IOP Publishing Reports on Progress in Physics
Rep. Prog. Phys. **78** (2015) 036602 (51pp) doi:10.1088/0034-4885/78/3/036602

Review Article

The unforeseen challenge: from genotype-to-phenotype in cell populations

Erez Braun

Department of Physics and Network Biology Research Laboratories, Technion, Haifa 32000, Israel

E-mail: Erez@physics.technion.ac.il

Received 27 June 2014, revised 4 December 2014
Accepted for publication 18 December 2014
Published 26 February 2015

Abstract

Biological cells present a paradox, in that they show simultaneous stability and flexibility, allowing them to adapt to new environments and to evolve over time. The emergence of stable cell states depends on genotype-to-phenotype associations, which essentially reflect the organization of gene regulatory modes. The view taken here is that cell-state organization is a dynamical process in which the molecular disorder manifests itself in a macroscopic order. The genome does not determine the ordered cell state; rather, it participates in this process by providing a set of constraints on the spectrum of regulatory modes, analogous to boundary conditions in physical dynamical systems. We have developed an experimental framework, in which cell populations are exposed to unforeseen challenges; novel perturbations they had not encountered before along their evolutionary history. This approach allows an unbiased view of cell dynamics, uncovering the potential of cells to evolve and develop adapted stable states. In the last decade, our experiments have revealed a coherent set of observations within this framework, painting a picture of the living cell that in many ways is not aligned with the conventional one. Of particular importance here, is our finding that adaptation of cell-state organization is essentially an efficient exploratory dynamical process rather than one founded on random mutations. Based on our framework, a set of concepts underlying cell-state organization—exploration evolving by global, non-specific, dynamics of gene activity—is presented here. These concepts have significant consequences for our understanding of the emergence and stabilization of a cell phenotype in diverse biological contexts. Their implications are discussed for three major areas of biological inquiry: evolution, cell differentiation and cancer. There is currently no unified theoretical framework encompassing the emergence of order, a stable state, in the living cell. Hopefully, the integrated picture described here will provide a modest contribution towards a physics theory of the cell.

Keywords: cell-state organization, adaptation, population dynamics, evolution, cell differentiation, cancer

(Some figures may appear in colour only in the online journal)

Contents

What was life, really? It was the existence of what… It was not matter, it was not spirit. It was something in between the two, a phenomenon borne by matter, like the rainbow above a waterfall, like a flame.

Thomas Mann, *The Magic Mountain,* 1924[1]

1. Cell-state organization: a phenomenological approach

1.1. The nature of living cells

Living cells, under specific conditions, exhibit well defined traits and long-term stability, even in fluctuating environments. However, these cells also present a broad spectrum of characteristics with flexibility that allows them to adapt to novel and severe challenges. Indeed, the development of a multicellular organism from the single-cell zygote, crucially depends on the process of cell differentiation; the stabilization within a lineage of specific cell states (types)—neuron, muscle, skin etc, at precise times and locations, forming the body-plan and functional tissues. This developmental process is highly reproducible and robust; so much so, that it is almost natural to regard development as an execution of a program, much like a computer program. However, the simultaneous existence of these seemingly contradicting capabilities, robustness and flexibility, is an essential property that enables evolution [1–3]. The roots of the organism's ability to evolve—the hallmark of the living world—to large extent may be traced to the level of the cell and manifested in cell plasticity, its ability to react to internal or external cues and constraints. Thus, the organization of cell states in development and the emergence of novel phenotypes in evolution are tightly connected complementary characteristics of the living cell and their coexistence is one of the greatest mysteries in biology.

The issue of an emergent order—stable cell states—in essence is that of the *genotype-to-phenotype* associations. The genotype of a cell is its genetic makeup while the phenotype encompasses its traits, such as morphology and function. Genotype and phenotype represent two separate cellular entities; while the former is the structure of the genome—the DNA sequence, the latter is the determination of the form, growth and interactions with the external world of the cell. The phenotype is a central concept in the description of a biological system. In principle, it can be any observable property of the living organism. However, in the context of the present article, we reserve this term to the composite of observables related to the growth, morphology, metabolism and functionality of the cell. The establishment of a phenotype, given a certain genotype, depends on the protein makeup of the cell. The set of expressed proteins, a subset of the entire genome potential, and their concentrations, are determined by regulatory systems at many levels. Thus, the emerging phenotype depends on the spectrum of *regulatory modes*—temporal profiles of expressed genes. However, a snapshot of the molecular content of a cell and the structure of its underlying interactions do not capture the spectrum of regulatory profiles (the intermolecular correlations and the stability of temporal modes) that define the *relevant* observables that determine the phenotype. As shown below, the protein content of each isolated gene is by itself not such a relevant observable. Therefore, inquiring into the genotype-to-phenotype associations requires a shift in focus from *structure to dynamics*, from the molecular *stuff* of the cell to its temporal *organization*.

The prevailing approach of modern molecular biology for studying the living cell relies on several tenets. First, a biological cell is the product of an evolutionary process; the neo-Darwinian framework regards existing genotypes as the result of the accumulation of random mutations in DNA sequences, shaped by selection processes. Second, the genotype dictates the phenotype. The Central Dogma of molecular biology [4] assumes a one-way mapping from DNA to protein through

[1] Translated from German by J E Woods 1995 (New York: Vintage International), 271

the processes of transcription (mRNA production from the DNA sequence) and translation (production of proteins from mRNA) without the possibility for proteins to affect the DNA sequence itself [5]. These molecular processes are regulated at different levels by protein–DNA and protein–protein interactions as well as by interactions of other molecules. The prevailing reductionist approach to molecular biology seeks the molecular cause underlying any cellular process, preferably due to a single or a few genes. The genotype-to-phenotype mapping is largely assumed to be deterministic in nature, accompanied by 'noise' by environmental influences and intracellular stochastic processes due to the small volume of the cell and the small number of molecules involved. The Darwinian evolution approach motivates the assumption of optimization; a sweeping view regarding the biological processes and cell functionality, including every molecular interaction, as optimized by strong selection towards a specific function.

There is currently ample evidence calling for an expansion of the prevailing molecular framework. Most importantly, this approach dictates the way biological cells are studied; seeking the molecular causation of a response following a specific and well-designed perturbation (either genomic or environmental), assumed to be weak enough to allow isolation of a specific molecular module from the rest of the cellular processes. This approach is highly successful in isolating molecular elements involved in specific responses and enabled molecular biology to become the leading avenue in almost all areas of biological inquiry and medical applications. However, this methodology does not allow the search for a wider understanding of the cell as a complex dynamical system. As shown in this article, the seemingly simplistic molecular picture is only part of the story. Applying a different methodology, studying the long-term dynamic response to a strong perturbation and diverting the cell far away from its relaxed state, leads to a very different picture of the genotype-to-phenotype associations and the processes underlying the stabilization of an adapted cell state. This methodology allows separation of the phenotype dynamics from the genotype structure. This approach to cell biology is in fact complementary to the molecular approach. While the latter allows the efficient construction of the catalogue of molecular processes, the former is essential in order to understand the underlying physical principles. Since the methodology presented here is unconventional, let us first expand on its underlying principles and then summarize to what extent it leads to deviations from the prevailing picture.

Studying the principles underlying *cell-state organization,* calls for an investigation of cells in their natural context. The lineages emerging from the zygote, as well as proliferating unicellular organisms, like bacteria or yeast, should be regarded as heterogeneous populations that can develop multiple coexisting phenotypes. In both types of cells, a developing embryo or unicellular asexual organisms, we are interested in the *potential* of a founder 'mother-cell' to produce a spectrum of phenotypes under constraining external and internal conditions. The living cell is a complex dynamical system, the underlying molecular interactions spanning a huge combinatorial space of possible temporal modes. The emphasis on studying the potential of the system, rather than merely its *end-state realizations*, enables us to expose the nature of the biological system that can support such seemingly contradictory behavior; the simultaneous capabilities of stability and evolvability [6]. However, this shift in attention towards the potential of the cell similarly demands a shift in emphasis, from *structure to dynamics*. Moreover, a biophysical understanding of proliferating cells requires us to link intracellular dynamics and gene regulation, via the process of cell division, to the level of the population. This is particularly essential since in many cases, the intracellular responses extend over time-scales longer than a cell generation, which is not well separated from the time-scales of protein production, degradation or metabolic processes. These in turn depend on the history of the population mainly through inheritance of molecules (e.g. proteins) and structures (e.g. DNA conformations or cellular organelles) for generations [7]. This article therefore is focused on the dynamics of cells in the context of a population, emphasizing the interrelations between these two levels of biological organization, the cell-and the population. Note that a population of cells is not a statistical ensemble of independent individuals; transgenerational inheritance and long-term dynamic modes imply that cells within a population are correlated. This important distinction between populations of living cells and physical ensembles requires the development of experimental methodologies to study cell-population dynamics, emphasizing the population aspects of living cells.

Central to our understanding of biological phenomena and diversity is the complex challenge of uncovering the potential of cells to evolve. Arguably, this avenue of research also sets a new frontier in the field of complex systems. Living cells are history-determined objects, whose precise evolutionary history is not known, so uncovering their intrinsic potential requires us to discriminate between necessity and contingency, between inevitable and accidentally instilled intracellular processes.[2] This is a non-trivial experimental challenge and often a source of confusion; the erroneous assignment of crucial functional roles to elements that have emerged in the system by mere historical accidents.[3] Thus, there is a need to develop a new experimental paradigm exposing the potential of cells and universal principles of organization. This is highly non-intuitive since the common motivation in biological research is usually the opposite; exposing specific mechanisms underlying a given cell state.

We took an experimental approach to bypass this obstacle of the cell's unknown history and tried to penetrate the heart of the ability of cells to evolve by measuring the dynamics of yeast cells facing an *unforeseen challenge*—a novel perturbation they had not encountered before in their evolutionary

[2] This notion that the living world presents both inevitable and accidentally instilled processes was also discussed by Gould and Lewontin (1979 *Proc. R. Soc. Lond.* B **205** 581), Jacob [94] and Koonin (2011 *The Logic of Chance* (Upper Saddle River: FT Press)).

[3] This feature of a biological cell is well captured by Rube Goldberg caricatures, where a functional machine is made out of what looks like a random pile of elements, each of which is nevertheless absolutely necessary for its functionality (see [129]). This is as far as it can get from optimal design of machines in the engineering world—a most commonly used and somewhat misleading metaphor in biology.

history. This experimental framework allows us to acquire an *unbiased* view of the cell dynamics. We let the experiments lead us, instead of testing hypotheses that reflect our prejudices. Evidence accumulated from our experiments revealed a coherent set of observations within this framework, painting a picture of the living cell that in many ways is not aligned with the conventional one. It is worth summarizing, even at this early stage, the main deviations from the prevailing picture. (i) Inherited adaptation of cells responding to a perturbation could result from processes other than mutations in DNA sequences and is not necessarily the result of selection. Thus, cellular response is consequently not an optimization process. (ii) The genotype does not directly dictate the phenotype and it is not possible to reduce the adaptation process to a simple set of molecular causes determining the stabilization of a cell state. (iii) There is a strong cross-talk between levels of organization, in particular between the intracellular and population processes mainly due to long-term correlations for generations. (iv) The environment plays an outstanding role by participating in the cellular dynamics rather than being a passive selection filter. (v) Population dynamics support the coexistence of a wide spectrum of metastable phenotypes, in contrast to the Darwinian picture in which an optimized phenotype takes over.

The consistency of our results has been verified in multiple ways, by different types of experiments and in laboratories other than ours. The somewhat surprising picture emerging from our endeavor can be easily dismissed as a mere case-study; *it is not!* The lessons learned from our experiments, open a wide vista on some fundamental issues in cell biology. Starting from experiments that aimed to study adaptation of cells to an unforeseen challenge, we have discovered that some basic concepts of cell biology need to be revisited. Of particular importance here is our finding that adapted cell-state organization is essentially an *exploratory dynamical process* [6, 8]. This observation has significant consequences for our understanding of the emergence and stabilization of a cell phenotype in diverse biological contexts.

This article is not a typical review summarizing the state of the art of a field. The first part presents an experimental framework, developed and elaborated by us over the last decade, focusing on adaptation of yeast populations to an unforeseen challenge. The discussion avoids technical details to make it accessible to non-experts. These technical details and other data omitted here for the sake of coherency and brevity can be found in the listed references. The lessons learned from the yeast experiments, motivate revisiting and re-interpreting published results from different branches of biology. The discussion in the second part of this article reflects on cell-state organization in three major areas of biological inquiry: *evolution, cell differentiation and cancer*. It does not mean in any way to be an exhaustive review of these broad fields, or even a representative overview of them. Rather, the discussion is focused on a limited set of examples that highlight essential principles and stimulates further inquiries. Finally, we summarize by noting that there is currently no unified theoretical framework encompassing the emergence of order, a stable phenotypic state, in the living cell. There are many beautiful pieces that until now have not provided the critical seeds for a collective effort towards such understanding. Hopefully the integrated picture painted in this article will provide a modest contribution towards a physics theory of the cell.

1.2. An experimental framework: genome rewiring and cell adaptation to unforeseen challenges

> *Organisms can do all types of things: they do fantastic things... Trying to make everything fit into set dogma won't work… So if the material tells you, 'It might be this', allow that. Don't turn it aside and call it an exception, an aberration, a contaminant…. That's what's happened all the way along the line with so many good clues.*
>
> Barbara McClintock[273]

We present now our experimental framework based on cell adaptation to an *unforeseen challenge*. It is important first to realize the significance of this concept by distinguishing two types of cellular responses to a perturbation. Cells react to a common perturbation by fast operation of existing 'hard-wired' functional modules. By contrast, they do not have such a 'pre-designed', specific response to a novel challenge—a perturbation leading to significant deviations from the current cell state and therefore requiring a very different type of operation. For example, a change in food ingredients (e.g. switching between two types of sugars) or an environmental stress (e.g. a reduction in oxygen level) familiar to the cell, usually result in a fast stereotypic response of dedicated genetic and protein networks having specifically determined functionalities. On the other hand, exposure to an unforeseen challenge in which no *a priori* response is 'instilled' in the cell, requires genuine adaptation. The two types of responses differ not only in their speed of reaction. As we shall see, adaptation to an unforeseen challenge is based on global reorganization of *regulatory modes*, determining the protein makeup of the cell.

Remodeling gene regulation has been recognized as playing an important role in evolution. The biodiversity observed in nature shows that the emergence of novelty is at the heart of the evolutionary process. But, how do such novelties emerge, and how are they utilized by organisms during evolution? These issues are still open and serve as subjects of inquiry at the forefront of evolutionary biology. Genomes are roughly composed of two functionally distinct parts; coding regions, dictating the amino-acid sequences of proteins (with the possibility of alternative splicing, combining exons—separated coding portions of the same gene—into different coding sequences) and regulatory regions. The latter are part of the genome 'dark-matter', classified in times misleadingly as 'junk' DNA, but lately revived as vastly containing functionally important regions [9, 10]. Many years ago, King and Wilson [11] suggested that developmental evolution involves changes in gene regulation rather than merely mutations in coding regions. Since their work, the significance of regulatory evolution has gained further support by numerous studies, relying on detailed comparative genomics enabled by the available genomic sequences across organisms and species.

Thus, the common understanding today is that the evolution of regulatory systems plays a crucially important role in the evolution of developmental systems, in particular in the development of body-plans [12–14]. Evolution of regulatory systems allows a great deal of flexibility in the generation of novel life forms by modifying the potential spectrum of regulatory modes, i.e. modifying the spectrum of expressed proteins, compared to mere changes in coding regions which are more limited [3]. This is not to say that changes in coding regions are of no significance. New enzymatic functions, for example, usually require changes in protein structures or the emergence of new proteins. The consequence of these two complementary facets of evolution, mutations in coding and non-coding regions with potential regulatory functions is far reaching. However, there is a crucial difference in the significance of these two parts. While a change in a coding sequence (e.g. a point mutation) could, at least in principle, be directly tested by studying the functionality of an isolated modified protein, this is not the case for changes in a non-coding regulatory region. Understanding the implications of a mutation in a regulatory, non-coding sequence, requires uncovering its effect on the modes of regulation which depend on the context, the history of the system and on the activity and interaction of different proteins. Surely, changes in protein structure can be highly complex and their functional significance can also be context dependent. In particular, changes in regulatory proteins, in contrast to structural ones, contribute to the regulatory modes and constraints. Nevertheless, the impact of a mutation in a coding region is much more apparent and amenable to laboratory testing and interpretation. Shifting attention from protein sequences to regulatory modes is again essentially a shift from *structure* to *dynamics*. We are familiar with a single genetic code enabling us to translate a change in nucleic-acids to a change in amino-acid. No similar codes are available for sequence changes in regulatory regions that do not code for proteins. In essence, this is exactly where mere sequence information does not reveal the organization of the cell and the organism. Given the genome sequence one could not 'compute' the organism, as sometimes misleadingly declared [15]. Clearly, even the organism itself does not compute its own development, which is a dynamical process and not a mere algorithmic translation from genotype-to-phenotype via a code [16].

Comparative genomics and advances in molecular biology teach us that there are multiple ways to affect the spectrum of regulatory modes of a cell which can be flexible and diverse. For example, modifications in the binding site of a transcription-factor could lead to localization of different sets of proteins at a given locus on the DNA. Alternatively, the insertion of a new regulatory sequence or structural changes in enhancer regions may lead to new forms of repression/activation of a gene. In the case that regulatory DNA sequences are modified, the proteins themselves might remain intact. Phenotypes then evolve by creating *new functional contexts* for existing proteins. A change in regulation modes is possible via genome 'rewiring' events, in which an existing gene becomes linked to a foreign (i.e. previously functionally unrelated) regulatory system. For example, mutations in *cis*-regulatory elements (promoter sequences) have enabled an existing protein to be used in new developmental processes [12, 17, 18]. Such events occur naturally in evolution and are well documented by comparative genomic studies, identifying them as significant drivers of the evolution of gene regulation in developmental systems [3, 13, 14] (see also discussions in [19–21]). Interestingly, comparison of enhancer regions, within and across species, shows the existence of different combinations of similar sets of sequences, identified as binding sites for transcription factors (figure 1) [3, 22]. It seems that in these cases, evolution proceeded by shuffling similar binding sequences across genomes, linking different genes to a combinatorial set of transcription factors, thus forming novel developmental patterns. From the evolutionary view point, this can explain (at least partially) the emergence of the complexity of pathways and networks in genomes [3]. However, this impressive set of comparative data is insufficient to explain how such novelties actually emerged and became established. Very little information exists on the dynamics of evolutionary processes, the role of phenotypic plasticity, and the intermediate states of evolving forms [1, 2]. In particular, the kind of cellular processes supporting such evolvability, remains elusive. Surprisingly, although recognized as key factors in evolution, the effects of imposed changes and constraints on regulatory modes have rarely been studied in laboratory experiments.

Genome rewiring might also emerge naturally in other contexts. An interesting example of massive genome-rewiring was demonstrated recently in the sequencing of the commonly used human *HeLa* cell-line [23]. Derived from cancerous cells, this first human cell-line has an interesting history of its own [24]. What the recent sequencing revealed, however, is a complete ***havoc*** in the genomes extracted from these cells. In particular, it was found that: '…countless regions of the chromosomes in each cell were arranged in the wrong order and had extra or fewer copies of genes' [23]. The extensive genomic rearrangements are indicative of catastrophic chromosome shattering, known as *chromothripsis*; a recently discovered phenomenon in which regions of the genome are shattered and then stitched together in a single devastating event [25]. Such an event has been associated with 2–3% of cancers, and whether it leads the cancer process or merely results from it is not known. What this finding implies is that even genomes that are chaotically chopped and shuffled, with genes 'cut and pasted' randomly, on a large scale, can still support viable phenotypes in metabolism, reproduction, functionality and morphology. In the case of the *HeLa* cells, notwithstanding the catastrophic rearrangement of the genome, these cells are viable and have the morphology of human cells. In fact, being cancerous cells, they are highly adaptive to their environment in the context of their original tissue. RNA analysis in these cells revealed that gene expression is strongly affected by the underlying genomic 'chaos'. The gene expression profiles dramatically differ from those measured in normal human tissues. The spectrum of regulatory modes in these cells, which is completely different from those in normal cells, is a remarkable demonstration for the breadth of the potential genotype-to-phenotype associations leading to viable cell states. It brings up an important issue:

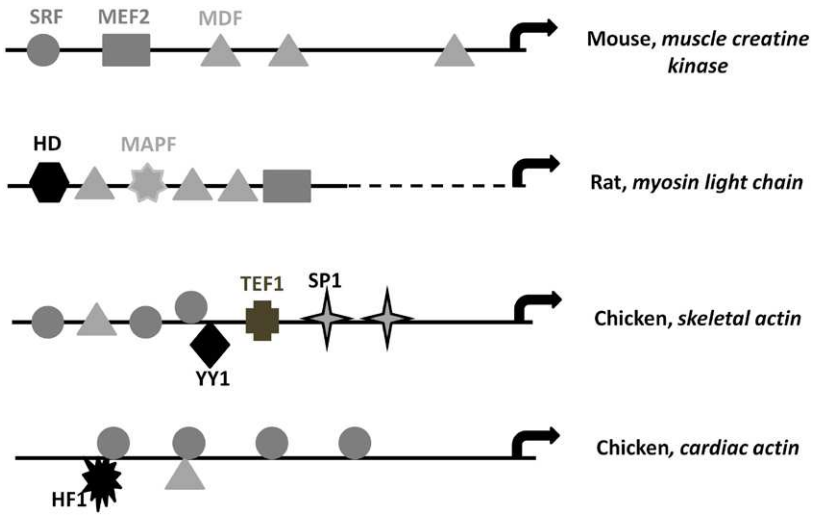

Figure 1. Genome rewiring in evolution. A group of genes under the regulation of a joint set of transcriptional regulators is termed a 'gene battery' (from [3]). Each line is the regulatory region of a different gene from such a gene battery participating in the development of striated muscles in various organisms. Each object represents a DNA sequence that can serve as a specific binding site for a regulatory protein. The figure, a rough partial sketch of a few examples (see [22] for a complete description and more examples), shows that the same binding site has been 'shuffled' or mutated in evolution across genes and species. None of the sets of binding sites completely repeats for the different genes, but they all share some binding sites. The labels of the binding sites represent DNA-binding domains: [3] MDF-Myogenic determination factor; MAPF-Muscle actin promoter factor-1; MEF2-Myocyte-specific enhancer binding factor-2; SRF-Serum response factor; HD-Unidentified HOX protein; TEF1-Transcription enhancer factor-1; YY1–YY1 factor; SP1–SP1 factor; HF1–Cardiac myocyte factor HF-1.

the combinatorial nature of regulatory constraints spans a high dimensional space of possible gene-expression profiles. The question then arises; how do such 'rewired' cells stabilize a viable phenotype under arbitrary conditions? Inquiring deeply into this question is at the heart of our experimental program and is the focus of this article.

Modifications of regulatory circuits may lead to *frustration*—incompatibilities between the spectrum of possible protein expression profiles and the metabolic and functional demands of the cell. Therefore, the potential of organisms to evolve depends crucially on the ability of cells to overcome these incompatibilities [6, 26]. The neo-Darwinian evolutionary framework attributes adaptation to the selection of advantageous phenotypes that exist in the population due to accumulation of genetic mutations that are rare, random, and occur independently of the selection process [27–29]. Mutations usually provide specific solutions to emerging challenges, but the multitude of possible unforeseen challenges raises the question of whether a more general mechanism, such as cellular plasticity relying on *exploration–exploitation* cellular processes, can provide an alternative strategy for evolution. Our experimental framework described next, aims exactly to inquire deeply into this possibility by utilizing a synthetic genome-rewiring event.

1.3. Genome rewiring of yeast cells

To mimic an event of *genome-rewiring* in a laboratory setup, we synthetically modified the genome of a strain of the unicellular budding yeast *Saccharomyces cerevisiae*. This *eukaryotic* microorganism is commonly used in laboratory experiments to study a wide range of biological phenomena and has played an important role in numerous discoveries. The gene *HIS3*, an essential enzyme from the biosynthesis pathway of the amino-acid histidine, was detached from its natural regulatory system and was placed exclusively under a promoter of the GAL system that is responsible for galactose (type of sugar) utilization (figure 2) [30]. Modern genetic engineering techniques allow this 'rewiring' by deleting the native *HIS3* coding sequence from the genome and 're-writing' it downstream of a copy of a GAL promoter either on a plasmid or integrated into the yeast genome (see [30] for technical details). The main results presented below are insensitive to the precise setup. The arbitrary regulatory linkage created here between these two evolutionarily conserved and highly specific modules (histidine biosynthesis and GAL) was stressful, challenging and created incompatibilities in gene expression for the 'rewired' cells [30]. Histidine is an essential amino-acid whose production is interconnected to that of the other amino-acids; all regulated in the cell to provide the essential substrates for its proper functionality. Therefore, if histidine is not supplied in the medium, the functionality of the *HIS3* gene is absolutely essential since it does not have a redundant backup. The GAL system also has its own elaborated regulatory system (a single specific transcription factor binding to the GAL promoters and other proteins that interact with each other to regulate the expression of themselves and other GAL enzymes) and it does not have a known functionality outside of galactose utilization. In particular, there are no known natural examples in which the two separate modules—histidine production or regulation of *HIS3* protein expression and the functionality of the GAL system—are associated. Thus, there are good reasons to believe that the exclusive linkage made here between them is novel. Based on the known nominal

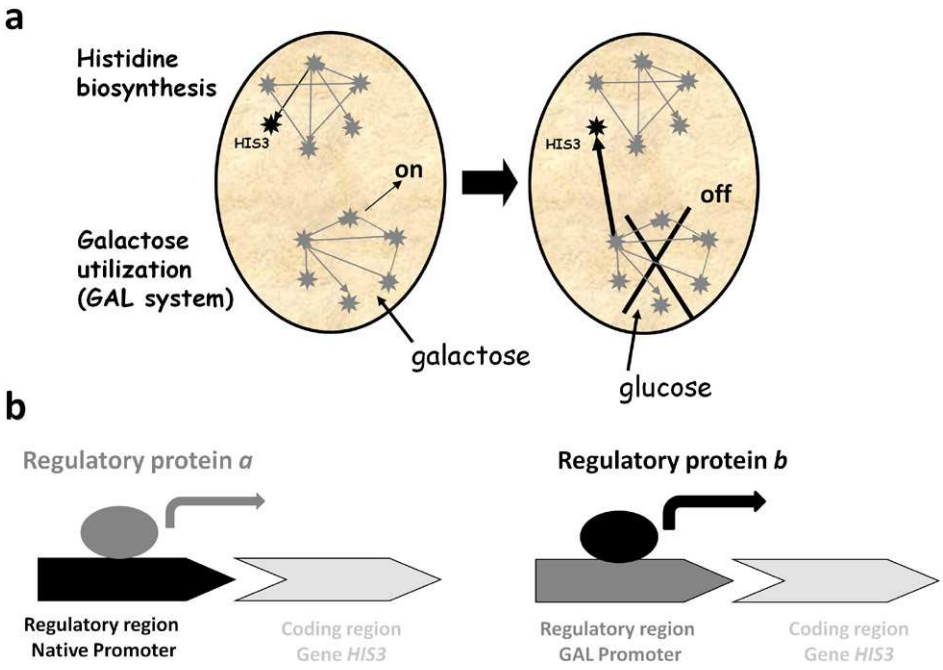

Figure 2. Genome rewiring in the budding yeast. (*a*) Left: A schematic representation of the two modules involved in our genome rewiring experimental setup, the histidine biosynthesis pathway and the GAL system responsible for galactose utilization. Right: In our rewiring setup, the essential gene *HIS3* from the histidine pathway is detached from its original regulatory linkage and placed exclusively under the promoter of the GAL system. The GAL system is highly induced when the sole sugar in the environment is galactose and is highly repressed when it is glucose. (*b*) In practice, the native regulatory region in front of the *HIS3* coding region is replaced with the GAL promoter (the promoter in front of the gene *GAL1*). In our experiments, the insertion was done by a standard genetic engineering approach in different ways, by a constructed plasmid as well as integration into the genome. The basic phenomena discussed in this article are insensitive to the mode of construction of the rewired strain.

functionalities of the histidine biosynthesis pathway and the GAL system, we can compose a tentative list of *a priori* challenges presented to such rewired yeast cells. First, in pure glucose-based medium (yeast's preferred sugar), the GAL system is highly *repressed* (i.e. its protein production is either shut-down or is maintained at very low levels) and since *HIS3* is exclusively linked to the GAL regulation, it is also repressed [31, 32]. Note that cells defective in *HIS3* protein production or lacking its coding sequence, could not survive in a medium lacking histidine [30]. Second, in pure galactose-based medium, the GAL system must simultaneously support the regulation of two metabolic functions; galactose utilization and histidine biosynthesis. This *multifunctionality* is non-trivial since under certain conditions it might present contradictory requirements to the system and overload the GAL regulatory system. Third, the dynamic range of the GAL system (1000-fold between induced and repressed states; one of the highest known in biological cells) [31–34] is much wider than that determined by the native amino-acid regulation (a few fold) [35, 36]. Thus, *HIS3* can be *over- or under-expressed* well beyond its natural operation levels. Since the functionality of *HIS3* is tightly connected to other enzymes in the biochemical metabolic network, its level of expression and its activity should be somewhat coordinated with the rest of the system. In particular, over-expression is as problematic to the cell as under-expression, since this might lead to accumulation of toxic intermediates. Finally, the recruited *HIS3* is *detached* from its natural feedback, limiting its response to relevant metabolic and regulatory demands [35]. The above list is merely a guess, relying on nominal functionalities of the histidine and GAL systems. In fact, we do not know the actual challenges faced by the cells under different conditions. Indeed, the aim of the experiments was to study, in detail, the dynamical response under different environmental conditions, without imposing any prior bias. The cells had never before performed regulation of *HIS3* based on carbon sources availability (different sugars) at any time in their evolutionary history, so a substantial adaptive response was required for them to survive in medium lacking histidine. We learned that the novelty of the challenge resulted in unexpected responses—mainly that the initial *local* perturbation (*HIS3*-GAL rewiring) resulted in a *global* (genome-wide) reorganization.

An unbiased view of the population dynamics requires large, proliferating cell populations to be measured over extended timescales, while at the same time covering the wide dynamic range of the intracellular and population processes. It demands the development of an experimental methodology allowing long-term high temporal resolution measurements of the evolving populations. Towards this end, we developed a home-made chemostat, a continuous culture device allowing measurements of a large population over many generations under stable environmental conditions [30, 37]. A chemostat is a reactor that grows cells by pumping fresh medium in at a constant rate, while balancing it by dilution of the culture at precisely the same rate to maintain a fixed volume. Under constant conditions, there exists a stable fixed-point of the dynamics, in which the average growth-rate of the cells equals the dilution rate of the chemostat [38]. The cell density then

reflects the maximal medium capacity to support this stable growth-rate. Therefore, when all parameters are held constant, a change in cell density reflects a change in their metabolism. Our unique setup enables continuous online measurements of the cell density (via the optical-density (OD) which is proportional to cell density) and the fluorescence intensity of *single cells* at high temporal resolution. Changes in OD reflect the mean population density dynamics while fluorescence signals of tagged proteins report expression levels, allowing us to correlate gene expression dynamics with metabolism. In addition, an automated cell collector constructed in-line with the chemostat setup, facilitates the collection and instantaneous freezing of cell aliquots at precise time points, so that the entire population history is available for later analysis (e.g. mRNA content and DNA sequencing) [30, 39, 40].

We turn now to the major phenomena observed when a population of rewired yeast cells, with *HIS3* exclusively linked to the GAL system, is followed over extended timescales in a medium lacking histidine. We mainly concentrate on an environmental switch from galactose to glucose-based medium since this experimental setting was the one that opened the widest vista on cell biology to us. We later discuss other challenges and phenomena observed under different settings and environments, allowing us to generalize and gain more insight into the behavior of cell populations.

1.4. Main phenomena observed

When rewired yeast cells with *HIS3* linked exclusively to the GAL system are switched from a galactose-based to a glucose-based medium, there are three major observed phenomena we wish to discuss here: (i) fast inherited adaptation, (ii) global gene expression response, and (iii) metastable population phenotype states.

1.4.1. Fast inherited adaptation. Detailed experiments have shown that a cell population carrying the *HIS3*-GAL rewired genome and switched from a galactose to glucose-based medium lacking histidine rapidly adapts (within ~10 generations) to grow competitively in this medium despite the strong initial repression of *HIS3* (figure 3(*a*)) [30]. Similar adaptation of genome-rewired cells to glucose has been shown for three different culture techniques: chemostats, batch cultures and cells grown on agar plates (figure 3(*b*)) [30, 39, 41]. An adapted cell has a distinct phenotype since it can grow into a mature visible colony on a glucose plate within 2–4 d (i.e. producing an exponentially growing lineage in this medium) while naïve cells take between 6–21 d to adapt and start growing on the plate. However, the precise population dynamics can only be measured in chemostat experiments, as all other techniques are invasive and interfere with the dynamics. Once established, the adapted population state can be stably propagated, with cells reproducing at a rate similar to that of wild-type cells, for hundreds of generations. Note that adaptation here is detected as a *population-average* response—the ability of the population as a whole to grow exponentially and maintain high cell density in the chemostat, which does not necessarily imply that every individual within the population has reached the same stable adapted state. In fact, as discussed below, individuals within the adapting population might alternate between transient metastable states, requiring a much longer time to stabilize their metabolism and growth condition in glucose. Our experiments show that the inherited adaptation is not due to selection of rare advantageous phenotypes; every cell in the population has, in principle, the potential ability to adapt [41]. We discuss now this adaptation phenomenon and the observed dynamical response in more detail.

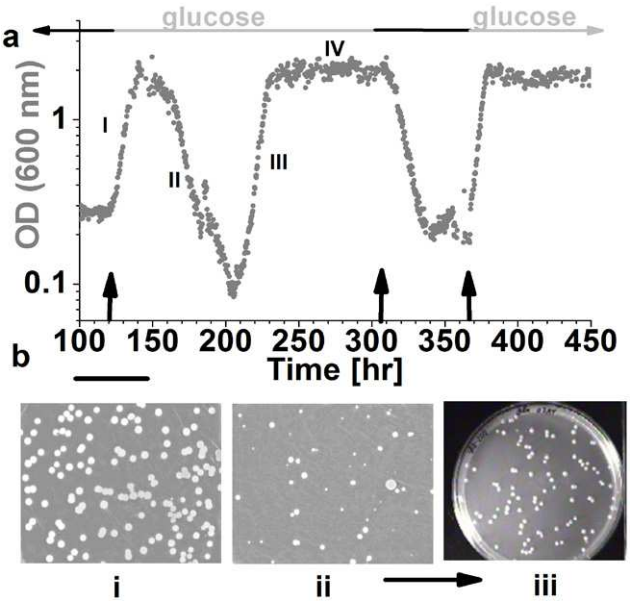

Figure 3. Fast inherited adaptation. (*a*) Typical population adaptation dynamics in the chemostat. The optical density (OD at 600 nm) measured in-line with the chemostat as a function of time. The OD is proportional to the population cell density. The population was first stabilized in a galactose medium lacking histidine. Next, the medium was switched from galactose to glucose at the left arrow, leaving all other nutrients the same. The medium was switched back from glucose to galactose at the middle arrow, and to glucose again at the right arrow. Four phases of the population dynamics are marked I–IV; fast exponential increase (I), followed by an exponential decline (II). The adapted population exhibits an exponential increase in density at (III) and the stabilization of a new steady state in (IV). At the second medium switch to glucose, the population density was immediately increased (exponentially growing at maximal rate) to the steady-state level of phase (IV), indicating that the population adapted state is inherited. The chemostat dilution time is 7 h. Bar, 10 cell generations (generation time equals chemostat dilution time × ln 2~5 h). Note the *y*-axis logarithmic scale. Repeated experiments show similar phases of the dynamics with variations in the population adaptation time. (*b*) The chemostat experiment in (*a*) was repeated on agar plates lacking histidine. Cells from a batch culture propagated in galactose were placed on an agar plate with (i) galactose (first visible colonies after 2–3 d, saturation after 4 d) and (ii) glucose (first visible colonies after ~6 d, image taken at day 14, showing ~40% of the colonies on galactose). Each colony was grown from a single cell on the plate. Both plates were initiated with the same number of cells. Note the uniform colonies in (i) and the variations in colony-size in (ii), indicating variation in adaptation times of the different lineages. A visible colony contains about 10^6 cells. Single cells from a colony from (ii) were re-distributed on a glucose agar plate in (iii), showing uniform colonies after 2 d of growth, indicating that adaptation is inherited. Reproduced with permission of the Genetics Society of America from [30]; permission conveyed through Copyright Clearence Center, Inc.

By studying cells with a rewired genome we demonstrated an adaptation process with the following characteristics: (i) adaptation occurs rapidly, within 10–30 generations; surprisingly shorter than adaptation processes usually encountered in laboratory experiments involving the fixation of mutations, which are on the order of hundreds to thousands of generations [30, 41]. (ii) The fraction of cells in the population switched from galactose whose lineages can adapt to grow on glucose is extremely high and variable, with an average of about 50% and as high as 80% [41]. For comparison, evolution of microorganisms in lab experiments usually shows a fraction of cells adapting to a severe stress in the range of 10^{-9}–10^{-5} [28, 42–47]. (iii) The adaptation of the population relies on many individual cells that independently develop the adaptive phenotype as a response to the challenging environment. The adapted phenotype is induced in numerous individual cells and does not result from selection of a rare subpopulation. (iv) In spite of its unusual characteristics, this adaptation is nevertheless genuine since the phenotype of adapted cells is distinct from that of naïve cells in the galactose phase and since this phenotype is stably *inherited* at the population level. Taken together, these characteristics signify an adaptation phenomenon that does not conform to the common view of the *neo-Darwinian* evolutionary process, based on selection of a rare pre-existing variant and thus, extends this framework and our understanding of how new phenotypes evolve.

At the onset of the glucose dynamics, despite of the strong repression of *HIS3*, the chemostat population exhibits an exponential increase in cell-density marking its ability to switch to the metabolically preferred glucose metabolism (marked phase I in figure 3(*a*)) [30]. Only after a few generations of exponential growth, the population shows signs of a crisis by a sharp decrease in its average cell density. The initial response of cells might be thought to reflect existing store of resources (e.g. histidine, *HIS3* mRNA molecules or proteins) allowing them to grow normally in spite of the *HIS3* repression, but in fact it reflects a richer behavior (see [41] for discussion). Indeed, the dynamics of the population immediately following the switch to glucose are somewhat unexpected when looking closely at the behavior of its individuals. When the initial phase of growth is further examined it becomes clear that the individuals composing this population progressively lost their adaptation capability namely, their ability to grow stable lineages in glucose (figure 4) [41]. In a conventional adaptation process based on selection, the fraction of adapted individuals is expected to increase with time in the selecting environment, or else it would be difficult for rare advantageous individuals to become established. Our experiment shows the opposite trend at the onset of exposure to the challenging environment. Thus, naïve cells, which were not exposed to glucose before, grow at a well defined single rate after the first encounter with glucose but do so only for a limited number of generations before either slowing down, or cease dividing. The size of a lineage emerging from a single cell extracted at a given time point in this phase, gets smaller and smaller with time, thus reducing the probability of its lineage to adapt [41]. In the next phase of the dynamics, the population-average density collapses, while many of its individual cells start developing

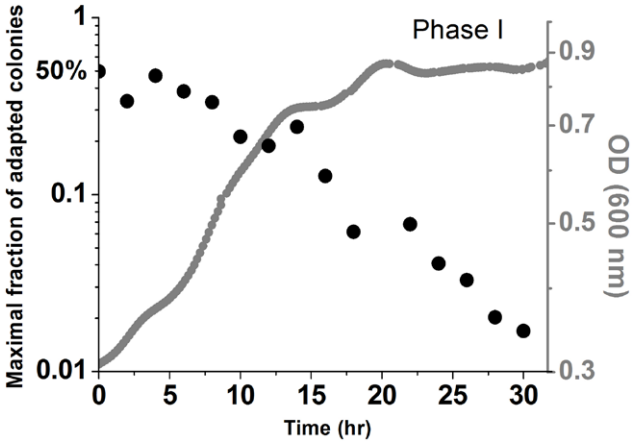

Figure 4. Population characteristics upon exposure to glucose. The maximal fraction of adapted colonies (dots) along phase I of the chemostat dynamics as marked in figure 3 (switching from galactose to glucose shifted to $t = 0$). Single cells extracted from the chemostat at different time points during phase I were placed on glucose agar plates lacking histidine. The number of mature (visible) colonies initiated from individual cells was counted after incubation for 20 d. Each mature colony represents an adapted lineage initiated from the plated ancestor cell. The maximal fraction of cells leading to adapted colonies at each time point was estimated by comparing the number of mature colonies to the total number of cells placed on the plates (estimated by placing the same sample on rich-medium plates). The maximal fraction is at the onset of switching from galactose to glucose ($t = 0$), around 50%. Note the logarithmic scale. The chemostat population density, measured by the chemostat OD, is shown by the gray line. Reproduced from [41] with permission (Taylor & Francis Ltd; www.tandfonline.com).

an adapted phenotype; turning on the population-average growth capability within a few more generations (phase II–III in figure 3(*a*)) and finally stabilizing the cell density at a high level (phase IV in figure 3(*a*)) indicating that the population on the average is adapted to the glucose-based medium. There are several indications that adapted cells have a very similar metabolism to that of naïve cells upon their exposure to the glucose medium in phase I, despite the challenge. The similarity in metabolism is reflected in the similar growth-rate of naïve (phase I) and adapted cells (phase IV), as measured by growing individual cells, extracted from the chemostat at the different phases, under similar conditions (the naïve cells growth eventually slows-down, but before that they show an exponential growth at a constant rate) [41]. Moreover, the chemostat population density of adapted cells in phase IV is similar to the peak density level of naïve cells at the end of phase I (figure 3(*a*)). The density of cells in the chemostat is sensitive to their metabolic state due to the limiting-nutrient condition. Thus, a similar density of cells in the chemostat under the same medium composition reflects similar metabolic yield. Therefore, given the familiar glucose-containing environment and the similarity in metabolism it is noteworthy that the adaptation phenomenon observed in our experiments is not so much about acquisition of a new metabolism but rather about *reorganization and stabilization* of a familiar growth capacity under *novel regulatory constraints* imposed by the genome rewiring event.

What are the possible processes involved in the adaptation phenomenon? To start answering this question, we can take advantage of the observed rapid inherited adaptation and utilize our genome-rewiring setup to inquire about the deep relationship between genetics and physiology, at the heart of the *genotype-to-phenotype* associations. Our experiments led to the identification of *multiple alternative trajectories* to adaptation. While in some of the cases mutations could be identified in the adapting populations, we have shown that adaptation can occur without mutations and even when mutations appear they require additional processes beyond a mere change in DNA sequence to stabilize the adapted state [48, 49]. If mutations are involved in the process, the fact that adaptation occurs simultaneously in numerous individuals, they should arise by one of the following processes. Either random mutations occur at the conventional low rate and rare advantageous ones propagate in the population prior to the first exposure to glucose, or cells acquire adaptive mutations (mutations directed toward advantageous phenotypes) at a very high rate after the exposure to glucose. Already the observation that more than 50% of the cells lead to an adapted lineage is not compatible with the picture of rare random mutations [41]. Methodologies in yeast genetics allow us to correlate specific phenotypes with genomic loci of interest. A diploid cell containing two sets of chromosomes is first made by mating two haploids each containing a single set of chromosomes. Such a diploid cell can be induced to undergo a process of meiosis, forming—after two cycles of division—four haploid cells. The genomes of these four haploids are recombinant mixtures of the original diploid genome. Conveniently, in yeast the four haploid cells resulting from a single diploid are arranged as a tetrad within a sac and so are easily identified as resulting from the same mother cell. Utilizing analysis of such tetrads, we next inquire about the phenotype of the four segregating haploid cells resulting from the meiosis of a single diploid. In our case, diploids were generated by mating two haploid cells, one adapted and the other naïve (figure 5(a)). We can then get statistics of phenotypes resulting from mating of a naïve cell with an adapted cell, looking for the fraction of adapted cells in the resulting four haploids. Large-scale statistics over many such tetrads, originating from mother-cells extracted from different adapting populations, show three outcomes, with a ratio between adapted and non-adapted cells in the tetrads of 2 : 2, 0 : 4 and 4 : 0 (figure 5) [48]. An adapted cell is practically identified as one that can grow into a mature visible colony on a glucose plate within 2–4 d (i.e. producing an exponentially growing lineage in this medium; recall that naïve cells take between 6–21 d to adapt). Remarkably, a population derived from the same mated diploid mother cell can show all three outcomes. Since the four haploid cells are random recombinants of the original diploid genome, a ratio of 2 : 2 (i.e. 50% of the cases) is what expected from the so called simple *Mendelian* process; modification in a single DNA locus, a result also supported by the absence of 3 : 1 or 1 : 3 cases from the statistics. Note that this analysis is not sensitive to whether the resulting phenotype, naïve or adapted, is due to a genetic (i.e. a mutation) or an epigenetic process (a process that could be associated with a DNA locus but without a DNA sequence modification). The common tendency of automatically assuming a mutation as the source of an emerging phenotype might be misleading in this case.

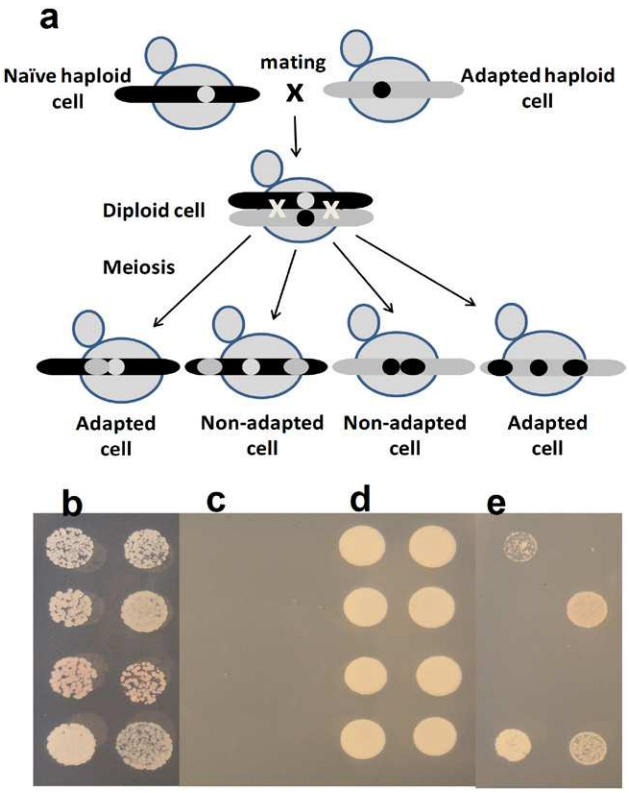

Figure 5. Multiple trajectories for cell adaptation. (*a*) A scheme of a 'tetrad' experiment: Mating two haploid cells (containing a single copy of the 16 chromosomes of the budding yeast); one adapted (extracted from phase IV of a chemostat (see figure 3), or a batch propagated population in glucose; black) and the other naïve (not exposed to glucose; gray) created a diploid cell containing two copies of the set of chromosomes, one from each mother cells. Chromosomes are marked schematically as a bar. The haploid 'mother' cells can be of the same strain or from two different polymorphic strains (containing single-base changes along their genomes—serving as markers), as used for linkage-mapping, correlating the DNA locus with the adapted phenotype (see main text for details). Following meiosis, a diploid formed a tetrad, four haploid cells. Each cell of the tetrad contains a single copy of the chromosomes which are recombinant mixtures of the genomes of the two mother cells. Growing the four cells of a tetrad, on glucose agar plates lacking histidine, reveals the different patterns of phenotypic segregation between adapted and non-adapted cells (*c*–*e*). (*b*) On a rich medium all four cells of a tetrad can grow into visible colonies within two days. Three types of segregation ratios between adapted and non-adapted cells were found on glucose medium lacking histidine: **0** : **4** (*c*), **2** : **2** (*d*), and **4** : **0** (*e*), respectively. Each column presents spotting of four cells of a single tetrad and two tetrads are shown in each case. Colony spots were imaged four days post plating and thus, visible spots contain colonies of adapted cells while absent spots represent non-adapted cells. Note that ratios of 3 : 1 or 1 : 3 were not detected in our experiments. Reproduced from [48] by permission of Oxford University Press.

Based on detailed analysis we conclude that genetic changes could not account for the entire spectrum of adaptation solutions [41, 48]. The cases of four non-adapted haploid cells (0 : 4), or all adapted (4 : 0), are of great significance since they show that the complexity of the adapted phenotype is not

manifested in a simple 'genetic rule'; the adapted phenotype can either 'infect' the naïve genome and then stably propagate through meiosis, or be completely eliminated. The 2:2 case on the other hand, looks simpler and in principle allows us to map the DNA locus of interest (again, for either a genetic or an epigenetic change). Mating two highly polymorphic strains (i.e. having numerous base differences, serving as markers, along their genomes), one naïve and one adapted, allows linkage-mapping of a specific locus in the recombinant genome using the polymorphic sequences as markers. We have performed three independent such mappings, by correlating the emerging adapted phenotype in a haploid cell of the tetrad, with a specific genomic locus (see figure 5 for the process and [48] for more details on the method). Sequencing the mapped loci lead to the identification of a mutation in one case, and *lack of mutations* in the two others. Interestingly, however, the emergence of a mutation does not by itself guarantee the stabilization of the adapted phenotype. High temporal-resolution experiments show that even when a mutation arises, the adapted phenotype is not shared by all cells in the lineage proliferating from the mutated cell, although all progeny do carry the same mutation. In other words, when propagating a population of cells all having the mutation, only a fraction is found to grow with the adapted phenotype on glucose plates. At first, such an incomplete inheritance seems to contradict the strong correlation between the adapted cells and the genomic locus, in the 2:2 adapted-naïve ratio cases. However, note that the mapping is based on a population-average phenotype; the emergence of a visible colony within a short time, which does not imply that every cell within the colony stabilizes an adapted state. This important distinction between the population-average behavior and the dynamics of single cells is not always appreciated in the context of genetics. Here, it suffices to point out that given a large population (in this case, a million cells in a visible colony of a haploid cell resulting from the meiosis process) with a distribution of phenotypes, the fast proliferating subpopulation at the tail of the distribution dominates (at least transiently) the population-average response. There is no requirement for all, or even the majority, of cells in the population to be stably adapted for designating the colony as adapted. The incomplete propagation of the adapted phenotype in a lineage means that intracellular processes beyond the mutation are necessary for stabilization of the phenotype.

The entire spectrum of mutations arising in our adaptation experiments is still under investigation. We have shown by direct experiments that when mutations arise, they are *induced* late in the process. In particular, mutations are not traces of preexisting genetic variability present in the galactose medium prior to the medium switch; *they are induced after the exposure to glucose at the end of phase I* (figure 3(*a*)) [49]. This is a clear demonstration that mutations can be induced during a dynamical process towards adaptation, rather than merely spontaneously emerging in a random fashion independent of the conditions (see [49] for details). The role of the emerging mutations is not clear. They may well be side-effects of the adaptation process. Two types of mutations arising in our experiments were analyzed in more detail, providing a glimpse at the complexity of the genotype-to-phenotype relationship [48]. In one case, mutations emerge in the coding region of the gene *GAL80*; the main repressor of the GAL system in glucose. Interestingly, several different mutations (misense or nonsense; i.e. introducing a change in amino acid or a premature stop codon, respectively) in the same coding region of *GAL80* were found to coexist in different cells of the same adapted chemostat population. Thus, several different mutations can occur independently in different progenitor cells and coexist in a single, proliferating population. A similar spectrum of mutations was found both in chemostat populations as well as in cells adapting on plates under non-competitive conditions, where ~50% of the naïve single cells adapt into visible colonies [48]. In all measured cases, no mutations were found in the regulatory regions, up or downstream of the *GAL80* coding sequence. Furthermore, detailed analysis suggests that the emerging mutations are not the result of a large-scale mutagenesis. High-resolution genome-wide sequencing of complete lineages arising from a single mother cell demonstrates this point as well as clear cases of adapted cells with no genetic mutations compared with the original naïve population [49]. Superficially, from the functional viewpoint, the emergence of a particular mutation in the *GAL80* coding sequence seems *a priori* sensible in the context of the rewiring challenge. A plausible picture might be that since the challenge imposed on the system involves the repression of the essential *HIS3*, due to repression of the GAL system in glucose, a mutation in the main repressor of the GAL system alleviates it. Our analysis however, shows that *this is not so*. A mutation in *GAL80* does not alleviate the *HIS3* repression in glucose. In fact, detailed expression measurements reveal that, in the presence of a *GAL80* mutation (even when the latter was introduced synthetically by replacement of the relevant DNA portion) while the GAL genes are indeed to large extent released from repression, *HIS3* remains repressed at a level similar to that without the mutation (see [48] for details). This is a remarkable demonstration that identical promoters in different loci and contexts can lead to very different modes of regulation, highlighting the non-trivial associations between genotypes and phenotypes. Moreover, the long-term expression level of *HIS3* in glucose is found in all cases of adapted rewired cells, either with or without a mutation in *GAL80*, to be similar to that of wild-type cells with *HIS3* under its native regulation. This result testifies to the success of the adaptation process in establishing *homeostasis*. This outcome only looks outstanding in light of our prejudice, but from another angle paints a perfectly reasonable picture of its own: without a repression mechanism, the expression of *HIS3* would be too high, reaching levels that are toxic to the cells. In reality, we have to admit that we simply don't know the actual 'economic reasoning' of the cell. From the regulatory point of view, it might be that weakening the constraints imposed by the GAL system on *HIS3* provides the necessary flexibility to reorganize regulation as part of the adaptation process. Independent of the precise interpretation, the take-home message is the apparent *degeneracy* of intracellular processes. In the absence of a major repressor protein, another repression process, of yet unknown origin, gets activated, tuning *HIS3* expression to proper physiological levels. Degeneracy, as we

further elaborate below, is the hallmark of the living cell; a property that stands at the basis of its ability to evolve.

A mutation in *GAL80* is not the only example of such degeneracy. The second type of mutation we found and examined in detail is the elimination of the entire GAL promoter in front of the *HIS3* coding sequence [48]. This deletion includes all the known binding sites for *GAL4*, the only transcription factor of the GAL system. Previous work has shown that *GAL4* and its binding sites are essential for significant expression levels of genes under the GAL promoter [50]. However, since in our case the loss of the *GAL4* binding sites do not hinder the adaptation process (or might even assist it), either *GAL4* can bind to other locations and activate *HIS3*, or another activation process overrides the natural *GAL* activator. In support of the possibility that transcription factors can bind to multiple locations outside of promoters, recent large-scale statistics on human cells showed that in all cell types examined transcription factors bind in a significant fraction inside the coding regions [51]. However, this large-scale study did not prove the functional significance of such bindings. In our case, functionality of *GAL4* binding is guaranteed by the essential role of *HIS3* and it remains for future work to uncover the molecular events leading to the activation of it in the absence of the *GAL4* binding sites. Since *HIS3* regulation (i.e. activation and repression) is required for the adapted phenotype, both the cases of *GAL80* mutations and promoter deletion seem to weaken the constraints imposed by the GAL system on the regulation of *HIS3*, suggesting that it facilitates reorganization of regulatory modes. We emphasize, that in all cases *HIS3* remained essential and there was no evidence for any genetic modification introducing a link to a different regulatory system other than the GAL. Thus, reorganization of regulatory modes did not involve new genetic 'rewiring' of *HIS3* beyond the one syntactically imposed by our setup. Moreover, we have shown that a double-mutant strain with both a non-functional GAL80p and a deleted GAL promoter, exhibits a similar adapted growth phenotype at the population level (i.e. a mutated cell grows into a mature colony within 2–4 d on a glucose plate) to strains with either of the mutations alone [48]. Thus, both mutations have a redundant phenotypic outcome, suggesting that indeed the main issues are the release of regulatory constrains and the degeneracy of the intracellular processes.

1.4.2. Global gene expression response. Intriguingly, underlying the adaptation process is a *global* reorganization of gene regulation: Following the switch from galactose to glucose, the adapting cell populations exhibit changes in genome-wide expression levels, involving a sizable fraction of the genome [52]. Genome-wide mRNA expression measurements show that the expression response presents strong correlations between genes across functional modules, with more-or-less symmetric profiles of two major clusters of genes that behave coherently, having enhanced and reduced levels of expression relative to that in the galactose steady-state. This symmetry might reflect limiting intracellular resources (RNA polymerase, other proteins or metabolic resources). Interestingly, the expression measurements show a simultaneous induction and repression of different genes within the same functional module, even genes that are nominally regulated by the same system and having similar promoters. A large fraction of the responding genes are nonspecific toward the challenge. Moreover, in spite of our extensive efforts, we could not find a biological 'logic' in the emerging gene expression response, with genes from the same functional groups behaving incoherently. At the same time, genes from different functional groups that do not have any *a priori* common functionalities, show coherent dynamics over extended time-scales. Thus, *co-functionality* does not necessarily imply *co-expression* and vice versa.

A series of experiments, quantifying the expression response of key genes at high temporal resolution, shows the emergence of a novel feedback between the histidine pathway and the GAL system. In the presence of the drug *3AT*,[4] a competitive inhibitor of the enzyme HIS3p, the mRNA expression levels of the GAL genes and *HIS3*, increase in the repressing glucose medium with increased concentration of the drug in the supplied medium. At high concentrations of *3AT*, both the GAL genes and *HIS3* are expressed in glucose at similar levels to that in galactose; i.e. *exploiting the full dynamic range of the GAL system* [30]. Therefore, inhibiting the activity of the essential histidine-production enzyme leads to an increased expression of the *a priori* unrelated structural GAL genes in glucose, due to the rewiring of *HIS3*, even though, the GAL system by itself is non-essential in this medium [30]. In fact, the response is graded by the amount of inhibitor in the medium, showing that the GAL system effectively 'reads' the demands from the foreign histidine metabolic module. This is another facet of reorganization of gene regulation in the adaptation process; the emergence of novel regulatory feedbacks between *a priori* functionally foreign modules.

We find that the global gene expression response is *irreproducible* between repeated experiments that nevertheless show similar population growth dynamics. This is a surprising result, since the irreproducibility in expression patterns is global and spans the entire set of metabolic genes that participate in the emergence and maintenance of a stable, adapted growth phenotype. Gene expression response and its relationship to the phenotypic cell-state depend both on the environment and on the history of the population. Indeed, the non-specific and irreproducible global transcriptional response is found to be sensitive to the level of applied pressure by the *HIS3* protein inhibitor *3AT* [52]. To better understand the significance of the irreproducibility in expression response, the richness of the gene expression spectrum and the sensitivity of the dynamics to the history of the population we compared the dynamics of gene expression between two populations with *identical histories*. Towards this end, we developed a novel chemostat setup in which two populations with a joint history could be

[4] *3-amino-triazole*, a known competitive inhibitor of the *HIS3* protein, shown to have negligible side effects on other genes in *S. cerevisiae* (Marton *et al* 1998 *Nat. Med.* **4** 1293). This inhibitor competes with its natural substrate on the active site of the *HIS3* enzyme, thus blocking a fraction of available functional proteins. Since *HIS3* does not have a redundant gene in the cell, this can only be compensated by increasing the protein copy number via gene regulation. Note that wild-type cells are severely damaged at concentrations of 10 mM of this drug, much below the doses used in our experiments.

separated at a defined time point and their dynamics as isolated populations could be examined under identical environmental conditions [40]. In the chemostat, populations are grown under severely challenging conditions in which cells compete for a limited resource. Thus, the relevant phenotype that integrates essential metabolic functions is that of growth-rate and proliferation. This phenotype is highly constrained for the adapting cells in our experiments, creating a well-defined history for the population. The experiments show that chemostat populations with identical histories ('twin' populations partitioned from a single steady-state mother-population prior to the switch from galactose to glucose) demonstrate variable mRNA expression dynamics of essential genes in spite of their similar metabolic response reflected in their population density dynamics (figure 6(*a*)) [40]. Our high temporal-resolution gene expression measurements show that the observed variable gene expression patterns are *not due to cellular 'noise'*. Rather, these patterns of expression reveal *collective dynamics* in the expression response of the cells within the population. Each population of genome-rewired cells develops a unique temporal pattern of mRNA expression profiles of essential metabolic genes, reflecting the collective population dynamics—an integrated outcome of intracellular and intercellular processes connected through transgenerational memory [7, 40, 53–55]. Thus, the population itself is a relevant level of organization affecting the cellular gene expression response via its collective dynamics. The emergent collective modes observed in the mRNA expression levels relax on extended time-scales of 10–20 generations (figures 6(*b*) and (*c*)) [40]. The coherency of the expression profiles on time-scales of many-generations demonstrates a non-trivial, trans-generational memory. Moreover, since we only measure a minute fraction of a large population,[5] coherent modes of activity over generations require a high degree of synchronization between the activities of individual cells. Thus, in our large population of 10^{10} cells, if the expression profile of each cell would be a random variable, the coherency of the population-average expression profile would decay within one generation.

Significant gene expression dynamics also emerge during periods of apparent steady growth (steady-state cell density in the chemostat), reflecting the complex relationship between gene expression and metabolism. Importantly, the observed expression profiles exhibit multimode dynamics in which each mode is populated with a group of coherently responding genes from different functional modules.[6] Deciphering the mapping between dynamic profiles of expression and a metabolic state requires the development of a statistical approach to populations. Note the shift in focus, from cell variation within a population to statistics over populations. The technology for laboratory experiments on statistical physics of populations, allowing the measurement of the dynamics of a large number of populations with identical initial conditions, does not exist yet. Our results motivate the development of such a research avenue. Indeed, a significant conclusion from our experiments is the need to connect the two levels of organization: intracellular processes to population dynamics.

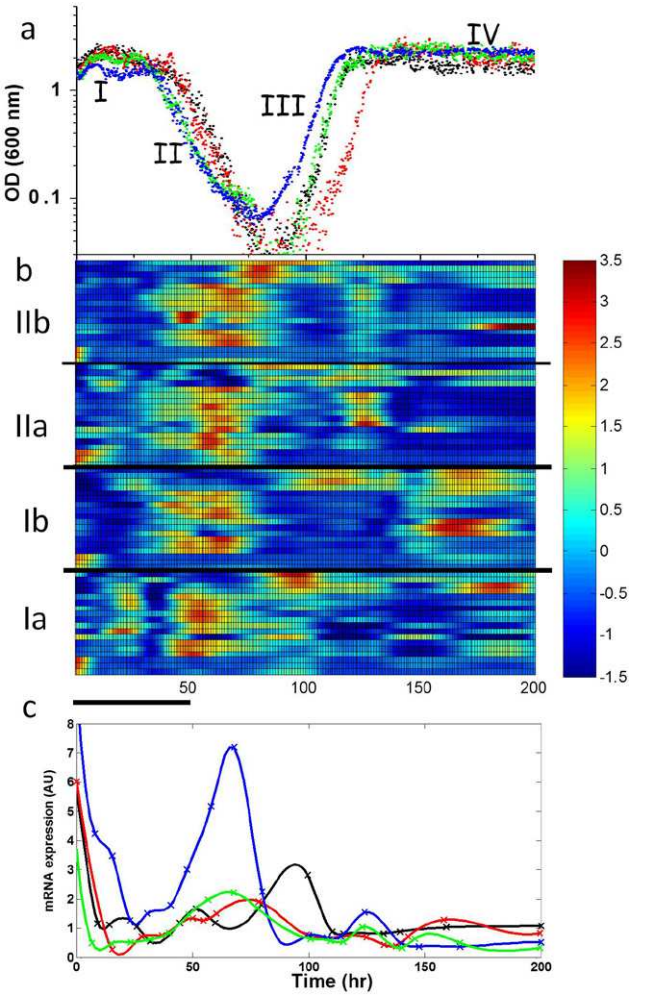

Figure 6. Phenotypes and gene expression profiles for 'twin' populations. (*a*) Cell density (OD at 600 nm) as a function of time for two pairs of 'twin' chemostats with populations of rewired cells (I*a*-black and I*b*-red are 'twin' populations and so are II*a*-blue and II*b*-green). The histidine lacking medium was switched from galactose to glucose as a sole carbon source at $t = 0$, leaving all other nutrients the same. A steady-state typical of galactose metabolism was first established for each pair of twin chemostats, mixed at rates faster than their dilution rate, therefore behaving as a single population. The two chemostats were decoupled prior to the medium switch into glucose. Thereafter, each population of the twin was grown in a separate chemostat. Note the *y*-axis logarithmic scale. Different phases of the dynamics are marked I–IV as in figure 3. (*b*) Color-coded raster plot of the mRNA expression profiles: I*a*–I*b* and II*a*–II*b* mark the same twin populations as in (*a*). The expression level of a gene as a function of time is marked by the color-variation across the horizontal line. Each line is a measurement of a different gene. The plot shows the measurements of 18 genes, belonging to different metabolic functional modules, repeated for each population. The measured expression levels were normalized for each gene to zero mean and unit standard deviation across its entire time profile. Bar—10 chemostat generations. (*c*) mRNA expression levels of *HIS3*, the rewired gene, for the four populations shown in (*a*) and (*b*). The data are normalized as in (*b*). Reproduced from [40].

[5] Typically we measure at each time point the average expression levels of 10^6 cells out of 10^{10} cells of the chemostat population.

[6] These measurements of a limited number of genes are at much higher temporal resolution than the global genome-wide measurements reported above. Therefore what appeared to be two large clusters of responding genes in the global measurements breaks now into a few modes.

There is a tendency to characterize variations in biological systems as resulting from noise, usually attributed to some intracellular molecular fluctuations. The irreproducibility in the population-average gene expression between populations demonstrates that this is not always the case. What about the expression variation between cells within the same population? We make now a short detour from our description of the adaptation phenomenology to discuss the general phenomenon of population-variation in gene expression. We later attempt to connect these fluctuations, in the case of our adapting yeast populations, to the observed collective population dynamics.

The protein content of a cell is thought to be a primary determinant of its phenotype, and the variation between the protein content of individual cells in a genetically uniform population has been the subject of intensive research in recent years [56–61]. It has generally been found that protein copy number widely varies even among genetically identical cells grown under uniform conditions ([62] and refs therein). Gene expression is generally coupled to all aspects of cell physiology, such as growth [63], metabolism [64], aging [65], division [66, 67] and epigenetic processes [68, 69], as well as gene location and function [70], all of which have in turn, been shown to affect variation in protein content. The emerging picture is of a plethora of interconnected mechanisms at different levels of organization. How they integrate to shape the total variability in protein content in a dividing population remains elusive. Motivated by this question, we studied the universal aspects of protein fluctuations in clonal (genetically identical) populations of microorganisms by comparing the population distributions over a wide range of biological contexts (for details see [62]): (a) Two archetypical microorganisms, bacteria and yeast, with two well-studied regulatory systems of essential metabolic pathways: the LAC operon in bacteria *Escherichia coli* [71] and the GAL system in yeast *S. cerevisiae* [34]. Both systems were studied under environmental conditions in which gene expression is strongly coupled to metabolism, namely they control the utilization of an essential sugar (lactose and galactose, respectively) as the sole carbon source. (b) Different metabolic growth conditions: the organisms were grown in chemostats–continuous culture in steady state and transients—as well as in batch cultures. (c) Highly regulated versus constitutive (not regulated) expression: the regulated LAC and GAL systems were compared to constitutively expressed proteins in both organisms. (d) Different promoter copy numbers: the same regulatory systems were placed on high-copy and low-copy number plasmids as well as integrated into the genome in a single copy. (e) A reporter Green Fluorescent Protein (GFP) was compared to an essential functional tagged-protein controlled by the same promoter. The spectrum of our experiments spans an array of '*control parameters*' that covers many of the essential processes affecting protein content in cells. The two organisms, *E. coli* and *S. cerevisiae*, are distinct in almost every aspect of their cell biology and life style, from gene regulation and expression to cell division and physical characteristics such as shape and volume. Remarkably, the broad protein distributions measured in this wide range of biological contexts were shown to collapse to a single *non-Gaussian* curve under scaling by the first two moments—subtracting the mean and dividing by the standard deviation [62]. Moreover, in all experiments the variance was found to depend quadratically on the mean, showing that in fact, *a single population-average degree of freedom* (e.g. the population-average expression level) determines the entire distribution of protein content.

Without a theoretical framework, the physical significance of a universal shape distribution of the protein content determined solely by a single population-average variable remains elusive. However, at this stage this finding has two major implications: First, protein variations do not reflect specific dominating molecular or cellular mechanisms, since if there was such a dominant mechanism it would be reflected in deviations from universality across the wide array of 'control parameters' varied in our experiments. Second, the observed distributions result from dynamical or highly-correlated processes rather than from statistics over fluctuations in independent processes which would converge to a Gaussian distribution by the central-limit theorem. The universality of the distribution and its insensitivity to external perturbations also implies that some buffering process masks the details of the intracellular molecular interactions; and that the intracellular degrees of freedom are somehow '*protected*' from external perturbations, much like in a physical system near a critical point [72]. In the same spirit, as shown above, the single population-average degree of freedom determining the protein content distribution, exhibits collective dynamics characterized by slow relaxation times (~10–20 generations), much slower than that of any known intracellular process. These population collective modes are further iscussed below. The picture arising from these measurements is of protein variations with fast universal dynamics (determining the protein distribution within the population) riding on a slower envelope of collective population-average expression modes (determining the population-average protein content).

To summarize this part, there are three lessons from our experiments on the expression dynamics: (i) Gene expression is not mapped to metabolism in a straightforward way; certainly the mapping between them is *not one-to-one* and it is hard to identify a biological *'logic'* directly connecting the hard-core biochemistry (e.g. in metabolism) to expression levels. (ii) The dynamics of gene expression show *collective* population modes exhibiting *slow relaxation* over extended time-scales and 'memory' over generations at the population level (see next section). Significantly, these modes survive the averaging over a large number (~10^{10}) of cells. (iii) The observed multi-mode dynamics are *idiosyncratic* to the population. Each population develops its own unique genome-wide expression profiles that are irreproducible in repeated experiments and even between 'twin' populations with identical initial conditions. This '*sensitivity to initial conditions*' at the population level (again, surviving population averaging) and the other points mentioned above call for a change in our conception of a cell population as further discussed below.

1.4.3. Metastable population states. The chemostat populations of rewired cells stabilize in glucose on a time-scale of

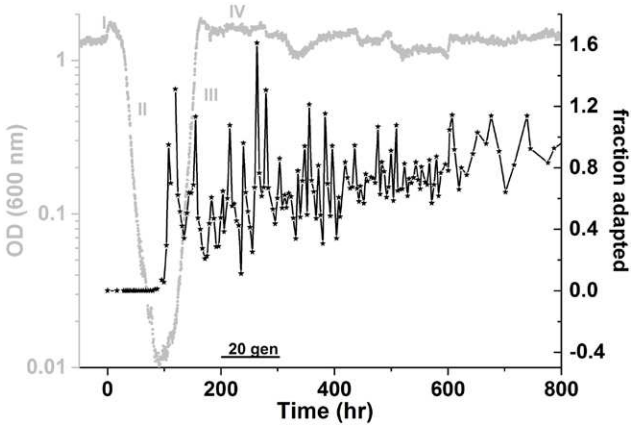

Figure 7. Fraction of adapted cells in the chemostat population dynamics. Cells extracted from the chemostat at different time points were dispersed as isolated single cells on glucose agar plates lacking histidine. The chemostat was switched from galactose to glucose at $t = 0$. Adapted cells grew into mature colonies within 3–4 d after plating. The black trace shows the number of cells that are able to grow a visible colony within this time period ('fraction' of adapted colonies) relative to the number of colonies grown on rich medium plates (and thus this 'fraction' can be larger than 1, since adapted cells can sometimes grow on minimal glucose medium better than on rich medium). The gray trace shows the population density (measured by the OD at 600 nm) as a function of time. Note the logarithmic scale. The four phases of the dynamics are marked I–IV, as in figure 3. Reproduced from [39].

~100 generations, long after their average growth capability adapted [39]. The populations exhibit nontrivial dynamics, as observed by several different probes, for hundreds of hours after the switch to glucose under constant conditions. In spite of the fact that the adapted population, on the average, can maintain a high cell-density in the glucose chemostat, the hallmark of an adapted state, a closer examination of the population dynamics reveals that it exhibits significant fluctuations both in its average cell density as well as in single-cell phenotype distributions (phase IV in figure 7) [39]. The fluctuations in the cell-density indicate that the average yield (biomass production per unit nutrient), and thus the average growth-rate and cellular metabolism, fluctuates on long time-scales of many generations. The slow relaxation of the population dynamics, in a constant environment during this period is most significantly observed by direct measurements of the phenotypic structure of the population. Samples of cells taken from the chemostat at high-resolution time intervals and spread on glucose plates show that this seemingly fully-adapted population exhibits fluctuations in the fraction of adapted cells, indicating that the adapted phenotype is not stably inherited for each individual propagated cell (figure 7) [39]. Note that the fluctuations in the adapted phenotype are significant; there are instances in which the fraction of adapted cells in the population drops to low values of ~10%. These measurements show that while the average population growth is exponential, *new cells are continuously being born in the chemostat that are not able to stably propagate the adapted phenotype* in glucose and thus to grow a colony on an agar plate. The fluctuations in the fraction of adapted cells seem to decay on a time-scale of ~100 generations, converging to values close to 100%. The contrast between the seemingly adapted state of the population as measured by its average behavior and the fluctuations in the fraction of adapted cells presents a puzzle, reminiscent of the one mentioned before in the context of the tetrad analysis. It again reflects a distinction between the picture arising from the state of individuals and that of the population, as measured by its average dynamics. A closer examination however, resolves this apparent paradox. Individual cells proliferate into lineages that can maintain growth *on the average,* while at the same time not every individual within the lineage sustains a constant rate of growth. Nevertheless, this growth mode can support the steady propagation of lineages and consequently the steady growth of the entire population. This situation is maintained in the glucose medium, until the adapted state is fully stabilized, ~100 generations after the switch to glucose. At this point, the majority of cells can grow exponentially at a constant rate, similar to that of wild-type cells. This means that adaptation, in the broad sense as observed in our experiments, involves a rich spectrum of *metastable* states. A transient metabolic state, enabling exponential growth for only a limited duration, cannot guarantee a steady growth of a cell over extended time-scales. However, the rich spectrum of phenotypes exhibited by its progeny supports the long-term propagation of the lineage. Interestingly, a plausible picture consistent with this population-structure dynamics is that cells that develop a transient growth state early after exposure to glucose also show a weaker transgenerational memory within their lineage, compared with cells that slowly adapt. Otherwise, if the fastest growing cells would have a long transgenerational memory they would take-over the population, eliminating the slow-growing lineages. This is not the case however, if the slow growers can more easily stabilize their state within the lineage, making it easier for their progeny to stabilize a growth-enabling metabolic state. In other words, it seems that the rate of developing a transient metastable state supporting growth is anti-correlated with the 'memory' of this state within the lineage. Population growth dynamics thus depend both on the instantaneous growth-rate and the stability of its long-term correlation within the lineage.

The existence of metastable states in phase IV of the dynamics in glucose (see figure 7), should be directly reflected in the distribution of growth-rates within the population and its dynamics. To test this prediction, we measured the distributions of instantaneous growth-rates[7] in batch cultures, switched from galactose to glucose and propagated in phase IV without ever diluting the populations to small numbers, avoiding bottlenecks [39]. We have shown that the adaptation dynamics including the four phases observed in chemostat experiments, are reproduced in batch cultures [41]. Importantly, by growing the populations in dilute batch cultures, with excess nutrients for the entire growth period, we can ensure that there are no stressed cells in the population due to nutrient limitations.

[7] The instantaneous growth rates were measured on single cells placed on a solid support by time-lapse microscopy. Each measurement was limited to a short duration of ~3 generations, allowing us to capture the phenotypic variability in the population while ensuring a single-exponent growth curve. Thus, this instantaneous growth-rate determines a 'local' variable, which in the population context varies on longer time-scales.

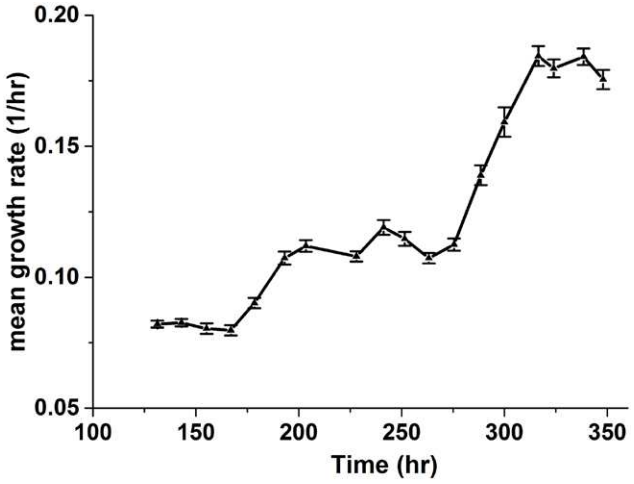

Figure 8. Mean growth-rate dynamics. Rewired cells with *HIS3* under the GAL system, were grown in batch culture lacking histidine in the medium and switched from galactose to glucose at $t = 0$. The populations were propagated throughout the adaptation phases by serial dilutions. We used a microscopy assay to measure the instantaneous growth rate of single cells throughout phase-IV; after the population on the average adapted to the glucose medium. For each experiment, at different time points throughout the propagation, the instantaneous growth rates of hundreds of cells were measured. Each measurement was limited to a short duration of ~3 generations, allowing us to capture the phenotypic variability in the population while ensuring a single-exponent growth curve. Thus, this instantaneous growth-rate determines a 'local' variable, which in the population context varies on longer time-scales. The mean growth-rate was estimated from these distributions (see [39] for the distributions). Note the plateaus in the mean growth-rate that persist for many generations, followed by decreases and sharp increases. Error bars are the standard error of the mean. Reproduced from [39].

Even after exponential growth resumed in glucose, we measure a very broad distribution of the instantaneous growth-rates of single cells (σ/μ is about three times that measured for wild-type cells). These broad distributions persist for very long durations (~100 generations), indicating that subpopulations with a wide range of growth-rates coexist for long periods. The mean growth-rate also does not increase monotonically, as expected from a simple selection process. Instead, it continues to fluctuate throughout the propagation time in glucose. Of note is that the fluctuations include plateaus of long durations and instances of reduction in the average growth-rate, indicating an overall decrease in the population fitness (figure 8) [39]. These interesting dynamics are also reflected in the fraction of the population that can grow exponentially. This fraction converges to 100% on a similar long timescale of ~100 generation observed by the other probes (fraction of adapted cells, growth-rate distributions and gene expression, shown below) [39]. It shows again that from the single-cell perspective, the adaptation process is not completed even when the population-average dynamics seem relatively stable. These results show that the fastest growing cells do not take over the adapting population on a time-scale dictated by the width of the growth-rate distribution as expected from a selection process. In fact, *subpopulations of non-growers, slow growers and fast growers coexist*, forming a dynamic, continuous, spectrum of growth phenotypes within a proliferating population. As discussed later, these observations shed new light on the dynamics of cell populations and on our understanding of the concept of *fitness*.

Another significant observable that probes the extended relaxation dynamics is the expression level of the rewired *HIS3* protein. When tagged with a fluorescent marker, the level of *HIS3* protein shows large fluctuations, with individual peaks relaxing on time-scales of ~20 generations and an overall envelope of activity relaxing over ~100 generations (figure 9) [39]. As with the mRNA expression dynamics discussed before, the dynamics of this essential protein reflect collective population modes; otherwise these fluctuations would be averaged-out in the large population of $\sim 10^{10}$ cells. As with other genes, the dynamics of the essential protein are not mapped directly to cell metabolism and are not directly correlated with the ability of the population to grow and proliferate. Recall that the entire population dynamics are supposedly triggered by the repression of *HIS3* linked to the GAL system. This demonstrates again the intricate and indirect relationship between the expression level of an essential protein and the metabolic state of the cell. As before, the dynamics of *HIS3* are measured to be *irreproducible* between repeated experiments (compare figures 9(*a*) and (*b*)). The emergence of long time-scales in the dynamics, much longer than the familiar intracellular processes involved in gene expression and metabolism, is a highly significant input for any framework dealing with the genotype-to-phenotype associations and the emergence of stable cell states.

Finally, single-cell measurements show that the level of the *HIS3* protein is broadly distributed within the population (figure 10(*a*)). Furthermore, although the mean expression varies as a function of time, the shape of the single-cell protein distributions is preserved throughout the entire relaxation period of ~100 generations (hundreds of hours) in glucose. All the distributions collected at high temporal resolution throughout this extended period, collapse to a single curve by subtracting the mean and dividing by standard deviation (figure 10(*b*)). Moreover, the variance throughout this period is a quadratic function of the mean, showing that throughout the adaptation process a single population-average variable determines the protein expression distribution [39]. We reach again the same conclusion as for the general case discussed before: a universal, single-parameter scaling of the essential *HIS3* protein distribution indicates that the population as an entity plays a critical role in the dynamics. The wide protein-content distribution among cells does not result from intracellular noise in gene expression but rather from fast correlated, intracellular dynamics [62]. The consistency of these results with our previous observations shows that gene expression in the case of an absolutely essential gene in an adapting population is not fundamentally different from other (essential or non-essential, highly regulated or constitutive) genes in cell populations grown under a wide range of conditions [62]. This result leads to a similar conclusion; the expression dynamics reflect the composition of fast universal intracellular processes riding on a slowly-varying envelope of collective population modes. The biological specificity is manifested in these slowly-varying gene activity modes underlying the adaptation process.

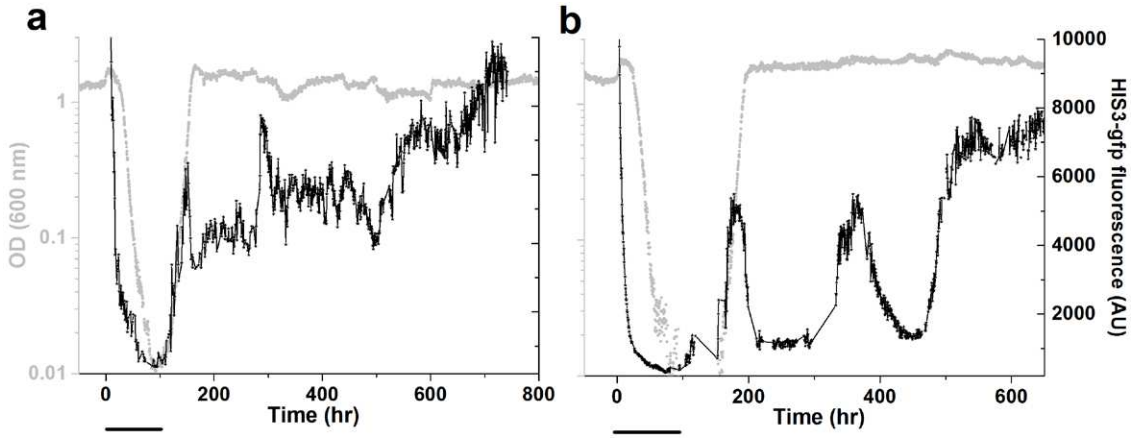

Figure 9. *HIS3*-protein dynamics in a chemostat population. The average protein level measured by the fluorescent level of a GFP tag fused to the *HIS3* protein (*HIS3*-gfp), in a chemostat population as a function of time (black curves). The chemostat was switched from galactose to glucose at $t = 0$ in a medium lacking histidine as in figure 3. The population-average fluorescence was extracted from statistics over single-cell measurements, utilizing our home-made cytometer in-line with the chemostat. The protein content fluctuates hundreds of hours into phase IV, showing collective modes that relax on a time-scale of 10–20 generations. (*a*) and (*b*) are two repeated experiments under nominally identical conditions, showing the irreproducibility of the expression response between populations. The population density dynamics measured by the optical density is shown in gray for reference. Bar—20 chemostat generations. Reproduced from [39].

1.4.4. The generality of adaptation to a rewiring challenge. How general are the characteristics of the adaptation dynamics and the phenomenology observed in the case of *HIS3* rewired to the GAL system? To gain insight on this question, we used a similar genome-rewiring methodology to link *HIS3* exclusively to different cell-cycle promoters [73]. The control of the cell-cycle is an important process affecting cell viability and functionality. Historically, the budding yeast had played an important role in understanding this process [74]. The cell-cycle progression was shown to be regulated by a set of promoters used in our experiments, mainly at the level of transcription (mRNA production), which advance the cell-cycle process through a set of well-defined phases [5]. Thus, our experiments also enable us to open a window on important dynamical aspects of the cell-cycle itself, in particular on the flexibility of its interface with the metabolic system. Harnessing cell-cycle regulators to directly control an essential metabolic process increases the load on the cell-cycle's regulatory network at a specific phase of the cycle and demands re-distribution of its resources. As in the case of the GAL system, such a perturbation introduces a complex challenge to the cells by requiring the cell-cycle regulators to operate outside their natural context and in concert with arbitrarily chosen metabolic demands.

The general characteristics of the adaptation dynamics observed for the GAL system rewiring repeat in the case of the cell-cycle. We have shown that yeast cell populations with *HIS3* 'rewired' to the cell-cycle promoters can rapidly (10–40 generations, depending on the promoter) adapt to grow at normal rates despite increased inhibition of the rewired metabolic gene by the drug *3AT* [73]. Moreover, the new adapted phenotype is stably inherited for generations. A significant fraction of the population has the potential to adapt to the severe unforeseen genome-rewiring challenge. Thus, as before, the ability to adapt is not a special property of a rare subpopulation. Furthermore, we showed that underlying the adaptation process there is a *non-specific* and *irreproducible* genome-wide transcriptional response. The cell-cycle system is regarded as a tightly regulated network, ensuring its robust temporal ordered dynamics and proper progression through its various phases. Our experiments show that it forms a flexible and adaptive interface with the metabolic cellular processes. This interface can support a *multifunctional* utility of the cell-cycle promoters and enables concurrent regulation of their native function and a foreign, essential metabolic gene. Since *HIS3* is essential for histidine synthesis and there is no alternative pathway for this synthesis in the budding yeast, adaptation to accommodate high levels of inhibition under the regulation of a cell-cycle promoter requires reorganization of the regulatory system enabling its expression at proper levels. This reorganization is manifested in a genome-wide transcriptional response similar to the one observed for the GAL system. Once again, a local perturbation linking an essential gene to a foreign system requires global reorganization of gene regulation to accommodate the challenge. These results prove once more the existence of a general and non-specific cellular process allowing adaptation to unforeseen challenges.

In spite of the similarities between the GAL and cell-cycle cases, the latter is different in certain important aspects from the former. While the GAL system does not have any particular functional role in glucose and is certainly not essential in this medium [30], the proper functionality of the cell cycle is absolutely essential under all environmental conditions. Thus, while the resources of the GAL system in glucose could be fully harnessed to meet the metabolic requirements of the rewired cells, this is not the case for the cell-cycle which must operate in a multi-functional mode. Moreover, in the case of the cell-cycle, each rewired promoter binds a number of transcription factors thus, requiring combinatorial re-distribution of resources. By contrast, the GAL system is regulated by a promoter binding a single transcription factor (*GAL4*), the activity of which is modified by interactions with other

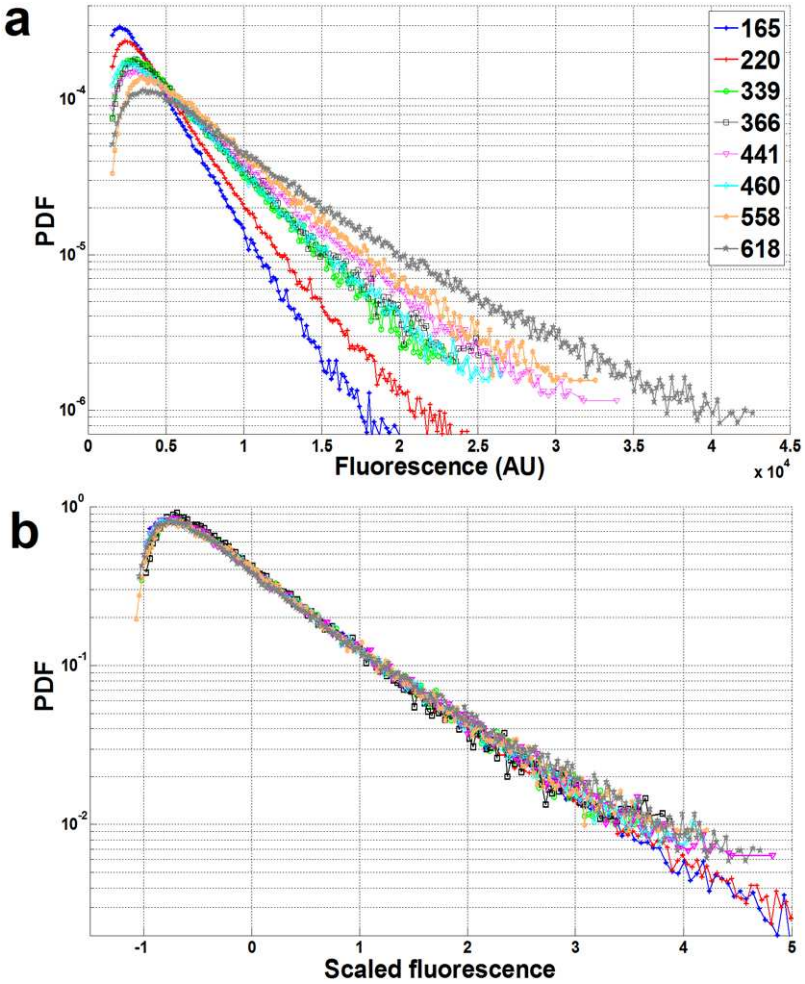

Figure 10. Population distributions of *HIS3*-protein expression levels. (*a*) Examples of distributions of the single-cell *HIS3*-protein expression levels (for the same population as in figure 9(*a*)) for several time-points during phase-IV of the dynamics. The legend shows the time in h from the switch of the chemostat from galactose to glucose. (*b*) All the distributions from (*a*) collapse to a single curve when subtracting the mean and dividing by standard deviation of each distribution. See [39] for the entire set of distributions, measured over hundreds of hours in phase IV of the chemostat dynamics. They all show a collapse to the same universal shape upon scaling. Reproduced from [39].

proteins (e.g. *GAL80*). The cell-cycle rewiring experiments also add a temporal constraint to adaptation that does not exist in the case of the GAL system. Since the transcription factors driving the cell-cycle are active only during specific phases and must be sharply degraded outside of these relatively narrow time windows, their rewiring to control an essential metabolic process adds further constraints on their functionality. Furthermore, in native cells these transcription factors are mostly expressed at low levels [75], so that the multifunctional challenge is expected to result in a 'costly' re-distribution of relatively limited resources (see a model in this spirit in [76]). Thus, the unforeseen rewiring challenge either forces the cell-cycle transcription factors to operate outside of their natural context (e.g. outside of their narrow time-window activity), or causes the rewired cells to utilize physiological processes to differentiate between the native cell-cycle promoter and the one that regulates *HIS3*.

Multifunctionality can also be examined in the case of cells with *HIS3* rewired to the GAL system grown in galactose, when both galactose utilization and histidine biosynthesis need to be supported by a single regulatory system. In fact, multifunctionality is widespread in cells even without recruitment of foreign genes, due to the diverse functions of many regulatory systems and the *pleiotropy*[8] of many genes. In the case of the GAL system, this phenomenon was studied by measuring the chemostat population dynamics while applying the *HIS3* competitive inhibitor drug *3AT* in galactose. The introduction of *3AT* applies pressure on the histidine metabolic pathway and due to the rewiring of *HIS3* to the GAL system, requires adaptive dynamics in galactose [30]. The galactose population exhibits a reduction in the population density at the onset of the application of the inhibitor to the medium, followed by an adaptive recovery. This adaptation occurs, similar to the adaptation in glucose, on the time-scale of 10–20 generations. Interestingly, below a critical concentration of *3AT* (~45 mM), the population reaches a final steady-state density that is higher than the one prior to the introduction of the inhibitor. In other words, cells can develop a better

[8] Pleiotropy—affecting multiple target genes or traits.

metabolic state under stressful conditions. Removal of the inhibitor after the establishment of the adapted state leads to a transient decrease in cell density followed by reestablishment of the same high-density steady state. Thus, as in the case of adaptation to glucose, also in galactose, the adapted state is memorized by the population of cells. Once cells adapt, the presence of the environmental trigger is not required for the maintenance the new phenotype.

Multifunctional behavior brings to light another issue that was observed in our experiments. When the GAL system is required to regulate both sugar metabolism and amino acid biosynthesis, the expression of the genes in the GAL system is not necessarily optimal in galactose, even though galactose is the inducer of this specific regulatory system. Extremely high concentrations of *3AT* (>80 mM) could still be tolerated by a steady-state population in glucose. However, that same population collapses upon a switch from glucose to galactose. Thus, multifunctionality can limit the utility of a system due to its limited resources—in the case of the GAL system, the copy-number of its regulatory proteins [53, 76]. Switching a low-density galactose population back to glucose, results in rapid recovery to the original steady-state density. This shows that the population memorizes its steady-state phenotype under the high concentration of inhibitor, even after metabolism became transiently poor in galactose.

1.5. Summary of main experimental results

1.5.1. Perturbations of regulatory modes. Genome rewiring, recruiting an essential gene to a foreign regulatory system, is a perturbation of *regulatory modes*. Therefore, the adaptation to this *unforeseen challenge* requires cells to surmount the incompatibilities arising between the metabolic demands and protein expression. This type of perturbation, reveals cell dynamics beyond the fast relaxations observed under more conventional perturbations, such as switching between familiar nutrients or applying familiar stresses [30, 52]. These dynamics seem *exploratory* in nature and involving timescales much longer than a generation time.

1.5.2. Inherited adaptation to an unforeseen challenge. A general process enables fast *inherited* adaptation of cells to the severe, *unforeseen* challenge. Adaptation is induced in the challenging environment simultaneously in numerous (~50%) cells in the population and does not result from selection of a rare subpopulation [41]. In our experiments, adaptation is carried out in a familiar glucose-containing environment but the regulatory response, due to the exclusive linkage of the histidine essential gene *HIS3* to the GAL system is novel. Thus, adaptation in this case involves reorganization of gene regulation, enabling single cells to stabilize a familiar metabolic state [30, 41, 48, 52]. This reorganization can be rapidly gained, at least transiently, by individual cells. Its long-term maintenance however, requires stable inheritance of the organized state for generations; a process that can be quickly established at the population level but spans a wide range of time-scales for individuals exhibiting highly variable adaptation dynamics.

1.5.3. Multiple trajectories in adaptation—genetics-physiology interrelations. There are *multiple alternative trajectories* to adaptation, involving both genetic and epigenetic-physiological processes, with an intricate relationship between physiology and genetics [48]. The existence of multiple trajectories is another manifestation of the exploratory dynamics underlying the adaptation process. Clearly, genetic changes in the form of mutations could not account for the entire spectrum of adaptation solutions. There are clear cases of inherited adaptation that do not involve mutations at all. However, in the cases where mutations appear, they are induced after the exposure to glucose [49], exhibit a redundant phenotypic outcome and cannot by themselves support the stable propagation of the adapted phenotype across generations without the assistance of other physiological processes in the cell. A puzzling situation arises, as the late-induced mutations in fact, do not show their expected functionalities (e.g. *GAL80* mutation alleviating the glucose repression of *HIS3*) facilitating the adaptation process. Thus, the role of these mutations is not clear and they may arise as side products of the adaptation dynamics. The emergence of multiple alternative trajectories in adaptation, reveals an important characteristic of the living cell; *degeneracy* of regulatory modes and responses.

1.5.4. From local perturbation to global reorganization of gene regulation. The adapting cell populations exhibit *genome-wide expression* dynamics involving a sizeable fraction of the genome and presenting strong correlations between genes across the entire system [52]. This indicates that underlying the adaptation process is a *global reorganization of gene regulation*. Thus, in lieu of a pre-instilled response to the challenge, a local constraint on gene regulation imposed by the genome rewiring event is resolved by a non-specific, global reorganization process. Certainly, tight regulation is not necessary for evolving an adapted phenotype. These non-specific, global expression dynamics are yet another signature of the exploratory dynamics underlying the adaptation process. In yeast, where reproduction is the major function in population experiments, metabolism is the underlying process, not gene expression. Surely, there are associations between metabolism and gene expression but our understanding of them is still lacking. What is clear is there is no direct mapping between these two types of processes. Our experiments expose the fact that there is no one-to-one mapping between the profiles of gene expression, both at the mRNA and protein levels, and the growth-rate of a cell population [39, 40]. Moreover, the temporal patterns of gene expression that emerge in repeated experiments are *irreproducible*. This is another manifestation of the *degeneracy* of the intracellular processes. Our yeast experiments show that 'twin' populations, derived from a single mother population and thus with identical initial conditions and similar metabolic responses (i.e. similar phenotypes), quickly diverge in their population-average gene-expression response [40]. For such large populations (~10^{10} cells), divergence at the population-average level does not represent random fluctuations among individual cells but rather population complex dynamics.

1.5.5. Slow collective population modes. In spite of the irreproducible gene expression response between populations, our experiments show that cells within a population, on average, have similar patterns of expression manifested in their collective dynamics [39, 40]. This contrast between the population-intrinsic coherent dynamics and the inter-populations irreproducibility shows that the population itself is a *relevant level of organization*. The response of a cell within a population is highly affected by the response of its sibling cells. Cells within the population thus exhibit *effective interactions* with each other [40]. This is manifested in the emergence of slow collective modes—expression profiles of sets of genes that coherently relax over timescales of 10–20 generations. These slow modes are observed both at the mRNA and the protein levels. As explained before, in a large population of 10^{10} cells, relaxation of the average gene expression profiles over many generations must be the result of *collective dynamics*. The emergence of such long relaxation times, detached from any known time-scale that characterizes intercellular processes, is a manifestation of the simultaneous collective action of numerous cells. Further experiments show that these collective dynamics are most probably not the result of specific intercellular signals; in particular, the insensitivity of the collective expression profiles to the cell density in the chemostat, which can vary over two orders of magnitude without losing the coherency of the expression profiles [40]. Other experiments show that when two separated chemostat adapting populations, each tagged with a different fluorescent marker, are mixed at the end of phase I (figure 3), each subpopulation preserves its own dynamics even in the presence of the other subpopulation in the same mixed chemostat environment for long durations. Each subpopulation showed collective modes of activity that relax on a time-scale of 10–20 generations, for extended durations of ~100 generations (results unpublished). Thus, while not a direct proof, this evidence strengths the conclusion that specific, intercellular molecular signaling is not involved in establishing the collective population dynamics.

1.5.6. Universal distributions of protein content. The protein level in a cell population, even for an essential functional protein (such as *HIS3*), is highly variable among cells, exhibiting a broad, *non-Gaussian* distribution. The shape of the protein-level distribution is *universal*, independent of the biological context and insensitive to changes in central processes affecting gene expression [62]. The universal scaling by a *single population-average variable* implies that the population-average response contains enough information to reproduce the entire population distribution. The protein content dynamics seem to reflect two separate types of processes; a fast universal response reflecting cell individuality which is insensitive to context, riding on slow collective dynamics, a population-average response that is specific to each population. Note that the observed population-average expression dynamics show a multimodal response, with different sets of genes exhibiting different modes [40]. Thus, it does not reflect a 'rigid-body' response like the one observed due to metabolic oscillations [77, 78]. Of note is the fact that different genes, while highly correlated in their response, are not necessarily functionally related. The opposite is also true; genes that are functionally correlated do not necessarily respond coherently [40, 52]. Thus, co-expression and co-functionality represent two separated types of dynamic correlations.

1.5.7. Metastable growth-rate phenotypes. The adapting populations exhibit long-term memories of their dynamical history with slow relaxation (~100 generation time-scales) towards a stable state. While the population can, on average, sustain growth, its individuals exhibit strong phenotypic fluctuations before stabilizing the adapted state. The population dynamics reveal *metastable growth-rate phenotypes* with a broad spectrum of subpopulations; fast-growers, slow-growers and non-growers *coexisting,* without exponential takeover by the fastest growing cells [39]. Our experiments show that over very long durations after the population on the average adapted, the single-cell growth-rate is not stably inherited within a lineage, exhibiting dynamics over a wide range of time-scales as measured by the broad growth-rate distribution within the population. The population-average growth-rate consequently, does not necessarily monotonically increase with time, showing wide plateaus and even decreasing phases. The expression levels of the essential rewired *HIS3* gene also exhibit slow (10–20 generations) collective modes that correspondingly with the other phenotypes (growth-rate and adapted state), relax after a long time of ~100 generations. These dynamics are again irreproducible between populations and there is no direct mapping between the expression levels of the *HIS3* protein and the metabolic response manifested in the cells' growth-rates.

1.6. Implications: main lessons from our yeast experiments

The results summarized above suggest a picture of the living cell that strongly deviates from the conventional picture of molecular cell biology. We briefly describe now the main implications following from our yeast experiments that necessitate revisiting some of the prevailing assumptions and dogmas.

1.6.1. Two types of responses. The distinction between a response to familiar perturbations (encountered before in the evolutionary history of the cell) and adaptation to novel constraints (never before encountered) on regulatory modes is fundamental to our understanding of cell biology. Biological systems often present the capability of response both to the familiar and to novel situations. The idea of novelty in biology characterizes a broad class of '*exploratory*' systems, the prototypes of which are the nervous and immune systems [6, 8]. Although much less recognized as such, the living cell belongs to this class. It exhibits two types of complementary processes with considerable overlapping capabilities: an '*envelope'* of signal-transduction processes that respond within a generation time-scale in a well-characterized, stereotypic manner to external or internal familiar stimuli, as well as deeper processes dealing with novelty, that are *non-specific toward a function.* The latter are able to 'learn and memorize' via exploratory dynamics, having a rich spectrum of behaviors

over a wide range of time-scales. In fact, it could well be that even the response to familiar stresses and challenges contains components of exploratory dynamics which are not well separated from the fast 'hard-wired' response, as in the case for unforeseen challenges and thus are not easily recognized.

1.6.2. Adaptation. The response observed in our experiments, where a large fraction of the population adapts to a severe challenge, marks a dynamical niche available to cells which is completely different both from the physiological response to familiar perturbations, as well as the one described by the conventional genomic evolutionary framework dealing with novelty. The notion of adaptation has different meanings in different contexts. In particular, physiological adaptation is a process in which an organism adjusts its internal conditions or behavior to a change in the environment. However, adaptation in the evolutionary context has to do with an inherited change of phenotype and it is customary to assign it to a permanent change in the genotype. The spectrum of adaptive behaviors of cells to the unforeseen challenge observed in our experiments extends the concept of adaptation to cover the entire range between the two extremes; fast physiological adaptation to external stimuli and slow genetic adaptation through the fixation of mutations. This wider scope of cell adaptation is manifested by its flexibility in response to external cues, the broad-range of timescales involved in the dynamics, the level of stability of the adapted state and its inheritance fidelity. This rich spectrum of behaviors reflects the associations of genotype and phenotype, carried out by physiological processes.

1.6.3. Epigenetics. Genetic determinism stems from the assumption that there are specific genes and specific pathways responsible for a given phenotype. Cell physiology complements genetics by bridging genotype and phenotype via flexible dynamical responses, including a rich array of epigenetic processes. The regulatory nature of the challenge, the abundance of adapted cells and the rapid dynamics in our experiments, make cellular plasticity supported by epigenetic processes a reasonable candidate to mediate the induced adaptation process and its stable inheritance [7, 55, 79–82]. The term cellular plasticity here refers to the dynamic response of cells to the challenge that enables the stabilization of a new phenotype by multiple ways [2]. When the stabilization of a new phenotype does not involve changes in DNA sequence it must involve epigenetic processes. However, our experiments indicate that the notion of epigenetics should be extended beyond the somewhat narrow definition, usually adopted in the literature—*molecular mechanisms leading to heritable changes in gene activity not involving changes in DNA sequences*. While this concept significantly extends the framework of genetics [83], we believe that its scope goes well beyond the collection of structural molecular processes. It is useful to regard epigenetics in the context discussed here, as a concept related to dynamics, and not to structural changes. It stands for the entirety of physiological organization processes over a wide range of time-scales, bridging genotype to phenotype; from the fastest molecular processes to those extending for many generations with different degrees of fidelity [84]. To avoid confusion with the array of molecular epigenetic processes, hereafter we utilize the term *systems-epigenetics* as the entirety of processes involving in the *genotype-to-phenotype* associations.

1.6.4. Gene activity phase-space. The lack of direct associations between metabolism and gene expression reflects the complexity of gene activity phase-space. One should realize that gene expression is by itself merely an '*auxiliary tool*'. Cell functionalities, in particular metabolism, depend only indirectly on the levels of gene expression. The irreproducible gene expression response in our yeast experiments, shows that the dimensionality of phase space of gene activity reflected in the expression levels of mRNA and proteins is extremely large. This observation has a deeper consequence; the organization of the regulatory system is not 'hard-wired' in the genome, but rather reflects the fluidity of gene interactions and their flexibility. On the other hand, the consistency in metabolic response between populations shows that gene-activity phase-space is highly degenerate; different expression profiles support similar metabolism. Thus, the main problem arising from the physics view-point is to understand the converging dynamics in large phase-spaces under internal and external constraints. In that respect, the genome is part of these constraints, serving as 'boundary conditions' of the dynamics, rather than as the determining source of expression profiles.

1.6.5. Levels of organization. An important aim in analyzing a biological phenomenon is to identify the relevant levels of organization. The current reductionist approach attempts to search for the lowest relevant level, preferably down to the molecular processes. Thus, when cells adapt and inheritance is involved, the usual level is that of the DNA—either a change in its sequence or an epigenetic process affecting its structure or protein binding to it. Other intracellular processes might be involved and be inherited, such as RNA molecules, proteins, other small molecules, or organelle structures. At any rate, these are all molecular intracellular processes. For gene expression, the current favorite level of organization is that of genetic networks, involving protein–DNA and protein–protein interactions. Our experiments suggest that focusing on isolated intracellular processes may not be enough. The population level also plays an important role in stabilizing the cellular response. In particular, the fact that each population exhibits its own *idiosyncratic* response means that cells within a population react in a coherent way. If, as argued above, specific intercellular molecular signaling is not involved in establishing the collective population dynamics observed in our experiments, an important conjecture arises: cells growing in a population and sharing the same environment, tend to *synchronize their dynamics*. This property might be enhanced by periods of stress or through an adaptation process. There are two ways in which cells within a population can affect each other without a direct specific molecular signaling. First, cells in a proliferating population are correlated due to transgenerational inheritance of molecules and structures. Second, the medium in which each population grows is dynamically evolved with the population, thus providing a

specific environment to the cells. It could be that the sensitivity of cell metabolism to small changes in the environment is such that cells within a population converge to similar metabolisms compared to those of cells in a different population. This might suggest a general non-linear population dynamics effect, reflecting *'sensitivity to initial conditions'* at the population level. Such dynamics have not been studied in detail until now.

Under the conditions in which the population level plays an important role, the dynamics cannot be understood by analyzing intracellular processes in isolation (e.g. analyzing the dynamics of a genetic network detached from population processes). It suggests that the cellular response reflects a triad of *cell-population-environment* dynamics, with the environment itself serving as a dynamic niche, both affected by and affecting the cells. The emergence of slow modes, relaxing on time-scales of 10–20 generations is another indication of population effects. It is not unique to gene expression, it was also observed for other phenotypes in our experiments, most notably in the fraction of adapted cells in the population which exhibits similar relaxation time-scales [39]. Thus, the stabilization of an adapted phenotype involves a wide range of processes and extended timescales, from the short ones characterizing local molecular processes, through the non-local response of large sets of different intracellular subsystems, up to population processes at the scales of many generations.

1.6.6. Relevant and irrelevant variables. The biological cell presents a complex dynamical system raising a major question: what are the relevant variables, or observables, determining its dynamics? The current tendency in biology is to follow the technological edge—every molecule whose level can be quantified is regarded as a relevant observable. This is particular true for the protein content of a cell. Under the assumption that the cell's phenotype is determined by its protein makeup, this is reasonable as a first-order approximation. Indeed, it may well be that the levels of some specific proteins, especially those that can switch activity above a sharp threshold, do determine the phenotype (e.g. in certain developmental processes). But our results suggest that this is not a generally valid assumption for the majority of proteins. Our experiments rather suggest that the universal distribution of protein content, taken together with the irreproducible population-average multimode expression dynamics, implies that the protein contents of single genes are by themselves not the *relevant variables* to determine the cell phenotype. We have to admit that the profiles of protein contents of genes in a cell, by themselves are not a unique determinant of the cell metabolism or other phenotypes. In that sense a single-gene expression profile is an *irrelevant* variable. Clearly, proteins do play central role in determining the phenotype; their precise individual concentrations nevertheless, seem like irrelevant observables. The vector of the entire set of expressed proteins contains the relevant information affecting the phenotype. However, it determines the phenotype through some yet unknown *coarse-grained* variables, reflecting the overall correlations underlying the multi-gene expression patterns [85, 86]. We are facing a serious situation, since without knowing the relevant variables determining the phenotype, it is impossible to understand the genotype-to-phenotype associations. Identification of the relevant variables requires the development of experimental techniques and a theoretical framework to close this gap of knowledge. This should include a systemic approach, concentrated on identification of the *coarse-grained* observables—correlations and patterns—rather than merely analyzing expression profiles of isolated genes. We currently *do not have a logical framework relating the protein expression levels with the phenotype*.

1.6.7. Population dynamics beyond genetics. The population dynamics observed in our experiments show the coexistence of metastable phenotype states; coexisting subpopulations of non-growers, slow-growers and fast-grower for very long durations. This situation is not consistent with conventional population dynamics models in which the fastest growing cells take-over on times scales determined by growth-rate variation between individuals composing the population; relying on the assumption that growth-rate itself is a more-or-less stable variable within a lineage. Conceptually, a framework describing the dynamics observed in our experiments, requires two types of variables. One describing the instantaneous growth-rate, while the other its stability over time within the lineage. Models of population dynamics in the biological context commonly rely on genetics; [87–90] a framework that does not seem compatible with the above observations. Population genetics is based on several fundamental tenets. The genome determines the *fitness*, representing the rate of reproduction, which in turn is usually treated as a single variable (a single degree of freedom). A clear distinction is made between the source of variability and the process of selection applied by the environment. In a given environment, individuals with higher fitness reproduce faster and are thus selected in that environment. The genome is inherited and the only driving force of variation in asexual reproduction is provided by mutations. These are relatively rare events, random and independent of environment, of physiological processes and of history. This approach practically circumvents the entire problem of *genotype-to-phenotype* associations. Selection, however, works on phenotypes rather than genotypes and the intracellular many-body processes play important role in the population dynamics. In population genetics, changes in growth-rates that depend on rare mutations, occur on long timescales well separated from the reproduction time which is the fastest timescale in the dynamics. This separation of time-scales, practically allows us to ignore the effect of physiological processes on the dynamics. Our yeast experiments show that population genetics, its success in other circumstances notwithstanding, is inadequate for describing the population dynamics revealed by adaptation to an unforeseen challenge [39]. Briefly, the following observations in particular call for revisiting the basic principles underlying population dynamics: (i) the wide phenotype variability observed in isogenic (genetically uniform) cell populations in the same environment; (ii) the involvement of a variety of physiological mechanisms in determining the phenotype and the maintenance of the variants for generations via epigenetic inheritance processes with variable degrees of

fidelity; [91] (iii) the complex connection between genetic and physiological processes manifested in the multiple trajectories to adaptation described above. Given the complex nature of cellular phenotypes, the multiplicity of epigenetic and physiological processes and the degeneracy in this space, it is realistic to assume that population dynamics should be extended beyond genetics to include the interrelations between physiology and genetics. In particular, as elaborated below, this calls for a conceptual framework broader than that of population genetics, one that is based on *systems-epigenetics*.

2. Reflections on the biology of cell-state organization

What is true for E. coli is also true for the elephant.

Jacques Monod, 1954

The first chapter presented a comprehensive view of the phenomenology arising from our experiments on yeast populations adapting to an unforeseen challenge. In this chapter we generalize the discussion to reflect on the biology of cell-state organization via examples from three branches of biological inquiry: evolution, cell differentiation and cancer. It is not our aim here to fully cover these broad areas but rather to use them, through specific examples, to discuss some fundamental concepts in cell biology. Revisiting this broad set of biological phenomena in the context of our lessons from the yeast experiments, allows us to appreciate these lessons on a wider scope and to expand the discussion of the principles underlying cell-state organization. In essence, this discussion sharpens the main features coming out of the yeast experiments, showing that in some cases, insisting on a dogma can lead to wrong interpretations of observed phenomena. From another perspective, with the proper methodology these natural phenomena are amenable to an in-depth inquiry in a laboratory setting, preparing the stage for a theoretical framework.

2.1. Evolution

But as my conclusions have lately been much misrepresented, and it has been stated that I attribute the modification of species exclusively to natural selection, I may be permitted to remark that in the first edition of this work, and subsequently, I placed in a most conspicuous position—namely at the close of the Introduction—the following words: 'I am convinced that natural selection has been the main but not the exclusive means of modification.' This has been of no avail. Great is the power of steady misrepresentation; but the history of science shows that fortunately this power does not long endure.

Darwin, *The Origin of Species*, 6th edition, 1872

2.1.1. Evolution of developmental systems. The marriage of development and evolution as a discipline of inquiry is a recent endeavor. It has been described in numerous publications and has been boosted by the discovery of remarkable universality of developmental processes and the participating molecules across the kingdom of animals, from arthropods to human [3, 13, 14].

Evolution of developmental systems presents a paradox: how the apparent molecular universality in gene regulation leads on the one hand, to the enormous diversity manifested by the broad spectrum of morphology, behavior and physiology, while on the other hand appears to follow a deterministic 'programmed' process in the development of each individual. As noted by others [3], it is easy to phrase this paradox, but the inquiry of developmental systems until now has not provided satisfactory answers, or even a research program, towards resolving it. The main reason is the lack of detailed information about the dynamics of the evolutionary processes and on the intermediates arising in evolution. Gene regulation and developmental dynamics can only be studied on extant species, a minute sample of the potential biodiversity. On top of that, laboratory model systems are not even representatives of this sample; after all, these models were selected because of experimental convenience, so each of them has its own peculiarities. Nevertheless, the great molecular universality is apparent even from this unrepresentative sample. The common knowledge in the area of development is that since all the cells (with a few exceptions) in a multicellular organism have identical genomes, their phenotypic differences arise because of differences in gene regulation. This led to an extrapolation: much of the phenotypic diversity across species was also caused by changes in gene regulation and diversity in gene expression. As mentioned in the introduction, King and Wilson in their well cited paper [11] concluded that developmental evolution involves *changes in gene regulation* rather than merely mutations in coding regions. Their work, as well as earlier speculations [92], specifically identified transcriptional-level regulation as the dominant source of changes in the evolution of developmental systems.

Our point of departure is that in order to understand the duality of diversity and stability in development, a property reminiscent of the tension between flexibility and robustness across biology, we need to inquire about the *potential* of the system to create diversity rather than merely on the details of its realizations. An important concept in that respect is the *'reaction norm'*; the set of phenotypic responses arising due to environmental changes [2]. The variation of reaction norms between organisms might in turn reflect differences in regulatory systems. Indeed, an important insight from the more recent progress made in the area of developmental evolution, is that the range of visible diversity of animal forms observed in nature reflects the diverse range of genetic regulatory systems. It requires the study of regulatory systems in the context of the emergence of *novelty* in evolution. We now look more closely at the universality and diversity in gene regulation and connect it to the emergence of novelty.

The universality underlying the coexistence of diversity and stability in developmental systems is best exemplified in the case of the *chordates* (animals possessing a notochord). Starting from single-cell eggs of diverse size in a wide range of environmental conditions, all the chordates converge

during the initial stages of their development into an intermediate body-plan shape (the *pharyngula*) that looks remarkably similar across species as far as fish, birds and mammals [6]. Beyond a certain time-point in development, however, quick divergence in development results in diverse body-plans, physiologies and behaviors. We now know that the apparent universality at the onset of development is the result of universality in gene regulation, the so called universal 'tool-kit' genes. Of course, the molecular universality goes well beyond the chordates. The emerging picture is that the same set of proteins, through their wide range of expression modes and interactions, results in the apparent diversity of the end-point phenotypes. Superficially, we can say that from the spectrum of regulatory modes of nearly identical molecular building-blocks, emerges a diverse set of species; and not less important, huge variations among individuals within the same species. Note the shift in focus: to understand the emergence of novelty in evolution and the diversity within and between species, rather than studying end-point realizations one needs to understand the potential of regulatory systems—the possible spectrum of regulatory modes under specific constraints and the ability of cell physiology to support them. Metaphorically, this situation resembles the diversity of existing inanimate materials from a universal, restricted ensemble of atoms. Excitations in the form of covalent and supramolecular interactions result in the entire spectrum of potential material forms. Which of those is actually realized in nature is a matter of history, environmental constraints and statistical sampling. Similarly, for the emerging biodiversity in evolution, we would like to know the potential of the underlying universal building blocks (genes) to interact and then understand from that the limited spectrum of observed forms in nature. The shift in focus from realizations to potential calls for a research program emphasizing the principles of dynamics. Rather than mapping structures, listing the molecules and their realized, observed interactions, we should aim to uncover the effects of constraints and boundary conditions on the dynamics.

2.1.2. Gene recruitment. An important and wide-spread process in evolution is that of *gene recruitment,* the linking of existing genes to new processes. It is widely accepted that gene recruitment has played an important role in the diversification of developmental systems, increasing the complexity of gene interactions along the way [3]. Gene recruitment can occur on different levels. An example often discussed in the metabolic-functional domain is that of the crystalline protein lens of the eye. The lens presents contradictory demands. High protein density is required for shaping of the refractive index to redirect light, while at the same time protein aggregation must be prevented since it distorts the imaging. Crystallins is a class of soluble proteins that have this property and thus was recruited in evolution for the purpose of forming eye lenses in diverse animals across species [3, 6]. The crystalins however, initially performed biochemical functions unrelated to vision and were recruited to this function. For example, *e-crystallins* found in birds and crocodiles, serves as an enzyme participating in metabolism, and has the right properties for forming high refracting index changes without aggregations and thus now also participates in forming the eye lens [6]. This is a repeated theme that becomes even more powerful when the recruited gene is a regulatory one [3, 6]. A well known example for universality and versatility in gene recruitment is *Pax6* which regulates important processes in eye formation across species[9] [6]. *Pax6* is an example of how developmental regulatory pathways can change while retaining conserved regulatory genes. This demonstrates the power of modifying regulatory modes. A small change in a regulatory system can reorganize the entire developmental process forming new organs and changing forms and functions. This case also demonstrates another important issue; in spite of being a central regulator, its modes of activity and impact of regulation depends on the context and on the activity of many other genes.

Despite the apparent modularity of biological processes [93], it is rare to find isolated, small subsystems because genes are massively interacting at a level of complexity that cannot be revealed by simply considering single elements. As we further discuss below, this goes much beyond the current trend of extending pathways to networks. Such examples raise a fundamental question: *what are the cellular mechanisms supporting flexibility in regulatory modes as manifested by re-linking existing genes*? If these are indeed major driving forces in evolution, we should be able to understand the implications of changes in regulatory constraints on cell dynamics. This is one of the motivations for our research program on yeast. In the majority of cases such genome rewiring events lead to *unforeseen challenges*. How flexible are the intracellular dynamics that allow to overcome such novel challenges? This is an important issue since the ability of developmental systems to form novel morphologies and functions, absolutely depends on the ability of cells to overcome modifications in regulatory modes. Beyond the selection for specific interactions and processes in evolution, there seems to be a selection for a *type of dynamical system* allowing the transformation of a non-specific and *a priori* imprecisely determined complex set of interacting molecules into organized states. Such a process cannot be simply traced to a molecule or a specific set of interactions but rather should be searched for and understood at the system level. In other words, imagine the following, somewhat simplistic, heuristic picture of evolution: a fixed set of genes (ignoring mutations in coding regions) forms a 'library' containing a combinatorial set of all possible interactions and modes of regulation. Different members of this set represent novel forms in development but only a subset of this library can develop into viable forms and in reality, only a minute part is eventually realized in nature. Our ultimate aim should be to understand the nature of this *library*, its potential to form novel structures and its sensitivity to physical, environmental and physiological constraints. As noted by others, gene recruitment is inherently a saltatory process, a jump in genetic organization, rather than a smooth graded modification of a preexisting structure [3]; it necessarily involves, at least initially, a novel challenge. It is interesting to recall in that respect the widely quoted view of Jacob—evolution is

[9] *Pax6* is a member of a family of transcriptional regulators.

a tinkering process [94], of which the recruitment of genes engaged in new processes is one example. These phenomena have mostly been examined by comparative genomic methods, emphasizing structures rather than dynamics. In only a few cases, the functional implications of gene recruitment have been tested, mostly looking at the level of a single gene. What is badly missing is a systemic view developed with the aid of controlled laboratory experiments. Based on our experience from yeast, we now know that cells are able to resolve arbitrary unforeseen regulatory challenges in a rather efficient way. The evolutionary implications of this potential have never been seriously studied.

The best demonstration for the systemic nature of gene recruitment in development is the widely discussed, universal set of molecules underlying its early stages—the *HOX* genes. Outside of duplications of the *HOX* clusters between species like fly and mouse, these transcription factors can be exchanged between species without losing their functionality [95, 96]. However, perturbing their operation at critical points in time, dramatically affects the course of development. In the old view, *HOX* genes were considered to be sufficient determinants of the developmental processes in which they are involved. This view is not compatible with the accumulating data showing that much of the specificity of the action of these factors actually resides in their capacity for distinct protein–protein interactions rather than specific DNA-sequence binding. Remarkably enough, *in-vitro* experiments with isolated *HOX* genes showed a complete *lack of specificity* in the DNA binding [97]. The unavoidable conclusion is that the functionality of the *HOX* proteins is context dependent; they serve as modifiers of transcriptional specificity of the other proteins interacting with them. In other words, they serve as *co-factors* of a particular developmental process rather than drivers of it. But this presents a serious question: if *HOX* gene functionality depends on context and interactions with other proteins, what actually determines this context? This is a *closure* problem; additional molecular processes do not necessarily lead to the stabilization of the phenotype, as the new processes require further processes for their own stabilization. Our understanding of the dynamics of developmental processes is lacking in this respect. Unavoidably, it suggests that development should be discussed as some form of a dynamical *organization* process. This in turn, cannot be understood by tracing the separate activity of each molecule without a systemic view. In this sense the biological cell is *irreducible*.

2.1.3. Degeneracy. There are ample examples for apparent redundancy in genetic systems in which an existing copy backups up a defective component [3]. Redundancy is usually thought of as an essential ingredient in error correction. von Newman, in his famous treatment of error-correction mechanisms in engineering and computation, discussed redundancy as a possible mechanism that allows the construction of *a reliable organism from unreliable elements* [98].[10] There is however, a fundamental distinction between redundancy and *degeneracy* [99]. While the former means mere duplication of components the latter describes the ability of elements that are structurally different and belong to different modules to replace the original components, supporting the same function and phenotype. In the context of genetic networks, degeneracy means that components operate outside of their normal domain due to dynamic reorganization of the system. While redundancy is a common approach in manmade machines, biological systems present wide-spread *degeneracy*. This characteristic is apparent in physiology at the level of systems such as the brain. A lesion or disability in part of the brain can cause a malfunction that can sometimes be repaired by the plasticity of the system, operating other parts to replace the damaged ones. It is less apparent that such degeneracy also exists at the cellular level. A specific, remarkable example is the *MAP Kinase* (mitogen-activated protein kinase) system which has been adopted for numerous developmental and physiological functions in diverse organisms from yeast to mammals and shows great versatility [3, 100]. It consists of a series of kinases (enzymes catalyzing the phosphorylation of other proteins) whose final member (MAPK) phosphorylates and activates one or more transcription factors. In yeast for example, combinations of the same genes participate in *MAPK* modules with diverse functionalities operated by specific stimuli—mating response to a pheromone, sporulation resulting from carbon and nitrogen starvation, osmolyte synthesis due to high osmolarity etc [100]. The activation of each *MAPK* pathway by a specific stimulus, leads to a unique response, raising a serious question: how do the shared components preserve their specific functionalities in two different pathways [101]? Here, we wish to elaborate on the other remarkable facet of degeneracy, demonstrated by these signal transduction pathways. For example, mating pheromone molecules in yeast activate the *MAPK* (the last kinase in the chain) Fus3p[11] while starvation activates Kss1p, through a chain of three common kinases, Ste20p, Ste11p and Ste7p, resulting in mating or filamentation phenotypes, respectively. Now, here is the point. Separately, deletion of either *Fus3* or *Kss1* does not affect the mating phenotype, while deletion of both eliminates it [101, 102]. Clearly, these pathways are degenerate since Fus3p and Kss1p do not perform the same biochemical function in wild-type cells, but the latter replaces the former only when it is not present. Given the apparent strong cross-talk between the components of the system on the one hand but the high specificity in response on the other hand, this is a remarkable outcome, but it is far from unique to this example. Degeneracy seems to be an essential property that enables cells to surmount unforeseen challenges and to develop novel forms, and it might be a natural property of genetic regulatory networks [103]. In the case of our yeast experiments, we observed degeneracy resulting in multiple alternative trajectories to adaptation and in restoring functionality when mutations in critical regulatory modules (*GAL80* and *GAL4* binding sites) emerged [48].

2.1.4. Canalization. Canalization, a term coined by Waddington [104, 105] and elaborated also by Schmalhausen [106], is

[10] Notes based on a lecture by von Newmann in 1952.

[11] The p stands for protein.

defined as the reduced sensitivity of an organism to genetic or environmental perturbations [1, 2]. The stability and almost deterministic characteristic of developmental processes in face of perturbations, was the prime motivation for defining this concept. It is usually discussed in connection with another term; genetic assimilation, which is connected to the selection of heritable traits that emerge following a perturbation in a developmental system (and in a broader sense also in behavioral systems—the so called *Baldwin effect*). Waddington, in his experiments on fly embryos exposed to heat-shock or ether, showed the emergence of phenotypes that could be selected and following inbreeding could even be fixated and thus be expressed in the absence of the external stimulus [1]. In one type of experiments, Waddington exposed a laboratory population of wild-type *Drosophila* pupae to heat-shock, observing that some of the adults developed an aberrant phenotype with a gap in their posterior crossveins [104]. When the aberrant flies were inbred, after some generations of selection, the phenotype started to appear in adult flies without the exposure to the heat-shock. Inbreeding those flies led eventually to 100% of the population exhibiting the phenotype without being exposed to heat-shock. In another type of experiments, fly embryos exposed to ether, developed a phenotype with four wings instead of two [107]. Again, when individuals showing this phenotype were selected and inbred, after 20 generations the selected flies start producing the aberrant phenotype even without being exposed to ether. The conclusion is that in these experiments, selection on existing genetic variations can stabilize the environmentally induced phenotype. Note that in these experiments the phenotype was artificially selected and did not confer advantage to the organism under the experimental conditions, i.e. it was non-adaptive. Waddington's work gained renewed interest in recent years due to its potentially important implications for evolutionary processes. The apparent contradictory characteristics of flexibility and stability, the hallmark of biological systems mentioned before, are once again a major issue. Phenotypic plasticity, the sensitivity of the phenotype produced by a particular genotype to variation in the environment, which is thought to play an important role in the evolution of developmental systems [1, 2], is in some sense complementary to environmental canalization. Together, environmental canalization and phenotypic plasticity describe the range of available responses of the phenotype to perturbations. The separation between genetic and environmental canalization is somewhat artificial due to the spectrum of epigenetic and physiological processes connecting the triad of *environment-physiology-genetics* in a dynamic way. Canalization is also connected in the current literature to other phenomena; *buffering* of hidden genetic variability (so called 'cryptic' variations), *epistatic* interactions, *pleiotropy* and *redundancy*. Sometimes, all these phenomena are gathered together under the title of robustness (see [108]). For example, it is assumed that canalization allows the buffering of hidden ('cryptic') genetic variations in a population of organisms. The exposure of these variations is manifested in laboratory experiments via *de-canalization* due to mutations in key genes, epigenetic processes affecting key genes or environmental perturbations. Epistatic interactions, the non-additive dependence of the effect of a gene on other genes, are thought to play an important role in canalization. Similarly, pleiotropy, the multiple effects of a single gene on two or more traits, underlies some of the important aspects of canalization [3]. Redundancy and degeneracy obviously can also contribute to canalization. Finally, the concept of canalization is connected to another commonly used concept, that of a *landscape*; a geometrical description of the dependence of a trait on underlying variables. Next, we discuss these issues in connection with the lessons from our yeast experiments.

Canalization as a systemic property. Recent research has attempted to focus on molecular mechanisms that lead to phenomena resembling canalization. However, canalization is a systemic property; a dynamical process of the cell in the context of development, rather than a molecule. This fact was well understood by Waddington, as is reflected in his insistence on *strategic principles* and described in his book: *The strategy of the genes* [109]. Thus, despite the potentially important implications of canalization on cell-state stabilization and phenotype determination in development and evolutionary processes, the concept remained obscure for several reasons. The attempt to tightly connect canalization with cryptic variations is problematic since it was demonstrated that such hidden variations exist without canalization of the wild-type (the latter is by itself not a well-defined term) [110–112]. Thus, experiments showing hidden variations do not provide evidence for canalization. Moreover, most experiments do not touch on the adaptive aspects of canalization and thus lack the essential contact between canalization in the developmental context and evolutionary processes such as selection. Modeling provides a picture of canalization as an intrinsic genetic property, a manifestation of the complexity of genetic networks and biochemical pathways, rather than an adaptive dynamic phenomenon. Therefore, although it is possible that organisms have evolved mechanisms making them robust against perturbations (genetic, physiological or environmental; [108]), there is a possibility that canalization is an *emergent phenomenon*—a non-adaptive essential feature of the complex dynamics leading from genotype to phenotype. Interestingly, degeneracy might as well be such an emergent phenomenon, as it is an inevitable property of certain types of complex dynamical systems. Note that regarding canalization as an adaptive phenomenon, leads to the conclusion that organisms develop mechanisms that might decrease their capacity to evolve. This seems to contradict the widely observed phenotypic plasticity in developmental systems, unless canalization allows the accumulation of hidden mutations which is not necessarily always the case [1, 2, 110]. It is also not obvious how canalization should be measured in laboratory experiments. Current work on the subject has mostly hypothesized the connection to canalization and speculated about its evolutionary importance, rather than convincingly demonstrating it. As for genetic assimilation, most of the experiments do not prove that the observed, emerged variation in phenotype is in fact associated with a genetic variation. Experimental procedures such as back-crossing of an emergent variant with the original population not exposed to the perturbation that show

reduction in variation cannot serve as proof for their genetic origin. Clearly, we are currently familiar with ample examples of non-genetic inheritance that can lead to such effects. Moreover, as demonstrated later, the origin of heritable variations can be due to epigenetic effects such as changes in the chromatin structure.

The many facets of canalization. The concept that the wild-type phenotype is less variable than its genetic mutants, the basis of Waddington's view of canalization, should first be revisited. Recent excitement from the canalization phenomenon came from experiments identifying specific genes whose mutation or protein inhibition lead to the emergence of a wide spectrum of phenotypes, apparently demonstrating de-canalization. A prime example in this context, is the chaperone *Hsp90* which seems to affect many other proteins and even chromosomal structures and thus its perturbation has a widespread effect, assumed to expose hidden genetic variation [113–120]. It is convenient to look for a molecule (a gene variant) responsible for such variation, but further work demonstrated that the nature of the perturbation underlying these 'cryptic' variations can be manifold; potentially involving any process in the intracellular hierarchy, including gene expression, RNA stability, protein structure and protein folding, metabolism and other physiological processes. Researchers usually assume that there must be a specific genetic variant underlying every trait that can be selected and enriched in a population. This is not necessarily true; the same genome can lead to a wide range of phenotypes. For example, experiments exploring the *reaction norm*, showed that the spectrum of phenotypes arising from identical genomes might be as wide or even wider than that arising from genomic variants [121]. Thus, without a direct proof of a genetic variation as the cause, one simply does not know the source of the phenotypic variation emerging following a perturbation. Note that Waddington himself, while showing the involvement of multiple genetic loci in the assimilation process, did not prove a genetic origin (i.e. changes in DNA sequences) of the phenomena observed in his experiments, he simply assumed it. This was very reasonable, given the knowledge of biology of his time, but not of today. In a recent review on the subject [116], the authors claimed: 'Together, the divergence of phenotypes, different heritability and dissimilar chromosomal contributions prove that expression of the trait depended upon multiple and existing polymorphism.' The lessons from our yeast experiments show that one needs to be cautious with such statements. Diverging phenotypes, different heritability and dissimilar chromosomal contributions can also arise in perfectly clonal (genetically homogeneous) populations. This is not to say that pre-existing genetic variation cannot be the source of a phenomenon; we simply note that there are other possibilities, extending the scope of canalization and de-canalization beyond genetic polymorphism. A remarkable example of de-canalization and heritability without selection, was recently demonstrated in flies that were rewired to regulate a foreign antibiotic resistance gene under several of their native developmental promoters [122, 123]. These experiments utilize genome-rewiring to study, for the first time, the response of multicellular organisms to unforeseen challenges during development. Antibiotics applied at the larva stage led to an array of fly phenotypes with variable degrees of heritability (namely, the phenotypes persist for generations without antibiotics applied beyond the first generation). In some cases, the fraction of flies preserving the modified phenotype was 100%, showing that no selection was involved. This example demonstrates again that canalization and de-canalization are concepts much wider than genetic polymorphism and selection of pre-existing variations. While some traits become more variable following a perturbation and thus, presumably de-canalized, some other traits remain intact. In particular, while some experiments show variations in morphology they fail to show, at the same time, similar variations in gene expression [110]. This raises a question we asked before at the level of the cell; *what are the relevant degrees of freedom underlying a phenotype?* If gene-expression profiles are robust under a perturbation while morphology varies, *what does actually determine a specific morphology?* The opposite effect is also puzzling. In some cases, morphology seems highly canalized while gene-expression is widely variable among individuals in a population [110]. Thus, the interpretation of observed variations following a gene perturbation in laboratory experiments seems paradoxical. As if buffering of the perturbation at the level of the phenotype may nevertheless result in higher variations at another level, for instance that of gene expression. We come back to this important question in a wider context below. Viewed from the lessons learned from our yeast experiments, we can formulate a more consistent view of emerging variations following genetic or environmental perturbations. A perturbation that presents a challenge to the cell population (e.g. cells composing the embryo during development), such as a perturbation in the activity of *Hsp90,* requires reorganization of gene regulation to ensure viability, similar to the response of rewired yeast to the unforeseen regulatory challenge. Enhanced population variation is then an inevitable consequence of an *exploratory process* underlying the ability of an organism to overcome a challenge. The issue of adaptive stabilization of the phenotype and its robustness is the result of this *exploration-exploitation* dynamical process; canalization is then a property resulting from the dynamics in the *phase-space of relevant variables*. The nature of this phase-space is indeed one of the important open aspects of genotype-to-phenotype associations.

A related phenomenon is manifested in gene deletion experiments that many times show no effect of a deleted gene on the observed phenotypes, in the absence of redundancy or degeneracy. For example, it has been demonstrated that the majority of genes in the budding yeast are non-essential; ~80% of them can be deleted separately without notable effect (under laboratory conditions in which the essential nutrients are externally supplied—rich medium) [124, 125]. Connected to this observation is the fact that out of the ~6,000 genes in the budding yeast only about 2,000 are known to have well characterized functions (and usually are multi-functional) all the rest are still open to interpretation (and this is in yeast! a 'simple' eukaryotic organism) [126]. Other eukaryotes show a similar phenomenon. Most of the genes do not show any

obvious mutant phenotypes. This widespread phenomenon seems to contradict the commonly accepted view of neutral evolution in which an organism should accumulate genetic mutations if not under selection pressure, including mutations leading to premature stop codons (i.e. translation cannot go through the entire coding sequence) leaving them as remnants in genomes as pseudo-genes. The common argument explaining the embarrassing lack of deleterious mutations in the absence of an essential functionality is that these genes do have essential functions in environments not tested in the laboratory or in response to stresses not usually encountered there. This appears to be a poor explanation, especially in the case of microorganisms like yeast, which have been propagated in laboratories for many thousands of generations without losing systems that seem nonessential under these conditions. For example, the GAL system in the yeast *S. cerevisiae* was proved nonessential in a glucose medium [30, 53] which is the common sugar in the majority of laboratory experiments. Yet, the GAL system in yeast remained intact in all commonly used strains even after exceedingly long propagation periods. Another approach claims, based on competition experiments between mutants and wild-type strains, that the majority of genes indeed are nonessential but do make small contributions to the efficiency of routine processes. This might be true, but seems like a circular argument. The high connectivity of genetic networks and their degenerate nature inevitably lead any gene to contribute some effect due to its interactions with the rest of the system. As clearly demonstrated in our yeast experiments, a mutation in a gene accompanied by a phenotypic change does not necessarily testify to the functionality of this gene. It should be clear that in a highly-connected complex system, any change can lead to reorganization of the interactions between its elements under these conditions. The observed macroscopic changes cannot be easily traced down to a microscopic cause. This connectivity is also manifested in the sensitivity of perturbations to the genetic background and the fact that in examined quantitative traits (QTLs; a quantitative phenotypic characteristic that is associated with multiple genes), the number of genes identified to affect the phenotype is usually very high.[12] In all the experiments using genetic or environmental perturbations, including the ones mentioned above in the context of canalization, there is a general sensitivity to the genetic background (i.e. to the unperturbed genome). We should add to this also sensitivity to the history of the population. In particular, adaptation of the type observed in our experiments leads to reorganization of gene regulation making the response of the system highly sensitive to the precise history of the population.

The landscape picture. The concept of canalization was pictured by Waddington in terms of a *landscape*. As a metaphor, he envisioned a marble rolling down a 3D surface containing hills and valleys; the height of the surface is a measure of a phenotypic trait while the other axes determine the activity of genes. In Waddington picture, the landscape describes organ formation during embryonic development. In a more modern view, the curvature of the surface determines the degree of canalization in which the loci with lowest slopes represent the maximal degree of canalization. Waddington envisioned that the underlying genes 'pull' on the surface thus setting its landscape. Clearly, non-additive interactions (e.g. due to epistasis or genotype-environment interactions) are responsible for curving the landscape. Additive interactions simply result in a flat surface. From a developmental system perspective, the landscape determines the stability of the process. Accordingly, a more modern modification of Waddington's original description utilizes this picture to determine cell differentiation during development. Starting from a stem cell at the top, the 'marble' rolls downhill to a valley representing a differentiated end-point (e.g. a skin cell, a muscle, a neuron etc; see below the section on cell differentiation). As often happens in science, an initially innocent metaphorical picture later gets its own tangibility and is adopted as a realistic description of the process. However, although an inspiring picture at the time, Waddington's metaphor is a far-fetched caricature of reality. The cell is an integrated dynamical system. The landscape itself is dynamically organized during development rather than fixed by the genes providing a fixed substrate for the 'rolling' developmental process (the marble). The rolling of the marble should simultaneously determine the landscape itself, through a *dynamic organization* process. There is no physical force shaping the kinematics of the system in an *a priori* determined landscape. The more recent attempts to view the landscape in a more physically realistic framework, as a set of attractors in a dynamical system [127–131] miss important features of the cell dynamics as well. In particular, its high-dimensional characteristics and the lack of a metric (measure of distance; see an example in [132–134] for the genotype-to-phenotype associations in a simple setting of a folding RNA, in which neighbor configurations are not necessarily at small distances from each other). Note also that the underlying hidden variables 'pulling' on the surface and determining the landscape are actually not known. These cannot actually be single genes—or even the collection of many isolated genes. They must be related to correlated variables reflecting protein content and affecting the phenotype. What is missing is a procedure to *coarse-grain* the microscopic degrees of freedom into *relevant* variables determining the dynamics.

2.1.5. Population dynamics. Population dynamics is a broad area of research. In the context of evolution it is usually regarded as synonymous to population genetics, based solely on the neo-Darwinian framework of evolution. Notwithstanding its success in applying statistical approaches to population phenomena, founding the dynamics solely on genetic principles narrows the scope of this field. To make contact with our yeast experiments, we focus here on the *population dynamics of asexual organisms*, excluding important aspects such as sex, recombination and separation between soma and germ

[12] This is manifested to the extreme, in attempts to search for 'missing heritability' in highly heritable traits such as height in humans (see [188]). For such traits, the number of loci found to be highly correlated with the variation is measured in tens. Moreover, different groups find different sets of genes for the same trait, meaning that the identified sets consist of only a partial list and in fact the phenotype is affected by the majority of genes in the system in varied degrees.

lines. In some circles, even mentioning evolution beyond genetics sounds like heresy, a smell of 'dangerous Lamarckism' [82, 135]. However, times have changed and there is now ample evidence that calls for extending the discussion of populations beyond the narrow prism of genetics. It should naturally involve the framework of *epigenetics* which is currently a very active area of research. However, the term *epigenetics* in this context carries the broadest possible meaning. In the evolutionary context, the subfield of *epigenetic inheritance* (non-genetic inheritance) sets the framework for transgenerational inheritance beyond genetics [7, 82], but has concentrated mostly on the processes and molecular mechanisms supporting inheritance without the involvement of changes in DNA sequences. As mentioned before, we should extend the framework of individual molecular mechanisms to the systemic level. The relationship between systems-epigenetics and genetics will be examined later from other perspectives of cell differentiation and cancer. Here, the focus on asexual organisms, allows us to stay away from the emotional, fierce tension over the role of epigentics in evolution compared to that of genetics, and relax the (usually negative due to misinterpretation) prejudice attached to the term *Lamarckism* due to the assumed separation of the soma and germ lines.

Population Genetics is based on several fundamental assumptions that were outlined before. Noteworthy are that the genome determines the *fitness*, which for asexual organisms stands for the rate of reproduction, and time-scale separation between the source of variability (rare and random mutations) and the process of selection applied by the environment. Recent developments require revisiting the basic assumptions underlying population genetics. In particular, epigenetic and physiological processes can be induced by the environment; their varying degree of heritability and their interplay with genetic processes make this a nontrivial point of departure from population dynamics based solely on genetics [7, 80, 136].

In the following, the discussion on population dynamics is based on several principles: (a) variability in the population necessarily involves *epigenetic and physiological processes* rather than only genetic ones [7, 82]. (b) the *cell-population-environment* poses a dynamical triad which should be understood by the collective dynamics of its components and their symbiotic inter-relations [137]. (c) The spectrum of cell phenotypes within the population is *continuous* and spans a broad range, rather than being discrete and composed of a small number of well-defined phenotypes. These three tenets have far-reaching implications for population dynamics. This view requires a change in perspective due to the deep differences between genetic and physiological processes. First, while genetic mutations in cells occur on long timescales much longer than typical cellular physiological times, systems-epigenetic processes occur on a wide range of timescales overlapping the physiological ones [2, 84]. Second, contrary to the case of random mutations, which are blind to the environmental demands, systems-epigenetic processes are part of the intracellular dynamic repertoire, can respond to environmental cues and are coupled to other intracellular processes. The environment thus plays an outstanding role not only as a selection agent but as a *co-evolving* player strongly interacting with the population in a two-way interaction, thus enabling the population to construct a dynamic history and context-dependent niche.

Systems-epigenetics—a bridge from genotype to phenotype. As discussed before, there is a gap between the concept of adaptation in physiology versus the one in genetics. We argue for, and existing evidence support, the existence of a wide spectrum of adaptation processes which can be classified according to the kinetics of response, the susceptibility to different types of perturbations and the level of stability. Populations of asexual organisms display the entire spectrum of adaptation phenomena, but for historical reasons reversible adaptation is considered to be merely a physiological process or phenotypic variation while 'true' adaptation is a term reserved for the fixation of mutations. However, the relationship between the fixation of a mutation and the process of adaptation (the latter necessarily a concept related to the space of phenotypes rather than that of genotypes) is a question of *causation*. Only in rare cases, can the appearance of a mutation be demonstrated to be the cause of adaptation. In the majority of cases only correlations are detected between the existence of a mutation and the population adaptive phenotype. But what does the term adaptation really mean in the context of evolution? In genetics, adaptation is tightly connected to the notion of *fitness*, by itself, a vague term whose meaning is context dependent. The search for a single, scalar variable characterizing the population can have misleading results, in particular when fitness is interpreted as the actual degree of suitability of the population to the environment. In fact, the population-average reproduction rate, a relevant scalar (dynamic) variable characterizing the population could, in principle, be identified with the term fitness. However, it is not necessarily connected to the question of how fit the population is to the environment. Following Ariew and Lewontin [138], we conclude the following: First, the Darwinian metaphor of organisms 'fitting' into an environmental niche should not be confused with the reproductive fitness. Second, we should abandon the attempt to make a general, quantitative dynamical theory of evolution based solely on a single scalar parameter—the population-average reproductive fitness. Finally, if we desire to differentiate between types within an evolving population and predict their fates, we need in each case to decide on the nature of the variable that characterizes the different types, not necessarily their instantaneous growth-rates. Our yeast experiments exemplify these issues by showing the coexistence of *metastable* growth-rate phenotypes over very long durations with a non-monotonic population-average growth-rate which can hardly be explained within a framework that assumes it to be the sole measure of the reproduction ability of the population. For example, a fast-reproducing cell may produce a lineage of slow growers due to rapid memory-loss of the adapted state, while a slow-grower can stabilize a better-growing lineage due to long memory; thus, the instantaneous growth-rate is not necessarily the sole parameter describing the population dynamics. Note, that adaptation as manifested in the yeast experiments is not simply an adjustment of the population to

a given static environment. It is rather a dynamical process involving *niche construction* [1, 139, 140]. This is the general situation in evolution, as further discussed below.

The recognition that adaptation should be extended beyond genetics, necessarily brings in the concept of *epigenetics* and in particular *epigenetic inheritance* processes. The issue from the evolutionary perspective indeed relies on *transgenerational inheritance,* in the form of transmission of phenotypic variations to subsequent generations that do not stem from variations in DNA sequence. Unfortunately, the majority of the current literature on epigentics extends genetic determinism by adding new molecular processes; additional layers of regulation such as DNA methylation, histone modifications, small RNA processes and others [80]. In that respect, molecular epigenetics is a direct extension of genetics, providing processes that are more flexible in responding to environmental demands. Adding more layers of molecular responses, however, cannot provide a satisfactory conceptual and explanatory framework for the emergence of phenotypes from genotypes and their stabilization without a systemic integrating concept. The *fluidic* nature of the genome and epigenome and their overlapping temporal dynamics through intracellular physiology, suggests that they form a continuum of processes with intimate and tight relationship between physiology and genetics. For example, two important molecular epigenetic processes, DNA methylation and histone modifications, while certainly playing some role in stabilizing a cell state (discussed in the chapter on cell differentiation below), do not encompass the wide-scope nature of systems-epigenetics. DNA methylation is not universal. Many organisms exhibiting systems-epigenetic processes, yeast, flies and worms among others, simply do not have any known machinery for stabilizing DNA methylation patterns. As for histone modifications, while chromatin structure is certainly important in stabilizing the phenotype, there are no clear indications for the existence of a process to stably propagate these modifications for generations of cell division as they rapidly turnover due to the lability of the nucleosomes [141, 142]. Moreover, even if such a mechanism exists, it cannot carry out this task alone and must depend on other molecular processes.

Another type of epigenetic mechanism that can support inheritance is that of autocatalytic regulatory circuits. However, these processes present their own problems as sources of heritability. They are susceptible to fluctuations and therefore rely on unrealistically sharp thresholds, buffered against protein-level variations. Moreover, autocatalytic circuits have only been demonstrated for very simple and small genetic systems, such as the *Lambda* phage.[13] It is an open question whether this simplified picture, can be carried over to the highly complex and widely interacting networks of hundreds and thousands of genes in eukaryotic cells. In fact, none of the molecular processes suggested as epigenetic mechanisms—DNA methylation, histone modifications, small RNA or protein-DNA binding and autocatalytic circuits—can by themselves stabilize a phenotype and hold *transgenerational inheritance*, without assistance from other components of cellular physiology. There are of course interactions between the various different epigenetic molecular processes. Nevertheless, there is currently no framework demonstrating a mechanism that ensures the convergence of these processes into a viable stable phenotype. This is the *closure* problem mentioned before. Moreover, there is a difference between propagating the process and initiating it. The initiation and the stabilization mechanisms of DNA methylation, histone modifications and profiles of small RNA molecules are not known. This means, that assuming these molecular processes as the dominant epigenetic inheritance mechanisms, simply sweeps the problem under the rug instead of solving it. At this stage, we have to admit that making an ever growing list of molecular processes, does not lead to progress towards resolving the fundamental problem of genotype-to-phenotype associations introduced at the onset; we simply do not understand epigenetics in a systemic way yet in the context of cell-state organization.

Irrespective of the precise mechanism, there is ample empirical evidence for the existence of a variety of inheritance mechanisms beyond genetics that operate alongside traditional Mendelian inheritance [7, 82, 143]. This is well accepted, but the debate becomes emotional over the *inheritance of acquired traits*, identified as *Lamarckism*, and the overall stability of these inheritance processes on evolutionary timescales [143–146]. This debate, the entire subject of non-genetic inheritance, and the historical misconception of Lamarck's ideas are well covered in the literature, including recent writings [82, 147], and there is no need to repeat them here. Waddington, who coined the term *epigenetics*, was concerned with causal processes by which genetic systems interact with the environment to materialize the phenotype, its plasticity and its ability to ensure robust developmental processes [109, 148]. Nanney, distinguished between genetic and epigenetic processes causing changes in the cell phenotype [149]. Historically therefore, heritability was not the main concern in the framework of epigenetics. This issue has only come into focus more recently [150]. Epigenetics in the systemic broad sense is therefore the set of processes that lead to change in gene activity and organized stable cell-states that persist under changing conditions. In particular, in the absence of the originating inducing conditions. It is thus a bridge from genotype-to-phenotype, precisely as Waddington envisioned it [148]. Although, originally discussed in the context of development, its scope is much wider and includes 'developmental' processes in the broad sense of the stabilization of multi-phenotype states both for prokaryotes, eukaryotes as well as unicellular and multicellular organisms. At another level, epigenetics is also the organizing principle between group of cells, tissues and even between the whole organism and its environment.

A comprehensive review of some currently known examples of epigenetic inheritance, both in mitosis and meiosis (therefore transmitted through the germline) can be found in [7, 83, 151], showing that this phenomenon is widespread. A more recent example was shown in experiments on rewired flies mentioned before [122]. This work demonstrates the multiple paths of epigenetic inheritance within the same organism and broadened the scope of epigenetics to encompass

[13] A virus parasite in bacteria.

symbiosis between host and parasites; between the developing fly and its bacteria microflora [152, 153]. While in some known epigenetic inheritance processes the adaptive significance is not known, in many others it is clear that epigenetic variants do participate in the adaptation process. In our yeast experiments, for example, genuine adaptation is achieved only when the adapted phenotype (ability to grow in glucose) is stabilized within lineages. As discussed below, from the population perspective this means that the dynamics are affected by two types of variables; one determining metabolism (manifested in growth-rate) and another 'memory'—i.e. encompassing the stability of the phenotype in the proliferating lineage through time. The open issue is the significance of these processes in the context of evolution. However, systems-epigenetics and epigenetic inheritance processes must play a crucial role in stabilizing adaptive phenotypes, even if eventually genetic changes must step-in to ensure long-term stability. In that respect, the attempts to distinguish between genetic and epigenetic processes are less significant, as long as we accept the need to broaden the scope of genetics beyond the neo-Darwinian dogma, by including concepts such as: sensitivity to environmental conditions, inducible processes, context and history-dependence and possibly others. There is ample experimental evidence for expanding this view.

In the context of developmental plasticity [2], it was already suggested that genes are *followers not leaders* in adaptive evolution (similar to the case of cancer [154], see below). Our current understanding of biology gives an opportunity to broaden the scope of neo-Darwinism. This is hardly new (see [155] and the other books in the series: Towards a Theoretical Biology edited by Waddington for discussions of this point already more than 50 years ago). The crosstalk between physiology, epigenetics and genetics is such that we can find examples for any possible order of events, or alternative dominance of processes, beyond the conventional order starting from a mutation. Our yeast experiments provide another relevant example; there is no simple way to explain the adaptation observed in our experiments by the traditional assumption that, mutation comes first, and phenotype follows.

Experimental evolution. Experimental evolution has attracted attention due to the development of techniques allowing the growth and measurement of cell populations over extended timescales. In these experiments, populations of organisms are maintained in supposedly controlled environments over many generations, attempting to correlate changes in phenotypes with those in genotypes. Two types of culture techniques have been utilized for evolution experiments with microorganisms which until now serve as the best opportunity to observe evolution in action; continuous culture (chemostat) and serial dilution of batch cultures, each with its own advantages and limitations.[14] The experimental methodologies are commonly based on the foundations of the neo-Darwinian picture and analyzed in light of the tenets of population genetics described above. Thus, the canonical experiments in this field concentrate on the fixation of specific mutations, assumed to determine the fitness of the population, while intermediate phenotypes emerging in the dynamics are usually overlooked. The focus is mainly on the endpoint realizations of the evolutionary process. The strong adherence to the mutation-selection paradigm has far-reaching implications that limit our ability to fully appreciate and understand the complexity of genotype-to-phenotype associations.

Fluctuation analysis of mutations. Historically, the breakthrough following the seminal work of Luria and Delbruck (LD) [29] marked a change in direction that still leads the thinking in the field despite the enormous knowledge gained in biology since then. The original work of LD aimed to understand the origin of variations, assuming that they solely result from mutations. In their experiments, LD analyzed the variance of survivors (fluctuation analysis) when parallel propagating populations of bacteria were switched at the same time to lethal conditions (e.g. exposure to phages—viruses prey on bacteria). The count of survivors from each population showed a strong deviation from a Poisson distribution (variance equals the mean); demonstrating that mutations emerged randomly at independent time points in the parallel cultures and not upon the exposure to the lethal environment (which due to the stochastic nature of the process would be expected to result in a Poisson distribution of survivors). The interpretation is that populations with early emerging rescuing mutations develop larger counts, due to the extended time in the favorite conditions, allowing the buildup of larger mutated lineages compared to those having them later, closer to the exposure to the lethal conditions. However, such a fluctuation analysis cannot distinguish between heritable changes occurring due to mutations or any other epigenetic process [156]. Thus, LD did not actually prove the emergence of mutations or even the relevance of mutations to the emerged variation in the fraction of survivors. They simply assumed that mutations are the only possible sources of variations. This assumption might be reasonable under their lethal selection conditions, but could hardly be generalized to other conditions of weaker selection. In the context of our yeast experiment, for example, the question of whether the fluctuation analysis can even be performed under general conditions has a simple answer, No! If the environmental conditions are non-lethal but only stressful for the majority of the population, the basic assumptions of this experiment are violated. This was recognized long ago, motivating attempts to identify extensions of the random mutations picture [157–159]. Specifically, our yeast experiments clearly demonstrate that adaptation can occur without mutations and that the latter, when appear, can be induced by the exposure to challenging conditions [49], limiting the generality of the LD approach. Nevertheless, it is insightful to analyze the LD framework in light of our current understanding of cell biology. (i) Each cell in the population was assumed to have an equal and constant probability per unit time to undergo a change (a mutation under their assumptions). It is now known that even for the case of mutations, the rate is dynamic and

[14] Isolation and propagation of colonies on plates is another useful technique. However, the information that can be extracted on the dynamics is much more limited than the one from batch or chemostat. On the other hand, plate cultures are indispensible for getting a snapshot picture of the population structure.

can be modified over time, specifically depending on environmental conditions. Moreover, LD assumed that the probability for change in an individual cell was directly related to its replication and therefore to its growth-rate. This limits the applicability of this type of experiments to *spontaneous* variations. Any induced change, either by physiological or chemo-physical sources would cause probabilities that might deviate from uniformity. Also, the assumption that cells either die or proliferate, with no intermediate states contributing to the variations, limits the scope of this experiment. Moreover, the possibility of subpopulations and transient effects, causing adaptation to be somewhat reversible, undermine the analysis of the fluctuation experiment. (ii) The fluctuation experiment is valid under the assumption that the growth of both variant and wild-type cells is exponential with equal rates (or at least with a constant ratio between rates). When the growth-rate itself becomes a dynamical variable, as demonstrated in our yeast experiments, this assumption is invalid. (iii) A time lag between the emergence of a variant and its manifestation in the phenotype might distort the analysis. In our yeast case, cells exhibit highly variable lags between changes in their intracellular processes and the change in their growth-rate phenotype. Since the growth-rate results from the integration of multiple processes in gene regulation and metabolism, including processes affected by cell division (e.g. dilution of protein content), these variations in lag times are expected and might play a considerable role in the fluctuations analysis. (iv) The resistance of cells to the drug (or phages) used for selection, or in our case adaptation to the medium switch, might be highly variable. The expected spectrum of responses is in fact very wide. LD simply assumed that when applying strong enough pressure, only 'real' variants can survive. The possibility of multiple routes to adaptation and the wide spectrum of potential responses limit the applicability of this analysis. In our experiments, the response of the *HIS3*-GAL rewired cells to the *3AT* drug, inhibiting the enzymatic activity of the *HIS3* protein, demonstrates this problem. In fact, it shows that even when applying doses that are nominally lethal to the cells, the multiple routes to adaptation still allows for variants with a wide range of responses and levels of inheritance. (v) Finally, long-tailed distributions (rather than a Gaussian distribution) cannot be solely characterized by two variables (mean and variance used in the LD analysis) requiring more information on higher moments of the fluctuations, that is hard to achieve due to limited statistics obtained in such experiments.

Evolving microorganisms. Estimates of the overall rate of spontaneous mutations from lab experiments on diverse types of organisms, bacteria, yeast, flies, worms and plants, show that such mutations are indeed rare. In microorganisms, which provide the best substrate for estimation because of their genome size and short generation time, the expectation is to have a single mutation per every hundred or even thousand generations [27, 159–162]. Moreover, it is believed that most of the new emerging mutations in various organisms are neutral and that deleterious mutations outnumber beneficial ones. Thus, the rate of beneficial mutations is extremely low [28, 163–168]. Even in cases where hypermutants emerge (i.e. variants with an elevated rate of mutations), the rates of beneficial mutations are still well separated from the reproduction rate. Most of the experiments on evolving populations take the timescale separation between the rate of mutations and the rate of reproduction as guidelines affecting the experimental design. Moreover, the experimental efforts are strongly influenced by dominating theoretical concepts. These prejudices constrain new discoveries in experimental evolution [169–171]. For example, the dynamics of adaptive populations are often visualized in terms of successive steps in which the populations 'climb' peaks in a *fitness landscape*; as if there were really a simple, low-dimensional topographic description of the dynamics, or that evolving populations can be regarded as particles moving under force in a potential field [172]. Another related theoretical prejudice guiding the experiments is that of *optimization*; as if there were functions that could be maximized in the dynamics. We further discuss these issues below by a few representative examples of experiments on evolving microorganisms.

In a long-term experiment started in 1988 that is considered a cornerstone in experimental evolution, multiple bacteria populations have been propagated in parallel, reaching by now over 50 000 generations [28, 163]. Each day, the populations grown to saturation (i.e. *starvation*), have been diluted by a factor of ~100 (approximately 7 generations per day) in a medium with glucose as the limiting nutrient. Samples of the populations, frozen at a resolution of hundreds of generations are used for the analysis, mainly, for estimating fitness via competition with the ancestor population and for DNA sequencing in the search for mutations. A perspective of these experiments can be found in a recent review [173]. Two of the significant results emerging from these experiments show: [174] (i) the parallel populations exhibit pretty repeatable evolution. When population fitness was compared to the initial population, it quickly improved in the first generations but then progressed at a much slower pace at later generations. Despite the improvement in growth-rate in comparison with the ancestor population, the propagated populations hardly diverged with respect to each other. (ii) Overall, *45 mutations* have been found after 40 000 generations, way beyond the initial phase of major fitness improvements. These mutations were found by whole genome sequencing, done at a resolution of approximately every 5,000 generations (including one at 2,000 generations) in comparison to the ancestor population. Mutations accumulated over time in almost a linear fashion. Looking closely on the results, it is clear that most of the changes in fitness occur within the first 2,000 generations (lack of higher resolution data prevents more precise timing of the changes), with its improvement considerably decelerating at later times. In contrast, the accumulation of mutations remained approximately unchanged during the entire period. Given the genome size and known rates of mutations it is clear that the mutations found by this crude analysis represent a minute sample of the huge combinatorial space of expected mutations. The conclusion based on the neo-Darwinian framework of natural selection and random drift is that most of this vast genotypic space has not been tested. Four of the populations showed the emergence of mutator phenotypes with mutation

rates ~100 fold higher than that of the normal populations. Other types of measurements showed a systematic increase in the average cell size with generations and a significant change in the population-average genome-wide gene expression (mRNA measured by arrays) after 20 000 generations in only ~59 genes, compared to the ancestor population; a relatively minor change compared to the potential of the genome [175].

These experiments are highly valuable in providing for the first time, a comprehensive view of long-term evolution under controlled laboratory conditions. Notwithstanding their importance, we next discuss some issues that limit their scope and the overall conclusions that can be drawn from them. The measurement resolution was guided by the neo-Darwinian paradigm according to the expected low rate of beneficial mutations. There are two problems with this assumption: First, it restricts the analysis of this paradigm, preventing the potential discovery of alternative routes. Second, the coarse-grained analysis might mask important modes in the dynamics and hide possible emerging intermediates which could shed light on the evolutionary dynamics beyond the one expected in the conventional picture. The analysis is based on the assumption that the measured changes in phenotype are due to mutations; correlating between fitness measured relative to the ancestor (by performing competition between the two populations, the evolving and ancestor, in batch cultures) and the mutations found after thousands of generations. In light of our current understanding of the complex relationship between genotypes and phenotypes, this is a narrow and constraining assumption. What are the relevant phenotypes and at what resolution do they vary? Namely, what is the spectrum of emerging phenotypes? A complete picture of the dynamics should relate different phenotypes over time. For example, comparing the gene expression response between an evolving and the ancestor populations after thousands of generations limits the rich spectrum of intermediate phenotypes affecting the dynamics, in particular in the initial phase when most of the change in fitness took place. Indeed, the relatively small changes found in a few genes (both a relatively small number of mutations and a small number of changes in gene expression) and the lack of correlations with the growth-rate phenotype, indicate that something deep has been missed in the attempt to build a comprehensive picture of the genotype-to-phenotype associations in evolving lab populations. Thus, both the limited temporal and phenotype resolutions have restricted our ability to discover alternative routes in evolution. Another important issue is the environment which is considered as a selective filter according to the natural selection paradigm. Does the experiment actually provide a controlled environment? The answer is definitely no! In these experiments the environment is dynamically determined by the evolving population. Thus, although the nominal medium at the time of each dilution is well controlled, the actual medium experienced by the cells should be considered as a dynamic niche constructed by the evolving population. It is an active player rather than a passive filter. In the serially-diluted batch experiments the populations are propagated through repeated periods of *starvation*, applying stresses that affect the population dynamics in uncontrolled ways. The alternations between starvation and fresh medium create specific histories for the populations that might affect their behavior and lead to the observed similarity between the parallel evolving populations. Moreover, these periodic alternations impose external time-scales on the populations which intervene in their intrinsic dynamics. Also, a competition experiment, estimating the relative fitness between the evolving and ancestral populations is problematic. While it provides a measured number, the significance of that number remains elusive. In light of this, it may not be surprising that all the action seems to occur within the initial phase of the experiment. It is interesting to note in that respect, that comparing different types of populations, say at different windows in time or parallel populations at a given time point, gives a different picture of fitness; it can also decrease with time compared to the ancestor population rather than monotonically increase as expected, since fitness is a relative measure rather than an absolute one [176]. The entire concept of fitness therefore should be revisited (see the discussion above). The points raised here are rarely discussed in the literature and they might seem negligible within the dogmatic mutation-selection framework. However, focusing on random mutations as the sole source of variation distorts the picture of evolution in lab experiments, leaving no room to discover alternative scenarios in the wider spectrum of time-scales enabled by systems-epigenetic mechanisms [84]. In particular, it could well be that in some cases mutations do not lead the process and by themselves do not determine the phenotype. Certainly, the type of adaptation observed in our yeast experiments, would be completely missed under the methodology of experimental evolution described here. But even when focusing on mutations, population dynamics might involve processes beyond random mutations, such as induced mutations [49], or synchrony of multiple mutations [177].

A different type of experiments on evolving yeast populations in chemostats, utilized drug resistance as a neutral marker allowing to follow population sweeps which were shown to occur on timescales of approximately 100 generations [46]. In contrast to the experiment discussed above, this experiment demonstrated evolution in action for a eukaryotic microorganism in a continuous culture, avoiding the serial dilutions and starvation periods that intervene in the dynamics. More recently, experiments on parallel yeast populations, serially diluted for 500 generations without starvation (i.e. always exponentially growing), showed that the speed of fitness increase (relative to the ancestor population) was indeed larger than expected from the fixation of a single advantageous mutation [170, 178]. The aim in this experiment was to complement the picture of clonal interference with that of multiple mutations. Clonal interference occurs in a large population, since mutations appear in different lineages before fixation, thus slowing down the rate of fixation. However, in this experiment it was demonstrated that multiple mutations occur in the same lineage,[15] before the first mutation fixated. This causes the maintenance of variation and accelerates the speed of fitness increase, showing no signs of decline in this

[15] Multiple mutations occur here in the same lineage while in the concept of clonal interference one assumes the occurrence of mutations in competing lineages.

rate within the 500 generations measured. Without minimizing the importance of the discussion on the precise theory of evolution, clonal-interference or multiple mutations, it is worth noting that all of these experiments apply crude resolution and most importantly do not monitor the distribution of fitness and other phenotypes within populations as they evolve, mostly estimating the distributions at the endpoint of the experiment rather than showing the stability of the fitness distribution throughout the course of evolution. Presumably, both clonal-interference and multiple mutations occur in large populations and the question of which of them dominates depends on the experimental conditions and perhaps type of organism. A more recent chemostat experiment on yeast, attempted to shed some light on clonal interference by continuously monitoring three co-growing marked populations [179]. The experiment showed rich dynamics between the populations and a timescale of ~500 generations for the dominance of one population over the others. Again, the fitness measurements assumed a genetic picture; isolating a small clone from the population and competing between clones within the chemostat while ignoring all together the history of the competing populations and the possibility for epigenetic effects. The assumption that a small colony represents the entire large population and moreover, growing this colony to a new chemostat population for the competition experiment, as if this expansion and the exact conditions in the chemostat are irrelevant to the issue of fitness, are highly problematic. This experiment highlights the problems characterizing the field of experimental evolution; *the experiments are designed and analyzed under a dogmatic theoretical picture that does not leave room for the populations themselves to tell their story*.

Two issues in our yeast experiments demonstrate this point. First, the fact that a large fraction of the population can simultaneously adapt means that adaptation involves processes beyond the conventional picture of random mutations and selection [41, 48, 49]. This shows alternative means of evolution, namely, a deviation from the neo-Darwinian route of selection of a rare mutation within the background population. Second, the evolution experiments described above concentrate on changes in fitness (more accurately, the differences in growth-rates in competition between the evolving and ancestor populations). Thus, a change in fitness means that within the familiar environmental condition and metabolism there might be an incremental increase in efficiency. In contrast, our yeast experiments highlight an aspect of evolution that is completely overlooked; the ability of a cell population to overcome an *unforeseen challenge*, and more generally to develop novel phenotypes. This is a different ability than mere increase in fitness within a familiar environment, since it requires adaptation to resolve a novel challenge, a process that one would *a priori* assume takes many thousands of generations to develop. Instead, our experiments demonstrate that cells can overcome a random severe novel challenge within a small number of generations. Although discussed within a particular framework, such a potential to evolve once demonstrated, should have an impact on epistemology and affect the methodology underlying lab experiments. Whether it sheds new light on our perspective of evolution in general, remains to be seen.

2.2. Cell differentiation

We turn now to discuss cell differentiation, by first asking: *is there a well-defined discrete set of cell types in a developing multicellular organism*? If so, *what characterizes a cell type*? In an adult animal it is rather easy to distinguish between different types of cells; for instance, between skin cells and neurons, which are different in their morphology, biochemistry and functionality. This by itself, however, does not prove that the spectrum of cell types is discrete. It could well be that there is a more-or-less continuous spectrum of variants spanning the entire range of morphologies, biochemical composition and functionalities between the extreme cases which we identify as specific types. In fact, the question of discreteness cannot have an absolute answer as the classification methods are essentially statistical; based either on clustering by similarity or on the so called *cladistics* classification, grouping according to shared unique characteristics that come from the group's last common ancestor (before lineages separated), as is done for species in evolution. The difficulty in classification of types, stems from the fact that no two cells are identical in appearance, generalized patterns of morphology, biochemistry, or even functionality. Despite the huge variability in characteristics of cells belonging to the same functional group, the number of patterns emerging during an embryo's development is rather limited. Cell types refer to these limited patterns of diversity and the problem of a type resembles in many ways the problem of identifying species in the biodiversity. The current consensus, however, is that the set of morphologies, cytoanatomy, molecular biology and biochemistry is not infinite. In humans, recent work identified ~400 cell types, 145 of which are different types of neurons [180]. Most of these cell types do have identical genotypes although in some cases (e.g. some neurons) there are genetic modifications due to retro-transposons, and in other cases due to genomic rearrangements. The current trend of emphasizing molecular characteristics, leads to attempts to identify cell types by their gene expression patterns. As further discussed below, this is problematic since gene expression levels are context and history-dependent, highly variable, degenerate and moreover, focusing on them neglects systems-epigenetics. Also, current research emphasizes the use of particular markers, usually in the form of a single gene whose fluorescent-tagging allows easy identification and separation of cells. However, in many cases such putative 'markers' are not unique and in others they do not mark the terminally differentiated state of the cell. Thus, basing the analysis on expression profiles usually requires simultaneous analysis of a large number of genes whose correlations and functional significance are elusive. In summary, the picture arising is that in spite of the large phenotypic variability, members of a cell type are more similar to each other in many characteristics than to non-member cells, but correlating types to expression profiles remains problematic. The remarkable conclusion is that in spite of the complexity of the genome and its potentially vast combinatorial

space, the emergence of a relatively small number of cell types in development shows that only a minute part of it is realized. From the physics perspective, the enormous reduction from the huge combinatorial space of microscopic degrees of freedom of intracellular molecular interactions to the minute number of stable cell states, is one of the most remarkable and fascinating properties of the living cell. This phenomenon should inspire a theoretical framework describing this class of dynamical systems.

Given the existence of well-defined cell types, the fascinating process of cell differentiation in multicellular organisms, the stabilization of a functionally-specific type during embryo development serves us here as a conceptual laboratory for critical examination of some lessons learned from the yeast experiments. In particular, we wish to discuss the following issues: First, what are the relevant observables (variables) or degrees of freedom that determine the identity and stabilization of a cell type? Second, how can a stable cell type even exist in light of the degeneracy and lack of uniqueness of the microscopic-macroscopic relationship in intracellular processes, with the one-to-many and many-to-one relations? Third, given the spectrum of widely-distributed expression levels found in cell populations and their universal characteristics, what are the relationships between protein expressions and cell types?

In the following, we divide the discussion into two parts: First, we discuss stem cells—self-renewing cells with unlimited potential to generate functionally-specific cell types within a lineage—and the process of their differentiation. Second, we discuss the process of trans-differentiation and reprogramming—the ability of cells to de-differentiate and change their type, including reversion all the way back to becoming stem cells.

2.2.1. Stem-cells and differentiation. *What exactly are stem cells?* From the functional point of view, stem cells are characterized by two essential properties: (i) the ability of self-renewal, and (ii) the potential to produce different cell-types of a lineage with various degrees of plasticity, depending on the type of stem cells; *totipotent* (in the zygote stage—potentially can produce all lineages), *pluripotent* (producing all lineages except the placental tissue in mammals), or more specific ones like hematopoietic (blood cells lineage), neuronal etc. It is the property of plasticity—the potential to generate a differentiated lineage, which is unique to a given type of stem cells [181]. It forces us to focus on the potential of the genotype-to-phenotype associations rather than the end-point realizations. Stem cells present in the sharpest possible way the duality characterizing biological systems mentioned before. Namely, the stability and robustness of a phenotype coexisting with a remarkable plasticity allowing it to evolve. It seems that *stemness* is a unique phenotype since it is hard to preserve this potential over extended timescales. Although from the functional viewpoint stem cells are well defined, the molecular definition remains elusive. The search for *stemness genes*, a common set of genes for all types of stem cells, has turned out to be futile. Although some genes have been identified and seem to play a role in some stem cells, the phenomenon of stemness remains a mystery. Different studies have identified sets of hundreds of over-expressed genes, in different types of stem cells; embryonic stem-cells (ES; pluripotent cells giving rise to all lineages of the inner embryo), blood-forming stem-cells and neural stem-cells [182, 183]. However, the overlap between the sets identified in different groups is minute. In one case there is only an overlap of a single gene. There might be technical reasons for this lack of genetic signature [184–186], but I do believe that the problem is more conceptual than technical [187]. Even ignoring the fact that regarding over-expression as a signature for a phenotype might be a wrong concept (under-expression might be as important but harder to measure), chasing after a small number of genes to characterizing a complex phenotype such as stemness, seems like chasing a phantom. In an important way, it resembles the problem of '*missing heritability*' mentioned before, in which in most cases genetic variation can only explain a small part of heritable traits [188].

Thus, despite extensive experimental work attempting to identify the regulatory networks governing stem-cell lineage specification and their operational principles, these principles remain elusive. The unifying signature arising in many measurements is the widespread, low-level expression of multitudes of transcription factors. The general notion of a '*program*' as the leading concept underlying lineage specification, has a much deeper problem in that lineage specification is not merely the switching on-or-off the activity of a few genes, similar to execution of an algorithm. Rather, it is a dynamical process in which the *balance of regulators* is upset; the complex process in which the up-regulation of some transcription factors and down-regulation of others leads to reorganization of expression profiles that somehow eventually converge to a stable differentiated phenotype. The notion that there is a 'program' is based on somewhat misleading assumptions. In particular: (i) regulatory network architecture, that there are '*hard-wired*' *networks* accepting external signals as inputs and executing the process, (ii) stochastic gene expression characterized as mere molecular 'noise'. The discussion below is in the spirit of the approach based on the lessons learned from our yeast experiments. We take the view that there are no 'hard-wired' networks 'computing' the process of differentiation, nor there are stochastic processes of individual genes determining the differentiation transitions.

The most commonly assumed architecture of genetic networks leading to differentiation and cell-state stabilization is that of *cross-antagonism* between two or more 'hub' genes [189]. The notion is that there are key regulatory genes, coined *master genes*, which regulate a large number of other genes which in turn control genes further downstream—a sort of a structural hierarchy within a well-defined network architecture. In humans, for example, *Oct4*, *Sox2* and *Nanog* are considered such master regulators for stemness, and there is a significant overlap in the sets of genes they regulate. For example, more than 90% of genes regulated by *Oct4* and *Sox2* are also regulated by *Nanog* [189]. It also seems as if the stem-cell transcription factors are sufficient to activate genes involved in cell differentiation, long before the cultivation of a lineage and cell-type specification actually occur. Further

along the differentiation process, the balance of active transcription factors is changed showing reduced activity of these specific regulators. A plausible picture that is removed from the 'hard-wired' architecture of a regulatory network paints differentiation and stemness as the result of a subtle interplay between cooperative and competitive processes among the transcription factors. An analogous dynamic picture arose in our experiments on the rewired cell-cycle. Our model, inspired by experimentally observed dynamics, regards transcription factors as a limited resource and the process of regulatory reorganization due to change in conditions, as a competition over their binding sites and applied regulation [76]. Namely, the transcription factors can potentially interact with numerous DNA sites but their number is limited. Hence, there is a competition and selection for the final binding patterns that determine the expression profiles. In the case of transcription factors, they compete and cooperate to bind DNA loci, through protein–protein and protein–DNA interactions. Even under the simplifying assumption of a mean-field approximation, basic physical-chemistry dictates the 'law of mass-action', namely the rates of interactions within the complex 'soup' of proteins and DNA cannot be merely determined by the structure of an interacting network (the topology of edges). The kinetic constants and concentration of particles are highly important. Unfortunately, network modelers often overlook the kinetics and rely solely on structure [14, 129]. Thus, although the DNA sequence provides constraints and serves as a *selection agent* for the patterns of protein binding, the dynamic process leading to cell-state organization is far from being determined by 'hard-wired' circuits. Rather, the network of interactions is highly *fluidic*, dynamically reorganizing towards cell-type determination, as much as it is dynamically reorganized towards overcoming a challenge in our yeast experiments. Furthermore, differentiation appears to be a selection process, at the intracellular molecular level, in which a subset of proteins enhances their interactions towards the stabilization of the cell type. There is no need for a specific signal for differentiation, just a *trigger* for such *exploratory dynamics*. In this view, the intracellular molecular interactions can be described as a population dynamics process.

Coming back to the identity of stem cells, there is no wonder then that the deterministic view failed. It has been suggested that in the case of embryonic stem cells the genome is transcriptionally, globally active [190], with large-scale and a wide-range levels of gene expression manifested both in mRNA and proteins. It seems as if, cell differentiation requires large-scale repression of the genome. Significantly, similar to yeast cells (and microorganisms in general), a stem cell population exhibits a wide range of transcriptional levels for the same gene across a population [191]. When a small fraction of the initial population is isolated and re-cultured, it resumes the original population pattern. Similar behavior was shown by us for yeast populations, allowing to conclude once more, that the wide spectrum of expression levels found in cells across the population does not reflect molecular 'noise' but rather a dynamical process [62, 64]. Whether this is the case for stem cells remains to be seen. It is not known however, whether this global transcription plays any role in maintaining the identity and potential of stem cells. It could be that the expression of a wide range of transcripts is a mere side product of an overall state of uncondensed chromatin resulting from large-scale systems-epigenetic processes. Furthermore, exploring in detail the population structure, reflected in a single marker gene (*Sca-1*; [191]) shows that indeed the appropriate picture to explain the slow regeneration of the parental heterogeneity from a small subpopulation, is that of coexisting *metastable* states, rather than noise. Specifically, stemness does not reflect a single cell state, but rather a spectrum of states. Differentiation is the cultivation of these states from high plasticity to specificity. Interestingly, the current understanding of protein interactions reflects similar ideas in which specific protein interactions are due to selection over their excitations within ensembles of folded metastable states, rather than 'lock-and-key' rigid interactions of molecules in their unique folded state [192].

2.2.2. Trans-differentiation and reprogramming. It has long been known that cell-state differentiation is reversible. Fully differentiated cells, can be triggered to trans-differentiate into another type [193] and remarkably enough, to de-differentiate back into stem cells with various degrees of plasticity—a process coined *reprogramming* [194, 195]. These processes seriously challenge our understanding of the stability of the differentiated state. It was shown already sometime ago, by cell fusion and nuclear transfer experiments in metazoan (eukaryotic multicellular organisms) that differentiated cells can be highly plastic and triggered to change type. It was just a matter of time before diverse ways to convert a wide spectrum of differentiated cells from one type to another were discovered. Trans-differentiation and reprogramming raises a question about the dual characteristics discussed before; *how the identity of mature cells is stably maintained in the face of perturbations while simultaneously enabling plasticity in the form of a dynamical process of trans-differentiation and reprogramming?* We will meet a similar dilemma in the case of cancer below. The emergence of cancer, from the cellular perspective, reflects de-differentiation and gain of plasticity. Thus, from the organism viewpoint, it is hard to accept the common evolutionary justification that assigns advantage to any observed biological process via selection when the ready ability to de-differentiate seems like a dangerous potential. It reflects flexibility in regulatory organization that can hardly be explained as a program-like process. Fusing cells of different types originating from two species also demonstrates that unidentified regulators of one cell can enforce the activation of genes in the nucleus of the other cell type [196]. Such reprogramming is rapid, highly efficient and does not require ongoing DNA replication. The latter is usually thought to enhance the resetting of epigenetic chromatin markers such as DNA methylation and histone modifications. In some cases, such as the neuronal lineages in the worm *C. Elegans*, it seems as if sequentially-acting regulatory inputs initiate the expression of neuron-type specific transcription factors. These factors, coined terminal selectors, maintain their activity in mature neurons through autocatalytic-like self-regulation. However, this seemingly simple logic does not seem to hold in more

complex organisms where usually a large spectrum of transcription factors is involved in determining and maintaining the differentiated state. The emerging picture is highly *combinatorial*, raising a serious difficulty in understanding the coexistence potential of robustness and plasticity.

The more recent discovery that stem cells, resembling embryonic stem cells, can be generated *in-vitro*, first by over expression of a combination of a few transcription factors, and later also by small molecules or even mechanical shear forces, excites the field mainly because of potential applications to tissue engineering [197–203]. It is not clear yet whether such a reprogramming process naturally occurs *in-vivo*. For blood cells for example, it does not seem to occur at any appreciable rate [193]. Also, totipotency does not seem to hold *in-vitro* and after a few cell divisions the potential of cells to generate lineages appears heterogeneous. The existence of reprogramming also constrains models that need to encompass, in a unified and integrated way, the entire spectrum of stable cell-types and the switch of types from fully differentiated ones back to stem cells.

Recent experimental efforts show that the expression of a few key transcription factors can convert one cell-type into a developmentally related cell-type inside the animal, even in the absence of cell proliferation. Thus, these conversions do not seem to need DNA replication to aid the resetting of chromatin modifications. This is in contrast to direct reprogramming of somatic cells to a pluripotent state *in-vitro* which requires cell proliferation and more than one round of DNA replication [200]. Furthermore, the efficiency of reprogramming, which is quite low, does not seem to depend on the stage of differentiation. Thus, cells that are more differentiated do not seem to be harder to reprogram. Reprogramming seems like a stochastic process in which within a population of cells exposed to the inducing stimuli, only a fraction can go all the way back to become stem cells, while others may be stuck at different stages in the process. The reprogrammed cells do not seem to arise from a special sub-population [204]. Despite years of research and continuously emerging recipes to improve yield, the basic understanding of the reprogramming process remains elusive.

What do trans-differentiation and reprogramming teach us about the underlying gene regulatory systems that on one hand ensure stability and robustness in development and on the other hand ensure plasticity of the epigenome? These processes challenge the somewhat naïve view of the genotype-to-phenotype associations and the dynamical processes underlying cell-state organization. The possibility of triggering trans-differentiation and in particular reprogramming by perturbing the activity of a small number of genes or by introducing a relatively minor chemical or mechanical perturbation, may give the illusion that the entire process is local, involving simple kinetics in a circuit containing a few components. One common paradigm is that of cross antagonistic transcription-factor interactions, a process reminiscent of a simple genetic circuit controlling the binary switch between lysogenic and lytic pathways in the *Lambda* phage mentioned before [205]. Adopting a relatively simple binary model from the case of virus to cell differentiation in multicellular organism is an oversimplification that dangerously ignores the complexity of the underlying eukaryotic gene interaction and systems-epigenetic processes. In particular, even if local interactions do trigger the process of differentiation, its long term stability and response to perturbations must also depend on the interactions of those modules with the rest of the intracellular processes. It should not be overlooked that such seemingly isolated circuits are completely embedded in a much larger system exhibiting non-trivial interactions among its components. Support for this view of global rather than local interactions, can be found in the fact that reprogramming by over-expression of a few (or sometimes, just one) transcription factors requires only transient exposure to this trigger. Soon after the perturbation, the endogenous expression takes over and the process continues via sole operation of intrinsic processes. Given what we know today about the complexity of gene interactions and the fluctuations in these systems, there is no simple way to ensure the stability of a differentiated cell by an isolated circuit. Waddington understood this and proposed that '*strategic*' principles, beyond individual molecular interactions, were required for bridging the gap between genotype and phenotype [109]. He believed that the phenomenon of canalization of developmental systems, originally aimed to explore organ formation, manifested as the stability of the differentiation process in this context, should be understood by introducing the concept of an epigenetic landscape. Despite the impressive progress in molecular biology over the last decades, the strategic relation between the process of gene expression and the emergence of specific phenotypes has remained elusive. Waddington's basic picture is widely adopted, but only as a caricature, usually missing the main message it conveys and mostly disregarding its role as a mere metaphor. As described before, in this picture, commonly used to describe cell differentiation, this process is regarded as a marble rolling downhill in a landscape determined by gene interaction. Taking this metaphor too seriously at the cellular level is highly misleading because there is no fixed landscape that serves as a potential by applying force on the system. More appreciation of the complexity of gene interaction results in the proposal to regard the differentiated cell as a dynamical system and the stable phenotypic states (differentiated types) as *attractors* in the phase-space determined by the concentrations of expressed proteins [130, 206]. Given a genetic network architecture (connectivity), the finite number of attractors guarantees the stabilization of specific phenotypes by dynamically directing the initial vector of expressed proteins into one of its stable steady states [129, 207–210]. This attractive concept, a modern version of Waddington's landscape metaphor, was developed theoretically within the framework of specific models, and for certain classes of networks it has been shown that attractors do emerge naturally in the system, i.e. they are properties of the network's connectivity and structure. In particular, the framework of Boolean networks developed by Kauffman was instrumental in that respect [129]. However, many questions related to the attractor idea remain open: what intracellular processes do actually determine the stable attractors and their basins of attraction? Do these attractors reflect the intrinsic dynamic response of genetic networks to environmental signals? What is the level

of degeneracy in the phase-space of expressed genes? Do many different attractors result in similar cell phenotypes? The experimental basis necessary to tackle these key issues is still lacking [128].

The recent realization of reprogramming by a spectrum of different perturbations, seriously challenges the simplified picture of cell states as basins of attraction of a low-dimensional dynamical system. Given the complexity of gene interaction, it opens our repeated question: *what kind of a dynamical system supports the stability and robustness of cell differentiation and at the same time allows complete de-differentiation following a perturbation?* Two issues need to be addressed: First, the high dimensionality of gene-activity phase-space dictates dynamic characteristics that are far from the type of attractors discussed in the framework of low-dimensional dynamical systems. Most of the examples analyzed in the literature thus far, adhere to the low-dimensional type of systems and therefore their results may not be applicable to the realistic high-dimensional cases presented by gene regulation systems in the living cell [211–213]. For example, in a high-dimensional phase-space, in contrast to a low-dimensional one [210], the system may not directly flow into a nearby attractor since there are an exponentially large number of directions away from it [213]. In fact, high-dimensional systems are highly non-intuitive and unfortunately, till now there is no theoretical foundations for these dynamics. In some rare cases, the effective dimensionality of the system is low and the search for a procedure for reducing dimensions is successful [214]. Second, most of the analyses done so far have been based on fixed-connectivity networks. In reality, the essence of systems-epigenetics and multiply interacting proteins, manifested in the degeneracy discussed above, is that gene interactions are highly *fluidic*, exhibiting a high degree of plasticity and depending on the environment, context and the history of the system. Some aspects of these features arise also in our yeast experiments. The lessons learned there, suggest that the reprogramming process should be regarded as *triggered* by the perturbation (e.g. over-expression of some transcription factors), rather than induced by it. If this holds true, then the term '*induced pluripotency*' is a bit misleading. Perturbations only serve as stimuli of large-scale dynamical process rather than changes in a 'pre-designed' program. An alternative view suggests that certain perturbations 'kick' cells out of their relaxed states, by stimulating a *large-scale response* similar to the one observed when cells are faced with an *unforeseen challenge*. The large-scale response in gene expression is non-specific and can lead in some cases to trans-differentiation and in other cases to complete reorganization of gene regulation resulting in stem cells. The evoked large-scale expression profiles do not easily relax and therefore stemness is preserved. In short, trans-differentiation and reprogramming reflect non-specific *exploration–exploitation* dynamics. The process then does not rely on the concept of master genes, the search for which has largely failed. In principle, no particular perturbations are necessary, although it might be that under certain conditions, e.g. *in-vitro* cultures, particular perturbations are more efficient in creating the substrate epigenome amenable for a stimulation. We do not understand the fundamentals of this process yet, since we are not even sure about the relevant variables underlying cell-state organization. As discussed before, the content and expression of each individual protein do not seem to be the proper observables. Moreover, all cells can in principle undergo the transition to stem cells, albeit with different efficiencies under different types of perturbations and conditions. The fact that every cell has the potential to change its type seems, on the one hand to support evolvability but on the other hand, carries the risk of improper development and cancer. These two features go hand-in-hand. The biological cell seems to reside on the delicate edge of stability. As far as it known to this author, no specific quantitative model of this spirit has been proposed yet outside of Boolean networks at the edge of chaos [129], which were discussed above and do not actually capture the complexity of the system. Certainly, from the modeling perspective, high-dimensionality [213] and fluidic interactions [215, 216] represent difficult and open issues at the forefront of the physics of complex systems.

2.3. Cancer

2.3.1. The nature of the phenomenon. Cancer is a terrible disease, currently one of the greatest threats to the public health and one of the primary challenges in medical research [217]. Cancer, however, also encompasses a fascinating set of phenomena,[16] which largely remain elusive. Despite the diversity among different types of cancers, there seems to be some universal features common to many types. If we detach from the clinical aspects and the relation between the spread of a tumor and the body-scale physiology (e.g. the immune system, blood flows etc) and concentrate, as we do throughout this article, on the cell and cell-population levels of biological organization, the cancer phenomenon allows an excellent laboratory for understanding aspects of genotype–phenotype associations and cell-state organization [218]. Despite a great progress in molecular and cell biology, there is no agreement on the basic principles underlying this phenomenon. In particular, examination of the integrated experimental data calls for a shift in conceptual thinking from molecular causations to a problem of *organization*, understanding the symbiotic relationship between intracellular organization, cell-populations and the environment [219]. Cell differentiation in development and aspects of population dynamics and evolution discussed above seem to be tightly connected to the initiation, progression and establishment of a malignant tumor [220]. Researchers working on cancer at the beginning of the 20th century understood this very well, emphasizing the strategic principles and connecting the emergent cancer phenotype with the rest of system-level understanding of biological phenomena. The molecular revolution, especially in the genomic era and following the advance in technology (e.g. DNA sequencing), eliminated most of this line of thinking. It diverted the study of cancer into a narrow avenue searching

[16] In fact, it is hard to discuss the many facets of cancer as a single phenomenon; the spectrum of phenomena (in particular in the clinical context) which collectively can be classified under the heading of cancer is huge.

for a simplistic picture—a molecular cause of the disease—thus losing the *organization aspect* of the phenomenon. No doubt, there is nothing simple in the initialization and emergence of cancer, nor there is an isolated molecular cause. Moreover, even the current trend of adding more molecules to the game, more genes connected in networks, does not help in understanding the basic principles. We cannot specify this better than Smithers: [221, 222] '*cancer is no more a disease of cells than a traffic jam is a disease of cars*. A lifetime study of the internal-combustion engine would not help anyone to understand our traffic problem. A traffic jam is due to a failure of the normal relationship between driven cars and their environment and can occur whether they themselves are running normally or not.' The frustration in understanding the basics of the cancer phenomenon is also reflected in the clinical side in the treatments of the disease. After decades of enormous investments it is sadly a fact that progress has been minor. The mortality rate of some cancers is growing, while in others that are not strongly affected by behavioral and environmental aspects (e.g. smoking and air pollution) it remains at a constant level as in 1930 (even when corrected for age) or somewhat declining due to early diagnostics [223, 224]. In reality, inherited cancers, carrying the same mutation in all cells of the organism occur in something like 2% of clinical cases [225]. The majority of cancers are sporadic, due to either natural processes or due to the exposure of the organism (before or after birth) to carcinogens, either chemical, physical (e.g. radiation), or biological (e.g. virus). Even mutations in genes that are assumed to be highly associated with cancer risk, such as *BRCA1* and *BRCA2* are found to account for only about 30% of inherited breast cancers [226, 227]. The discrepancy between the expectation from a somewhat straightforward genetic interpretation of cancer and the apparent complex reality could be the result of the involvement of numerous interacting genes, or because conceptually the understanding lies somewhere else, in a lesson similar to the case of '*missing heritability*' [188]. More recent, high-throughput measurements reveal the complexity of the problem. Gene expression measurements have identified hundreds of genes as involved in cancer with no consensus on their functionalities [226]. The search for mutations is also problematic; a recent study looked for the presence of 238 known mutations in 17 genes that are supposed to be highly involved in tumor initiation (*oncogenes*; see below) across 1000 tumor samples. Mutations were found in only 30% of the cases [228]. Nevertheless, in spite of the low rate of clear genetic origins, the main view of the field, in particular towards clinical applications, is that cancer is a *genetic disease*. Abundant epidemiological and other experimental evidence, indicate however that the proximate causes of sporadic cancers are in fact environmental. This requires a considerable paradigm shift; putting the environment, context and history of the cells at central stage while moving genes backstage. Moreover, it is clear that *cancer is a process not stuff* and as such requires understanding of the dynamics rather than merely 'hunting' for molecular causes. The above is true even if somatic mutations play important role in the cancer process. These mutations, as we discuss below, might be part of the complex dynamics responsible for tumor initiation and progression.

2.3.2. Cancer—a process of organization. Here, detached from the disease aspects and in line with Smither's description, we take the viewpoint that cancer is a *process of organization*. Limiting ourselves to cell-population aspects, we shall see the similarity between cancer and cell differentiation, stem cells and reprogramming. This viewpoint suggests that cancer is an essential property, reflecting the intrinsic *potential* of every cell, the unavoidable consequence of pluripotency and evolvability. This phenomenon reflects the tension between robustness and flexibility discussed before and thus is deeply connected to the adaptation to an unforeseen challenge observed in our yeast experiments. We therefore focus the discussion parallel to the lessons from our yeast experiments, ignoring large parts of the history and the many facets of cancer [217].

The leading, mainstream dogma of the cancer process is the *somatic mutation theory* (SMT), which states that cancer emerges due to the occurrence of somatic mutations, and then progresses by an ordered sequence of more mutations. Viewed like this, cancer is considered a genetic disease. Some cancers are thought to be inborn and genetically inherited, while others (most) are sporadically occurring but still due to genetic failures. It is now also accepted that molecular epigenetic processes can play important role in cancer. These are largely viewed as surrogates of genetic modifications and thus while adding more mechanisms do not fundamentally change the conceptual picture. We come back to discuss epigentics in the context of cancer below. Genetic modifications are rare events that, following the initiation process, evolve much like in the Darwinian scenario for evolution; they over-compete the 'healthy' cells in the tissue, eventually stabilizing as a malignant tumor which can later also metastasize. This is the clonal evolution model for cancer [229]. The main problem is that although the late stages of cancer have been extensively characterized, the initiation of the process remains elusive. But even for the late stages, there is no coherent picture. There is ample evidence that mutations occur in many types of cancer, but this by itself does not prove that a mutation is the cause of the phenomenon nor that it is responsible for its initiation. In many cases it is not even clear which mutations drive the changes in phenotype ('driver' mutations) and which are side-effects of other processes ('passenger' mutations). Certainly, finding mutations at the endpoint of tumor growth does not necessarily advocate for their central causal role in the phenomenon. They may well be followers, mere side-effects of the intracellular dynamical processes. To get a better sense of the problem, let us look at the example of colorectal cancer, one of the most prevalent causes of cancer mortality [230]. Recent statistics show that 50% of individuals in the west develop a colorectal tumor by the age 70, and 10% of these it will progress to malignancy. Epidemiological studies have identified 15% of colorectal cancer incidence with a pattern of dominant inheritance. With such a large percentage, the assumption would be that there are certain *single* genes

causing colorectal cancer in a dominant fashion. Many studies have established, however, that at least seven mutations are required and they must appear in a certain order [230]. Such a sequence of specific genetic events must be extremely rare; each of these mutations by itself is a rare event, the probability of occurrence of such a sequence of mutations is extremely small [231, 232]. The analysis of candidate genes that lead to two forms of hereditary colorectal cancer shows that these can hardly be directly blamed for the chain of events leading to this cancer; detailed analysis of this case is beyond the scope of this article [222, 230]. However, in at least this specific case, it is clearly not the mutation but rather alteration of the organizational field of the colorectal epithelium into the form of polyps that seems to be the first essential step towards tumor progression.

The large heterogeneity observed among cells within a tumor, has led to the concept of the *cancer stem cell* [233]. In this view, the growth of tumors progresses via a limited number of cells capable of self-renewal. The initial genetic (or epigenetic) modifications then either occur in cells that have the stem-type or alternatively, they actually transform cells into this type. The process then proceeds by lineage differentiation and the tumor progresses as tissue organization during embryonic development. Unfortunately, many issues of the cancer stem-cell model remain unexplored due to technical difficulties. Even if stem cells are the basis of tumor growth, their number is very hard to estimate. Heterogeneity in cell types and cell behavior might reflect fluctuations between types. The detection of subpopulations might well reflect a dynamical process of *metastable* states [234]. In particular, as explained above, stemness and the stem-cell type are basically not understood, so evoking this concept in the case of cancer does not advance our fundamental understanding of the phenomenon. Given the recent recognition of the plasticity of cells during development, the relatively ease of trans-differentiation and reprogramming, it is tempting to think of cancer as a *pathological de-differentiation,* leading to the formation of a new branch of tissue formation [235]. In that respect, cancer is an essential consequence of cellular plasticity; the 'dark' side of evolvability and innovative potential. It is important to recall that also in the case of stem cells, large fluctuations in gene expression and other phenotypes, may indicate the existence of metastable states and dynamic transitions between them [191, 236]. A similar situation may exist in the case of tumor formation.

2.3.3. The classical picture and its alternatives. We need to get somewhat deeper into the details of this picture in order to understand why assuming a specific mutation causing the initiation of the cancer process is fundamentally flawed and consider alternative explanations. Following the discussion in Moss [222], we start from the summary of the mainstream view of cancer, as summarizes by Smithers himself in 1962 and largely remained the same until recently. His points are of particular interest for us, since their analysis highlights the analogy with the lessons learned from our yeast experiments. The conventional picture according to Smithers goes as follows: [221] (i) Cancer is a special disease of cells; (ii) a cancerous cell reflects a permanent change; (iii) cancer cells multiply without restraint; (iv) cancer cells grow at the expense of normal tissue; (v) there is a cause of cancer that if could be discovered the problem might be resolved; (vi) treatments fail due to the remaining viable cancer cells. In contrast, this is the more realistic picture (already in 1962): (i) many cancers appear to arise in parallel in more than one cell; (ii) histologists never see a radical transition in cancer, but rather a gradual change over time; (iii) cancer is age dependent and varies by geography—it is thus context dependent; (iv) progression and regression are observed, spontaneous regression is of particular interest; (v) some tumors appear to be continually dependent on environmental conditions; (vi) the status of a cell as a cancer cell may be hormone-dependent. The summary of these points is that instead of thinking of cancer as an intracellular problem, attention should be shifted to the interaction with higher levels of organization—our emphasis here is on the cell-population level and its symbiotic dynamics with the environment. It calls for a shift from the somatic mutation picture to *developmental-organizational* perspective, summarized as follows: First, cancer is an issue of organization. Second, there is no such thing as a cancer cell—only cells behaving or bearing a phenotype that show the cancerous process. Third, there is no molecular cause of cancer to be found inside the cell, and finally, a tumor reflects a change in organization and it can progress or regress according to the dynamic behavior of the cell-population and its relation to the environment. Smithers, suggested that carcinogenesis is a gradual process rather than the all-or-none type of phenomenon stressed by the somatic mutation theory. Certainly, the intracellular processes cannot be separated from the external world. For example, chemical carcinogens affect tissues according to the following characteristics: (i) experimentally proven carcinogens are actually highly variable with respect to their mutagennicity; (ii) chemical carcinogens do not promote cell growth but rather *inhibit* it; (iii) cancer develops due to the exposure to carcinogens (as well as radiation and viruses), it is a prolonged process that in many cases takes a large chunk of the organism's life-span; (iv) unrestrained growth of cancer is observed only at a very late stages of the process of tumorgenesis. Following this discussion [222] and in accordance with our lessons from yeast we ask: *could it be that carcinogens and other environmental agents serve as triggers of an adaptive destabilizing response of cells in their population (tissue) context?* It is a response of many cells exposed to the stimulus; different cells take different courses and re-organize differently leading to heterogeneous populations. For example, experiments showed that by exposing cells derived from mouse prostate to a potent carcinogen, all exposed cells resulted in clonal populations out of which some minority of cells give rise to transformed foci [237]. Studies showed that the same alteration had taken place in 100% of the exposed cells (in a non-treated control population only 6% were transformed spontaneously); i.e. not a rare minority. Each and every exposed cell becomes capable of giving rise to progeny cells, out of which a smaller subset then produces tumorigenic colonies. This situation is reminiscent of the adaptation observed in our rewired yeast cells; every cell has the potential to adapt via multiple heterogeneous processes [41, 48].

Certainly, a population-wide 100% response does not fit the expectation that a mutation causes the effect, since mutation is a rare event. It rather suggests a physiological response that propagates through systems-epigenetic processes for generations. Genetic damage could follow due to genomic destabilization as part of the reorganization processes. The main problem in the case of cancer is that it is extremely hard to follow the entire process from initiation through progression to the emergence of a malignant tumor. There is ample information on late stages in tumor formation but relatively little is known about the initiation process due to its rarity. Analyzing the end result of a mature tumor might be highly misleading as was already discussed in the context of evolution above.

To get further insight into the phenomenon, let us come back to the classical approach to cancer and explore further how it meets the experimental reality. The premises of the somatic mutation theory are: cancer is derived from a single somatic cell that has accumulated multiple DNA mutations which lead to a change in cell state from quiescence to proliferation; the default state of most cells is quiescence so the mutations affect the control of proliferation and that of the cell-cycle [238, 239]. Viewed in that way, cancer seems like a gain-of-function; the assumption being that the mutated genes are special *oncogens*. Later experiments however, have proved that when cancer cells are fused with normal cells the process halts, leading to the concept of *tumor suppressor genes*. Mutations in those genes are required for cancer to proceed, thus this process is built on a loss-of-function. Currently, the picture is quite complicated as there have been more than 100 oncogenes and more than 30 tumor suppressor genes identified thus far. In fact, analysis shows a highly intricate complex network involving numerous interacting factors, replacing the simplistic view of a dominant single or few factors.

To get a bird's-eye view of why the classical picture is problematic, we follow a discussion by Harris, one of the discoverers of the tumor suppression genes [240]. He writes: 'During the last half century, cancer research has not delivered an agreed explanation of how malignant tumors originate; the models reflect waves of fashion and time revealed their inadequacy'. He summarizes sharply: 'cancer is (1) not caused by the direct action of oncogenes, (2) not fully explained by the impairment of tumor suppressor genes, (3) not set in motion by mutations controlling the cell-cycle, (4) not governed by the dependence of malignant tumors on an adequate blood supply, (5) not triggered by a failure of programmed cell death'. His main point is that cancer is not a phenotype initiated by cell multiplication but rather a *deviation in the trajectory of differentiation*. In this view, the basic tendency of a cell is to multiply and it stops doing so, when differentiated into a specific type. In the language based on the lessons from our yeast experiments discussed above in the context of differentiation, the cell realizes its basic *potential* once triggered into an *exploratory* mode; within the broad spectrum of possible realizations, parallel to differentiated states in normal developmental trajectories there are other trajectories that lead to cancer.

It worth briefly listing the waves of fashion discussed by Harris [240] as they reflect the *hallmarks of cancer* [239]. This might teach us how a research field, caught in its own dogma, can only be freed by cutting the *Gordian Knot*. These fashions arose in the following order: starting with the *oncogenes* and later displaced by *tumour suppressor genes*; discovery of genes governing the *cell-cycle* in yeast and their homology in human led to focus on mutations affecting these genes; the idea of *angiogenesis*—the dependence of the tumor growth on blood supply; the idea that *apoptosis*—programmed cell death, may halt tumor progress led to a search of mutations impeding this process; finally, the more recent idea of *aneuploidy*—the multiplication in the number of chromosomes is the current basis for debate as to whether it is a leader or follower of the phenomenon. The point is that while each of these processes by itself is important and interesting, none of them leads to an integrative understanding of cancer initiation. From the modern biology point of view, one can safely conclude that: oncogenes do not form tumors directly but at best establish a predisposition to tumor formation. In only a minority of cases in which recessive genes were classified as tumor suppressor genes, it was actually shown that they suppress the growth of malignant tumors. The evidence that cell-cycle mutants are responsible for the onset of tumor formation is absent. The apoptosis process does not provide a clue of how tumor originates, although the appearance of aneuploidity seems to correlate with tumor formation, it could well be that it is merely a secondary effect of tumor growth. The main conclusion is that an alternative picture should be sought. Harris sketches such a picture, which can be summarized as follows: The inherent steady state of all cells is exponential multiplication while deviation from this state is due to cell differentiation. Thus, the disordered cell multiplication seen in malignant tumors is due to an 'error' in differentiation [240].

Experiments on Drosophila showed that indeed during the development of the organism, tumors arose at specific times and at specific sites when mutations blocking the process of normal differentiation occurred [241, 242]. In these experiments, mutations in genes, estimated to be around 100 in number, led to aberrations in many tissues, overgrowth of cells that lost their normal function and differentiated state, a spectrum of plastic changes with a wide range of effects including invasive tumors and at sites where overgrowth was not observed leading to developmental abnormalities. These experiments therefore, established a tight connection between the process of tumor initiation and distorted paths of differentiation. In principle, there is nothing special about mutating these genes. Any perturbation either environmental-epigentic or genetic that is strong enough may lead to similar effects, simply recognized many times as abnormal developmental processes. The arising picture makes a tight connection between cancer, stemness and the differentiation, trans-differentiation and reprogramming discussed in the previous chapter. The point in cancer is that in the context of the adult organism, initiating stemness and novel trajectories of differentiation may lead to disastrous consequences in an inappropriate context while in the evolutionary context it serves as a potential for innovation and facilitates evolvability. At any rate, the context plays important role whether the spectrum of realized differentiated states of a cell remains

unnoticed, leads to innovation, or to malignant tumors that eventually kill the organism.

2.3.4. Cancer as an adaptation process. There are ample experimental indications that the tissue condition plays a major role in defining the course of cancer. For example, when cancer cells from a rat's liver were transplanted in young and old rats, they formed cancer at high probability in the old ones and at low probability in the younger rats [243]. Interestingly, the same cells re-differentiated to normal cells in the younger rats. The following experiments on liver cancer of rats give some perspective how problematic it is to study the end-points rather than following the process [222, 244]. Exposure of rat's liver to 75 different chemical carcinogens led to the observation that a small number of cells acquire a new phenotype coined '*resistant*', which could be characterized in three ways: (i) resistant cells can be induced to grow, while the majority of liver cells (hepatocytes) are growth-inhibited by the carcinogen. (ii) Resistant cells exhibit highly enhanced tolerance for cytotoxins, and (iii) they have a distinctive profile of enzymes consistent with their ability to withstand cytotoxic challenge [245]. The importance of developing this assay was in the appearance of resistant cells with the initiation of the cancer process. The exposure to carcinogens by itself does not lead to an aberrant phenotype, but rather to an *adaptive response to the challenge*. The point is that following this initiation, the next set of processes is not inevitable but sometimes does occur. The next step could be promotion—the development of focal proliferation of cells that act as precursors for subsequent steps in the carcinogenic process [244]. In his experiments, Farber noticed that nodule formation constitutes a highly organized developmental process which serves a physiologically adaptive function in protecting the organism from exposure to toxins. This adaptation manifests itself in the acquired ability of the organism to withstand high doses of hepatoxins. Doses that are lethal for 100% normal rats are completely non-lethal for rats with the hepatocyte nodules. It is highly tempting to mention here our analogous experience with the drug *3AT*, which inhibits the enzymatic activity of the *HIS3* protein, the rewired enzyme essential for histidine production in yeast. Our adapted rewired cells can sustain two orders of magnitude higher doses of this drug at levels that are 100% lethal for wild-type or naïve rewired cells [30]. Further support for the physiologically adaptive nature of the hepatocyte nodules is in the ability of the majority of their cells to undergo re-differentiation, becoming normal adult-like hepatocytes. This is strong evidence that the formation of resistant phenotype, with the ability to proliferate, is part of a 'normal' developmental spectrum of the liver cells. When the balance is shifted and this 'developmental' mode is detached from the environmental demands, slowly evolving to hepatocellular carcinoma and therefore out of the organismal context of normal development, the end result is a disaster for the organism.

Similar indications for spontaneous transformation in culture and the effect of the environment are routinely obtained due to metabolic stress and especially when cells grow under conditions of crowded postconfluent conditions [246]. The apparent population response is again not consistent with the expectations of a specific mutation—necessarily a low probability event. However the effect is heritable, suggesting that postconfluent cultures have characteristics of epigenetic adaptation to the stressful conditions, followed by or accompanied by genetic alterations. Propagation of the altered cells leads eventually to monoclonal cultures of transformed cells. This summarizes a chain of events suggested also in our yeast experiments. Adaptive epigenetic-physiological changes can sometime lead to genetic instabilities. The discussion in the literature of experiments in the context of cancer raises again and again the complex interrelations between systems-epigenetics and genetics. Rubin and colleagues experiments can be summarized saying that all the evidence points to cancer originating from a *field of altered unstable but normal-appearing cells* rather than from isolated mutants residing among otherwise unaltered cells [246].

2.3.5. The role of the microenvironment. The recognition that the cell microenvironment plays a crucial role is critical to our understanding of cancer (as for our understanding of the adaptation phenomenon) but for some reasons has not yet become part of mainstream research [247]. The appreciation of the importance of the tissue environment for cancer led to the proposition of the *tissue organization field* theory (*TOFT*; [225, 248, 249]) which challenges the core premises of the somatic mutation theory. In this approach, the focus of attention is shifted from the intracellular processes to the tissue (cell population) level of biological organization and the default state of the cell is assumed to be that of proliferation rather than quiescence (see also [250]). Additionally, cell mobility within the tissue is a property that every cell can acquire. Thus, an altered tissue state facilitates these basic capabilities, proliferation and mobility, the prerequisite for malignant tumor.

We discuss now some aspects of this picture in light of the lessons learned from our yeast experiments. The search for an alternative level of biological organization outside of the genome is motivated by the difficulties encountered by genome-centered approaches. Besides the somatic mutation theory, mentioned above, the other two extensions are molecular epigenetic processes and aneuploidity (chromosomal destabilization). It is pretty evident, given the discussion above and in particular in view of the stem-cell approach to cancer that systems-epigenetic processes play an important role both in cancer initiation as well as progression; in reality, of course both genetic and molecular epigenetic processes are interconnected in diverse and complex ways during the progression of different types of cancers. For our discussion here, as mentioned before, genetics and molecular epigenetics do not necessarily present radically different concepts, and similar to the somatic mutation picture are mainly DNA-based mechanisms. This does not mean that epigenetic processes do not add important aspects to the mutation-based approach. Certainly, epigenetic processes in particular are important due to their susceptibility to environmental perturbations, fast dynamics that results in broadening of the temporal dimension and tight coupling with physiology and

metabolism; all aspects that are either missing or weak in the case of mutations. Aneuploidity might be a secondary effect due to an instability from the emergence of other processes. Other extensions, like transposons [251], illuminate additional mechanisms that broaden the range of intracellular responses. The tissue (population) organization field is conceptually different since it brings to the table the crosstalk between two levels of organization. Irrespective of details, the huge heterogeneity within a single tumor calls attention to the external organization field. Indeed, recent studies show that any measured aspect, reflects this heterogeneity. For example, gene expression profiles of different regions within a tumor lead simultaneously to 'good' and 'bad' prognosis [252]. The same study also found with 26 of 30 samples from four tumors had ploidy heterogeneity (different number of sets of chromosomes) and divergent allelic-imbalance. From another angle, there are strong indications for the field organization coming from cell transplant experiments. For example, tissue recombination of stroma (connective supporting tissue) exposed to a carcinogen with normal unexposed epithelial cells resulted in neoplasm (tumor) [253]. Interestingly, induction of neoplasm was also shown when embryonic cells were misplaced in adult tissues, and reverted to normalcy when placed into early embryo [254]. This point was already discussed above. Interrupting the normal developmental trajectory or misplacing cells developing along a certain differentiation trajectory into the wrong context (e.g. adult tissue) is enough to trigger a change in type leading to cancer. The opposite process leading to reversion of cancer is also observed. Indeed, reversion of the tumor phenotype was found when cells from rat mammary-gland tumor were inoculated into rats of different ages; in adult rats these tumor cells generated phenotypically normal mammary ducts [255, 256]. Similarly, when renal carcinoma cells in frog were transplanted into enucleated and activated ova, they developed and reached the swimming tadpole stage [238, 257]. The transplantation of tissues from these tadpoles into normal recipients generated normal tissues that were indistinguishable from those of the host [258]. Reversion of different neoplastic phenotypes when the cancer cells are transplanted into a normal tissue is a well documented phenomenon found in many studies (see for example, [259, 260]). The plasticity of the neoplastic phenotype shows the involvement of epigenetic processes [261] and great susceptibility to the tissue organization field—features not well aligned with of the concept of mutations as the basis of cancer. The reversion of malignant cells by embryonic environments has been suggested to be due to common regulatory signals of embryonic and tumor stem cells [260, 262]. Transplanted human melanocytes and metastatic melanoma cells into zebrafish blastula-stage embryos showed that these cells could survive and *participate in embryo development* without forming tumors. The melanoma cells lost their tumorigenic phenotype [263]. All these examples indicate that the embryo microenvironment is able to affect cancer cell and change their phenotypes, supporting the dynamic field-view presented above.

The above discussion suggests that we should think of cancer as an *adaptation* phenomenon. In other words, the emergence of cell states resulting from exploratory dynamics triggered by a perturbation. This suggests that the *adaptation of yeast cells to unforeseen challenge belongs to the same class of phenomena*. The dependency on context, history of the cells, their environment and all other constraints eventually specify the dynamic trajectory and the realization of various cell states [218]. In the case of yeast, the population level of organization was found to play a crucial role, similar to the tissue-level in the case of solid tumors. Recall that cells within a population respond by expressing genes in a coherent way while the correlation between responses of different populations is weak [40]. Thus, even though the profiles of gene expression are non-specific and irreproducible between populations, even between 'twin' populations, cells within the population exhibit a highly coherent response. The lesson from these experiments is that the population 'field' affects the gene expression response of individuals, which does not result solely from autonomous intracellular processes. There is no need for specific, intracellular molecular signaling to achieve coherence among cells in a population. Their common environment serves to converge their response through nonlinear dynamical processes, i.e. the growth of cells affects other cells by their uptake and extraction of ingredients changing their common environment. Under certain conditions, cell metabolism can become very sensitive to small environmental changes that are then amplified by the response of other exponentially growing cells. Similarly, in the case of cancer, there is no need for special signaling; the tissue provides the organizing field. Therefore, understanding cell-state organization and the associations of genotype and phenotype requires one to go beyond genetic networks and 'information' approaches. In short, one needs to understand the type of dynamic organization presented by the living cell.

3. Summary and outlook

> *A society that permits biology to become an engineering discipline, that allows that science to slip into the role of changing the living world without trying to understand it, is a danger to itself.*
>
> C R Woese, 2004 [274]

Starting from experiments aiming to explore the adaptation of yeast cell populations to an unforeseen challenge, a set of concepts at the basis of cell-state organization and the genotype-to-phenotype associations have been developed. Extending the discussion by revisiting three major branches of biological inquiry—evolution, cell differentiation and cancer—gives a wider vista of these concepts and demonstrates their generality. The common theme is the question: *How does phenotypic order emerge from molecular disorder in the living cell?* This is a shift of focus from the famous concept of '*order from order*' developed by Schrodinger in his influential book *What is Life*? [264]. The focus of this article is the emergence of phenotypic order, cell-state organization, from the disordered molecular makeup of the cell which at the same time maintains its flexibility to evolve. The genotype, an apparently

ordered entity, is not translated directly to a phenotypic order. Rather, cell-state organization is a dynamical process in which the molecular disorder manifests itself in macroscopic order. The genotype, notwithstanding its important role, participates in this process but does not fully determine it. The genome in this view provides a set of constraints on the spectrum of regulatory modes, analogous to boundary conditions in physical dynamical systems.

This article attempts to sketch an organizing framework for a reader interested in lifting the veil from this fascinating issue of cell organization in biology. It provides only an outline, necessarily bringing up more open questions than answers. We have a long way to go towards resolving this issue, but the emphasis on *dynamic organization* taken here, which I believe should occupy a central stage in modern biophysics, might help to crystallize it. Paraphrasing William James statement on the mind, *the emerging order in the living cell is a process not a stuff* [275]. This emphasis is hardly new. However, as this article attempts to show, accumulation of experimental evidence makes the time ripe to bring it back to center stage.

One should be aware of the deep difference between the analysis of the molecular makeup of a cell and the organization principles underlying the emergence of order. The contrast between process and stuff, in the context of organization, has been discussed in the past. It could be crudely illustrated in analogy to the insight leading Kepler to write his marvelous little book, *On The Six-Cornered Snow-Flake*; [276] attempting to explore the principles underlying organization of a snow-flake into well-defined patterns. This is perhaps one of the earliest scientific attempts to understand the emergence of natural patterns, beyond the specificity of their underlying material content. In the introduction Kepler writes: *I crossed over the bridge, mortified by my incivility in having appeared before you without a New Year's gift… Just then, by a happy occurrence, some of the vapor in the air was gathered into snow by the force of the cold, and a few scattered flakes fell on my coat, all six-cornered… Here, indeed, was a most desirable New Year's gift for the lover of* ***Nothing***. Kepler indeed emphasizes that organization principles are about '*Nothing*'. They are not stuff, and therefore are a proper gift to his mentor who is rich, famous and clever and so could not be easily impressed by material objects. At the end he remarks: *But I am getting carried away foolishly, and in attempting to give a gift of almost Nothing, I almost make Nothing of it all. For from this almost Nothing, I have very nearly recreated the entire universe, which contains everything*. Organization of patterned snow-flakes is an issue altogether detached from the catalog of underlying atoms forming a water molecule; the same combination of atoms, indeed the same water molecules, could just form a structureless fluid. Surely, there is no realization of a snow-flake without the water molecules. But the existence of these molecules, by itself, could not explain the highly symmetrical macroscopic six-cornered pattern of the snow-flake, which is an *emergent* phenomenon. Today, scientists are well aware of the issue of emergence, most notably when driven by *symmetry breaking* processes [265], either in space, time or space-time. The argument is that in trying to understand cell-state organization, we should put more emphasis on the 'nothing'—the principles of dynamic organization rather than relying solely on the material stuff—trying to build order from the specificity of molecules and their interactions.

Biological cells present a challenge for our quest to perceive and comprehend organization in natural phenomena, far beyond the ones presented by physical systems; in particular, due to their microscopic heterogeneity and the multiple types of interactions of their underlying constituents. Unfortunately, the complexity of biological systems seems to mask this conceptual separation between stuff and process, between the underlying molecular material objects and the dynamic process of organization. In principle, we know three ways in which systems can develop macroscopic order from their underlying molecular disordered constituents. Thermal equilibrium in thermodynamic systems is the most common way and stands at the basis of equilibrium statistical mechanics. Next, constraints can limit the number of possible modes in a non-equilibrium driven system. Examples are lasers in optics and patterns emerging in fluid dynamics. Finally, there are cases in which dynamical rules dictate order. This is the class of self-organized systems, manifested more recently also in examples in the form of random organization [266]. In biology, the first option is ruled out; the living cell is far from thermal equilibrium. There are certainly constraints applied on the cell, both physical [267] and genetic ones dictated by the composition of the genome and its organization. However, note that in a system spanning a microscopically combinatorial large phase-space, if constraints are responsible for the macroscopic order they need to eliminate a huge number of possible dynamic modes in a very effective way. As far as I know, there is no example for such a process in amorphous heterogeneous materials. At this stage, we cannot entirely rule out this possibility that in fact reflects the common picture in biology, based on selection in the evolutionary process. However, as demonstrated in our yeast experiments and further discussed in the examples from other branches of biology, this approach is not compatible with the ability of cells to efficiently adapt to arbitrary *unforeseen challenges*, nor it is compatible with the observed plasticity and evolvability that are the hallmarks of the biological cell. We are left with the concept of *dynamic organization* based on exploratory processes. Note however, that random exploration in a large combinatorial space spanned by the living cell is highly inefficient and can hardly lead to the emergence of order discussed in this article. Processes like random organization mentioned above [266] work away from thermal equilibrium and are not limited by exhaustive scans of a huge number of microscopic configurations. They rely on stabilization of a many-body system at an absorbing state, one out of many possible degenerate states available to the system. Thus, such processes can in principle serve as a basis for a dynamical theory of cell-state organization. Unfortunately, until now a theory connecting such processes to biology is lacking. In particular, a significant obstacle along the way is the gap in our ability to identify the *relevant* variables underlying the dynamics in the living

cell. As demonstrated by our work, as well as of others, the huge combinatorial phase-space spanned by the intracellular microscopic degrees of freedom, is highly degenerate. It has been recognized that many complex systems belong to the same class, roughly identified as 'sloppy systems', in which most directions of change in variable space do not affect the macroscopic behavior, while a small number of 'stiff' directions do affect it [86, 268, 269]. I strongly believe that this is not a problem of parameter estimation in multivariate complex models. Rather, it is a central property reflecting the nature of biological systems in general and the living cell in particular—enabling the duality of robustness and flexibility and underlying the exploration-exploitation dynamics. In other words, the biological cell itself is a *sloppy* system. Finding the 'stiff', *relevant* variables, is indeed one of the most urgent and important problems in our quest for a theory of the biological cell.

A plausible sketch of an *exploratory* adaptation process includes the following ingredients: (i) a driving force resulting from stress due to the mismatch between the inner state of the cell and the external and internal demands (e.g. inner metabolic fluxes not compatible with the environmental demands). This driving force is global, non-specific and works like 'heat', by causing large-scale changes in gene expression profiles. (ii) Exploratory dynamics, involving a plethora of emerging modes, which compete over the limited resources of the cell (e.g. modes of gene expression competing over limited numbers of polymerases, ribosomes, protein–DNA binding sites etc). These dynamics are based on the labile protein–DNA and protein–protein interactions due to their weak (compared to k_BT) intermolecular forces. (iii) Physiological selection of a set of compatible modes, which does not need to be unique, thanks to the degeneracy of the system. (iv) Finally, a drive-reduction mechanism [270] which alleviates the stress and leads to relaxation and stabilization of a cell state. A theoretical framework along these lines should go beyond the program formulated by D'arcy Thompson in his famous book *On Growth and Form* [271], which highlights respectable mathematical principles behind growth and form of biological systems, but remains relatively silent about the biology itself (not surprisingly, given the biology state of the art at his time). Unfortunately, without a specific theoretical framework, these ideas leave the concept of dynamic organization quite empty at this stage. A collective experimental and theoretical effort is required in order to crystallize such a framework.

In 1966, Waddington organized the first of four yearly meetings, aiming to discuss a sketch towards theoretical Biology [155]. Unfortunately, this effort to inquire into the principles underlying the organization of living matter did not proceed far beyond those meetings. It was soon overshadowed by the fast and swamping progress in molecular biology; the quest for universal principles was displaced by the 'hunting' of molecules. Notwithstanding the impressive advance in molecular biology, the last 50 years have taught us that progress in understanding biology, which is not synonymous with progress in medical applications, is severely impeded by the reductionist approach, focusing solely on cataloguing an ever increasing list of molecular processes, without a complementary effort in unraveling the system-level organization principles. What seems to be missing is indeed a unifying concept of organization. Regrettably, very little is left from Waddington's spirit [109]. The recent intense interest in living systems of people coming from disciplines outside of biology, in particular physicists and computer scientists has raised hope for reviving such a theoretical-based program. Indeed, the previous wave in the 1930s of physicists moving into biology, made a great impact [272]. The current multi-disciplinary effort is still on, so it might be too early to judge. However, until now it has not yet developed an original view, raising its own voice regarding the origin, evolution and development of biological systems as *natural phenomena,* independent of the tyranny of the molecular approach. I hope that the experimental framework and discussions presented in this article will stimulate readers to meet the challenge of developing a physics-based framework of cell-state organization.

For newcomers to biology, it helps to appreciate the formidable mission of facing the essential complexity of biology. This is echoed in the words of the famous physicists-becoming-biologist Max Delbruck[17] in describing his attempts to solve the '*riddle of life*' (virus replication), assuming initially that it '*... so simple a phenomenon that the answers cannot be hard to find; In a few months we will know.*' He later admits: '*Well, I made a slight mistake, I could not do it in a few months. Perhaps it will take a few decades, and perhaps it will take the help of a few dozen other people. But listen to what I have found, perhaps you will be interested to join me.*' We have to face the reality presented by the complexity in biology. It might take more than what Delbruck could have imagined to have a theory of the living cell. But whether it can be done at all, we will never know without trying. Physics can provide a fruitful framework and essential tools. There is no better way to state this than in the words of the great physicist Leo Szilard [277]: '*The mysteries of biology are no less deep than the mysteries of physics were one or two generations ago, and the tools are available to solve them provided only that we believe they can be solved*'. This article reflects my personal journey into the living cell, insisting on a physics approach. It has opened perspective on biology that has changed my view of some fundamental issues of biological organization. I invite the interested reader to join the effort. We are still far from solving the riddle of biology but the journey is certainly worth the effort, as the road itself is not less exciting than the target.

Acknowledgments

I am grateful to Shimon Marom, Kinneret Keren, Naama Brenner, Lindsay Moore, Yitzhak Rabin and Yoav Soen for discussions and remarks on the manuscript. I am also grateful to the referees that encouraged me in the process of publication and helped to improve the manuscript. This research was supported by the Israel Science Foundation and the Binational USA-Israel Foundation.

[17] 1946, in a lecture to the Harvey Society. See [272] p33–4.

References

[1] Gilbert S F and Epel D 2009 *Ecological Developmental Biology* (Sundeland, MA: Sinauer Associates, Inc.)
[2] West-Eberhard M J 2003 *Developmental Plasticity and Evolution* (Oxford: Oxford University Press)
[3] Wilkins A S 2002 *The Evolution of Developmental Pathways* (Sunderland, MA: Sinauer Associates, Inc.)
[4] Crick F 1970 Central dogma of molecular biology *Nature* **227** 561–3
[5] Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P 2008 *Molecular Biology of the Cell* 5th edn (New York: Garland Science)
[6] Gerhart J and Kirschner M 1997 *Cells, Embryos, and Evolution* (Malden, MA: Blackwell Science)
[7] Jablonka E and Raz G 2009 Transgenerational epigenetic inheritance: prevalence, mechanisms and implications for the study of heredity and evolution *Q. Rev. Biol.* **84** 131–76
[8] Edelman G M 1987 *Neural Darwinism: the Theory of Neuronal Group Selection* (New York: BasicBooks)
[9] Consortium E P 2012 An integrated encyclopedia of DNA elements in the human genome *Nature* **489** 57–74
[10] Graur D, Zheng Y, Price N, Azevedo R B R, Zufall R A and Elhaik E 2013 On the immortality of television sets: 'function' in the human genome according to the evolution-free gospel of ENCODE *Genome Biol. Evol.* **5** 578–90
[11] King M C and Wilson A C 1975 Evolution at two levels in humans and chimpanzees *Science* **188** 107–16
[12] Carroll S B 2005 Evolution at two levels: on genes and form *PLoS Biol.* **3** e245
[13] Carroll S B, Grenier J K and Weatherbee S D 2001 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Oxford: Blackwell Science)
[14] Davidson E H 2006 *The Regulatory Genome* (Amsterdam: Elsevier)
[15] Brenner S 2010 Sequences and consequences *Phil. Trans. R. Soc.* B **365** 207–12
[16] Noble D 2006 *The Music of Life: Biology Beyond Genes* (Oxford: Oxford University Press)
[17] Carroll S B 2008 Evo–Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution *Cell* **134** 25–36
[18] Wray G A 2007 The evolutionary significance of cis-regulatory mutations *Nat. Rev. Genet.* **8** 206–16
[19] Alonso C R and Wilkins A S 2005 The molecular elements that underlie developmental evolution *Nat. Rev. Genet.* **6** 709–15
[20] Stern D L and Orgogozo V 2008 The loci of evolution: how predictable is gene evolution? *Evolution* **62** 2155–77
[21] Erwin D H and Davidson E H 2009 The evolution of hierarchical gene regulatory networks *Nat. Rev. Genet.* **10** 141–8
[22] Arnone M I and Davidson E H 1997 The hardwiring of development: organization and function of genomic regulatory systems *Development* **124** 1851–64
[23] Landry J *et al* 2013 The genomic and transcriptomic landscape of a HeLa cell line G3 **3** 1213–24
[24] Skloot R 2010 *The Immortal Life of Henrietta Lacks* (New York: Random House)
[25] Maher C A and Wilson R K 2012 Chromothripsis and human disease: piecing together the shattering process *Cell* **148** 29–32
[26] Kirschner M and Gerhart J 1998 Evolvability *Proc. Natl Acad. Sci.* **95** 8420–7
[27] Drake J W, Charlesworth B, Charlesworth D and Crow J F 1998 Rates of spontaneous mutation *Genetics* **148** 1667–86
[28] Elena S F and Lenski R E 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation *Nat. Rev. Genet.* **4** 457–69
[29] Luria S E and Delbruck M 1943 Mutations of bacteria from virus sensitivity to virus resistance *Genetics* **28** 491–511
[30] Stolovicki E, Dror T, Brenner N and Braun E 2006 Synthetic gene recruitment reveals adaptive reprogramming of gene regulation in yeast *Genetics* **173** 75–85
[31] Carlson M 1999 Glucose repression in yeast *Curr. Opin. Microbiol.* **2** 202–7
[32] Johnston M, Flick J S and Pexton T 1994 Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression *Sacharomyces cerevisiae Mol. Cell. Biol.* **14** 3834–41
[33] Biggar S R and Crabtree G R 2001 Cell signaling can direct either binary or graded transcriptional responses *EMBO J.* **20** 3167–76
[34] Johnston M and Carlson M 1992 Regulation of carbon and phosphate utilization *The Molecular and Cellular Biology of the Yeast Saccaromyces: Gene Expression* ed E W Jones *et al* (New York: Cold Spring Harbor Laboratory Press) p 193
[35] Hinnebusch A G 1992 General and pathway-specific regulatory mechanisms controlling the synthesis of amino acid biosynthesis enzymes in S. cerevisiae *The Molecular and Cellular Biology of the Yeast Saccharomyces Gene Expression* vol 2, ed E W Jones *et al* (New York: Cold Spring Harbor Laboratory Press)
[36] Struhl K and Davis R W 1981 Transcription of the his3 gene region in *S. cerevisiae J. Mol. Biol.* **152** 535–52
[37] Novick A and Szilard L 1950 Experiments with the chemostat on spontaneous mutations of bacteria *Proc. Natl Acad. Sci. USA* **36** 708–19
[38] Smith H L and Waltman P 1995 *The Theory of the Chemostat Dynamics of Microbial Competition* (Cambridge: Cambridge University Press)
[39] Moore L S, Stolovicki E and Braun E 2013 Population dynamics of metastable growth-rate phenotypes *PLoS One* **8** e81671
[40] Stolovicki E and Braun E 2011 Collective dynamics of gene expression in cell populations *Plos One* **6** e20530
[41] David L, Stolovicki E, Haziz E and Braun E 2010 Inherited adaptation of genome-rewired cells in response to a challenging environment *HFSP J.* **4** 131–41
[42] Balaban N Q, Merrin J, Chait R, Kowalik L and Leibler S 2004 Bacterial persistence as a phenotypic switch *Science* **305** 1622–5
[43] Heidenreich E 2007 Adaptive mutation in *S. cerevisiae Crit. Rev. Biochem. Mol. Biol.* **42** 285–311
[44] Lenski R E and Travisano M 1994 Dynamics of adaptation and diversification: a 10 000-generation experiment with bacterial populations *Proc. Natl Acad. Sci. USA* **91** 6808–14
[45] Levin B R and Rozen D E 2006 Non-inherited antibiotic resistance *Nat. Rev. Microbiol.* **4** 556–62
[46] Paquin C and Adams J 1983 Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations *Nature* **302** 495–500
[47] Storchova Z, Gil A P R, Janderova B and Vondrejs V 1998 The involvement of the RAD6 gene in starvation-induced reverse mutation in *S. cerevisiae Mol. Gen. Genet.* **258** 546–52
[48] David L, Ben-Harosh Y, Stolovicki E, Moore L S, Nguyen M, Tamse R, Dean J, Mancera E, Steinmetz L M and Braun E 2013 Multiple genomic changes associated with reorganization of gene regulation and adaptation in yeast *Mol. Biol. Evol.* **30** 1514–26
[49] Moore L S, Wei W, Stolovicki E, Benbenishty T, Wilkening S, Steinmetz L M, Braun E and David L 2014 Induced

mutations in yeast cell populations adapting to an unforeseen challenge *PLoS One* **9** e111133
[50] Johnston M and Davis R W 1984 Sequences that regulate the divergent GAL1–GAL10 promoter in *Saccharomyces cerevisiae Mol. Cell. Biol.* **4** 1440–8
[51] Stergachis A B *et al* 2013 Exonic transcription factor binding directs codon choice and affects protein evolution *Science* **342** 1367–72
[52] Stern S, Dror T, Stolovicki E, Brenner N and Braun E 2007 Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge *Mol. Syst. Biol.* **3** 106
[53] Braun E and Brenner N 2004 Transient responses and adaptation to steady state in eukaryotic gene regulation system *Phys. Biol.* **1** 67
[54] Kaufmann B B, Yang Q, Mettetal J T and Oudenaarden A V 2007 Heritable stochastic switching revealed by single-cell genealogy *PLoS Biol.* **5** e239
[55] Lachman M and Jablonka E 1996 The inheritance of phenotypes: an adaptation to fluctuating environment *J. Theor. Biol.* **181** 1–9
[56] Kaern M, Elston T C, Blake W J and Collins J J 2005 Stochasticity in gene expression: from theories to phenotypes *Nat. Genet.* **6** 451–64
[57] Maheshri N and O'Shea E K 2007 Living with noisy genes: how cells function reliably with inherent variability in gene expression *Ann. Rev. Biophys. Biomol. Struct.* **36** 413–34
[58] Raj A and Oudenaarden A V 2008 Nature, nurture, or chance: stochastic gene expression and its consequences *Cell* **135** 216–26
[59] Booth I R 2002 Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress *Food Microbiol.* **78** 19–30
[60] Avery S V 2006 Microbial cell individuality and the underlying sources of heterogeneity *Nat. Rev. Microbiol.* **4** 577–87
[61] Eldar A and Elowitz M B 2010 Functional roles for noise in genetic circuits *Nature* **467** 167–73
[62] Salman H, Brenner N, Tung C-K, Elyahu N, Stolovicki E, Moore L, Libchaber A and Braun E 2012 Universal protein fluctuations in populations of microorganisms *Phys. Rev. Lett.* **108** 238105
[63] Tsuru S, Ichinose J, Sakurai T, Kashiwagi A, Ying B-W, Kaneko K and Yomo T 2009 Noisy cell growth rate leads to fluctuating protein concentration in bacteria *Phys. Biol.* **6** 036015
[64] Brenner N, Farkash K and Braun E 2006 Dynamics of protein distributions in cell populations *Phys. Biol.* **3** 172–82
[65] Bahar R *et al* 2006 Increased cell-to-cell variation in gene expression in ageing mouse heart *Nature* **441** 1011–4
[66] Friedlander T and Brenner N 2008 From cellular properties to population asymptotics in the population balance equation *Phys. Rev. Lett.* **101** 018104
[67] Huh D and Paulsson J 2011 Non-genetic heterogeneity from stochastic partitioning at cell division *Nat. Genet.* **43** 95–100
[68] Becskei A, Kaufmann B B and Oudenaarden A V 2005 Contributions of low molecule number and chromosomal positioning to stochasic gene expression *Nat. Genet.* **37** 937–44
[69] Raser J M and O'Shea E K 2004 Control of stochasticity in eukaryotic gene expression *Science* **304** 1811–4
[70] Newman J R S, Ghaemmaghami S, Ihmels J, Breslow D K, Noble M, DeRisi J L and Weissman J S 2006 Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise *Nature* **441** 840–6
[71] Muller-Hill B 1996 *The Lac Operon: a Short History of a Genetic Paradigm* (Berlin, New York: Walter de Gruyter)
[72] Sornette D 2006 *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* 2nd edn (Berlin: Springer)
[73] Katzir Y, Stolovicki E, Stern S and Braun E 2012 Cellular plasticity enables adaptation to unforeseen cell-cycle rewiring challenges *Plos One* **7** e45184
[74] Hartwell L H 2002 Nobel lecture. Yeast and cancer *Biosci. Rep.* **22** 373–94 Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, DI-060 Seattle, WA 98109-1024, USA
[75] Spellman P T, Sherlock G, Zhang M Q, Iyer V R, Anders K, Eisen M B, Brown P O, Botstein D and Futcher B 1998 Comprehensive identification of cell cycle-regulated genes of the yeast *S. cerevisiae* by microarray hybridization *Mol. Biol. Cell* **9** 3273–97
[76] Katzir Y, Elhanati Y, Averbukh I and Braun E 2013 Dynamics of the cell-cycle network under genome-rewiring perturbations *Phys. Biol.* **10** 066001
[77] Murray D B, Beckmann M and Kitano H 2007 Regulation of yeast oscillatory dynamics *Proc. Natl Acad. Sci. USA* **104** 2241–6
[78] Tu B P, Kudlicki A, Rowicka M and McKnight S L 2005 Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes *Science* **310** 1152–8
[79] Ahmed S and Brickner J H 2007 Regulation and epigenetic control of transcription at the nuclear periphery *Trends Genet.* **23** 396–402
[80] Allis C D, Jenuwein T and Reinberg D (ed) 2007 *Epigenetics* (Cold Spring Harbor: Cold Spring Harbor Laboratory Press)
[81] Brickner D G, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee P-C, Widom J and Brickner J H 2007 H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state *PloS Biol.* **5** e81
[82] Jablonka E and Lamb M J (ed) 2005 *Evolution in 4D* (Cambridge, MA: MIT Press)
[83] Jablonka E 2012 Epigenetic variations in heredity and evolution *Clin. Pharmacol. Ther.* **92** 683–8
[84] Rando O J and Verstrepen K J 2007 Timescales of genetic and epigenetic inheritance *Cell* **128** 655–68
[85] Alter O, Brown P O and Botstein D 2000 Singular value decomposition for genome-wide expression data processing and modeling *Proc. Natl Acad. Sci. USA* **97** 10101–6
[86] Daniels B C, Chen Y-J, Sethna J P, Gutenkunst R N and Myers C R 2008 Sloppiness, robustness, and evolvability in systems biology *Curr. Opin. Biotechnol.* **19** 389–95
[87] Fisher R A 1930 *The Genetical Theory of Natural Selection* (Oxford: Oxford University Press)
[88] Dobzhansky T 1970 *Genetics of the Evolutionary Process* (New York: Columbia University Press)
[89] Templeton A R 2006 *Population Genetics and Microevolutionary Theory* (Hoboken, NJ: Wiley)
[90] Gillespie J H 2004 *Population Genetics: a Concise Guide* (Baltimore, MD: Johns Hopkins University Press)
[91] Hunter B, Hollister J D and Bomblies K 2012 Epigenetic inheritance: what news for evolution? *Curr. Biol.* **22** R54–6
[92] Britten R J and Davidson E H 1971 Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty *Q. Rev. Biol.* **46** 111–38
[93] Hartwell L H, Hopfield J J, Leibler S and Murray A W 1999 From molecular to modular cell biology *Nature* **402** C47–52
[94] Jacob F 1977 Evolution and tinkering *Science* **196** 1161–6
[95] Hueber S D and Lohmann I 2008 Shaping segments: Hox gene function in the genomic age *BioEssays* **30** 965–79
[96] Foronda D, Navas L F D, Garaulet D L and Sanchez-Herreo E 2009 Function and specificity of Hox genes *Int. J. Dev. Biol.* **53** 1409–19

[97] Carr A and Biggin M D 1999 A comparison of *in vivo* and *in vitro* DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos *EMBO J.* **18** 1598–608

[98] Neumann J V 1958 Probabilistic logics and the synthesis of reliable organisms from unreliable components *Automata Studies* ed C E Shanon and J McCarthy (Princeton: Princeton University Press)

[99] Tononi G, Sporns O and Edelman G M 1999 Measures of degeneracy and redundancy in biological networks *Proc. Natl Acad. Sci. USA* **96** 3257–62

[100] Schaeffer H J and Weber M J 1999 Mitogen-activated protein kinases: specific messages from ubiquitous messengers *Mol. Cell. Biol.* **19** 2435–44

[101] Madhani H D and Fink G R 1998 The riddle of MAP kinase signaling specificity *TIG* **14** 151–5

[102] Schwartz M A and Madhani H D 2004 Principles of map kinase signaling specificity in *S. cerevisiae Annu. Rev. Genet.* **38** 725–48

[103] Ciliberti S, Martin O C and Wagner A 2007 Innovation and robustness in complex regulatory gene networks *Proc. Natl Acad. Sci. USA* **104** 13591–6

[104] Waddington C H 1953 Genetic assimilation of an acquired character *Evolution* **7** 118–26

[105] Waddington C H 1942 Canalization of development and the inheritance of acquired characters *Nature* **150** 563–5

[106] Schmalhausen I I 1949 *Factors of Evolution: the Theory of Stabilizing Selection* (Philadelphia, PA: Blakiston)

[107] Waddington C H 1956 Genetic assimilation of the bithorax phenotype *Evolution* **10** 1–13

[108] Wagner A 2005 *Robustness and Evolvability in Living Systems* (Princeton, NJ: Princeton University Press)

[109] Waddington C H 1957 *The Strategy of the Gene* (London: George Allen & Unwin Ltd)

[110] Flatt T 2005 The evolutionary genetics of canalization *Q. Rev. Biol.* **80** 287–316

[111] Gibson G and Dworkin I 2004 Uncovering cryptic genetic variation *Nat. Rev. Genet.* **5** 681–90

[112] Hermisson J and Wagner G P 2004 The population genetic theory of hidden variation and genetic robustness *Genetics* **168** 2271–84

[113] Lindquist S 2009 Protein folding sculpting evolutionary change *Cold Spring Harb. Symp. Quant. Biol.* **74** 103–8

[114] Queitsch C, Sangster T A and Lindquist S 2002 Hsp90 as a capacitor of phenotypic variation *Nature* **417** 618–24

[115] Rutherford S L and Lindquist S 1998 Hsp90 as a capacitor for morphological evolution *Nature* **396** 336–42

[116] Sangster T A, Lindquist S and Queitsch C 2004 Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance *BioEssays* **26** 348–62

[117] Sangster T A, Salathia N, Undurraga S, Milo R, Schellenberg K, Lindquist S and Queitsch C 2008 HSP90 affects the expression of genetic variation and developmental stability in quantitative traits *Proc. Natl Acad. Sci. USA* **105** 2963–8

[118] Specchia V, Piacentini L, Tritto P, Fanti L, D'Alessandro R, Palumbo G, Pimpinelli S and Bozzetti M P 2010 Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons *Nature* **463** 662–6

[119] Taipale M, Jarosz D F and Lindquist S 2010 HSP90 at the hub of protein homeostasis: emerging mechanistic insights *Nat. Rev. Mol. Cell Biol.* **11** 515–28

[120] Salathia N and Queitsch C 2007 Molecular mechanisms of canalization: Hsp90 and beyond *J. Biosci.* **32** 457–63

[121] Griffiths A J F, Miller J H, Suzuki D T, Lewontin R C and Gelbart W M 2000 *An Introduction to Genetic Analysis* 7th edn (New York: W H Freeman)

[122] Stern S, Fridmann-Sirkis Y, Braun E and Soen Y 2012 Epigenetically heritable alteration of fly development in response to toxic challenge *Cell Rep.* **1** 528–42

[123] Stern S, Snir O, Mizrachi E, Galili M, Zaltsman I and Soen Y 2014 Reduction in maternal Polycomb levels contributes to transgenerational inheritance of a response to toxic stress in flies *J. Physiol.* **592** 2343–55

[124] Giaever G *et al* 2002 Functional profiling of the *S. cerevisiae* genome *Nature* **418** 387–91

[125] Winzeler E A *et al* 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis *Science* **285** 901–6

[126] Thatcher J W, Shaw J M and Dickinson W J 1998 Marginal fitness contributions of nonessential genes in yeast *Proc. Natl Acad. Sci. USA* **95** 253–7

[127] Huang A C, Hu L, Kauffman S A, Zhang W and Shmulevich I 2009 Using cell fate attractors to uncover transcriptional regulation of HL60 neutrophil differentiation *BMC Syst. Biol.* **3** 1–13

[128] Huang S, Eichler G, Bar-Yam Y and Ingber D E 2005 Cell fates as high-dimensional attractor states of a complex gene regulatory network *Phys. Rev. Lett.* **94** 128701–4

[129] Kauffman S A 1993 *The Origins of Order; Self-Organization and Selection in Evolution* (New York: Oxford University Press)

[130] Serra R, Villani M, Barbieri A, Kauffman S A and Colacci A 2010 On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types *J. Theor. Biol.* **265** 185–93

[131] Wang J, Zhang K, Xu L and Wang E 2011 Quantifying the Waddington landscape and biological paths for development and differentiation *Proc. Natl Acad. Sci. USA* **108** 8257–62

[132] Fontana W and Schuster P 1998 Shaping space the possible and the attainable in RNA genotypephenotype mapping *J. Theor. Biol.* **194** 491–515

[133] Fontana W and Schuster P 1998 Continuity in evolution: on the nature of transitions *Science* **280** 1451–5

[134] Stadler B M R, Stadler P F, Wagner G P and Fontana W 2001 The topology of the possible: formal spaces underlying patterns of evolutionary change *J. Theor. Biol.* **213** 241–74

[135] Jablonka E and Lamb M 1995 *Epigenetic Inheritance and Evolution: the Lamarckian Dimension* (New York: Oxford University Press)

[136] Furrow R, Christiansen F B and Feldman M W Environment-sensitive epigenetics and the heritability of complex diseases *Genetics* **189** 1377–87

[137] Lewontin R C 2001 Gene, organism and environment *Cycles of Contingency: Developmental Systems and Evolution* ed S Oyama *et al* (Cambridge, MA: MIT Press)

[138] Ariew A and Lewontin R C 2004 The confusions of fitness *Br. J. Phil. Sci.* **55** 347–63

[139] Laland K N, Odling-Smee J and Gilbert S F 2008 EvoDevo and Niche construction: building bridges *J. Exp. Zool.* **310B** 549–66

[140] Odling-Smee F J, Laland K N and Feldman M W 2003 Niche construction: the neglected process in evolution (Princeton, NJ: Princeton University Press)

[141] Ptashne M 2007 On the use of the word 'epigenetic' *Curr. Biol.* **17** R233–6

[142] Ptashne M 2013 Faddish stuff: epigenetics and the inheritance of acquired characteristics *FASEB J.* **27** 1–2

[143] Jablonka E and Lamb M J 2006 The changing concept of epigenetics *Ann. N Y Acad. Sci.* **981** 82–96

[144] Jablonka E and Lamb M J 2008 Soft inheritance: challenging the modern synthesis *Genet. Mol. Biol.* **31** 389–95

[145] Kirschner M and Gerhart J 1998 Comment on 'Epigenetic inheritance in evolution' *J. Evol. Biol.* **11** 213–7

[146] Richards E J 2006 Inherited epigenetic variation —revisiting soft inheritance *Nat. Rev. Genet.* **7** 395–401
[147] Gissis S B and Jablonka E 2011 *Transformations of Lamarckism: From Subtle Fluids to Molecular Biology* (Cambridge, MA: MIT Press)
[148] Waddington C H 1942 The epigenotype *Endeavour* **1** 18–20 Reprinted in 2012 Int. J. Epidemiol. **41** 10–13
[149] Nanney D L 1958 Epigenetic control systems *Proc. Natl Acad. Sci. USA* **44** 712–7
[150] Haig D 2012 Commentary: the epidemiology of epigenetics *Int. J. Epidemiol.* **41** 16–20
[151] Jablonka E 2013 Epigenetic inheritance and plasticity: the responsive germline *Prog. Biophys. Mol. Biol.* **111** 99–107
[152] Fridmann-Sirkis Y, Stern S, Elgart M, Galili M, Zeisel A, Shental N and Soen Y 2014 Delayed development induced by toxicity to the host can be inherited by a bacterial-dependent, transgenerational effect *Front. Genet.* **5** 27
[153] Soen Y 2014 Environmental disruption of Host-Microbe co-adaptation as a potential driving force in evolution *Front. Genet.* **5** 168
[154] Prehn R T 1994 Cancers beget mutations versus mutations beget cancers *Cancer Res.* **54** 5296–300
[155] Waddington C H 1968 Towards a theoretical biology *The Origin of Life* ed C H Waddington (New Brunswick: Transaction)
[156] Kendal W S and Frost P 1988 Pitfalls and practice of Luria–Delbrück fluctuation analysis: a review *Cancer Res.* **48** 1060–5
[157] Cairns J and Foster P L 1991 Adaptive reversion of a frameshift mutation in *E. coli Genetics* **128** 695–701
[158] Cairns J, Overbaugh J and Miller S 1988 The origin of mutants *Nature* **335** 142–5
[159] Roth J R, Kugelberg E, Reams A B, Kofoid E and Andersson D I 2006 Origin of mutations under selection: the adaptive mutation controversy *Annu. Rev. Microbiol.* **60** 477–501
[160] Marini N, Matmati N and Morpurgo G 1999 Starvation in yeast increase non-adaptive mutation *Curr. Genet*. **35** 77–81
[161] Perfeito L, Fernandes L, Mota C and Gordo I 2007 Adaptive mutations in bacteria: high rate and small effects *Science* **317** 813–5
[162] Leea H, Popodib E, Tanga H and Fosterb P L 2012 Rate and molecular spectrum of spontaneous mutations in the bacterium *E. coli* as determined by whole-genome sequencing *Proc. Natl Acad. Sci. USA* **109** E2774–83
[163] Barrick J E and Lenski R E 2013 Genome dynamics during experimental evolution *Nat. Rev. Genet.* **14** 827–39
[164] Sniegowski P D and Gerrish P J 2010 Beneficial mutations and the dynamics of adaptation in asexual populations *Phil. Trans. R. Soc.* B **365** 1255–63
[165] Kondrashov F A and Kondrashov A S 2010 Measurements of spontaneous rates of mutations in the recent past and the near future *Phil. Trans. R. Soc.* B **365** 1169–76
[166] Lang G I and Murray A W 2008 Estimating the per-base-pair mutation rate in the yeast *S. cerevisiae Genetics* **178** 67–82
[167] Martincorena I and Luscombe N M 2013 Non-random mutation: the evolution of targeted hypermutation and hypomutation *BioEssays* **35** 123–30
[168] Lynch M 2010 Evolution of the mutation rate *Trends Genet.* **26** 345–52
[169] Orr H A 2005 The genetic theory of adaptation: a brief history *Nat. Rev. Genet.* **6** 119–27
[170] Desai M M, Fisher D S and Murray A W 2007 The speed of evolution and maintenance of variation in asexual populations *Curr. Biol.* **17** 385–94
[171] Fisher D S 2011 Leading the dog of selection by its mutational nose *Proc. Natl Acad. Sci. USA* **108** 2633–4
[172] Weissmana D B, Desai M M, Fisher D S and Feldmana M W 2009 The rate at which asexual populations cross fitness valleys *Theor. Population Biol.* **75** 286–300
[173] Lenski R E 2011 Evolution in action: a 50 000-Generation Salute to Charles Darwin *Microbe* **6** 30–3
[174] Philippe N, Crozat E, Lenski R E and Schneider D 2007 Evolution of global regulatory networks during a long-term experiment with *E. coli BioEssays* **29** 846–60
[175] Cooper T F, Rozen D E and Lenski R E 2003 Parallel changes in gene expression after 20 000 generations of evolution in *E. coli Proc. Natl Acad. Sci. USA* **100** 1072–7
[176] Paquin C E and Adams J 1983 Relative fitness can decrease in evolving asexual populations of *S. cerevisiae Nature* **306** 368–71
[177] Lang G I, Rice D P, Hickman M J, Sodergren E, Weinstock G M, Botstein D and Desai M M 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations *Nature* **500** 571–6
[178] Zeyl C 2007 Evolutionary genetics: a piggyback ride to adaptation and diversity *Curr. Biol.* **17** R333–5
[179] Kao K C and Sherlock G 2008 Molecular characterization of clonal interference during adaptive evolution in asexual populations of *S. cerevisiae Nat. Genet.* **40** 1499–504
[180] Vickaryous M K and Hall B K 2006 Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest *Biol. Rev*. **81** 425–55
[181] Leychkis Y, Munzer S R and Richardson J L 2009 What is stemness? *Stud. Hist. Philos. Biol. Biomed. Sci.* **40** 312–20
[182] Ivanova N B, Dimos J T, Schaniel C, Hackney J A, Moore K A and Lemischka I R 2002 A stem cell molecular signature *Science* **298** 601–4
[183] Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan R C and Melton D A 2002 'Stemness': transcriptional profiling of embryonic and adult stem cells *Science* **298** 597–600
[184] Fortunel N O *et al* 2003 Comment on '"stemness": transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature' (I) *Science* **302** 393b
[185] Evsikov A V and Solter D 2003 Comment on '"stemness": transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature' (II) *Science* **302** 393c
[186] Ivanova N B, Dimos J T, Schaniel C, Hackney J A and Moore K A 2003 Response to comments on '"stemness": transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature' *Science* **302** 393d
[187] Vogel G 2003 'Stemness' genes still elusive *Science* **302** 371
[188] Maher B 2008 The case of missing heritability *Nature* **456** 18–21
[189] Halley J D, Winkler D A and Burden F R 2008 Toward a Rosetta stone for the stem cell genome: stochastic gene expression, network architecture, and external influences *Stem Cell Res.* **1** 157–68
[190] Amariglio N *et al* 2005 Design principle of gene expression used by human stem cells: implication for pluripotency *FASEB J.* **19** 147–9
[191] Chang H H, Hemberg M, Barahona M, Ingber D E and Huang S 2008 Transcriptome-wide noise controls lineage choice in mammalian progenitor cells *Nature* **453** 544–8
[192] Tokuriki N and Tawfik D S Protein dynamism and evolvability *Science* **324** 203–7
[193] Zhou Q and Melton1 D A 2008 Extreme makeover: converting one cell into another *Cell Stem Cell* **3** 382–8
[194] Yamanaka S 2012 Induced pluripotent stem cells: past, present, and future *Cell Stem Cell* **10** 678–84
[195] Yamanaka S and Blau H M 2010 Nuclear reprogramming to a pluripotent state by three approaches *Nature* **465** 704–12

[196] Holmberg J and Perlmann T 2012 Maintaining differentiated cellular identity *Nat. Rev. Genet.* **13** 429–39
[197] González F, Boué S and Belmonte J C I 2011 Methods for making induced pluripotent stem cells: reprogramming à la carte *Nat. Rev. Genet.* **12** 231–42
[198] Graf T and Enver T 2009 Forcing cells to change lineages *Nature* **462** 587–94
[199] Cohen D E and Melton D 2011 Turning straw into gold: directing cell fate for regenerative medicine *Nat. Rev. Genet.* **12** 243–52
[200] Hanna J H, Saha K and Jaenisch R 2010 Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues *Cell* **143** 508–25
[201] Wilmut I, Sullivan G and Chambers I 2011 The evolving biology of cell reprogramming *Phil. Trans. R. Soc.* B **366** 2183–97
[202] Okita K and Yamanaka S 2011 Induced pluripotent stem cells: opportunities and challenges *Phil. Trans. R. Soc.* B **366** 2198–207
[203] Efe J A and Ding S 2011 The evolving biology of small molecules: controlling cell fate and identity *Phil. Trans. R. Soc.* B **366** 2208–21
[204] Yamanaka S 2009 Elite and stochastic models for induced pluripotent stem cell generation *Nature* **460** 49–52
[205] Ptashne M 2004 *A Genetic Switch: Phage Lambda Revisited* 3rd edn (Cold Spring Harbor: Cold Spring Harbor Laboratory Press)
[206] Huang S 2011 Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways *Phil. Trans. R. Soc.* B **366** 2247–59
[207] Huang S 2009 Reprogramming cell fates: reconciling rarity with robustness *BioEssays* **31** 546–60
[208] Kaneko K 2006 *An Introduction to Complex Systems Biology* (Berlin: Springer)
[209] Ribeiroa A S and Kauffman S A 2007 Noisy attractors and ergodic sets in models of gene regulatory networks *J. Theor. Biol.* **247** 743–55
[210] Corson F and Siggia E D 2012 Geometry, epistasis, and developmental patterning *PNAS* **109** 5568–75
[211] Stein D L and Newman C M 2013 *Spin Glasses and Complexity* (Princeton, NJ: Princeton University Press)
[212] Stein D L and Newman C M 2012 Rugged landscapes and timescale distributions in complex systems *Emerg. Complex. Comput.* **5** 65–79
[213] Kurchan J and Laloux L 1996 Phase space geometry and slow dynamics *J. Phys.* A: *Math. Gen*. **29** 1929–48
[214] Jordana D, Kuehna S, Katifori E and Leibler S 2013 Behavioral diversity in microbes and low-dimensional phenotypic spaces *Proc. Natl Acad. Sci. USA* **110** 14018–23
[215] Parshani R, Dickison M, Cohen R, Stanley H E and Havlin S 2010 Dynamic networks and directed percolation *EPL* **90** 38004
[216] Schwartz I B and Shaw L B 2010 Rewiring for adaptation *Physics* **3** 17
[217] Mukherjee S 2010 *The Emperor of All Maladies: a Biography of Cancer* (New York: Scribner)
[218] Zhang Q and Austin R H 2012 Physics of cancer: the impact of heterogeneity *Annu. Rev. Condens. Matter Phys*. **3** 363–82
[219] Huang S, Ernberg I and Kauffman S 2009 Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective *Semin. Cell Dev. Biol.* **20** 869–76
[220] Lambert G, Estévez-Salmeron L, Oh S, Liao D, Emerson B M, Tlsty T D and Austin R H 2011 An analogy between the evolution of drug resistance in bacterial communities and malignant tissues *Nat. Rev. Cancer* **11** 375–82
[221] Smithers D W 1962 Cancer an attack on cytologism *Lancet* **279** 493–9
[222] Moss L 2003 *What Genes Can't Do* (Cambridge, MA: MIT Press)
[223] Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T and Thun M J 2008 Cancer statistics, 2008 *CA Cancer J. Clin.* **58** 71–96
[224] Bailar J C and Gornik H L 1997 Cancer undefeated *N. Eng. J. Med.* **336** 1569–74
[225] Sonnenschein C and Soto A N 2013 The Vestigial Remains of a Fallen Theory *Genetic Explanations Sense and Nonsense* ed S Krimsky and J Gruber (Cambridge, MA: Harvard University Press)
[226] Gottlieb B, Beitel L K and Trifiro M 2007 Will knowledge of human genome variation result in changing cancer paradigms? *BioEssays* **29** 678–85
[227] Lewontin R 2009 Where are the genes? *GeneWatch*
[228] Thomas R K *et al* 2007 High-throughput oncogene mutation profiling in human cancer *Nat. Genet.* **39** 347–51
[229] Fearon E R and Vogelstein B 1990 A genetic model for colorectal tumorigenesis *Cell* **61** 759–67
[230] Kinzler K W and Vogelstein B 1996 Lessons from hereditary colorectal cancer *Cell* **87** 159–70
[231] Jackson A L and Loeb L A 1998 The mutation rate and cancer *Genetics* **148** 1483–90
[232] Lynch M 2010 Rate, molecular spectrum, and consequences of human mutation *Proc. Natl Acad. Sci. USA* **107** 961–8
[233] Clevers H 2011 The cancer stem cell: premises, promises and challenges *Nat. Med.* **17** 313–9
[234] Roesch A, Fukunaga-Kalabis M, Schmidt E C, Zabierowski S E, Brafford P A, Vultur A, Basu D, Gimotty P, Vogt T and Herlyn M 2010 A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth *Cell* **141** 583–94
[235] Abollo-Jiménez F, Jiménez R and Cobaleda C 2010 Physiological cellular reprogramming and cancer *Semin. Cancer Biol.* **20** 98–106
[236] Kalmar T, Lim C, Hayward P, Munoz-Descalzo S, Nichols J, Garcia-Ojalvo J and Arias A M 2010 Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells *PLoS Biol.* **7** e1000149
[237] Mondal S and Heidelberger C 1970 *In vitro* malignant transformation by methylcholanthrene of the progeny of single cells derived from C3H mouse prostate *Proc. Natl Acad. Sci. USA* **65** 219–25
[238] Soto A M and Sonnenschein C 2004 The somatic mutation theory of cancer: growing problems with the paradigm? *BioEssays* **26** 1097–107
[239] Hanahan D and Weinberg R A 2011 Hallmarks of cancer: the next generation *Cell* **144** 646–74
[240] Harris H 2005 A long view of fashions in cancer research *BioEssays* **27** 833–8
[241] Mechler B M 1994 When the control is lacking—the role of tumour suppressor genes in cancer development *J. Biosci*. **19** 537–56
[242] Strand D, Raska I and Mechler B M 1994 The drosophila lethal(2)giant larvae tumor suppressor protein is a component of the cytoskeleton *J. Cell Biol.* **127** 1345–60
[243] McCullough K D, Coleman W B, Smith G J and Grisham J W 1997 Age-dependent induction of hepatic tumor regression by the tissue microenvironment after transplantation of neoplastically transformed rat liver epithelial cells into the liver *Cancer Res.* **57** 1807–13
[244] Farber E 1987 Experimental induction of hepatocellular carcinoma as a paradigm for carcinogenesis *Clin. Physiol. Biochem*. **5** 152–9
[245] Farber E 1991 Clonal adaptation as an important phase of hepatocarcinogenesis *Cancer Biochem. Biophys*. **12** 157–65

[246] Rubin H, Yao A and Chow M 1995 Heritable, population-wide damage to cells as the driving force of neoplastic transformation *Proc. Natl Acad. Sci. USA* **92** 4843–7
[247] Park C, Bissel M J and Barcellos-Hoff M H 2000 The influence of the micro environment on the malignant phenotype *Mol. Med. Today* **6** 324–9
[248] Soto A M and Sonnenschein C 2011 The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory *BioEssays* **33** 332–40
[249] Sonnenschein C and Soto A M 2008 Theories of carcinogenesis: an emerging perspective *Semin. Cancer Biol*. **18** 372–7
[250] Kupiec J-J 2009 *The Origin of Individuals* (Singapore: World Scientific)
[251] Wilkins A S 2010 The enemy within: an epigenetic role of retrotransposons in cancer initiation *BioEssays* **32** 856–65
[252] Gerlinger M *et al* 2012 Intratumor heterogeneity and branched evolution revealed by multiregion sequencing *N. Eng. J. Med.* **366** 883–92
[253] Maffini M V, Soto A M, Calabro J M, Ucci A A and Sonnenschein C 2004 The stroma as a crucial target in rat mammary gland carcinogenesis *J. Cell Sci.* **117** 1495–502
[254] Stewart T A and Mintz B 1981 Successful generations of mice produced from an established culture line of euploid teratocarcinoma cells *Proc. Natl Acad. Sci. USA* **78** 6314–8
[255] Maffini M V *et al* 2005 Stormal regulation of neoplastic development: age-dependent normalization of neoplastic mammary cells by mammary stroma *Am. J. Pathol.* **167** 1405–10
[256] Booth B W *et al* 2011 The normal mammary microenvironment suppresses the tumorigenic phenotype of mouse mammary tumor virus-neu-transformed mammary tumor cells *Oncogene* **30** 679–89
[257] DiBerardino M A, Orr N H and McKinnell R G 1986 Feeding tadpoles cloned from Rana erythrocyte nuclei *Proc. Natl Acad. Sci. USA* **83** 8231–4
[258] McKinnell R G *et al* 1993 Genomic plasticity of the Lucke renal carcinoma: a review *Int. J. Dev. Biol.* **37** 213–9
[259] Weaver V M, Petersen O W, Wang F, Larabell C A, Briand P, Damsky C and Bissell A M J 1997 Reversion of the malignant phenotype of human breast cells in 3D Culture and *in vivo* by integrin blocking antibodies *J. Cell Biol.* **137** 231–45
[260] Diez-Torre A, Andrade R, Eguizabal C, Lopez E, Arluzea J, Silio M and Arechaga J 2009 Reprogramming of melanoma cells by embryonic microenvironments *Int. J. Dev. Biol*. **53** 1563–8
[261] McCullough K D, Coleman W B, Ricketts S L, Wilson J W, Smith G J and Grisham J W 1998 Plasticity of the neoplastic phenotype *in vivo* is regulated by epigenetic factors *Proc. Natl Acad. Sci. USA* **95** 15333–8
[262] Abbott D E, Bailey C M, Postovit L M, Seftor E A, Margaryan N, Seftor R E and Hendrix M J 2008 The epigenetic influence of tumor and embryonic microenvironments: how different are they? *Cancer Microenviron.* **1** 13–21
[263] Lee M, Seftor E A, Bonde G, Cornell R A and Hendrix M J 2005 The fate of human malignant melanoma cells transplanted into zebrafish embryos: assessment of migration and cell division in the absence of tumor formation *Dev. Dyn*. **233** 1560–70
[264] Schrodinger E 1944 *What is Life?* (Cambridge: Cambridge University Press)
[265] Anderson P W 1972 More is different *Science* **177** 393–6
[266] Corte L, Chaikin P M, Gollub J P and Pine D J 2008 Random organization in periodically driven systems *Nat. Phys.* **4** 420–4
[267] Muller G B and Newman S A (ed) 2003 Origination of organismal form: beyond the gene in developmental and evolutionary biology (Cambridge, MA: MIT Press)
[268] Gutenkunst R N, Waterfall J J, Casey F P, Brown K S, Myers C R and Sethna J P 2007 Universally sloppy parameter sensitivities in systems biology models *PLoS Comput. Biol.* **3** 1871–88
[269] Brown K S and Sethna J P 2003 Statistical mechanical approaches to models with many poorly known parameters *Phys. Rev. E* **68** 021904
[270] Shahaf G and Marom S 2001 Learning in networks of cortical neurons *J. Neurosci.* **21** 8782–8
[271] Thompson D W 1942 *On Growth and Form* (Cambridge: Cambridge University Press)
[272] Judson H F 1996 *The Eighth Day of Creation: Makers of the Revolution in Biology* Expanded edn (New York: Cold Spring Harbor Laboratory Press)
[273] Keller E F 1983 *A Feeling for the Organism* (New York: Henry Holt)
[274] Woese C R 2004 A new biology for a new century *Microbiol. Mol. Biol. Rev.* **68** 173
[275] Edelman G M 1992 *Bright Air, Brilliant Fire* (New York: BasicBooks)
[276] Kepler J 1611 *Strena Seu de Nive Sexangula* (Frankfurt, Germany: Gottfried Tampach) Reprinted in 2010 *A New Year's Gift: On The Six-Cornered Snowflake* (transl. by J Bromberg) (Philadelphia, PA: Paul Dry Books)
[277] Lanouette W 2013 *Genius in the Shadows* (New York: Skyhorse) p 480