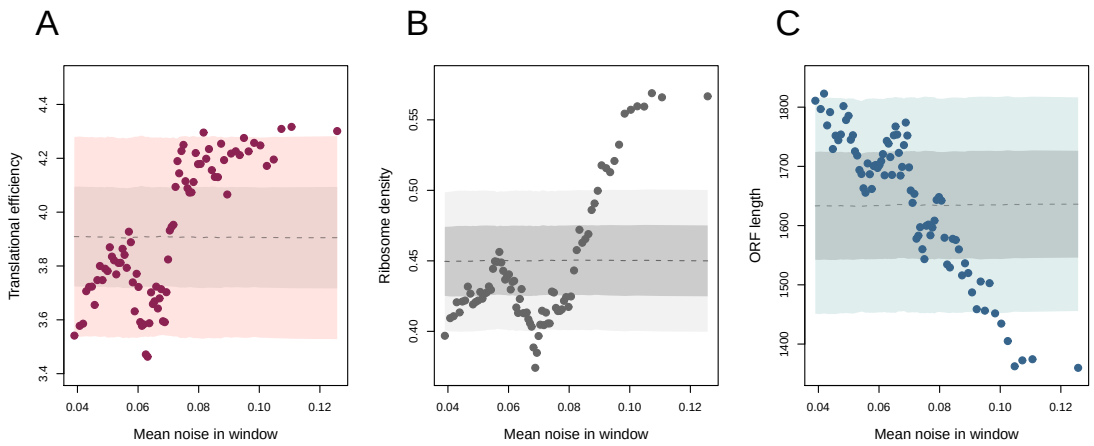


Figure 4.6: Translation modulates noise in low plasticity genes. Noise in LP genes is related to translational efficiency, which in turn is related to ORF length. We ordered LP genes by increasing noise. We performed a sliding window analysis of translational efficiency (A), ribosomal density (B) and ORF length (C). Shaded regions represent the mean and two standard deviations at each point obtained with the same sliding window analysis over randomized data; the process was repeated 10000 times. See also main text.

with ORF length ($\rho = -0.58, p < 10^{-15}, n = 327$) than with frequency of optimal codons (FOP, $\rho = 0.31, p = 9.9 \times 10^{-9}, n = 327$). Consistently, noise correlated with ORF length ($\rho = -0.18, p = 5.5 \times 10^{-5}, n = 513$ and 4.6) but not so with FOP ($\rho = 0.07, p = 0.12, n = 513$). On the other hand, noise correlated with ORF length in an opposite way in HP genes ($\rho = 0.20, p = 5.6 \times 10^{-6}, n = 309$) which probably reflects complementary constraints on gene length (e.g., ribosomal genes being small proteins as their action relies on multiprotein complexes).

4.4 Genomic neighborhood modulates noise-plasticity coupling

What other mechanisms could modulate the noise-plasticity coupling? The specific architecture of the genomic neighborhood of a gene appears as a potential candidate. Indeed, it was shown recently how bi-directional promoters (bipromoters) can reduce noise by favoring nu-

cleosome depletion[230,237]. We confirm this result for the full dataset (mean noise bipromoter $= 0.0796$, not bipromoter $= 0.0965$, Wilcoxon test $p = 1.2 \times 10^{-8}$) and validate as well that noise-sensitive gene classes are enriched in bipromoters: essential $(465/1062, 43.8\%)$ compared to nonessential $(1651/4577, 36.1\%, p = 3.4 \times 10^{-6}$, Fisher's Exact test), and genes coding for protein complex subunits $(666/1565, 42.6\%)$ compared to the rest $(1486/4195,$ $35.4\%, p = 8.1 \times 10^{-7}$, Fisher's Exact test).

In the case of low-plasticity genes, noise is modulated mostly at the translational level, and consequently should not be affected by the presence of bipromoters. Although depletion is observed independently of plasticity, it does not seem to affect noise in the LP class. In contrast, the bipromoter effect of noise reduction is observed in HP genes, as expected[230] (table 4.1). This is further corroborated by the enrichment of bipromoters observed in LNHP genes $(27/66, 41\%)$ compared to HNHP $(47/236, 20\%, p = 10^{-3}$, Fisher's Exact test). Consistently, this enrichment is not significant in LNLP genes $(75/145, 52\%)$ compared to HNLP $(38/67, 57\%, p = 0.6$, Fisher's Exact test). This pattern is confirmed in noise-sensitive genes: while they are enriched in bipromoters in the HP group $(79/237, 33.3\%,$ compared to $67/274, 24.5\%$ in non noise-sensitive, i.e., noise-tolerant, $p = 0.03$, Fisher's Exact test), in the rest (low and medium plasticity) there are no observable differences in bipromoter frequency (noise-sensitive $348/745, 46.7\%$; and noise-tolerant $348/789, 44.1\%, p =$ 0.33, Fisher's Exact test). However, this latter group does exhibit a difference in noise (noise-sensitive genes conform $257/442, 58.1\%$ of the LN, but only $109/252, 41.6\%$ of the HN subgroup, $p = 2.4 \times 10^{-5}$, Fisher's Exact test). If not to bipromoters, this difference could be attributed, as expected, to differences in ORF length (mean length noise-sensitive: 1732.0, noise-tolerant: $1517.2, p = 1.9 \times 10^{-4}$, Wilcoxon test).

Table 4.1: Genomic neighborhood influence on proximal nucleosome occupancy (PNO), noise and plasticity Differences in noise and plasticity between genes transcribing or not from a bi-directional promoter (bi-promoter). p-values correspond to Wilcoxon tests.

	LP			HP		
	Bipromoter	Not bipromoter	p	Bipromoter	Not bi-promoter	p
PNO	-1.79	-1.59	8.7×10^{-4}	-1.54	-1.13	1.3×10^{-9}
Noise	0.074	0.075	0.82	0.102	0.134	1.3×10^{-5}
Plasticity	0.023	0.023	0.26	0.111	0.135	1.6×10^{-4}

Beyond reducing noise, do bipromoters actually uncouple it from plasticity? We hypothesized that the answer is negative, because plasticity requires greater genomic space to accomodate a more complex regulatory architecture in the promoter. We found several evidences for this. Firstly, intergene distance correlates with plasticity ($\rho = 0.19, p < 10^{-15}, n = 5102$), and this effect is especially strong if we consider only transcripts with divergently oriented upstream partners ($\rho = 0.27, p < 10^{-15}, n = 3271$; fig. 4.7). Second, bipromoters also affect plasticity, and not only noise, in the HP group (table 4.1). Third, we detect a strong bipromoter-independent effect in noise-sensitive HP genes (ANNEX). While the noise-sensitive group has significantly lower noise (mean noise in bipromoter noise-sensitive $= 0.083$, mean noise in bipromoter noise-tolerant $= 0.124$, Wilcoxon $p = 1.0 \times 10^{-4}$), the difference in plasticity is not significant (mean plasticity in bipromoter noise-sensitive $= 0.105$, mean plasticity in bipromoter noise-tolerant $= 0.12$, Wilcoxon $p = 0.62$). Does selection favour then other mechanisms that do effectively reduce noise without affecting plasticity? Evidence suggests so: if we consider only bipromoter HP genes, we find that noise-sensitive ones tend to be TATAless ($80\%, n = 79$) compared to noise-tolerant ($61\%, n = 67$, $p = 0.017$, Fisher's Exact test). Intergenic distances also suggest that even when controlled from a bipromoter, noise-sensitive genes tend to maintain plasticity by maintaining higher

intergenic distances (234 bp for noise-sensitive bipromoters and 190 bp noise-tolerant bipromoters, Wilcoxon test, $p = 0.019$).

4.4.1 Noncoding transcripts and modulation

The above can be complementary analyzed if we consider all possible local genomic architectures around a focal gene (fig. 4.7), i.e., parallel, divergent and bipromoters with a coding or non-coding transcript as upstream partner (noncoding partners include "cryptic unstable transcripts", CUTs, and "stable untranslated transcripts", SUTs, see[239] and Methods. Bipromoters involving CUTS were recently associated with low noise[230]). Thus, we computed the coupling between noise and plasticity for each architecture. Coupling is strong for genes with divergent transcripts (independent of the type of upstream partner) and weak for those with a bipromoter with a coding partner (fig. 4.7). This further validates the observed absence of bipromoters in HNHP genes and their enrichment in the other three classes (bipromoters are the most commonly found architecture in LNLP, HNLP and LNHP) where they are associated, of course, to short intergenic distances (fig. 4.7). Interestingly, bipromoters of plastic genes with low noise are the ones with the biggest (relative) intergenic distance (with respect to LNLP and HNLP), which suggests again the requirement of a minimal distance to locate the regulatory demands associated to enhance plasticity (mean distance bipromoters of LNHP: 252 bp, in the LNLP and HNLP groups: 178 bp, Wilcoxon test $p = 1.1 \times 10^{-3}$). Overall, this emphasizes bipromoters as noise-abating architecture only when noise and plasticity are transcriptionally modulated.

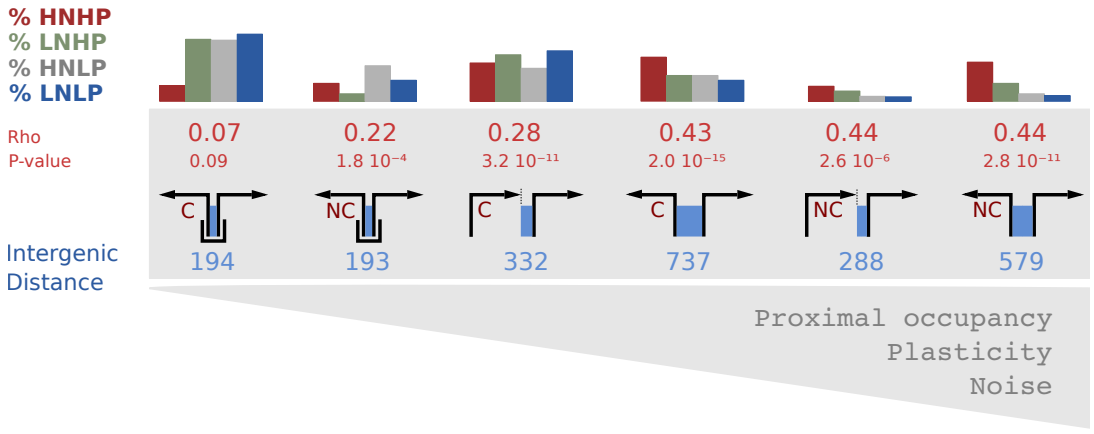


Figure 4.7: Noise-plasticity coupling is modulated by genomic neighborhood. A cartoon depicting the different genomic structures (bipromoter, parallel, divergent) upstream of coding genes is shown in ascending order of proximal nucleosome occupancy, plasticity and noise (which coincide). For each structure, we show the average intergenic distance in blue. In red is shown the Spearman ρ coefficient for the observed noise-plasticity correlation. We also show the percent within each class of a given upstream structure , e.g., HNHP mostly exhibit parallel/divergent coding (C) and divergent non-coding (NC) transcripts.

4.5 Discussion

We analyzed in this chapter the molecular determinants that modulate the balance between gene expression plasticity and noise in *S. cerevisiae*. Noise was confirmed to be linked to plasticity when genes exhibit particular modes of transcription initiation (and re-initiation) related to the presence of TATA boxes at the promoter and strong chromatin regulation [24,14,122,215,38,39]. This could suggest a model in which intrinsic noise is a by-product of the need for plasticity [124]. We show, however, that noisy expression can be observed also in genes with low plasticity. These genes are poorly regulated at the chromatin level, usually displaying a nucleosome-depleted promoters (fig. 4.3) without initiation enhancing mechanisms (e.g. TATA boxes). From such promoters, transcription likely occurs in single, isolated in time, initiation events [104], resembling bacterial transcriptional dynamics. In such scenario, noise is anticipated to de-

pend on translational efficiency[161,176] (but see[84]) and this is indeed what we notice.

In addition, ORF length appears as a strong determinant of ribosome occupancy in this class, and thus of translation efficiency (potentially due to the lack of post-transcriptional regulation , see fig 4.8)[120]. Following this model we expect essential genes –usually of low plasticity– to be large as we observe (size essential genes: 1646 bp, size nonessential: 1468 bp, Wilcoxon-test $p < 0.001$). ¡

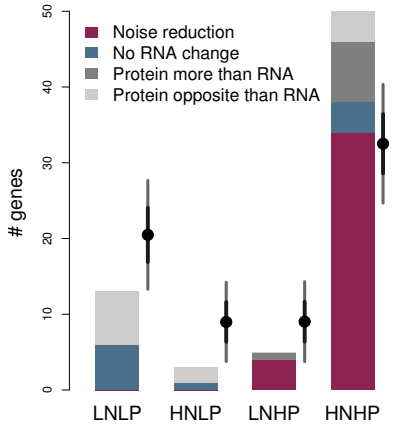


Figure 4.8: Post-transcriptional regulation The number of genes in each group that is subjected to post-transcriptional regulation (PTR) under osmolarity stress conditions is presented. Blades depicted next to each bar represent the expectance (mean $\pm$ 2SD) from a randomization test. HNHP group is enriched in PTR in these conditions, while groups with low plasticity are less regulated than expected. Additionally, different colors represent the different modes of PTR. Noise-reducing PTR is observed only in highly plastic genes.

On the other hand, we also show that chromatin regulation can enhance plasticity without necessarily coupling it to noise. This is observed in a group of genes enriched in growth-related functions (such as ribosomal protein genes, RPs) which depend strongly on `TAF1`, an essential subunit of TFIID, but less strongly on most of the more specific chromatin regulators (as compared to noisy and plastic genes, fig. 2). This indicates that they respond to general, rather than gene specific, regulatory strategies. This is consistent with the high degree of co-regulation previously observed among these genes[116]. Beyond this, the group exhibits a characteristic pattern of low nucleosome occupancy in both proximal and distal promoter regions[215] pos-

sibly caused by the strong enrichment in activating histone modifications[116], and particularly acetylations. We hypothesized that rather than a promoter-localized open-chromatin state, these genes could be located at broader open chromatin domains. Indeed, we distinguished that RP genes tend to be localized on broad open-chromatin domains, which extend up to at least 40Kb (fig S11), supporting this view.

The low noise but highly plastic expression is consistent with two (not mutually exclusive) models previously proposed. In first place, it confirms the detailed model proposed by Raser and O'Shea[175] in which low nucleosome occupancy at the promoter indicates a stable open state, allowing the high expression levels exhibited by these genes (fig. S12). The concomitant noise reduction would not be possible if the high expression level would be reached by increases in transcription or translation efficiency[175]. At a broader level, the localization in open chromatin, and consequently low noise, genomic regions could also contribute to the small level of noise detected[16].

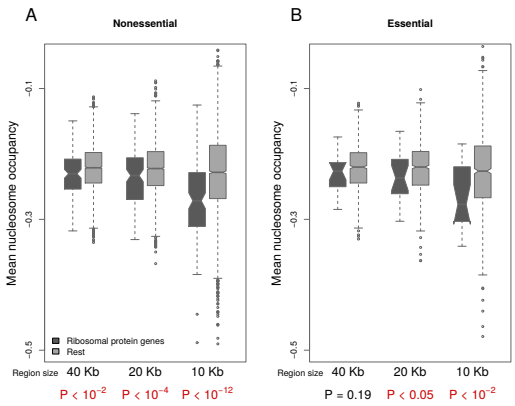


Figure 4.9: Ribosomal genes are located in broad open-chromatin genomic regions. Taking each ribosomal protein gene TSS as the focal point, we measured the mean nucleosome occupancy in 5, 10 and 20kb upstream and downstream (making up regions of 10, 20 and 40kb). This is true even controlling for essentiality (as essential genes are known to localize in open chromatin regions).

A further enhancement of plasticity comes, however, at the cost of coupling it to a strong intrinsic noise. Mechanistically, this involves two additional promoter features: a TATA box and a well-positioned nucleosome at the proximal region[122,215,38]. The presence of this nu-

cleosome is consistent with the enrichment of these genes in stress-related functions, thus likely to be "repressed-by-default". Its repressive function is emphasized by a tendence to destabilize when the gene is activated[116]. Indeed, we found that upon histone depletion, genes with noise coupled to plasticity increase their expression (fig 4.4). The presence of this nucleosome also allows a fine-tuned –and gene-specific– control of open and close promoter states by a variety of *trans*-regulators. Our results indicate that most of this regulation is repressive (fig. 4.3), and one could hypothesize that this repression is achieved by stabilizing the promoter nucleosome, favouring an inaccessible chromatin state. Indeed, an example of such control mechanism was discovered quite recently for the first time ??.

The second element crucial for coupling noise to plasticity are TATA boxes. During the timelapses the promoter is in open state, the presence of a TATA box allows the pre-initiation complex to stay assembled at the promoter, firing continuous initiation events. This increases the sensibility to changes (during the time the promoter stays in open state produced by chromatin remodeling) thereby allowing an increase in plasticity. Even when the nucleosome is not removed by regulation, it can occasionally be destabilized allowing strong bursts of transcription, producing gene expression noise. The fact that coupling is related to efficient transcription (re)initiation is confirmed by the strong coupling found in SAGA but TATAless genes (fig. 4.2), which confirms the model previously proposed[175] to a genome-wide scale.

Moreover, the critical promoter-covering nucleosome is likely stabilized also by particular DNA properties, such as high bendability[114,216,38]. This could potentially increase the number of phenotypically relevant mutational targets. This could explain the increased expression divergence observed in these genes in these genes[215,39] without the need of invoking selection (although of course it might be present). Therefore, coupling between different types of

variability is mechanistically a consequence of the sophisticated regulatory strategy involving promoter nucleosomes and TATA boxes; is responsible for a higher sensibility to chromatin regulation (leading to plasticity); to stochastic nucleosome fluctuations (leading to noise) and to mutational effects (leading to expression divergence).

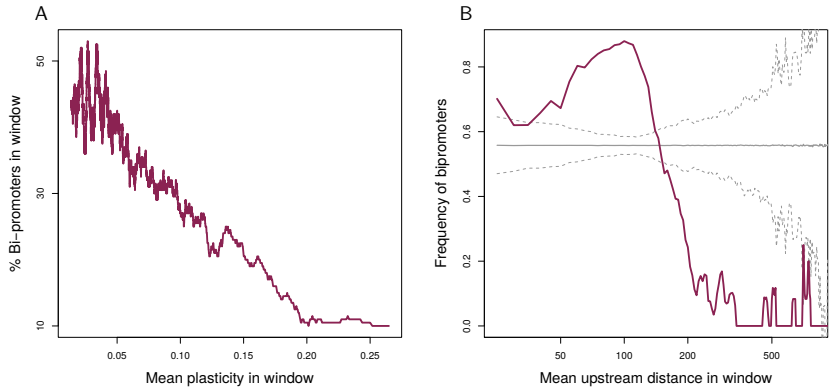


Figure 4.10: Evolution of bi-directional promoters. A) We ordered genes by ascending plasticity ($n = 5111$), and measured average plasticity (x) and frequency of bipromoters (y) in sliding windows of 200 genes. B) A sliding window analysis was performed over upstream intergene distance (window size of 50bp, step-size of 10bp) using $n = 4856$ ORFs (including noncoding) with divergently oriented upstream partners and available data of whether they start at bipromoters. A randomization test was conducted; for each window of size s, we took randomly s genes 10000 times and computed the mean frequency of bi-promoters (solid line) $\pm$ two standard deviations (dashed lines). Bidirectional promoters appear enriched at intergene distances up to ~ 150bp, approximating the distance potentially covered by a nucleosome (144bp, or a NDR for that matter).

These strategies to modulate noise coupling to plasticity are further emphasized by the structure of the genomic neighborhood. Of course, this only applies to genes where noise has a transcriptional basis (i.e. high plasticity genes, fig. 5). But although short intergenic distances – and particularly, bipromoters – are known to reduce noise (and we confirm this), they do not seem to effectively uncouple it from plasticity, as they reduce promoter space and limit regulation. We observed indeed a strong bias in bipromoter frequency towards genes with low plasticity (fig 4.10A), which indicates that satisfying these regulatory needs

could represent a major force determining their *absence* in highly plastic genes (fig 4.10A). On the other hand, the high frequency of bipromoters among LP genes can be explained without invoking selection. *S. cerevisiae* has a highly gene-dense genome, averaging one gene each ∼2 Kb with a median intergene distance (considering noncoding transcripts) of 204 bp, which indicates that bi-promoters could appear neutrally when there is no constraint against it. Indeed, the frequency of bipromoters displays an expectable distribution in view of intergenic distances (fig. 4.10A).

In sum, the data presented here confirms then important functional and structural consequences of balancing noise and plasticity. While transcriptional modulation allows *S. cerevisiae* to control the basic growth/stress programs[14,133,116], both transcriptional and translational modulation anticipate genomic and gene anatomy. Future analysis of additional questions, e.g., role of post-transcriptional regulation, potential presence of condition-dependent variation[120], or extent and relevance of coupling in higher eukaryotes, should ultimately expose the many aspects of gene expression variation.

4.6 Methods

Gene expression variability. To quantify gene expression plasticity, i.e. responsiveness to environmental change, the variability in mRNA levels among > 1500 different growth conditions was measured. as the sum of squares of the $\log_2$-ratios over all these conditions[218]. Noise, or stochastic variability, was measured by proteomic analysis[156]. We used the "distance to the median" (DM) score, which rules out confounding effects of protein abundance, allowing protein-specific noise levels to be compared. Evolutionary divergence in gene expression was measured as the variation of gene expression between orthologs in in four related

yeast species and 32 different conditions[218]. These three gene expression variability measures we scaled between 0 and 1. After scaling, mean values are 0.06166, 0.08877, and 0.1863 for plasticity, noise and expression divergence, respectively. For each of these measures, we define three categories "high", "medium" and "low", using percentiles 25 and 75 in each case as boundaries. Thus we obtain groups of genes with high noise and high plasticity (HNHP), high noise and low plasticity (HNLP) and so on (see also Table S4). Finally, mRNA level in rich media was obtained from[102] (mean=3.915 mRNA copies/cell).

Genomic localization and neighborhood. The coordinates of each transcript (coding ORFs–ORF-T; and noncoding, which can be in turn "cryptic unstable transcripts"–CUTs, and "stable untranslated transcripts"–SUTs) were obtained from a high resolution transcriptomic analysis[239]. For each of these transcripts, we used chromosomic coordinates of transcription start sites (TSS) and transcription end sites (TES), and orientation (strand). This data allowed us to characterize the genomic neighborhood of each transcript, in terms of distance to its upstream partner in bp, orientation of this upstream partner (which can be divergent or parallel). As well, from[239] we obtained data describing for each gene if it is transcribed from a bi-directional promoter or not based on the existence of a shared nucleosome depleted region (NDR). For the genomic neighborhood analysis, in order to maximize its reliability, we removed from the dataset gene whose upstream partner was a "pseudogene" or a "dubious ORF", as well as a few confounding cases where adjacent transcripts were overlapping. For some genes upstream distance could not be calculated as TSS and/or TES coordinates could not be accurately determined in the original source (see[239]).

Promoter characterization and regulation. The presence/absence of TATA boxes at the promoters was obtained from[14]. Nucleosome occupancy data for the whole genome was

obtained by DNA digestion with micrococcal nuclease and identification of nucleosome-protected fragments by high resolution microarray analysis[122]. We use the $\log_2$-ratios provided in the reference. As suggested in[215], we obtained two different nucleosome occupancy values for each promoter. Taking as reference the TSS, proximal nucleosome occupancy was the average in the –100 to obp region, while distal nucleosome occupancy corresponds to the -400 to -150 bp region. For an idea, the highest occupancy for a proximal region in our dataset was 0.27, and the lowest -3.64.

Transcription regulation data. To explore chromatin regulation, we used a compendium, assembled in[205], consisting of 170 expression profiles for chromatin regulation related mutations (expressed in $\log_2$-ratios). We classified these mutations in three classes (see table S1). "Chromatin" tag was assigned to mutations in histone acetyltransferases, deacetyltransferases, methylases, demethylases, ubiquitinating and deubiquitinating enzymes, chromatin remodellers and silencing factors. "General" tag was assigned to genotypes involving at least one mutation in essential, general transcription factors. "Histone" tag was assigned to mutations in the very histones. As suggested in[38], we normalized each dataset from the compendium to unit variance. The absolute value of the normalized $\log_2$ ratios represented responsiveness measures; the mean responsiveness of each gene represented its "chromatin regulation effect" (CRE) or "histone regulation effect", HRE). We used also this normalized dataset without taking the absolute value to analyze the sense of the observed regulation. Data for nucleosome-normalized, chromatin modification states at promoter were obtained from ChromatinDB (http://www.bioinformatics2.wsu.edu/chromatindb) which unifies several experimental genome-wide datasets measuring levels of different histone modifications. For dependence of each gene on general TFs, we used categorical data from[104]

defining for the expression of each gene if it is dominated by TFIID or SAGA complex.

Translation related measures. We used a measure of translation efficiency obtained in[136] and based for each gene in percent of each transcript in polysomes, its ribosome density, and the relative transcript level (mean$= 4.35$, SD$= 1.72$). We used an additional dataset of ribosome density obtained from[7] (mean$= 0.53$, SD$= 0.31$).

Noise-sensitive genes. We considered essential genes from the *Saccharomyces* Genome Deletion Project and genes especifying proteins in complexes[231]. Due to a big reduction in sample size, we excluded from this group haploinsufficient genes[53]. However, note that they are virtually not excluded, since $44/46$ (95%) of the identified haploinsufficient genes are labeled as either essential or as complex-forming; indeed, 100% of the haploinsufficient genes located in the HP group are so.

5

Discussion

A WIDELY APPLIED STRATEGY to attack complex systems is to sub-divide them in simpler, more operative "compositional units" with a reduced complexity, which can be then more easily understood. It is the case with living systems as well: we subdivide ecosystems into species, populations into individuals, and genomes in genes. However, neither the bigger system can be defined solely by its composition, nor the components can be fully understood without considering their context. Indeed, the genotype-to-phenotype mapping is a direct consequence of how information contained in the components of the system is combined and transferred to produce a phenotype. In this dissertation we attempted to understand

some specific aspects of this very broad question.

One of the central topics in this dissertation was the study of how genetic changes affect the phenotypic outcome of genomes. Genetic perturbations have represented possibly the most widely used approach to the characterization of biological components and systems. But perturbations in different genes often combine in unexpected ways to produce a phenotype. This *genetic interaction* can reflect different types of underlying functional link between the genes involved. In Chapter 1, we asked how genetic perturbations modify in turn other interactions between the remaining components of the network.

We tackled this question using *constraint-based* metabolic modeling. This methodology has several advantages (in addition to offering a relatively accurate prediction of experimentally observations, see below)??. Firstly, it allows automated measure of a large number of mutant phenotypes. Note that our analysis implied all possible single, double and triple mutants, which roughly amounts to n^3 measured phenotypes for a genome containing n genes. Besides the methodological difficulty in experimentally obtaining triple mutants in a high throughput manner, if such analysis is to be systematic and minimally exhaustive, it involves an enormous number of genotypes and measures[220]. Second, the constraint based modeling allows not only measuring phenotypes, but also the underlying distribution of metabolic fluxes. Finally, such a modeling framework eliminates all possible, known and unknown, confounding factors (e.g. regulation or environmental fluctuations). This could represent a problem if the goal is to accurately predict the exact phenotypes observed experimentally. In our case, we were most interested in the general principles underlying stability of genetic networks. From this perspective, the model is quite valuable, as it enables an univocal interpretation of the observed phenomena in terms of the underlying network structure and flux

distributions.

Overall, our analysis distinguished two broad subsystems exhibiting highly contrasting behavior: catabolic pathways on the one hand, and biosynthetic on the other. Catabolic pathways – including central metabolic ones – incorporate the carbon source (glucose) and break it down to produce energy in different forms (ATP, NADH, NADPH) and diverse small molecules that act as biosynthetic precursors. Two main characteristics of this "catabolic subsystem" resulted crucial. Firstly, its different functionalities are *degenerate*. Many of these can be performed in several, biochemically distinct ways. For instance, ATP can be synthesized either in respiration or in several glycolytic or fermentative reactions at the substrate level. Another example is pentose phosphate pathway, which can take over some steps of the initial glucose breakdown. This degeneracy[234] is intimately linked with pleiotropy at the metabolite level, i.e. the multi-purpose nature of some hub metabolites such as acetyl-CoA or pyruvate, but most notably, the so-called currency metabolites.

These features lead to an abundance of compensatory possibilities per gene, reflecting a high degree of functional versatility. As a result, the mostly nonessential genes in these pathways form a densely connected genetic network, enriched in weak and positive interactions, and highly prone to rewiring when one component is deleted. This is, of course, associated to flux re-distributions to these alternative, compensatory pathways. These observations also highlight also that more connected genes, or "hubs", are more prone to rewiring. This could be useful to understand why they often act as phenotypic capacitors[217,130].

In contrast, neither degeneracy nor pleiotropic metabolites are found in biosynthetic pathways. These are generally linear, and involve molecules with a very specific structure with the sole purpose of serving as intermediates in the corresponding pathway. Biosynthetic enzymes

are often essential, and possible compensatory mechanisms are usually based on redundancy (not degeneracy). Thus, genetic networks associated to biosynthetic pathways are smaller, more sparsely connected, and mostly composed of stable synthetic lethal interactions.

We also examined another class of genetic backgrounds, defined by trajectories of accumulation of neutral deletions. These genotypes are relevant from the point of view of evolution. Firstly, neutral evolution can reveal cryptic variability[65]. Indeed, we show that these genotypes generally show smaller, but more dense networks than the original one. Moreover, rewiring can unveil new, background-dependent interactions. This stresses that genetic interaction networks, when considered in a single background, do not reflect a big part of the *potential* functional associations.

The accumulation of neutral deletions also important in phenomena such as genome streamlining[81,162]. For example, evidence suggests that intracellular parasitic bacteria have lost many unnecessary genes in the adaptation to this constant-environment lifestyle. We observed in fact that neutral trajectories displaying stronger genetic network rewiring (in fixed conditions) were also associated to a heavier loss of environmental plasticity. The underlying mechanism is again related to the versatility of central metabolism. Pathways that can serve as less-efficient compensatory alternative for one carbon source could be the most efficient (or even essential) in another. Then, their inactivation produces both genetic network rewiring and a reduction in environmental plasticity. Ultimately, this observation provides a mechanistic rationale for the link between genetic and environmental robustness[212].

Finally, one of the most important evolutionary consequences of epistasis is the emergence of multiple-peaked fitness landscapes. But populations in the "wild" show a remarkable genetic diversity[57]. It is therefore an open question how the malleability of genetic networks

affects possible evolutionary paths remain an open question. From a more applied (although related) perspective, some recent studies have attempted to predict the order in which mutations appear during the development of cancer. The question of how the instability of genetic interaction networks impacts the evolutionary process and its predictability represents an interesting perspective for future research.

Epistatic interactions and networks can be highly informative on how genes combine to produce a phenotype. Indeed, one of their main values is in revealing phenotypic outcomes of sometimes quite obscure mechanisms operating at the molecular level[125]. But even in a situation when the components of a system are well characterized, the particular way these interactions combine (the "architecture") can lead to quite unintuitive dynamics. This is especially the case in regulatory systems incorporating feedbacks[5]. Indeed, different feedback architectures can lead to very interesting outcomes from the adaptive point of view; as a result, some of them are highly enriched as building blocks of regulatory networks. Among these, especially intriguing is the motif incorporating both a positive and a negative feedback, because these two lead in principle to opposed effects. In Chapter 2, we characterized both mathematically and experimentally such architecture using the *Escherichia coli mar* operon as a model.

Crucially, the strong negative feedback is coupled in this system to a weak and monomeric (linear) positive one. All these features appear as crucial to understand its effects. Firstly we show that the dominant effect of this dual feedback is to speed-up the response. This is expected from the negative feedback (marR, which is as well the signal sensor), but unexpected from the positive one (marA). Instead, the weak and linear effect of marA serves to further increase the promoter strength in de-repressed promoters, contributing to its speed-up. How-

ever, the system is buffered in absence of signal and quite homogeneous in the induced state, which should not be observed in case of a strong and/or nonlinear positive feedback.

Additionally, the motif does not act in isolation, but is embedded in the broader regulatory network. We thus investigated how additional inputs integrate and affect the response. Importantly, one of these inputs is cyclic AMP (through CRP), a major homeostasis regulator. It is not clear why would such an input be beneficial in the case of a resistance system. We show however that an additional regulatory input (Rob) limits potentially undesired cross-talks to cAMP signaling. We suggest then that the activation by CRP might be interpreted as an evolutionary contingency

Finally, we theoretically examined the effect of a more compact genetic implementation (by means of a bifunctional regulator). This hypothetical alternative is predicted to exhibit a smaller input range (i.e., the reaction to a signal gradient is more sigmoidal), a wider output range (what enables, for example, to differentially regulate large regulons, which could be the case of CRP), and a slower response (with no speedup at high dosages). In sum, the dual regulatory architecture is thus quite distinctive in its organization and function in order to trigger a response that is fast, graded, sensitive, robust, and coherent across the population.

Regulatory systems such as *mar* appear as a evolutionary fine-tuned, sophisticated and specific response mechanisms. But regulatory mechanisms exist as well that hardly fit in this paradigm. One might consider, for example, bacterial σ factors, or eukaryotic chromatin-based global regulation systems. Such mechanisms appear to share several characteristics. The first is a frequent partial redundancy. Different σ factors are not mutually exclusive in *E. coli*, and many promoters can use them indifferently. There is also abundant evidence of functional redundancy for eukaryotic chromatin effectors[127,116] or general transcription fac-

tors[121]. Second, these mechanisms often elicit global responses, involving large sets of genes that appear frequently as functionally heterogeneous. Thus, one can envisage a scenario in which, apart from the evolved and fine-tuned response to "known" stimuli, some sort of "coarse-grained" genomic responses are as well available.

We hypothesized that one strategy to cope with situations to which a cell is not prepared could be the evolution of stereotypical expression programs, a sort of "generic tools" to deal with situations in which no specific and fine-tuned mechanism is available. In Chapter 3, we presented evidence for the existence of tens of such programs in yeast. These not only appear recurrently in response to many different deletions, but also in response to environmental perturbation or after compensatory evolution. The existence of such a "generic" regulation mechanism could partly explain why stress-response genes are not only highly regulated in *trans*, but also encode elements influencing gene expression variability in *cis* in a general way, such as TATA boxes[14]. Indeed, the stereotypical responses appear greatly enriched in such genes.

The possibility of global, unspecific and sub-optimal responses as an evolutionary design principle has been scarcely explored in the context of genomic expression. This is possibly due to the common view of cells as highly optimized systems by billions of years of evolution. In other fields, however, the adaptive value of such strategies is readily recognized. For instance, humans often rely on stereotypical responses (heuristics) to make decisions in an environment that is often complex, *computationally intractable* or uncertain[80]. In such setting, rational responses can be highly costly, if not impossible. A fast response can be then highly beneficial, even if far from optimal. Our results suggest that yeast cells could also elicit stereotypical responses, as a sort of "educated guess" when no specific mechanism (in analogy to a rational

response) is available. Confirming an characterizing such possibility represents an exciting perspective for future research.

Finally, in Chapter 4 we explored the coupling between gene expression noise and plasticity (arising from regulation) in eukaryotes[124,38]. Unlike plasticity, noise can be generally considered a detrimental feature[123]. However, it can be modulated by several parameters during the process of gene expression. The main strategy to achieve high plasticity without noise is linked to a regulation by general transcription factors, most notably `TAF1`. This is strongly linked to the growth related expression program, highly enriched in ribosomal protein genes. In these genes, there has been arguably a strong selective pressure against noise because they are highly expressed and part of an essential protein complex.

In contrast, the cost of increasing noise seems to be compensated by the benefit of achieving a more versatile regulation (and thus a more diversified phenotipic abilities). This strategy involves a well-positioned promoter nucleosome (presumably enhancing sensitivity to chromatin regulation) and a TATA box (increasing re-initiation when the promoter is permissive). Moreover, this group is enriched in stress-related genes, with markedly lower expression level – and consequently lower cost of noise – than growth-related ones. An interesting possibiity is that noise in stress-related genes could enable the presence of sub-populations pre-adapted to stress, conforming an effective "bet hedging" strategy. Finally, we also show how noise and plasticity can be balanced through genomic architecture features[230], and how these features are shaped according to the need of specific types of noise-sensitive genes.

In summary, we believe this dissertation presents interesting insights into the mechanisms and strategies that shape genomic output. Perhaps most importantly, it proposes new hypotheses and questions, opening exciting perspectives for future research regarding the gen-

eral principles that shape the evolution of the living.

6

Conclusion

- Degeneracy associated to catabolism and central metabolism leads to highly connected, but unstable, genetic interaction networks.
- Neutral evolution trajectories can lead to an extensive rewiring, which is associated to a reduction in environmental plasticity. This mechanistically links genetic and ennvironmental robustness.
- Linear positive feedback can speed-up the response of a dual auto-regulatory circuit, and produces an homogeneous response
- Additional feedback inputs can help minimize the effect of regulatory cross-talk with

potentially undesired signals.

- A polycistronic implementation of dual autogenous control shows higher sensitivity and faster response.
- We found evidence of global stereotypical gene expression responses of yeast in response to a variety of different perturbations.
- In the regulation of eukaryotic genes, strate plasticity can be achieved without noise at the cost of reduced regulatory versatility
- Noise in genes with low plasticity is modulated at translation and correlates to gene length.
- Neighboring genome architecture offers limited abilities as a mechanism to uncouple plasticity from noise.

7

Conclusión

- La degeneración funcional asociada al catabolismo y metabolismo central conlleva una red genética altamente conectada, pero inestable.
- La evolucón neutral puede producir una significativa reorganización de la red genética, reflejando mecanismos comunes a la robustez genética y la ambiental.
- La retroalimentación positiva lineal acelera la respuesta en un circuito con autoregulación dual y produce una respuesta homogénea.
- Retroalimentaciones positivas adicionales ayudan a minimizar el efecto de señales potencialmente no deseadas.

- La implementación policistrónica de la autoregulación dual tiene como consecuencia una respuesta más rápida y una mayor sensibilidad a la señal.
- Hemos encontrado evidencia de programas de expresión gobales y estereotípicos en respuesta a varios tipos diferentes de perturbación.
- En la regulación de los genes eucarióticos, es posible obtener altos niveles de plasticidad evitando una expresión ruidosa, con el coste de una pérdida de versatilidad regulatoria.
- El ruido en genes poco plásticos se modula a nivel traduccional y está correlacionado con la longitud génica.
- La influencia del entorno genómico ofrece posibilidades limitadas como mecanismo para desacoplar la plasticidad del ruido.

A

Experimental and mathematical methods related to Chapter 2

A.1 Experimental Methods

A.1.1 Strain construction

We engineered a two-color fluorescent reporter strain (*Escherichia coli*) to measure the activity of the *marRAB* promoter, following the work by Miyashiro & Goulian (2007)[150*]. The strain (named as `IE01`) contains a chromosomal copy of the *yfp* gene (Yellow Fluorescent Pro-

*We thank Tim Miyashiro for strains, and Mark Goulian for strains and experimental advice.

tein) under the control of the marRAB promoter, and the *cfp* gene (Cyan Fluorescent Protein) expressed with a constitutive tetA promoter at the attachment sites in *E. coli* of lambda and HK022 phages, respectively.

To construct the YFP reporter, we amplified the promoter region of the *marRAB* operon (1616883-1617144) by PCR. The primers used were: `5'- TAGCAGAATTCCGGCAGCAAC` (forward), and `5'- GCATAGGGATCCTGGCAAGTAAT` (reverse). The PCR product was then cloned it into a plasmid (`pTM74`[150]) that contains a multicloning site upstream of the promoter-less *yfp* gene to yield the plasmid *pJFP01* (this work). This plasmid was integrated as single copy into the lambda phage attachment site of *E. coli* `MG1655` using the helper plasmid `pInt-TS`[233]. The residual *cat* gene linked to the *marRAB* promoter was replaced by *kan* by electroporating a PCR product created from the template plasmid `pKD4` into the appropriate strain carrying plasmid `pKD46` to yield the strain `JFP01` (this work). The two-color reporter strain `JFP02` was created by moving the previous marker into strain `TIM64` (a strain derived from a `MG1655` that constitutively expresses CFP; `TIM64` does not present any antibiotic resistance marker) by P1 transduction. Removal of the *kan* marker using plasmid `pCP20` originated *E. coli* strain `IE01`. This was verified by PCR. In addition, *E. coli* strain `IE02` was constructed by deletion of the rob gene in `IE01` with the application of a Datsenko & Wanner knockout protocol followed by the removal of the kan marker u ing plasmid `pCP20??`. The primers used were: `5'- ATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCGCTGGAGCTGCTTCGAA` (forward), and `5'- ATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATGAATATCCTCCTTAG` (reverse). The sequence of the resulting strain was verified by PCR. We also constructed a strain deleting the marB gene (named as `DB01`)[†]. This

[†]We thank Jerónimo Rodríguez-Beltrán for practical assistance.

was done from strain IE01 using a variant of the Datsenko & Wanner[51] knockout protocol, with the insert (kan marker with *marB* flanking regions) obtained by PCR from a KEIO collection strain[9]. Strain IE01 containing the pKD46 plasmid was electroporated with this insert in presence of arabinose, allowing recombination in the *marB* locus. The primers used to obtain the insert were: 5'- GACCAATATGCAGGGCGAATCG (forward), and 5'- ATGTATTTGGCTTGCGGTGG (reverse). Correct substitution of the *marB* gene by kan marker was checked by PCR and sequencing.

A.1.2 Culture media and reagents

Medium LB was always used for overnight cultures. Minimal medium M9 (M9 salts $1x$, $MgSO_4$ 2mM, $CaCl_2$ 0.1mM, glucose 0.4%, casamino acids 0.05%, vitamine B1 0.05%) was used to grow cells during characterization experiments. To induce the mar circuit, we used different concentrations of Salicylate (Sigma), resulting in a gradient ranging from 0 to 5mM, and also different concentrations of cAMP (Sigma), resulting in a gradient ranging from 0 to 20mM. When appropriate, we used kanamycin at a concentration of 50μg/mL.

A.1.3 Quantification of fluorescence in a cell population

Experiments of induction were carried out to study the dynamics of the system. Cultures (volume of 2mL) inoculated from single colonies (three replicates) were grown overnight in LB medium supplemented with glucose 0.4% at temperature of 37°C and shaking of 170rpm. Cultures were then diluted 1:200 in M9 minimal medium (10μL of culture into a final volume of 2mL) and were grown for 2h at temperature of 37°C and shaking of 170rpm. Cultures were then used to load the multiwell plate (Thermo) with final volumes of 200μL. Per well, we added directly a volume of the culture, a volume of salicylate stock, and a volume

of cAMP stock (when appropriate)[‡]. The set up for the Victor $\times 2$ was as follows. OD_600 measured with absorbance filter of 600nm (0.5s for reading). YFP measured with excitation filter of $497/16$nm, and emission filter of $535/40$nm. CFP measured with excitation filter of $434/17$nm, and emission filter of $479/40$nm. For both YFP and CFP, we took 8000 for energy lamp, and 0.1s for counting. Temperature was at $37°$C. The program started first with OD_600 , then YFP, and finally CFP, followed by 30s of shaking in orbital mode. Then it waited for 5min, and it started again.

A.1.4 Analysis of fluorescence data

We collected time-course data of fluorescence (YFP and CFP) and absorbance. Background values of absorbance and fluorescence, which corresponded to M9 minimal medium, were subtracted to correct the signals. The normalized fluorescence (for both YFP and CFP) was calculated as the ratio of fluorescence and absorbance. Similar values of normalized fluorescence were reported for `MG1655` cells and for M9 minimal medium, which indicated that the auto-fluorescence of cells was negligible in this case. The growth rate of cells was calculated as the slope of the linear regression between the log of corrected absorbance and time in exponential phase. Time-dependent promoter activity, defined as the instantaneous production rate of normalized YFP fluorescence (magnitude per cell), was calculated for each time point using the derivative of the normalized fluorescence. Promoter activity in steady state, defined as the stationary production rate of normalized YFP fluorescence (magnitude per cell) in exponential phase, was calculated as the average over time (for $t > 2$h) of normalized fluo-

[‡]We also used mineral oil for characterization. Cultures were grown as previously indicated, and then were used to load the multiwell plate (Thermo) with final volumes of 250μL. Per well, we added directly 200μL of culture (counting the volume of inducers), and 50μL of mineral oil (Sigma). However, we found that the mar circuit does not respond to salicylate in anaerobic conditions.

rescence times growth rate. The error associated to that measure was obtained by calculating the standard deviation over replicates in all time points, then squaring all these deviations to average them, and finally getting the root square. Data were analyzed with Matlab (MathWorks).

A.1.5 Quantification of mRNA levels by qRT-PCR

Cultures (volume of 2mL) inoculated from single colonies (three replicates) were grown overnight in LB medium at temperature of 37°C and shaking of 170rpm. Cultures were then diluted $1:200$ in M9 minimal medium (10μL of culture into a final volume of 2mL) and were grown for 2h at temperature of 37°C and shaking of 170rpm. Then, aliquots of 1.5mL were quickly pelleted by centrifuging 2min at 13000rpm and resuspended in 50μL of TE buffer (10mM Tris-HCl, pH 8.0, 1mM EDTA). Bacteria were broken by adding 50μL of a mix $1:1$phenol:chloroform (pH8.0) and vortexing thoroughly. Bacterial RNA from each sample was recovered in the aqueous phase by centrifuging 5min at 13000rpm. A subsequent re-extraction was done by adding 50μL chloroform followed by centrifugation of 5min at 13000rpm, recovering the aqueous phase. Samples were then passed through a silica-based, DNA-clean column (Zymo) and were eluted in 8μL of TE buffer. DNase I (Fermentas) was added (1U) for 1h at 37°C, which was then thermally inactivated (30 min at 65°C) by adding EDTA. Total RNA eluted was quantified in a NanoDrop. One-step SYBR PrimerScript RT-PCR Kit II (Takara) was used for detection, following the Kit protocol for preparing the reaction volumes. 16S rRNA was used as housekeeping gene to normalize RNA quantity in each reaction. The primers for qRT-PCR were synthesized by IDT for marA and Sigma for 16S rRNA. The primer sequences for amplifying *marA* were taken from [225]) and for 16S rRNA

from[31]. Reactions in triplicate were carried out using a Step One Plus Real-Time PCR System (Applied Biosystems). The thermal cycling program for amplification was 5min at 42°C, 10 s at 95°C, and 40 cycles of 5 s at 95°C and 34 s at 60°C (Shuttle PCR); following with default melting curves[§].

A.1.6 Quantification of fluorescence in single cells

Experiments of induction were carried out to study the heterogeneity of the dynamic response. Culture (volume of 2 mL) inoculated from a single colony was grown overnight in LB medium supplemented with glucose 0.4% at temperature of 37°C and shaking of 170rpm. Culture was then diluted 1 : 200 in M9 minimal medium (10µL of culture into a final volume of 2mL) and was grown for 4h at temperature of 37°C and shaking of 170rpm. The different cultures diluted 1 : 10 were then used to load the agarose pads. Agarose pads were prepared with a volume of 5mL of M9 minimal medium and 0.075g of agarose. It was dissolved by vortexing and microwaving. The pads were then allowed to solidify for about 1h at room temperature before seeding bacteria. 2µL of each culture were then used to load the pads. They were kept for about 15min at room temperature so that cells can be absorbed into the agarose. Just before characterization with the microscope, 2.5µL salicylate from a solution 0.1M was used to induce cells in solid medium, having estimated the volume of the agarose pad in 50µL (resulting concentration of salicylate about 5mM). In each pad, fields with an adequate initial density of cells were chosen, the first photo was taken, and salicylate was added. Photos were taken for each field every 12 minutes. Agarose pads were monitored in an inverted microscope Axiovert200 (Zeiss) with objective 100 × /1.45 oil Plan-Fluar at

[§]We thank Eszter Majer for help with the RNA extraction and Anouk Willemsen for help with the qRT-PCR.

temperature of $30°$C. The microscope was equipped with a digital camera C9100-02 (Hamamatsu), a Xenon lamp XBO 75W/2, an optical filter changer Lambda 10-2 (Sutter), a motorized stage (Marzhauser), and a temperature controller (Cell Observer-Zeiss). Moreover, the microscope was automated by the commercial software MetaMorph (Universal Imaging). Cell images were acquired from the bright-field and fluorescence channels. We used the fluorescence filters yellow FP ($490 - 510$ nm, $510 - 560$ nm) and cyan FP ($426 - 446$ nm, $460 - 500$ nm).

A.1.7 Analysis of single cell images

Microscopy photos were segmented and analyzed using the EBImage package for R from Bioconductor. Segmentation allowed us to identify sets of pixels belonging to individual cells, and measure the average apparent YFP and CFP intensities for each of these sets. The ratio YFP/CFP was used as proxy for system activity.

A.1.8 Quantification of fluorescence by flow cytometry

Cultures (volume of 2mL) inoculated from single colonies were grown overnight in LB medium without and with 5mM salicylate at temperature of $37°$C and shaking of 170rpm (to have uninduced and pre-induced cultures). These cultures were subsequently diluted $1 : 200$ in 2mL of M9 minimal medium with different concentrations of salicylate (0, 0.05, 0.25, 0.5, 2.5 and 5mM). After incubation for 4h at the same conditions, cells were spun down and resuspended in PBS buffer. YFP (525nm) intensity was measured in the flow cytometer Gallios (Beckman-Coulter). YFP distributions were obtained with a standard flow cytometry analysis (software Kaluza).

A.2 Mathematical methods

A.2.1 Empirical model of the dynamic response

Our reporter protein (YFP) monitors the promoter activity of the *marRAB* operon (P_{mar}). Then, we can write

$$\frac{d[YFP]}{dt} = \Pi_{mar} - \mu[YFP] \tag{A.1}$$

where μ is the cell growth rate and Π_{mar} the activity of promoter P_{mar} . Empirically, we proposed the following exponential model

$$\left(\frac{m\lambda}{e^{\lambda t}-1}+\mu\right)\left(1-e^{\lambda t}\right) \tag{A.2}$$

where λ and m are two parameters that control the dynamics of the response, and $[YFP]$ is the concentration in steady state. We fitted with our own experimental data by nonlinear regression the parameters λ and m. Combining Eqs. A.1 and A.2, promoter activity was calculated as

$$\Pi_{mar} = \left(\frac{m\lambda}{e^{\lambda t}-1}+\mu\right)\left(1-e^{-\lambda t}\right)^{m} \tag{A.3}$$

The time to reach the 50% of the steady state value (t 50) upon induction with salicylate reads

$$t_{50} = \frac{-1}{\lambda} ln\left(1-\frac{1}{2^{1/m}}\right) \tag{A.4}$$

We also considered the response speed, here defined as $\frac{1}{[YFP]}\frac{d[YFP]}{dt}$ and evaluated at t_{50}. This

gives $ln(2)\lambda$ for eq. A.2 and $m \gg 1$. In the case of a non-regulated gene whose dynamics is governed by $(1 - e^{\mu t})$, the response speed is directly μ. In addition, differentiating Eq. A.3 we calculated the time when promoter activity is maximal (τ), given by

$$\tau = \frac{1}{\lambda} ln \left(\frac{m\lambda - \mu}{\lambda - \mu} \right) \tag{A.5}$$

We used this model to fit either population or single-cell experimental data.

A.2.2 Bottom-up mathematical model

We constructed a system of ordinary differential equations (ODEs) based on the topology of the circuit. MarR, MarA and MarB form an operon controlled by promoter P_{mar}[2,40]. MarR represses P_{mar}, which can be modulated by salicylate[41]. MarB also represses P_{mar}[224], whereas MarA activates it[142]. On the other hand, Rob is controlled by promoter P_{rob} , and it activates P_{mar}[2]. MarA and Rob repress P_{rob}[191]. Finally, in our system YFP models a downstream gene controlled by promoter P_{mar}. Although MarR, MarA and MarB are transcribed from the same promoter, the corresponding protein expressions may be different each other due to distinct translation rates. By analyzing the 5' untranslated regions of MarR, MarA and MarB with RBS calculator[188], considering the 30 nucleotides upstream and the 7 nucleotides downstream of the start codon, we found that translation rates of MarA and MarB are about 30-fold and 20-fold, respectively, higher than the translation rate of MarR. This is in tune with previous experimental observations[145]. In addition, promoter P_{mar} is regulated by CRP-cAMP. Therefore, we could write

$$
\begin{aligned}
\frac{d[MarA]}{dt} &= \beta\Pi_{mar} - (\mu+\delta)[MarA] \\
\frac{d[MarR]}{dt} &= \Pi_{mar} - \mu[MarR] \\
\frac{d[MarB]}{dt} &= \beta^{*}\Pi_{mar} - \mu[MarB] \\
\frac{d[Rob]}{dt} &= \Pi_{rob} - \mu[Rob] \\
\frac{d[YFP]}{dt} &= \Pi_{mar} - \mu[YFP]
\end{aligned} \tag{A.6}
$$

where μ is the cell growth rate, δ the degradation rate of MarA ($\delta \ll \mu$), noting that MarA is degraded by protease Lon[91], β (and β^{*}) the fold increase of MarA (MarB) translation rate, and Π_{mar} and Π_{rob} the activity of promoters P_{mar} and P_{rob}, respectively.

The equations for Π_{mar} and P_{rob}, knowing that MarA and Rob act as monomers whereas MarR as a dimer[142], could be approached by Hill functions[22]. However, we knew that the effect of MarA on P_{rob} is not observable in physiological conditions[40], so we neglected this regulation to simplify the system of equations. We also demonstrated in this work by qRT-PCR that MarB does not repress Π_{mar} in presence of salicylate, whereas MarB indeed represses Π_{mar} in absence of salicylate. This also led us to eliminate MarB from the model. Therefore, it turned out

$$
\Pi_{mar} = \Pi_0 \frac{1 + f_a \frac{[MarA]}{K_A} + f_b Q([Rob],[cAMP])}{1 + \frac{[MarA]}{K_A} + Q([Rob],[cAMP])} \frac{1}{1 + \left(\frac{[MarR_{free}]}{K_R}\right)^2} \tag{A.7}
$$

where K_A and K_R are the effective dissociation constants for transcription regulation, and f_a and f_b the activation fold changes. Π_0 is the basal protein synthesis rate, and Q a regulatory

function that accounts for the activation of Rob and CRP-cAMP. Phenomenologically, and assuming competitive binding between the three transcriptional activators (MarA, Rob and CRP-cAMP) as their operators overlap, we can write

$$Q([Rob],[cAMP]) = \frac{[Rob]}{K_B} + frac[cAMP]K_C \tag{A.8}$$

where K_B and K_C are the effective dissociation constants for transcription regulation. This model can explain that, in presence of Rob or cAMP, MarA does not increase significantly the occupancy of polymerase at the marRAB promoter, which agrees with previous experimental and theoretical data[229]. In addition, we have

$$[marR_{free}] = \frac{1+\alpha\left(\frac{[Sal]}{\theta_S}\right)^{\nu_S}}{1+\left(\frac{[Sal]}{\theta_S}\right)^{\nu_S}}[MarR] \tag{A.9}$$

where θ_S is the effective dissociation constant between salicylate (Sal) and MarR, ν_S the Hill coefficient, and α the minimal fraction of free MarR.

A.2.3 Simplification of the mathematical model

Our bottom-up mathematical model can be simplified for a better analysis of the dynamic response. By noting $x = [MarA]/K_R$, $y = [MarR]/K_R$, $y_0 = [MarR_{free}]/K_R$ (the concentration of MarR is equal to the one of YFP), $\rho = f_a = f_b$ (assuming for simplicity equal induction by the three transcriptional activators), $\pi_0 = \Pi_0/K_R$, and also $\kappa = K_R/K_A$, we could write a simplified system of ODEs. We also denoted $z = ([Rob]/K_B)$ and $c = ([cAMP]/K_C)$ the normalized concentrations of Rob and cAMP, respectively (and we took $K_B = K_A$). Thus, we obtained

$$\begin{aligned} \frac{dx}{dt} &= \beta\pi_{mar} - \delta x \\ \frac{dy}{dt} &= \beta\pi_{mar} - \mu y \\ \pi_{mar} &= \pi_0 \frac{1+\rho(\kappa x+\kappa z+c)}{1+\kappa x+\kappa z+c}\frac{1}{1+y_0^2} \end{aligned} \tag{A.10}$$

where MarA could be approached to a quasi-steady state ($x \propto \pi_{mar}$, a function of time). And also we had

$$\frac{y_0}{y} = \frac{1+\alpha\left(\frac{[Sal]}{\theta_S}\right)^{\nu_S}}{1+\left(\frac{[Sal]}{\theta_S}\right)^{\nu_S}} \tag{A.11}$$

In case of maximal induction of the system with salicylate, α modulates the regulatory role of MarR. To obtain dimensional parameters, see Table S1 for values of K_R and K_A.

A.2.4 Analytical solutions of the model I: temporal behavior

In this section, we considered that a very strong the repressor acted on the system, MarR (i.e., $y \gg 1$). We also assumed that the system was induced with high levels of salicylate, so that y $0 = \alpha y$, and that the activation term was simply reduced to the fold change (ρ). This allowed us to simplify the model to just one ODE, given by

$$\frac{dy}{dt} \approx \frac{\pi_0\rho}{\alpha^2 y^2} - \mu y \tag{A.12}$$

The steady state of this ODE was

$$y_\infty = \left(\frac{\pi_0 \rho}{\mu \alpha^2}\right)^{1/3} \tag{A.13}$$

This has the following solution, after integration by separation of variables

$$\frac{y}{y_\infty} = \left(1 - e^{-3\mu t}\right)^{1/3} \tag{A.14}$$

However, this analytical solution gave an over-estimation of the dynamics when solving numerically the ODE. We then considered a situation at short times (i.e., $\alpha y \ll 1$) to write

$$\frac{dy}{dt} \approx \pi_0 \rho \left(1 - \alpha^2 y^2\right) \tag{A.15}$$

The steady state of this ODE was $y_\infty = 1/\alpha$. This has the following solution (after integration by separation of variables and partial fraction decomposition)

$$\frac{y}{y_\infty} = tanh\left(\pi_0 \rho \alpha t\right) \tag{A.16}$$

We verified with numerical simulations that this analytical solution is a good approximation for $t < t_{50}$. With this model, we calculated the response time of the system as $t_{50} = \frac{ln(3)}{2\pi\rho\alpha}$ (i.e. $y(t_{50}) = \frac{1}{2} y_\infty$).

A.2.5 Analytical solutions of the model II: dynamic range

Our model to describe the dynamics of a self-repressed, self-activated operon (y), implemented with two regulatory genes (repressor and activator), can be written as

$$\frac{dy}{dt} = \frac{\pi_0}{1+(\alpha_S y)^2}\frac{1+\rho\kappa' y}{1+\kappa' y} - \mu y \tag{A.17}$$

where α_S describes the effect of the inducer molecule (Eq. A.11), which acts on the repressor ($1 \geq \alpha_S \geq \alpha$). Here, MarA was assumed to be a stable protein ($\delta = \mu$). This assumption leads to $x = \beta y$, and does not change the dynamics of our protein of interest, MarR (y). This model can be solved in steady state for a strong repressor ($y \gg 1$) and high activation fold ($\rho >> 1$) as

$$y_\infty \approx \left(\frac{\pi_0 \rho}{\mu \alpha_S^2}\right) \propto \alpha_S^{-2/3} \tag{A.18}$$

For saturating levels of salicylate, we estimated with this model an output dynamic range of $R_{out} = \alpha^{2/3}$ (i.e. the ratio between the highest and lowest values of y_∞ when varying salycilate), and input dynamic range of $R_i n = 9^{2/n_{in}}$ (i.e. the ratio between the salycilate values at which we have $y_\infty = min(y_\infty) + 0.1\Delta y_\infty$ and $y_\infty = max(y_\infty) - 0.1\Delta y_\infty$, where $\Delta y_\infty = max(y_\infty) - min(y_\infty)$; n_{in} is an effective Hill coefficient (see also Goldbeter and Koshland, 1981[83]). For $\alpha = 0.05$, it turned out that $R_{out} = 7.37$ and $R_{in} = 28.32$ ($n_{in} = 1.31$).

For a constitutively expressed regulator (y), the dynamics of its regulated operon (y) can be written as

$$\frac{d\bar{y}}{dt} = \frac{\pi_0 \rho}{1+(\alpha_S y)^2} - \mu \bar{y} \tag{A.19}$$

where y is constant. The steady state solution for y is straightforward to obtain. Moreover, a model to describe the dynamics of a self-repressed gene (y), with equal production rate as before, can be written as

$$\frac{dy}{dt} = \frac{\pi_0 \rho}{1 + (\alpha_S y)^2} - \mu y \tag{A.20}$$

This model can be solved in steady state to obtain the same expression as before.

Finally, a model to describe the dynamics of a self-repressed gene (y) that becomes self-activator in presence of the inducer can be written as

$$\frac{dy}{dt} = \pi_0 \frac{1 + \rho \left((1 - \alpha_S)\, y\right)^2}{1 + (\alpha_S y) + \left((1 - \alpha_S)\, y\right)^2} - \mu y \tag{A.21}$$

where we have assumed competitive binding between the repressor and activator. This model can be solved in steady state, also for a strong repressor ($y \gg 1$) and high activation fold ($\rho \gg 1$), as

$$y_\infty \approx \frac{\pi_0 \rho}{\mu} \theta \left(1 - \alpha_S\right) \tag{A.22}$$

where θ is teh Heaviside function.

A.2.6 Signal transduction and de-repression

Our model also explains why the translation rate of MarR is limited. For the repressor to work as an efficient sensor, it has to reach an appropriate concentration. First, the concentration has to be higher than the effective dissociation constant (K_R , interaction protein-DNA). In this way, the regulator can exert its repressive action on the promoter (in absence of inducer). In the case of MarR, K_R is of the order of nM (Table S1), which allows having $[MarR] > K_R$ (lower bound). Second, at maximal concentration of inducer (salycilate), the promoter has to be de-repressed. This imposes the condition $[MarR_{free}] < K_R$. In the

marRAB operon, the environmental molecule (antibiotic or salicylate) does not interact directly with the repressor (contrary, for example, to the lac operon). Instead, Cu^{2+} is responsible for MarR oxidation[97], and the intracellular concentration of this cation is up to 10 μM[179]. This clearly imposes a maximal concentration of repressor. Knowing the oxidation curve of MarR by Cu^{2+} (Fig. S1G, and taking the effective oxidation constant as $K_{ox} \approx 1$), it turns out $[MarR] < (K_R[Cu^{2+}]^2)^{1/3}$ (upper bound). Therefore, we propose that an inefficient translation rate of MarR might have evolved to appropriately deal with copper signaling.

A.2.7 Apparent transfer function and stress cross-talk

The Hill-like model

$$\Pi_{mar} = \frac{\Pi_{mar}^{min} + \Pi_{mar}^{max} \left(\frac{[Sal]}{K_S}\right)^{n_S}}{1 + \left(\frac{[Sal]}{K_S}\right)^{n_S}} \tag{A.23}$$

was adjusted independently to the experimental data of induction of the wild-type and Δrob systems, obtaining the corresponding apparent transfer functions ($[Sal]$ is the concentration of salycilate in mM). We got $\Pi_{mar}^{min} = 2,260 \pm 746$ A.U., $\Pi_{mar}^{max} = 20,000$A.U. (fixed), $K_S = 0.88 \pm 0.12$mM, and $n_S = 1.28 \pm 0.19$ for the *wt* system, and $\Pi_{mar}^{min} = 1,089 \pm 1011$ A.U., $\Pi_{mar}^{max} = 12,500$A.U. (fixed), $K_S = 0.72 \pm 0.19$mM, and $n_S = 1.45 \pm 0.48$ for the Δ*rob* system. With the inferred sigmoidal models, we calculated the input and output dynamic ranges: $R_{in} = 20(30)$ and $R_{out} = 9(11)$ for *wt* (Δ*rob*).

Apart from salicylate, the marRAB operon can also be induced by CRP-cAMP. We predicted computationally and verified experimentally the two-dimensional transfer function of the system. Then, we defined a degree of stress cross-talk as

$$\left| \frac{[YFP]_{\infty}|_{cAMP}}{[YFP]_{\infty}|_{cAMP=0}} - 1 \right| \tag{A.24}$$

A.2.8 Stochastic modeling

Using a Langevin approach, our ODE-based mathematical model can be extended to account for the inherent stochasticity of biological systems. Because in bacteria noise has two components (intrinsic and extrinsic) and is predominantly generated at the transcription level (Swain et al., 2002), we considered a stochastic process ξ (where μ is the inverse of the correlation time) with statistics $\langle \xi_{\mu}(t) \rangle = 0$ and $\langle \xi_{\mu}(t')\xi_{\mu}(t'+t) \rangle = \frac{\mu}{2} e^{-\mu|t|}$. Thus, and having assumed similar mRNA degradation and MarR translation rates, we can write

$$\begin{aligned} \frac{dx}{dt} &= \beta\pi_{mar} - \delta x + \beta\sqrt{\frac{2\pi_{mar}}{K_R}}\xi_{\delta}(t) + \beta q\xi_{\mu}(t) \\ \frac{dx}{dt} &= \pi_{mar} - \mu y + \sqrt{\frac{2\pi_{mar}}{K_R}}\xi_{\delta}(t) + q\xi_{\mu}(t) \end{aligned} \tag{A.25}$$

where q is the extrinsic noise magnitude. This system can be solved numerically by following the method described in Rodrigo et al. 2011[180].

A.2.9 Nominal parameter values

Table A.1: Nominal parameter values used in the mathematical model.

Parameter	Value	Note	Reference
δ	0.5min^{-1}	Protein half-life of ~ 1 min.	[91]
μ	$0.007 - 0.01$ min^{-1}		This work
π_0	0.1min^{-1}	Assuming transcription rate of $\sim$ 1nM/min, translation rate of $\sim 0.05 - 0.1$ min^{-1}, and mRNA degradation rate of ~ 0.1 min^{-1}	[128]
β	30	With RBS calculator	[188]
ρ	10	Experimental data show $\sim$ 6-fold increase in expression due to the direct effect of MarA.	[143,40]
κ	0.02	Having $K_A = 1000$--1500nM and $K_R = 5-10$ nM (as $\kappa = K_R/K_A$)	[141,194]
θ_S	0.13mM	Adjusted to get the experimental transfer function. Estimated from $K_S \approx 0.9$mM, knowing that $\theta_S \approx K_S(2\alpha)^{1/\nu_S}$	[41]
ν_S	1.4		This work
α	0.05	Assuming [Cu^{2+}]/[MarR]$\approx$ 5 upon induction with 5mM salicylate	[97]
z	100	The number of Rob molecules per cell is ~ 10000, but they are inactive due to aggregation. The fraction that is free is assumed $\sim 10\%$	[90]
K_C	5mM		[195]
q	0.3	Knowing that $q = \eta_{ex}\pi mar(2/\mu)^{1/2}$	[209]

B

Chapter 2 Appendix

B.1 Singular value decomposition

Singular value decomposition is a form of matrix factorization that takes the form

$$\mathrm{M} = \mathrm{U}\Sigma\mathrm{V}^{\top} \tag{B.1}$$

where M is the data matrix, U and V are orthonormal matrices and Σ is a diagonal matrix. Columns of V represent the eigenvectors of $M^T M$, while columns of U are the eigenvectors of MM^T (in decreasing order). Σ contains the square roots of the eigenvalues of $M^T M$ (or MM^T). Effectively, U represents the orthonormal basis for the column space of M (its columns

are the right singular vectors). V represents the orthonormal basis for the row space of M (left singular vectors). These bases completely uncouple rows and columns of M from one another, diagonalizing it to Σ.

The relation between SVD and principal component analysis (PCA) is direct if we center the matrix M (i.e. subtract to each column its mean). Then, the matrix $M^T M$ is proportional to the covariance matrix of the columns of M.

SVD can be also understood as a decomposition of the matrix into a weighted sum of "simpler" matrices (called *components* or *modes*), where σ acts as a weighing factor:

$$M = \sum_{k=1}^{n} \sigma_k \mathrm{U}_k \mathrm{V}_k^{\top} \tag{B.2}$$

If we are interesnted in the fraction of total variation in M that is explained by mode i, this can be obtained either as the ratio between the corresponding single values, or as the ratio of the norm of the i-th projection to the norm of the entire dataset:

$$FV_i = \frac{\|\mathrm{u}_i \sigma_i \mathrm{v}_i^{\top}\|^2}{\|\mathrm{M}\|^2} = \frac{\sigma_i^2}{\sum_j \sigma_j^2} \tag{B.3}$$

We can use the same approach to compute the percentage of a genotype k that is explained by mode i:

$$R^2 = \frac{\|\mathrm{u}_i \sigma_i \mathrm{v}_{ik}\|^2}{\|\mathrm{M}_k\|^2} \tag{B.4}$$

Values of R^2 for all genotypes and modes are contained in the attached dataset R2.csv, that can be read with excel/libreoffice spreadsheets, or alternatively with the `read.csv()` function in R.

B.2 Log-linear modeling of the deleteome

We will consider that the expression level R_{ij} of gene i in genotype j, is the realization of a random Poisson variable with parameter $\lambda_{i,j}$[154]. In order to obtain the parameter value for each cell, we will use a loglinear model (which is a generalized linear model with a log link and Poisson distributed error):

$$log\, \lambda_{ij} = \eta + \theta_i + \delta_j \tag{B.5}$$

In this model, θ_i accounts for gene (row) effects, δ_j for genotype (column) effects, and η is a constant. We obtain the parameters using the "iterative proportional fitting" algorithm as implemented in the R function `loglin()`, which gives the maximum likelihood estimates of λ_{ij}. Then, one can sample a random value for each cell, obtaining random matrices that will maintain the overall structure of the original dataset but break nonrandom associations between genes or genotypes.

Two things must be pointed out. First, we are not exactly preserving the amount of *variation* per gene and genotype (i.e. the marginal sums of the M matrix). Rather, as we work with the R matrix, we are preserving the *total gene expression* per row and column. This could be a reason to explore alternative models (distribution assumptions) which would truly preserve the row and column amount of variation.

Second, besides the randomization, loglinear modeling also provides specific coefficients for the gene and genotype effect (θ and δ). Although we are using them only implicitly in this project, studying them could also provde some biological insight. For instance, if $\theta_i \gg \delta_j$ for a given gene, this could mean that the *cis* effect is much stronger than the *trans* one, and

vice-versa.

The data

Given the assumption of Poisson distribution, we need to work with positive values. For this reason we can not work with the M matrix ($log_2 R/G$ ratio, where G is the wt and R the mutant expression level). Therefore, we work with the R matrix. To obtain the G and R matrix, we used the M and A matrices available in the original publication (the A matrix being the mean value of R and G):

$$R = \frac{2A\,2^M}{1+2^M} \tag{B.6}$$

$$G = 2A - R \tag{B.7}$$

After randomizing the R matrix, we combine it with the original G matrix to build a randomized M.

B.3 Validation: prediction of differentially expressed genes

One way to validate the randomization method is to try to predict genes that are differentially expressed according to the original publication. We do this by constructing a log-linear model of the G matrix. The fitted (λ) values will tell us the expected Poisson distribution for each cell. Then, we can ask which observed R values fall in the tail extremes (i.e. are differentially expressed). Figure B.1 shows that, indeed, the log-linear model predicts quite well which genes are up and which are down-regulated.

Figure B.1: Prediction of differentially expressed genes by the log-linear model. Overlapping between up and down regulated genes as predicted from the wt log-linear model and from the original publication. Genes significantly down-regulated in the log-linear model are those in the lower 2% tail of the expected Poisson distribution for that gene and condition; analogously, upper 2% tail for the up-regulated. As in the original publication, for a gene to be significantly up or down-regulated, it was also required a fold change $FC > 1.7$.

Can we do better? To compute which genes are differentially expressed, we use two criteria: a fold-change higher than 1.7, and a $p < 0.05$. The Poisson distribution imposes a constraint in the fact that the mean is always equal to the variance. This implies that, for higher values of the parameter λ, we will detect smaller fold changes in the upper and lower 2% tails of the distribution, which we consider significant (figure B.2A). Therefore we attempted to multiply both the G and the R matrices by a factor ξ. This preserves exactly the fold changes but modulates how big they should be to be significant for our null model. Indeed, there is a value of ξ that maximizes the predictive ability of the log-linear model. (figures B.2B and B.3).

From these insights, we learn firstly that the log-linear model quite good in predicting known things about the data. Second, the need of a trick such as ξ indicates (again) that there could be better choices than assuming that values are Poisson distributed.

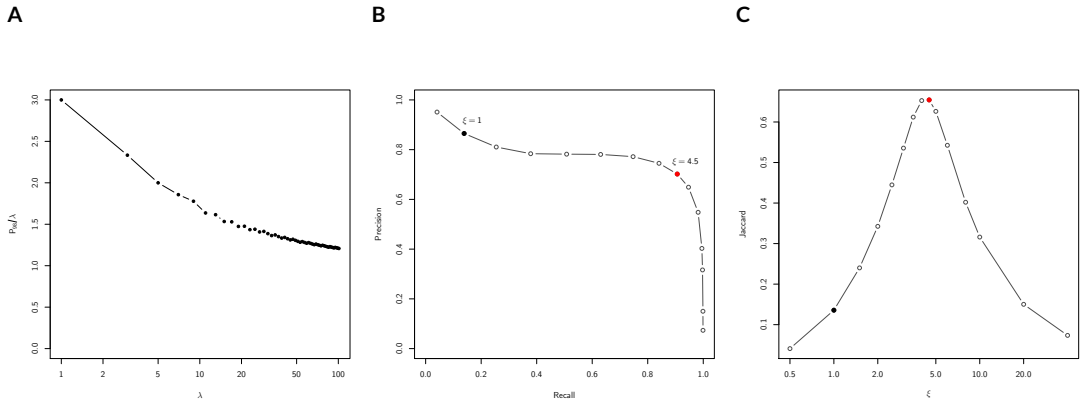
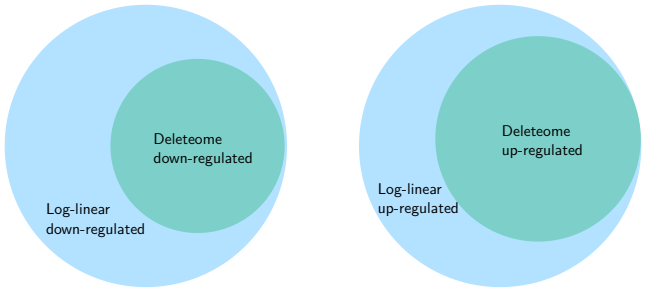


Figure B.2: Evaluation of predictive power of log-linear model A) Value of the 98 percentile of a Poisson distribution in units of lambda, as a function of lambda. B) We tried to predict the ``original'' differentially expressed genes using the log-linear model of the G matrix (wt). Given that les than 0.6% of the expression values in the entire matrix are significantly up or down-regulated, we used a precision-recall curve instead of a ROC curve to assess the predictive ability of the model, for different ξ values. C) We can also compute a Jaccard index, as a measure of how much the set of differentially expressed genes in the original paper overlaps with those predicted as differentially expressed by the log-linear model. We show the value of the Jaccard index for different values of ξ. In B and C panels, we colored in black the point for $\xi = 1$, i.e. using the original G and R matrices, and in red, the ξ value with the highest Jaccard index.

A ($\xi = 4.5$)



B ($\xi = 6$)

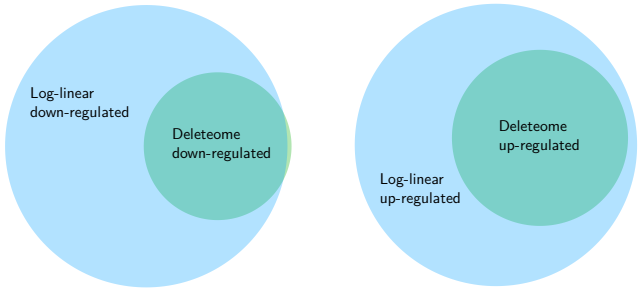


Figure B.3: Additional Venn diagrams showing the effect of ξ. Venn diagrams simmilar to those in the figure B.1, but mutliplying both G and R by two different ξ factors (4 above and 6 below).

References

[1] Aguilera, L., Campos, E., Giménez, R., Badía, J., Aguilar, J., & Baldoma, L. (2008). Dual role of LldR in regulation of the lldPRD operon, involved in l-lactate metabolism in Escherichia coli. *J. Bacteriol.*, 190, 2997–3005.

[2] Alekshun, M. N. & Levy, S. B. (1997). Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. *Antimicrobial agents and chemotherapy*, 41(10), 2067.

[3] Alon, U. (2003). Biological networks: the tinkerer as an engineer. *Science*, 301(5641), 1866–1867.

[4] Alon, U. (2007a). *An introduction to systems biology*. Chapman and Hall/CRC Press.

[5] Alon, U. (2007b). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8, 450–61.

[6] Angus-Hill, M. L., Schlichter, A., Roberts, D., Erdjument-Bromage, H., Tempst, P., & Cairns, B. R. (2001). A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. *Molecular cell*, 7(4), 741–51.

[7] Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., & Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), 3889–94.

[8] Avery, L. & Wasserman, S. (1992). Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends in genetics*, 8(9), 312–316.

[9] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Molecular systems biology*, 2(1).

[10] Bahn, Y.-S., Xue, C., Idnurm, A., Rutherford, J. C., Heitman, J., & Cardenas, M. E. (2007). Sensing the environment: lessons from fungi. *Nature Reviews Microbiology*, 5(1), 57–69.

[11] Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.-k., Chuang, R., Fiedler, D., Dutkowski, J., Guénolé, A., & Attikum, H. V. (2010). Rewiring of Genetic Networks in Response to DNA Damage. *Science*, (December), 1385–1389.

[12] Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., Hibbs, M., Hess, D., Gingras, A.-C., Bader, G. D., Troyanskaya, O. G., Brown, G. W., Andrews, B., Boone, C., & Myers, C. L. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods*, 7(12), 1017–24.

[13] Baryshnikova, A., Costanzo, M., Myers, C. L., Andrews, B., & Boone, C. (2013). Genetic interaction networks: toward an understanding of heritability. *Annual review of genomics and human genetics*, 14, 111–33.

[14] Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116(5), 699–709.

[15] Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T., & Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17), 6355–6360.

[16] Batada, N. N. & Hurst, L. D. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature genetics*, 39(8), 945–9.

[17] Bateson, W. (1979). *Mendel's Principles of Heredity*. Cambridge University Press, Cambridge.

[18] Baumberg, S. (1999). *Prokaryotic gene expression*. Oxford University Press.

[19] Becskei, A., Séraphin, B., & Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO journal*, 20(10), 2528–2535.

[20] Becskei, A. & Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405(6786), 590–593.

[21] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp. 289–300.

[22] Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., & Phillips, R. (2005). Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15, 116–124.

[23] 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. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell*, 24(6), 853–65.

[24] Blake, W. J., Kaern, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932), 633–7.

[25] Blank, L. M., Kuepfer, L., & Sauer, U. (2005). Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome biology*, 6(6), R49.

[26] Braun, E. (2015). The unforeseen challenge: from genotype-to-phenotype in cell populations. *Reports on Progress in Physics*, 78(3), 036602.

[27] Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., & Kondrashov, F. A. (2012). Epistasis as the primary factor in molecular evolution. *Nature*, 490(7421), 535–538.

[28] Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., & Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS biology*, 6(12), 2928–39.

[29] Camas, F. M., Blázquez, J., & Poyatos, J. F. (2006). Autogenous and non autogenous control of response in a genetic network. *Proc. Natl. Acad. Sci. USA*, 103, 12718–12723.

[30] Carlborg, Ö. & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, 5(8), 618–625.

[31] Carothers, J. M., Goler, J. A., Juminaga, D., & Keasling, J. D. (2011). Model-driven engineering of rna devices to quantitatively program gene expression. *Science*, 334(6063), 1716–1719.

[32] Carrozza, M. J., Utley, R. T., Workman, J. L., & Côté, J. (2003). The diverse functions of histone acetyltransferase complexes. *TRENDS in Genetics*, 19(6), 321–329.

[33] Cases, I., de Lorenzo, V., & Ouzounis, C. A. (2003). Transcription regulation and environmental adaptation in bacteria. *Trends in microbiology*, 11(6), 248–253.

[34] Cağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., & Süel, G. M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139, 512–522.

[35] Chandler, C. H., Chari, S., & Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in genetics*, 29(6), 358–366.

[36] Chari, S. & Dworkin, I. (2013). The Conditional Nature of Genetic Interactions: The Consequences of Wild-Type Backgrounds on Mutational Interactions in a Genome-Wide Modifier Screen. *PLoS Genetics*, 9(8), e1003661.

[37] Chen, C.-Z., Li, L., Lodish, H. F., & Bartel, D. P. (2004). Microrнas modulate hematopoietic lineage differentiation. *science*, 303(5654), 83–86.

[38] Choi, J. K. & Kim, Y.-J. (2008). Epigenetic regulation and the variability of gene expression. *Nature genetics*, 40(2), 141–7.

[39] Choi, J. K. & Kim, Y.-J. (2009). Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nature genetics*, 41(4), 498–503.

[40] Chubiz, L. M., Glekas, G. D., & Rao, C. V. (2012). Transcriptional cross talk within the mar-sox-rob regulon in Escherichia coli is limited to the rob and marRAB operons. *J. Bacteriol.*, 194, 4867–4875.

[41] Cohen, S. P., Hächler, H., & Levy, S. B. (1993a). Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *J. Bacteriol.*, 175, 1484–1492.

[42] Cohen, S. P., Levy, S. B., Foulds, J., & Rosner, J. L. (1993b). Salicylate induction of antibiotic resistance in Escherichia coli: activation of the mar operon and a mar-independent pathway. *J. Bacteriol.*, 175, 7856–7862.

[43] Cohen, S. P., McMurry, L. M., Hooper, D. C., Wolfson, J. S., & Levy, S. B. (1989). Cross-resistance to fluoroquinolones in multiple-antibiotic-resistant (mar) Escherichia coli selected by tetracycline or chloramphenicol: decreased drug accumulation associated with membrane changes in addition to OmpF reduction. *Antimicrob. Agents Chemother.*, 33, 1318–1325.

[44] Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C. G., & Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437, 699–706.

[45] Combarros, O., Cortina-Borja, M., Smith, A. D., & Lehmann, D. J. (2009). Epistasis in sporadic alzheimer's disease. *Neurobiology of aging*, 30(9), 1333–1349.

[46] Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R. P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.-Y., Liang, W., Marback, M., Paw, J., San Luis, B.-J., Shuteriqi, E., Tong, A. H. Y., van Dyk, N., Wallace, I. M., Whitney, J. a., Weirauch, M. T., Zhong, G., Zhu, H., Houry, W. a., Brudno, M., Ragibizadeh, S., Papp, B., Pál, C., Roth, F. P., Giaever, G., Nislow, C., Troyanskaya, O. G., Bussey, H., Bader, G. D., Gingras, A.-C., Morris, Q. D., Kim, P. M., Kaiser, C. a., Myers, C. L., Andrews, B. J., & Boone, C. (2010). The genetic landscape of a cell. *Science*, 327(5964), 425–31.

[47] Cox, N. J., Frigge, M., Nicolae, D. L., Concannon, P., Hanis, C. L., Bell, G. I., & Kong, A. (1999). Loci on chromosomes 2 (niddm1) and 15 interact to increase susceptibility to diabetes in mexican americans. *Nature genetics*, 21(2), 213–215.

[48] Crespi, B. J. (2010). The origins and evolution of genetic disease risk in modern humans. *Annals of the New York Academy of Sciences*, 1206(1), 80–109.

[49] Croft, L. J., Lercher, M. J., Gagen, M. J., & Mattick, J. S. (2004). Is prokaryotic complexity limited by accelerated growth in regulatory overhead? *Genome Biology*, 5(1), P2–P2.

[50] Daniel, J. A. & Grant, P. A. (2007). Multi-tasking on chromatin with the saga coactivator complexes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 618(1), 135–148.

[51] Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA*, 97, 6640–6645.

[52] De Crombrugghe, B., Perlman, R. L., Varmus, H., & Pastan, I. (1969). Regulation of inducible enzyme synthesis in escherichia coli by cyclic adenosine 3', 5'-monophosphate. *Journal of Biological Chemistry*, 244(21), 5828–5835.

[53] Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., Nislow, C., & Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, 169(4), 1915–25.

[54] Deutscher, D., Meilijson, I., Kupiec, M., & Ruppin, E. (2006). Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature genetics*, 38(9), 993–8.

[55] Dieci, G. & Sentenac, A. (2003). Detours and shortcuts to transcription reinitiation. *Trends in biochemical sciences*, 28(4), 202–209.

[56] Dixon, S. J., Fedyshyn, Y., Koh, J. L. Y., Prasad, T. S. K., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.-L., Kim, D.-U., Park, H.-O., Myers, C. L., Pandey, A., Durocher, D., Andrews, B. J., & Boone, C. (2008). Significant

conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43), 16653–8.

[57] Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Number 11. Columbia University Press.

[58] Dovey, O. M., Foster, C. T., & Cowley, S. M. (2010). Histone deacetylase 1 (hdac1), but not hdac2, controls embryonic stem cell differentiation. *Proceedings of the National Academy of Sciences*, 107(18), 8242–8247.

[59] Duarte, N. C., Herrgård, M. J., & Palsson, B. O. (2004). Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Research*, 14(7), 1298–1309.

[60] Durant, M. & Pugh, B. F. (2006). Genome-Wide Relationships between TAF1 and Histone Acetyltransferases in Saccharomyces cerevisiae. *Society*, 26(7), 2791–2802.

[61] Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., Ferguson, C., Haga, K., Sato, T., Liu, J. S., & Losick, R. (2004). The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. *PLoS Biol.*, 2, e328.

[62] Eldar, A. & Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, 467(7312), 167–173.

[63] Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G. M., Bontrop, R. E., & Pääbo, S. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science (New York, N.Y.)*, 296(5566), 340–3.

[64] Englesberg, E., Sheppard, D., Squires, C., & Meronk, F. (1969). An analysis of "revertants" of a deletion mutant in the c gene of the l-arabinose gene complex in escherichia coli b/r: Isolation of initiator constitutive mutants (i c). *Journal of molecular biology*, 43(2), 281–298.

[65] Félix, M.-a. & Wagner, a. (2008). Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, 100(2), 132–40.

[66] Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(2), 399–433.

[67] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon, Oxford.

[68] Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network. *Genome research*, 13(2), 244–253.

[69] Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D. C., & Thompson, C. R. (2004). Pleiotropy as a mechanism to stabilize cooperation. *Nature*, 431(7009), 693–696.

[70] Fralick, J. A. (1996). Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of Escherichia coli. *J. Bacteriol.*, 178, 5803–5805.

[71] Fraser, D. & Kaern, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular microbiology*, 71(6), 1333–40.

[72] Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J., & Eisen, M. B. (2004). Noise minimization in eukaryotic gene expression. *PLoS biology*, 2(6), e137.

[73] Frost, A., Elgort, M. G., Brandman, O., Ives, C., Collins, S. R., Miller-Vedam, L., Weibezahn, J., Hein, M. Y., Poser, I., Mann, M., Hyman, A. a., & Weissman, J. S. (2012). Functional repurposing revealed by comparing S. pombe and S. cerevisiae genetic interactions. *Cell*, 149(6), 1339–52.

[74] Fuhrer, T. & Sauer, U. (2009). Different biochemical mechanisms ensure network-wide balancing of reducing equivalents in microbial metabolism. *Journal of bacteriology*, 191(7), 2112–21.

[75] Furlong, L. I. (2013). Human diseases through the lens of network biology. *Trends in genetics*, 29(3), 150–9.

[76] Gagen, M. & Mattick, J. (2005). Accelerating, hyperaccelerating, and decelerating networks. *Physical Review E*, 72(1), 016123.

[77] Garcia-Bernardo, J. & Dunlop, M. J. (2013). Tunable stochastic pulsing in the Escherichia coli multiple antibiotic resistance network from interlinked positive and negative feedback loops. *PLoS Comput. Biol.*, 9, e1003229.

[78] Gasch, a. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., & Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell*, 11(12), 4241–57.

[79] George, A. M. & Levy, S. B. (1983). Gene in the major cotransduction gap of the Escherichia coli K-12 linkage map required for the expression of chromosomal resistance to tetracycline and other antibiotics. *J. Bacteriol.*, 155, 541–548.

[80] Gigerenzer, G. (2008). Why heuristics work. *Perspectives on psychological science*, 3(1), 20–29.

[81] Giovannoni, S. J., Thrash, J. C., & Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *The ISME journal*, 8(8), 1553–1565.

[82] Goldberger, R. F. (1974). Autogenous regulation of gene expression. *Science*, 183, 810–816.

[83] Goldbeter, A. & Koshland Jr, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. USA*, 78, 6840–6844.

[84] Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6), 1025–36.

[85] Gould, S. J. & Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, 205(1161), 581–598.

[86] Grant, P. A., Schieltz, D., Pray-Grant, M. G., Steger, D. J., Reese, J. C., Yates, J. R., & Workman, J. L. (1998). A subset of taf ii s are integral components of the saga complex

required for nucleosome acetylation and transcriptional stimulation. *Cell*, 94(1), 45–53.

[87] Greenspan, R. J. (2001). The flexible genome. *Nature Reviews Genetics*, 2(5), 383–387.

[88] Greenspan, R. J. (2009). Selection, gene interaction, and flexible gene networks. *Cold Spring Harbor symposia on quantitative biology*, 74, 131–8.

[89] Grewal, S. I. & Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *science*, 301(5634), 798–802.

[90] Griffith, K. L., Fitzpatrick, M. M., Keen, E. F., & Wolf, R. E. (2009). Two functions of the c-terminal domain of escherichia coli rob: mediating "sequestration–dispersal" as a novel off–on switch for regulating rob's activity as a transcription activator and preventing degradation of rob by lon protease. *Journal of molecular biology*, 388(3), 415–430.

[91] Griffith, K. L., Shah, I. M., & E. Wolf, R. (2004). Proteolytic degradation of Escherichia coli transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. *Mol. Microbiol.*, 51, 1801–1816.

[92] Guantes, R., Estrada, J., & Poyatos, J. F. (2010). Trade-offs and noise tolerance in signal detection by genetic circuits. *PloS One*, 5, e12314.

[93] Guantes, R. & Poyatos, J. F. (2006). Dynamical principles of two-component genetic oscillators. *PLoS Comput. Biol.*, 2, e30.

[94] Guénolé, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S., Krogan, N. J., Ideker, T., & van Attikum, H. (2013). Dissection of DNA damage responses using multiconditional genetic interaction maps. *Molecular cell*, 49(2), 346–58.

[95] Guet, C. C., Elowitz, M. B., Hsing, W., & Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science*, 296, 1466–1470.

[96] H, K. (1938). Enzymatische adaptation bei mikroorganismen. *Ergenbisse der Enzymforschung*, (7), 350–376.

[97] Hao, Z., Lou, H., Zhu, R., Zhu, J., Zhang, D., Zhao, B. S., Zeng, S., Chen, X., Chan, J., He, C., & Chen, P. R. (2014). The multiple antibiotic resistance regulator MarR is a copper sensor in Escherichia coli. *Nat. Chem. Biol.*, 10, 21–28.

[98] Harrison, R., Papp, B., Csaba, P., Oliver, S. G., & Delneri, D. (2007). Plasticity of genetic interactions in metabolic networks of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 104(7), 2307–2312.

[99] Hasty, J., Dolnik, M., Rottschäfer, V., & Collins, J. (2002). Synthetic gene network for entraining and amplifying cellular oscillations. *Phys. Rev. Lett.*, 88, 148101.

[100] Helmann, J. D. & Chamberlin, M. J. (1988). Structure and function of bacterial sigma factors. *Annual review of biochemistry*, 57(1), 839–872.

[101] Hibbs, M. A., Hess, D. C., Myers, C. L., Huttenhower, C., Li, K., & Troyanskaya, O. G. (2007). Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics*, 23(20), 2692–2699.

[102] Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., & Young, R. a. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5), 717–28.

[103] Huang, L. S. & Sternberg, P. W. (2006). Genetic dissection of developmental pathways.

[104] Huisinga, K. L. & Pugh, B. F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular cell*, 13(4), 573–85.

[105] Ideker, T. & Krogan, N. J. (2012). Differential network biology. *Molecular systems biology*, 8, 565.

[106] Ihmels, J., Collins, S. R., Schuldiner, M., Krogan, N. J., & Weissman, J. S. (2007). Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Molecular systems biology*, 3(1), 86.

[107] Ishihara, S., Fujimoto, K., & Shibata, T. (2005). Cross talking of network motifs in gene regulation that generates temporal pulses and spatial stripes. *Genes to Cells*, 10(11), 1025–1038.

[108] Ishizuka, H., Hanamural, A., Inada, T., & Aiba, H. (1994). Mechanism of the down-regulation of cAMP receptor protein by glucose in Escherichia col: role of autoregulation of the crp gene. *EMBO J.*, 13, 3077–3082.

[109] Jacob, F. & Monod, J. (1961a). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3(3), 318–356.

[110] Jacob, F. & Monod, J. (1961b). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, 3, 318–356.

[111] Ji, N., Middelkoop, T. C., Mentink, R. A., Betist, M. C., Tonegawa, S., Mooijman, D., Korswagen, H. C., & van Oudenaarden, A. (2013). Feedback control of gene expression variability in the Caenorhabditis elegans Wnt pathway. *Cell*, 155, 869–880.

[112] Johanssen, W. (1911). The genotype conception of heredity. *The American Naturalist*, 45(531), 129–159.

[113] Kadonaga, J. T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell*, 92(3), 307–313.

[114] Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236), 362–6.

[115] Kaplan, S., Bren, A., Zaslaver, A., Dekel, E., & Alon, U. (2008). Diverse two-dimensional input functions control bacterial sugar genes. *Mol. Cell*, 29, 786–792.

[116] Kim, S. C. & Choi, J. K. (2011). Controlling transcriptional programs for cellular adaptation by chromatin regulation. *Molecular bioSystems*, 7(5), 1713–9.

[117] Kramer, B. P. & Fussenegger, M. (2005). Hysteresis in a synthetic mammalian gene network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), 9517–9522.

[118] Kussell, E. & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science (New York, N.Y.)*, 309(5743), 2075–8.

[119] Landry, C. R., Lemos, B., Rifkin, S. a., Dickinson, W. J., & Hartl, D. L. (2007). Genetic properties influencing the evolvability of gene expression. *Science (New York, N.Y.)*, 317(5834), 118–21.

[120] Lee, M. V., Topper, S. E., Hubler, S. L., Hose, J., Wenger, C. D., Coon, J. J., & Gasch, A. P. (2011). A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology*, 7, 514.

[121] Lee, T. I., Causton, H. C., Holstege, F. C., Shen, W. C., Hannett, N., Jennings, E. G., Winston, F., Green, M. R., & Young, R. a. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature*, 405(6787), 701–4.

[122] Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., & Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics*, 39(10), 1235–44.

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

[124] Lehner, B. (2010). Conflict between Noise and Plasticity in Yeast. *PLoS Genetics*, 6(11), e1001185.

[125] Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends in genetics*, 27(8), 323–31.

[126] Lehner, B., Crombie, C., Tischler, J., Fortunato, A., & Fraser, A. G. (2006). Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nature genetics*, 38(8), 896–903.

[127] Lenstra, T. L., Benschop, J. J., Kim, T., Schulze, J. M., Brabers, N. a. C. H., Margaritis, T., van de Pasch, L. a. L., van Heesch, S. a. a. C., Brok, M. O., Groot Koerkamp, M. J. a., Ko, C. W., van Leenen, D., Sameith, K., van Hooff, S. R., Lijnzaad, P., Kemmeren, P., Hentrich, T., Kobor, M. S., Buratowski, S., & Holstege, F. C. P. (2011).

The specificity and topology of chromatin interaction pathways in yeast. *Molecular cell*, 42(4), 536–49.

[128] Levine, E., Zhang, Z., Kuhlman, T., & Hwa, T. (2007). Quantitative characteristics of gene regulation by small RNA. *PLoS biology*, 5(9), e229+.

[129] Levine, M. & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, 424(6945), 147–151.

[130] Levy, S. F. & Siegal, M. L. (2008). Network Hubs Buffer Environmental Variation in Saccharomyces cerevisiae. *PLoS Biology*, 6(11), e264.

[131] Levy, S. F., Ziv, N., & Siegal, M. L. (2012). Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS-Biology*, 10(5), 952.

[132] Lewontin, R. C. et al. (1974). *The genetic basis of evolutionary change*, volume 560. Columbia University Press New York.

[133] López-Maury, L., Marguerat, S., & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature reviews. Genetics*, 9(8), 583–93.

[134] Losick, R. & Desplan, C. (2008). Stochasticity and cell fate. *science*, 320(5872), 65–68.

[135] Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., & Hearst, J. E. (1995). Genes acrA and acrB encode a stress-induced efflux system of Escherichia coli. *Mol. Microbiol.*, 16, 45–55.

[136] MacKay, V. L., Li, X., Flory, M. R., Turcott, E., Law, G. L., Serikawa, K. A., Xu, X. L., Lee, H., Goodlett, D. R., Aebersold, R., Zhao, L. P., & Morris, D. R. (2004). Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Molecular & cellular proteomics : MCP*, 3(5), 478–89.

[137] Madar, D., Dekel, E., Bren, A., & Alon, U. (2011). Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli. *BMC Sys. Biol.*, 5, 111.

[138] Maeda, Y. T. & Sano, M. (2006). Regulatory dynamics of synthetic gene networks with positive feedback. *J. Mol. Biol.*, 359, 1107–1124.

[139] Mangan, S. & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100, 11980–11985.

[140] Mangan, S., Itzkovitz, S., Zaslaver, A., & Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *Journal of molecular biology*, 356(5), 1073–1081.

[141] Martin, R. G., Bartlett, E. S., Rosner, J. L., & Wall, M. E. (2008). Activation of the Escherichia coli marA/soxS/rob regulon in response to transcriptional activator concentration. *J. Mol. Biol.*, 380, 278–284.

[142] Martin, R. G., Jair, K.-W., Wolf, R., & Rosner, J. L. (1996a). Autoactivation of the marrab multiple antibiotic resistance operon by the mara transcriptional activator in escherichia coli. *Journal of bacteriology*, 178(8), 2216–2223.

[143] Martin, R. G., Jair, K. W., Wolf Jr, R. E., & Rosner, J. L. (1996b). Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in Escherichia coli. *J. Bacteriol.*, 178, 2216–2223.

[144] Martin, R. G. & Rosner, J. L. (2002). Genomics of the marA/soxS/rob regulon of Escherichia coli: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol. Microbiol.*, 44, 1611–1624.

[145] Martin, R. G. & Rosner, J. L. (2004). Transcriptional and translational regulation of the marRAB multiple antibiotic resistance operon in Escherichia coli. *Mol. Microbiol.*, 53, 183–191.

[146] Martin, R. G. & Rossner, J. L. (1997). Fis, an accessorial factor for transcriptional activation of the mar (multiple antibiotic resistance) promoter of Escherichia coli in the presence of the activator MarA, SoxS, or Rob. *J. Bacteriol.*, 179, 7410–7419.

[147] Mattick, J. S. (2004). Rna regulation: a new genetics? *Nature Reviews Genetics*, 5(4), 316–323.

[148] Meiklejohn, C. D. & Hartl, D. L. (2002). A single mode of canalization. *Trends in Ecology & Evolution*, 17(10), 468–473.

[149] Mendel, J. G. (1866). Experiments in plant hybridization. *Verhandlungen des naturforschenden Vereines in Brunn*, 4, 3–47.

[150] Miyashiro, T. & Goulian, M. (2007). Stimulus-dependent differential regulation in the escherichia coli phoq–phop system. *Proceedings of the National Academy of Sciences*, 104(41), 16305–16310.

[151] Monod, J. (1941). Recherches sur la croissance des cultures bactériennes. *Paris: Université de Paris*.

[152] Moore, J. H. (2003). The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Human heredity*, 56(1-3), 73–82.

[153] Munsky, B., Neuert, G., & van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. *Science*, 336, 183–187.

[154] Navarro-Alberto, J. A. & Manly, B. F. (2009). Null model analyses of presence–absence matrices need a definition of independence. *Population ecology*, 51(4), 505–512.

[155] Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., & Balázsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. USA*, 106, 5123–5128.

[156] Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., & Weissman, J. S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095), 840–6.

[157] Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K. M., Lee, K. J., Wong, A., Shales, M., Lovett, S., Winkler, M. E., Krogan, N. J., Typas, A., & Gross, C. A. (2011). Phenotypic landscape of a bacterial cell. *Cell*, 144, 143–156.

[158] Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3), 245–248.

[159] Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, 317(5844), 1544–1548.

[160] Österberg, S., Peso-Santos, T. d., & Shingler, V. (2011). Regulation of alternative sigma factor use. *Annual review of microbiology*, 65, 37–55.

[161] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature genetics*, 31(1), 69–73.

[162] Pál, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., & Hurst, L. D. (2006). Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 440(7084), 667–70.

[163] Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S., & Boeke, J. D. (2006). A dna integrity network in the yeast saccharomyces cerevisiae. *Cell*, 124(5), 1069–1081.

[164] Papp, B., Pál, C., & Hurst, L. D. (2004). Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, 429(6992), 661–4.

[165] Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11), 855–867.

[166] Plate, L. (1866). Genetics and evolution. *Festschrift zum sechzigsten Geburtstag, Richard Hertwigs. Fischer, Jena, Germany*, (pp. 536–610).

[167] Plumbridge, J. & Pellegrini, O. (2004). Expression of the chitobiose operon of Escherichia coli is regulated by three transcription factors: NagC, ChbR and CAP. *Mol. Microbiol.*, 52, 437–449.

[168] Povolotskaya, I. S. & Kondrashov, F. A. (2010). Sequence space and the ongoing expansion of the protein universe. *Nature*, 465(7300), 922–926.

[169] Poyatos, J. F. (2011). The balance of weak and strong interactions in genetic networks. *PloS one*, 6(2), e14598.

[170] Price, N. D., Reed, J. L., & Palsson, B. O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature reviews. Microbiology*, 2(11), 886–97.

[171] Ptashne, M. (2004). *A genetic switch: phage lambda revisited*, volume 3. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY:.

[172] Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, 386(6625), 569–577.

[173] Ptashne, M. & Gann, A. (2002). *Genes and Signals*. Cold Spring Harbor Laboratory Press.

[174] Ranea, J. A., Grant, A., Thornton, J. M., & Orengo, C. A. (2005). Microeconomic principles explain an optimal genome size in bacteria. *Trends in Genetics*, 21(1), 21–25.

[175] Raser, J. M. & O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science (New York, N.Y.)*, 304(5678), 1811–4.

[176] Raser, J. M. & O'Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science (New York, N.Y.)*, 309(5743), 2010–3.

[177] Reed, J. L., Famili, I., Thiele, I., & Palsson, B. O. (2006). Towards multidimensional genome annotation. *Nature reviews. Genetics*, 7(2), 130–41.

[178] Reid, J. L., Iyer, V. R., Brown, P. O., & Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Molecular cell*, 6(6), 1297–307.

[179] Rensing, C. & Grass, G. (2003). Escherichia coli mechanisms of copper homeostasis in a changing environment. *FEMS Microbiol. Rev.*, 27(27), 197–213.

[180] Rodrigo, G., Carrera, J., & Jaramillo, A. (2011). Computational design of synthetic regulatory networks from a genetic library to characterize the designability of dynamical behaviors. *Nucleic acids research*, 39(20), e138–e138.

[181] Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S. R., Qu, H., Shales, M., Park, H.-O., Hayles, J., Hoe, K.-L., Kim, D.-U., Ideker, T., Grewal, S. I., Weissman, J. S., & Krogan, N. J. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, 322(5900), 405–10.

[182] Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 323, 785–793.

[183] Rosner, J. L. (1985). Nonheritable resistance to chloramphenicol and other antibiotics induced by salicylates and other chemotactic repellents in Escherichia coli K-12. *Proc. Natl. Acad. Sci. USA*, 82(24), 8771–8774.

[184] Ruiz, C. & Levy, S. B. (2010). Many chromosomal genes modulate MarA-mediated multidrug resistance in Escherichia coli. *Antimicrob. Agents Chemother.*, 54, 2125–2134.

[185] Rutherford, S. L. & Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, 396(6709), 336–342.

[186] Ryan, C. J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S. R., Kliegman, J. I., Jiang, L., Kuo, D., Tosti, E., Kim, H.-S., Edelmann, W., Keogh, M.-C., Greene, D., Tang, C., Cunningham, P., Shokat, K. M., Cagney, G., Svensson, J. P., Guthrie, C., Espenshade, P. J., Ideker, T., & Krogan, N. J. (2012). Hierarchical modularity and the evolution of genetic interactomes across species. *Molecular cell*, 46(5), 691–704.

[187] Salgado, H. & et al (2013). RegulonDB (version 8.0): Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.*, 41, D203–213.

[188] Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology*, 27(10), 946–950.

[189] Sanchez, A., Choubey, S., & Kondev, J. (2013). Regulation of noise in gene expression. *Annual review of biophysics*, 42, 469–491.

[190] Savageau, M. (1974). Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 252, 546–549.

[191] Schneiders, T. & Levy, S. B. (2006). Mara-mediated transcriptional repression of the rob promoter. *Journal of Biological Chemistry*, 281(15), 10049–10055.

[192] Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J. F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, 123(3), 507–519.

[193] Segrè, D., Deluna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature genetics*, 37(1), 77–83.

[194] Seoane, A. S. & Levy, S. B. (1995). Characterization of marR, the repressor of the multiple antibiotic resistance (mar) operon in escherichia coli. *J. Bacteriol.*, 177, 3414–3419.

[195] Setty, Y., Mayo, A. E., Surette, M. G., & Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proceedings of the National Academy of Sciences*, 100(13), 7702–7707.

[196] Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1), 64–8.

[197] Sheppard, D. & Englesberg, E. (1966). Positive control in the l-arabinose gene-enzyme complex of escherichia coli b/r as exhibited with stable merodiploids. In *Cold Spring Harbor symposia on quantitative biology*, volume 31 (pp. 345–347).: Cold Spring Harbor Laboratory Press.

[198] Siryaporn, A., Perchuk, B. S., Laub, M. T., & Goulian, M. (2010). Evolving a robust signal transduction pathway from weak cross-talk. *Mol Syst Biol.*, 6, 452.

[199] Smits, W. K., Kuipers, O. P., & Veening, J.-W. (2006). Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.*, 4(4), 259–271.

[200] Snitkin, E. S., Dudley, A. M., Janse, D. M., Wong, K., Church, G. M., & Segrè, D. (2008). Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome biology*, 9(9), R140.

[201] Snitkin, E. S. & Segrè, D. (2011). Epistatic Interaction Maps Relative to Multiple Metabolic Phenotypes. *PLoS genetics*, 7(2), e1001294.

[202] Soyer, O. S. (2012). *Evolutionary systems biology*, volume 751. Springer Science & Business Media.

[203] St Onge, R. P., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R. W., Nislow, C., Roth, F. P., & Giaever, G. (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature genetics*, 39(2), 199–206.

[204] Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, 186(3), 767–773.

[205] Steinfeld, I., Shamir, R., & Kupiec, M. (2007). A genome-wide analysis in Saccharomyces cerevisiae demonstrates the influence of chromatin modifiers on transcription. *Nature Genetics*, 39(3), 303–309.

[206] Sternberg, P. W. & Horvitz, H. R. (1989). The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in c. elegans. *Cell*, 58(4), 679–693.

[207] Storz, G. & Hengge, R. (2011). *Bacterial Stress Responses 2nd edition*. American Society for Microbiology Press.

[208] Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., & Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature*, 456, 516–519.

[209] Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20), 12795–12800.

[210] Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M. J., Jelasity, M., Myers, C. L., Andrews, B. J., Boone, C., Oliver, S. G., Pál, C., & Papp, B. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature genetics*, 43(7), 656–62.

[211] Szekely, P., Sheftel, H., Mayo, A., & Alon, U. (2013). Evolutionary tradeoffs between economy and effectiveness in biological homeostasis systems. *PLoS Comput. Biol.*, 9, e1003163.

[212] Szöllősi, G. J. & Derényi, I. (2009). Congruent evolution of genetic and environmental robustness in micro-rna. *Molecular biology and evolution*, 26(4), 867–874.

[213] Thieffry, D., Huerta, A. M., Pérez-Rueda, E., & Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in escherichia coli. *Bioessays*, 20(5), 433–440.

[214] Thomas, R. & D'Ari, R. (2000). *Biological Feedback*. CRC Press.

[215] Tirosh, I. & Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. *Genome research*, 18(7), 1084–91.

[216] Tirosh, I., Berman, J., & Barkai, N. (2007). The pattern and evolution of yeast promoter bendability. *Trends in genetics : TIG*, 23(7), 318–21.

[217] Tirosh, I., Reikhav, S., Sigal, N., Assia, Y., & Barkai, N. (2010). Chromatin regulators as capacitors of interspecies variations in gene expression. *Molecular Systems Biology*, 6(435), 1–9.

[218] Tirosh, I., Weinberger, A., Carmi, M., & Barkai, N. (2006). SUPL-A genetic signature of interspecies variations in gene expression. *Nature genetics*, 38(7), 830–4.

[219] To, T.-L. & Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327, 1142–1145.

[220] Tong, A. H. Y., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294(5550), 2364–2368.

[221] Tong, A. H. Y., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. *Science*, 303(5659), 808–813.

[222] van Nimwegen, E. (2006). Scaling laws in the functional content of genomes. In *Power Laws, Scale-Free Networks and Genome Biology* (pp. 236–253). Springer.

[223] Venters, B. J. & Pugh, B. F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research*, 19(3), 360–71.

[224] Vinué, L., McMurry, L. M., & Levy, S. B. (2013). The 216-bp marB gene of the marRAB operon in Escherichia coli encodes a periplasmic protein which reduces the transcription rate of marA. *FEMS Microbiol. Lett.*, 345, 49–55.

[225] Viveiros, M., Dupont, M., Rodrigues, L., Couto, I., Davin-Regli, A., Martins, M., Pagès, J.-M., & Amaral, L. (2007). Antibiotic stress, genetic response and altered permeability of e. coli. *PloS one*, 2(4), e365.

[226] Wade, M. J., Winther, R. G., Agrawal, A. F., & Goodnight, C. J. (2001). Alternative definitions of epistasis: dependence and interaction. *Trends in Ecology & Evolution*, 16(9), 498–504.

[227] Wagner, A. (2005). *Robustness and Evolvability in Living Systems*. Princeton University Press.

[228] Wagner, G. P. & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, 12(3), 204–213.

[229] Wall, M. E., Markowitz, D. A., Rosner, J. L., & Martin, R. G. (2009). Model of transcriptional activation by mara in escherichia coli. *PLoS Comput Biol*, 5(12), e1000614.

[230] Wang, G.-Z., Lercher, M. J., & Hurst, L. D. (2011). Transcriptional Coupling of Neighboring Genes and Gene Expression Noise: Evidence that Gene Orientation and Noncoding Transcripts Are Modulators of Noise. *Genome Biol Evol*, 3, 320–31.

[231] Wang, H., Kakaradov, B., Collins, S. R., Karotki, L., Fiedler, D., Shales, M., Shokat, K. M., Walther, T. C., Krogan, N. J., & Koller, D. (2009). A complex-based reconstruction of the Saccharomyces cerevisiae interactome. *Molecular & cellular proteomics : MCP*, 8(6), 1361–81.

[232] Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences*, 24(11), 437–40.

[233] Weiss, D. S., Chen, J. C., Ghigo, J.-M., Boyd, D., & Beckwith, J. (1999). Localization of ftsi (pbp3) to the septal ring requires its membrane anchor, the z ring, ftsa, ftsq, and ftsl. *Journal of bacteriology*, 181(2), 508–520.

[234] Whitacre, J. & Bender, A. (2010). Degeneracy: a design principle for achieving robustness and evolvability. *Journal of theoretical biology*, 263(1), 143–53.

[235] Williams, G. C. (1957). Pleiotropy, natural selection, and the evolution of senescence. *Evolution*, 11(4), 398–411.

[236] Woltereck, R. (1909). Weitere experimentelle untersuchungen über artveränderung, speziel über das wesen quantitativer artuntershiede bei daphniden. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, (pp. 110–172).

[237] Woo, Y. H. & Li, W.-H. (2011). Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), 3306–11.

[238] Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 1, 356–366.

[239] Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., & Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232), 1033–7.

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

[241] You, L. & Yin, J. (2002). Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage T7. *Genetics*, 160(4), 1273–81.

[242] Zenklusen, D., Larson, D. R., & Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, 15(12), 1263–71.

[243] Zhao, R., Davey, M., Hsu, Y.-C., Kaplanek, P., Tong, A., Parsons, A. B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., et al. (2005). Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*, 120(5), 715–727.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The template to format this Thesis was downloaded from github.com/asm-products/Dissertate