Special Issue, Single-cell Analysis: Epistemological Inquiries

Vol. 6, No. 2 (2023)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/18306

# The Contrasting Role of Single-cell Studies in the Theoretical Debate on Determinism in Molecular Biology

*Thomas Heams*[ab*]

[a] *AgroParisTech, département SVS, Palaiseau, France*

[b] *INRAE, UMR Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France*

***Corresponding author:** Thomas Heams, Email: thomas.heams@*agroparistech.fr*

**Abstract**

In experimental biology, the last three decades have seen a flood of techniques dedicated to study biological phenomena at the single-cell level, and this article aims to reflect on how these technical advances can contribute to the renewal of theoretical perspectives in biology. The case studied here is that of the critique of the genetic determinism of molecular biology. The demonstration of unpredictability in gene expression at the single-cell level, a phenomenon known as stochastic gene expression, even in clonal populations, initially appeared to be a decisive indication that cells do not actually behave as predicted by deterministic frameworks. However, single-cell techniques have also revealed other sources of genetic variation that nuance this picture. The role of single-cell studies thus appears contrasted, and can be used to support or challenge the paradigm of genetic determinism (GDP). This opens up a more general debate on the practical ability of molecular biologists to criticize their own paradigms.

**Keywords:** single-cell studies, molecular biology, genetic variation, genetic determinism paradigm.

**Citation:** Heams, T 2023, "The contrasting role of single-cell studies in the theoretical debate on determinism in molecular biology", *Organisms: Journal of Biological Sciences*, vol. 6, no. 2, pp. 31–42. DOI: 10.13133/2532-5876/18306

## Introduction

The aim of this article is to observe and analyze the convergence of two different phenomena in biology: on the one hand, a long-running debate on the relevancy of determinism in molecular biology, and, on the other hand, the rise of single-cell studies, based on techniques allowing to analyze cells not just at the population level, but actually one by one. Although these techniques are often adaptations at the cellular level of previously mastered molecular biology techniques, they have not spontaneously appeared as neutral and inevitable technological improvements of the former ones: they have also, and probably mainly, responded to a growing theoretical interest for the single-cell scale in organisms. The desire to know each cell more and more precisely in its context, and the intuition that certain biological questions would need to achieve such a level of precision, are fundamental and not merely technical questions. They also have played a role in—and thus form the first link between—theory and practice in molecular biology. Further, these techniques appeared crucial for criticizing the theoretical soundness of a key dimension of molecular biology, namely its deterministic foundations. This article aims to explore the extent to which single-cell studies have been

mobilized to produce this critique, and the extent to which they have also shown their limitations in doing so. This will in turn raise questions about the nature of plastic thinking in science, as in the case of the currently dominant paradigm in molecular and cellular biology. Is its ability to resist and even metabolize the various criticisms a sign of its enduring relevance? Or is it rather a worrying symptom of a way of doing science that prefers to construct narratives and feed them, rather than consistently keeping self-critical and aware of its own aporias?

## 1. Thirty Years of Single-cell Approaches in the Era of Genetic Determinism

It is worth clarifying at the outset what this article means by single-cell techniques. These techniques have been developed for some thirty years. Of course, observations at the single-cell level have been made for a long time: microscopic observations and histological sections, among many other approaches, are sometimes as old as biology. Single-cell techniques are in fact a set of molecular and cellular biology techniques originally designed for use on populations of molecules or cells. Their advancement in precision and efficiency allowed use on individual cells. This is the case, for example, of PCR, which was developed in the early 1980s, and which, thanks to technical improvements, has been used on isolated cells in the following decade, as reviewed by Kehr (2003). In this context, single-cell approaches are understood as those that enable the identification of content and expression of genomes on a single-cell scale. In concrete terms, these are mainly genetic material amplification techniques such as PCR (for DNA) and RT-PCR (for RNA), and the various -omics approaches: genomics (Gawad *et al.* 2016), transcriptomics (Longo & Hasty 2006; Kolodziejczyk *et al.* 2015), proteomics, metabolomics and epigenomics (Bheda & Shneider 2014), and all these combined (Wang & Bodovitz 2010). Single cell techniques can also rely on fluorescent markers on living cells (Elowitz *et al.* 2002) and/or take advantage of recent development in flow cytometry, a decades-old technique able to sort cells one by one, now upgraded with new analysis markers and methods (Di Carlo & Lee 2006), and of the use of microfluidics (Templer & Ces 2006).

Presented this way, we can already see that single-cell techniques are tools that molecular biology research has used to reinvest in a long-neglected scale. To better understand such negligence, it appears necessary to explicit some important epistemological driving forces that are at stake in this discipline. Indeed, molecular biology, as a discipline, is based on an instructionist and deterministic paradigm, which is also the starting point of its research program. It can be stated as follows: in multicellular organisms, cells are seen as sending and responding to intracellular, intercellular or environmental *instructions*, *determined* through precisely regulated molecular reactions, and the functioning of the multicellular organism relies on intense intercellular coordination through the proper integration of these signals. Obviously, molecular biology first focuses on *molecules*, but with the goal of integrating these molecular interactions into a broader picture at the cellular and multicellular level, for which this paradigm is the consensus framework. Here, the somatic cells of the multicellular organism are assumed to be genetically identical, and the evolutionary rationale behind this coordinated functioning is that it is a profitable strategy, in Dawkins' terms, for maximizing the diffusion of each cell's genes via those that will be transmitted by the gametes emitted by the organism.

That this starting point should be considered as a paradigm, as will be the case in the remainder of this article, may seem obvious to some. However, others may see it as a strong stance that needs to be justified—even more so as it also implies underlining what we mean by molecular biology. Indeed, molecular biology has at least two facets: (1) on the surface, it is a practical, experimental discipline, characterized by the level at which it proposes answers to biological questions—that of biomolecules. In this sense, molecular biology has a pragmatic dimension that may seem at odds with the existence of rigidly fixed paradigms, especially if they are not explicit. This facet of molecular biology undeniably exists, and in fact, most molecular biologists do not engage in theoretical debates on the fundamental motivations of their discipline, which are rather restricted to a small number of research groups. Such is the case with the debate that will be the subject of this article, whose audience is as limited as its importance is crucial. Seen under this light, the very idea of molecular biology existing under the imperium of a paradigm may seem critical, even misleading.

In fact, molecular biology is not just that. It is also (2) a DNA- (and RNA-) centric vision of biology with a specific history, and its own sequencing and modifications. It postulates *de facto* that DNA is the organizing principle of living organisms, as evidenced by its lexicon imported from computer science: genetic code, genetic program, etc. It is therefore intrinsically deterministic (Noble 2006). This vision of biology, which has produced countless experimental results, is structured around a strong assumption: the search for precise explanations in biology is supposed to find its answers in the precise functioning of organisms. However, this assumption articulates two very different ideas: while it is logical that a discipline should seek precise answers to explain phenomena, there is nothing to imply that the objects studied provide these answers insofar as they themselves are necessarily precise. The indissoluble link between these two levels of precision indicates that molecular biology relies on strong but implicit theoretical presuppositions, the foundation of which is the aforementioned starting point. Molecular biology is also the dominant vision of biology at present, so we feel that it deserves to be called the Genetic Determinism Paradigm (GDP), even if its determinism is probably less rigid than those of other scientific disciplines. The fact that many practitioners do not necessarily experience it as deterministic is not in itself a contradiction or a counter-argument. In the Kuhnian sense of the term, the establishment of a discipline paradigm is followed by a phase of normal science that the majority of scientists exploit and take for granted without questioning or even imagining that it could be questioned, until the paradigm finally enters into crisis.

Acknowledging this deterministic framework clarifies why, in this context, the observation of individual and/or single cells has long been regarded as anecdotal: apart from the technical challenge involved, all cells of a given organism were assumed to be genetically identical and all cells of a given tissue were assumed to behave identically on first approximation. Indeed GDP was based on a postulate of homogeneity: all cells in the same tissue, receiving the same signals, react in a similar way because they possess the same genes. Hence, in order to find out how much RNA or protein is produced in a cell, this theoretical framework measures such a quantity in a large sample of cells and deduces the individual quantity by a simple division. In so doing, GDP largely overlooks any consideration of intercellular variability in gene expression other than residual.

Nevertheless, single-cell observations made sense in certain areas of experimental biology, notably in the study of the early stages of embryonic development, of precisely located small groups of neurons, which by definition involve minimal cell numbers. The 1990s saw the first significant wave of publications on single-cell approaches, notably by single-cell PCR (amplification of DNA enabling an approach to the genetic content) and then by RT-PCR (amplification of RNA, enabling an approach to the genetic content use) (Kumazaki *et al.* 1994). In the context of molecular biology's general research program, these techniques were primarily designed to increase precision at a scale that had long been inaccessible. While the majority of these articles were mostly technical, some others, using complementary techniques as fluorescent in situ hybridization, soon started to challenge experimentally the deterministic nature of gene regulation (Wijgerde *et al.* 1995).

Over the past thirty years, increasingly comprehensive atlases and databases documenting hundreds of cell types have been made available, providing access not only to genomes, but also to the transcriptomes, proteomes and epigenomes of particular cells, representing their cell type (Elmentaite *et al.* 2022). This is the continuation of one of the great projects of molecular biology, which is the mapping of living organisms on all scales, from the genomes of species, to the transcriptomes of cell types, to the microbiota of various environments. Concentrated on the so-called informational molecules (DNA and RNA), this trend towards producing collections, which is as old as biology itself, has found a new lease of life in the era of Big Data, where the single-cell scale appears as an additional dimension in the completeness of this undertaking. In this context, single-cell studies have continuously addressed an ever-wider range of biological issues, either fundamental or applied, from microbiology and plant sciences to medicine (where an intense focus of research is devoted to cancer research and tumor heterogeneity [Liang & Fu 2017]).

## 2. Challenging GDP

These techniques appeared in a period of GDP's triumph and expansion, which could pragmatically be

called the era of the genetic program (for and in-depth critical analysis, see [Noble 2006]). The very existence of a genetic program (etymologically: "written in advance") encoded in the genome of organisms is the dominant idea at the time and the GDP core, with strong or weak nuances to this largely consensual principle. The more nuanced versions admit that this program is either more flexible and can be reprogrammed (see the research on stem cells), or that it is not strictly genetic (see the research on epigenetics), or that it is open to external influences (see the criticism of the "all-genetic" approach). However, the fact remains that, at the time, and most probably even today, it was difficult to propose approaches that would radically dispense with the notion of any genetic program. In fact, single-cell approaches soon produced results that cracked the epistemological edifice of GDP. Further, in the same period, Elowitz and colleagues published a paper that, right from its title, caused a stir in the scientific community working on gene expression (Elowitz *et al.* 2002). Thanks to fluorescent markers on live bacterial cells, they showed that the default genetic expression state of a clonal bacterial population, in a controlled homogenous environment, was diverse and unpredictable. Mainly thanks to this paper, the issue of stochastic gene expression, postulated more than twenty years before (Spudich & Koshland 1976 ; Kupiec 1981, 1983), made a dramatic entrance into the scientific debate, as if it had been discovered on this occasion. Single-cell approaches showed decisive (McAdams & Arkin 1999 ; Raj & van Oudenaarden 2008) to unravel this long-hidden dimension of gene expression, that hand long been obscured by average values when measured at the cell population level, based on the postulate of homogeneity.

Thinking began to unfold around this counter-intuitive phenomenon in GDP, since cells no longer seemed to respond to each other in a coordinated fashion. On the contrary, they randomly explored avenues of behavioral adaptation to their immediate environment. Thanks to single-cell approaches, mechanistic causes were highlighted, such as macromolecular crowding (Ellis 2001), which creates topological differences between cells. The postulate of availability has also been challenged: the overwhelming majority of proteins are found to be present in less than a hundred copies per cell on average (Guptasarama 1995). For these, the law of large numbers does not apply, and this intercellular variability creates sampling and threshold effects, even within functionally homogeneous and genetically identical cells of the same tissue (this will be questioned below). Topological competition exists for certain supposedly regulatory molecules: they cannot be present on all their potential molecular targets, which also generates variability. In this context, it has also been documented that maintaining the biological order, regularity and precision of genetic regulations requires a correlated expenditure of energy (Lestas *et al.* 2010). The evolutionary rationality of GDP, which relies on these regulations, thus turned out to depend on the cost/benefit ratio of maintaining this order, and in so doing, partially lost its self-evident character. At this point, the aforementioned progresses in molecular biology allowed to start making sense of stochastic gene expression, both mechanistically and statistically. That being said, considering that "nothing makes sense in biology except in the light of evolution" (Dobzhansky 1964) a crucial question remained: if it can be explained mechanistically, what is the evolutionary rationality of this unpredictable variability of gene expression? In other words, what is its biological logic (Pearson 2008)?

## 3. Making Sense of Stochastic Gene Expression: From Damage Control to Radical Rethinking

The awareness about a non-programmed dimension of gene expression leads to, broadly speaking, three main classes of hypothesis.

(i) The first concedes the existence of this intercellular variability of expression, but relegates it to a status of parasitic background noise. This hypothesis is compatible with defending GDP, confining this variability to the status of a margin of uncertainty. This point of view has many supporters, notably in synthetic biology. This young multidisciplinary approach is often presented as the cutting edge of experimental biology. It aims to reconfigure living organisms radically for the purposes of both fundamental knowledge and varied applications: if stochastic variability in gene expression is widely studied here, it is with the main concern of taming it, of reducing it to enable small cellular chassis to function reliably and reproducibly. Under the guise of being disruptive, synthetic biology is above all the new

garb of GDP taken in its most literal sense, where living entities, whether cellular or multicellular, are explicitly compared to fine-tuned machines, and in which chance can be little more than a disturbance.

(ii) The second hypothesis is that the unpredictable intercellular variability of gene expression is a form of valve, an opportunity to relax the genetic program considered too schematic, yet without disqualifying it. This is where classical determinism is described as being able, in local situations, to accommodate or even make use of random gene expression. Biological literature describes numerous cases of bistable equilibria in genetic networks, which can produce diversity. One example is the different versions of rhodopsin (Wernet 2006), a wavelength-sensitive pigment, in the compound eyes of Drosophila, whose units are small groups of cells called ommatidia. These organs have two types of rhodopsin that enable them to capture two complementary wavelength spectra, and the relative proportion of ommatidia producing one or the other is a biological parameter of adaptation to a given environment, and therefore subject to natural selection. This is not a highly complex regulatory system controlling the proportions of expression of each version of the rhodopsin molecules in the ommatidium concerned. In fact, it is a simple molecular regulation—the bistable equilibrium—favoring the expression of one at the expense of the other in each of the ommatidium, with a differential affinity that may be the product of natural selection. On the scale of the individual cell/ommatidium, we cannot predict which rhodopsin will be synthesized. We can measure probability, but on the scale of the cell population composed of all ommatidia, this probability becomes a proportion which is, in turn, predictable. The proportion adapted to the environment is thus achieved without instructional coordination between cells. As we can see, this approach stems from a desire to reconcile GDP with the flagrant manifestations of phenomena that challenge it. The solution adopted is in fact to consider that these phenomena are a complex genetic trait (Ansel 2008), i.e. a trait itself driven by a certain number of loci in a given genome, just as an individual's height or weight might be. In this reading, stochastic gene expression is no longer opposed to GDP, but it becomes one of its possible outcome.

(iii) The third hypothesis brings in a more radical challenge. It consists in opposing GDP with an alternative framework, actually based on stochastic gene expression. We are going to call this framework as the probabilistic alternative framework (PAF). Here, stochastic gene expression is a fundamental biological parameter, as opposed to its status as a margin or valve in the previous classes of hypotheses. PAF explains what we take to be coordinated responses in GDP in terms of the exploratory behavior of cells. This means that, in a given context, cell exploit differentially, and largely blindly, a genome that is nonetheless common (hence the stochastic expression of genes). The genome is no longer an instruction manual as depicted by GDP, but rather a reservoir of possibilities. Response accuracy is not achieved by the docile obedience of cells to a rigid program, but by the fact that some of these cells find adapted solutions in this probabilistic exploration of their genome's possibilities, and are thus selectively favored in their local environment, in a sort of Darwinian process based on gene expression differences rather than actual genetic differences.

PAF relies on cell selection, which may seem odd in the context of GDP but has indeed a long history. Its beginnings can be found in Denis Diderot's *D'Alembert's Dream* (1769) in which the famous thinker portrays the physician and philosopher Théophile de Bordeu, to whom he lends the idea that each organ in the organism has its own will, and hence its own particular interests. A century later, competition between parts is at the heart of the seminal work of the founder of German experimental embryology, biologist Wilhelm Roux, *The Struggle of the Parts in the Organism* (Roux, 1881, Heams, 2012). Roux, after a Darwinian reading not free of ambiguities, explores the hypothesis that cellular subsets, cells, tissues and organs are characterized by natural selection dynamics at their respective scales. The revolutionary Darwinian hypothesis of the creation of a biological order based not on a superior will, but on the dynamics of chance and selection, is here transposed to the interior of organisms. In the course of the twentieth century, several theories concerning certain major physiological functions incorporated a selective component that is no longer debated, such as clonal selection for immunity, or the selective stabilization of synapses for brain development. These and others were analysed in a 1993 review by James Michaelson, outlining a landscape in which selective dynamics at least questioned the primacy of GDP. As mentioned, biologist Jean-Jacques Kupiec first laid the theoretical foundations of PAF in the early 1980s

(Kupiec 1981, 1983). He would refine his study for over thirty years, based on the dynamics of chance and selection as a modality of cell differentiation, embryonic development, and regulation. It is in the light of these all-too-brief historical reminders that we should appreciate the return of the theme of cell competition in the 2010s (de Beco *et al.* 2012) as revisited by new single-cell techniques. The portfolio of techniques now available has provided new life to the theme of cell competition, which is based precisely on differences between cells. This, until a resounding publication made it to Nature's cover with the headline "Battle Lines—Life-and-death Competition Between Cells in the Mouse Embryo" (Volume 500 Issue 7460, 1 August 2013). No doubt this is a sign that, at the very least, the journal considered this theory as innovative.

In addition to having a strong genealogy, which after all is not in itself a proof of validity, PAF also has some undeniable epistemological complementary forces of different natures. First (i) it gives a primary biological meaning to the phenomenon of stochastic gene expression, which pervasiveness has been confirmed decade after decade in particular thanks to single-cell techniques observations. Secondly, (ii) it is economical in terms of hypothesis, since it obviates the systematic need of devising biological explanations for complex, energy-intensive regulations. What is more, (iii) it is all-encompassing, since it can accommodate apparently deterministic phenomena (i.e. highly reproducible, systematically observed or almost so) by considering them as probabilistic with a probability of occurring close to 1. Furthermore, (iv) it is supported by general observations that GDP instructionist lenses cannot explain but poorly, such as the significant cell death observed during cell differentiation, or the transient increases in gene expression variability that precede cell differentiation phases (Buganim *et al.* 2012 ; Dussiau *et al.* 2022, Parmentier *et al.* 2022) predicted by proponents of this new theoretical framework (Heams 2004). Finally, (v) it proposes a unifying perspective of biological phenomena, since it is based on Darwinian-type dynamics of chance and selection: by proposing to import them into multicellular organisms, it relativizes the need to base these dynamics on sometimes murky and arbitrary additional principles of higher organization, such as the predicate that the organism is at the service of its genome.

## 4. The Probabilistic Alternative Framework (PAF) and its Critics

As with any theoretical proposition targeting the core of a dominant and productive discipline, PAF had to prepare to face substantial criticism. In this situation, its proponents, actively engaged in proposing a new framework through experimental demonstrations, were not in a symmetrical situation with the vast majority of biologists who were taking GDP for granted and, rather than explicitly defending it, were mostly validating it by default. At its very roots, this was an imbalanced situation. Further, GDP had already faced waves of criticism, due to other theoretical frameworks, such as organicism, or due to new trends in experimental research, such as the aforementioned epigenetics research that, among others, challenged a gene-centric view of biology.

One strong criticism is that PAF takes the stochastic aspect of genetic expression for granted, without being able to prove it, and it cannot deny that an underlying order may be lying behind this apparent disorder. This objection is not, however, likely to shake its foundations (Heams 2014). First of all (and leaving aside the general fact that asking for a proof of non-existence is not generally considered a valid scientific critic), the question of the existence of true randomness is metaphysical. Like so many others before him, Charles Darwin himself insisted on how cautious one has to be when dealing with the "chance", and when himself was doing so, he strongly stressed he did not exclude that the "chance" could be linked to the ignorance of certain causes (Darwin 1859, introduction of Chapter 5). But this did not detract from his main point: what is important in the mechanism of chance and selection, called natural selection, is not so much that the chance is "true", but that the variations due to this chance are independent, uncorrelated with the following selection. This also applies to PAF within the multicellular organism itself. PAF does not aim to discuss whether the unpredictable variability of stochastic gene expression is an ontological disorder or an appearance of disorder based on a subjacent order yet to be discovered, but to challenge the very logic of GDP, where, at the very scale at which it is described, genetic regulation is presented as precisely ordered. In this context, the hundreds of scientific articles providing experimental evidences of stochastic gene expression have delivered a clear

verdict: stochastic gene expression is observed across the board (Sood & Misteli 2022). While this does not prove that PAF is a more convincing framework than GDP, it does push GDP and its logic based on fine-tuned regulation of gene expression to their limits.

Another objection to PAF logic, rarely stated explicitly, is in fact so decisive that it deserves careful consideration. It consists in pointing out that any framework based on the stochastic behavior of individual cells goes against one of the most important presuppositions of evolutionary biology. This is important because PAF is drawing inspiration from Darwinian dynamics. The context of this objection is the following: in a conceptual framework where it is generally accepted that unicellular organisms compete for resources, one of the key questions of biology is to explain how a cooperative behavior can emerge in a cellular collective that we call a multicellular individual. Yet—and this is essential to understanding the extent to which the program paradigm is a pillar of both molecular and evolutionary biology—the consensus in evolutionary biology is that the emergence of multicellularity required the repression of the collective cells' "selfish" (i.e. uncoordinated for the benefit of the organism) behaviors. This leads to the strong implicit objection that the cells of a multicellular organism cannot constitutively behave in a stochastic, thus selfish manner. In the words of evolutionary biologist Richard E. Michod:

> "For multicellular organisms to emerge as a new unit of selection, the selfish tendencies of their component cells had to be controlled. Theoretical results indicate organisms may regulate this internal conflict and competition (...) by directly reducing the benefits to cells of defecting" (Michod 1996).

This is the keystone of the evolutionary explanation: the evolutionary trend towards multicellularity must be accompanied by a reduction in the unpredictable behavioral variability of individual cells, in favor of coordinated collective behavior. In other words, if cell behavioral stochasticity exists, then it can only be residual, probably belonging to an ancestral subsistence (Lehner 2008).

As crucial as this initial presupposition may be, it is not immune to criticism. Here, we will focus on one major objection. This stems from the very structure of the theoretical demonstration that leads to the decree that the rise of cellular cooperation is necessary for the emergence of multicellular individuality. Its formal model compares two categories of theoretical cellular individuals that differ in particular at a given locus, being either a defector or a cooperator. This model shows that, in this framework, defective non-cooperative individuals are at a disadvantage. But this fictitious situation has nothing to do with stochastic gene expression. The defective/cooperative locus hypothesis is based entirely on the deterministic functioning of the gene in question, where this locus appears as a switch that causes a bifurcation in the genetic circuit: it is a GDP-based hypothesis that cannot therefore, by construction, be used to decide between a GDP and a stochastic (and even any) alternative. Even supposing that the modelling would have led to the opposite result (advantage of the defective allele over the cooperative allele) it would still have been useless in deciding between GDP and PAF. In short, the two approaches are incommensurable. It follows that the consensus about the development of the multicellular state requiring the emergence of intercellular cooperation—whether one agrees with it or not—only makes sense within GDP, and cannot be used as an argument to weight the merits of this paradigm against others, nor a fortiori to disqualify the latter. Moreover, its influence is not absolute. The idea that cells must cooperate entirely within the higher unit that is the organism is open to debate. Evolutionary biologist Leo Buss proposed to explain multicellularity not as the eradication of all non-cooperative behaviors, but as a more subtle balance between different tendencies. He points out, for example, that

> "(cell) variants that favour both the proliferation of the cell lineage and the organism harbouring them were sequentially incorporated in an increasingly sophisticated epigenetic program. In contrast, variants that favour the replication of the cell lineage at the expense of the individual were eliminated and ultimately favoured the fixation of variants that limited the production and/or expression of subsequent variation, creating a stable developmental system" (Buss 1987).

It is therefore not inevitable that cells should have only one possible (cooperative) behavior; they can have a more unpredictable and less coordinated component, provided that this component is not such as to compromise, in return, the proliferation of the organism containing them. This suggests that it is

possible to envisage a range of ways of producing an appearance of cooperation or coordination, a spectrum of possibilities, and not just one modality that would be the strict unconditional cooperation of every cell, at every moment, for the benefit of the organism.

Here, it seems relevant to note that these two main objections to PAF do not stem from molecular biology itself and its tools meant to refute or reinforce a hypothesis: the former concerns the metaphysics of chance, the latter a key aspect of the evolutionary theory. This situation is not due to the impossibility of subjecting PAF to experiment: as mentioned above, not only the evidence for generalized stochastic gene expression keeps growing stronger, but also theoretical predictions specific to PAF and hardly compatible with GDP, such as the transient variation in the intensity of stochastic gene expression during cell differentiation, have been documented. Still, GDP remains dominant: the question facing molecular biology, then, is the level at which criticism of its conceptual underpinnings must take place so that a critical examination of these underpinnings can be undertaken in a demanding manner.

## 5. When Genetic Variation in Clonal Community Comes into Play

The back-and-forth between theory and experimentation was further complicated by another turn of events. This is the realization that genetic variability exists within clonal cell populations, in particular somatic cells derived from the fertilized egg of a multicellular organism (O'Huallachain *et al.* 2012; Ogawa *et al.* 2022). Contrary to what was long been thought, to the extent that the term "clonal" has become synonymous with "genetically identical", the cells of a multicellular organism, although clonal in the sense that they originate from the same egg cell, can in fact exhibit considerable genetic variation between themselves. As we shall see, this will challenge the relevance of the two frameworks, but for different reasons.

Both frameworks implicitly assume that two cells in the same organism are genetically identical, and that all other things being equal, it is on the basis of this similar gene endowment that we must explain the emergence of difference, namely cell types and their apparently coordinated functioning. GDP states that biological order is achieved by coordinating cells via their response to intercellular or environmental signals, while PAF proposes that this order is at least partly achieved by a dynamic of chance and selection. The reliability of the former is based on the precision of regulations, while the reliability of the latter is based on statistical reproducibility derived from the principle of the law of large (cell) numbers and the recurrence of certain micro-environmental constraints at certain stages of embryonic development and cell differentiation.

Against this backdrop of competing explanations, awareness of the significant mutability of somatic cells, while not new, is becoming, for both frameworks, an issue at the heart of this debate as biologists become increasingly aware of its magnitude. The inescapable potential for cell mutation has long been considered to derive from the residual error rate in the precision of genetic duplication during mitosis. The enzymatic apparatus controlling and correcting the appearance of "errors" in the copying of the new DNA strand, while remarkably reliable (and everything suggests that this is a parameter of natural selection), nevertheless admits a residual error rate whose order of magnitude is maximum one mutation per cell division in a genome. These exceptions were often analyzed as a source of possible explanations for the appearance of cancerous dynamics within an organism, where a cell would mutate and adopt a selfish behavior contrary to the default coordination existing between cells sharing the same genome. Advances in molecular and cellular biology have overturned this order. First, there has been a growing awareness of the multiplicity of sources of genetic differentiation between clonal cells. In addition to the residual mutation rate described below, a series of phenomena have been added that, although disparate, all contribute to the creation of genetic variation between clonal cells. These include transposable elements, variations in copy number, traces of viral infections and horizontal exchanges (Ogawa *et al.* 2022). The immune system, for example, produces lymphocytes that are all genetically different at certain loci, in line with the broadest possible capacity to detect the widest possible range of antigens and activate the immune response. This generation of diversity obviously reaches its peak in the context of gametogenesis, which produces haploid cells that are all genetically different from one another.

Secondly, the vision of an eukaryotic genome composed of a few functional sequences drowned in an ocean of "neutral" or "useless" sequences (according to the old dichotomy of coding DNA versus non-coding DNA) has been shattered by at least two major discoveries. These are: (1) the genome's significant

expression activity well beyond the three major classical RNA families (mRNA, rRNA and tRNA), with entire sections of the genome long considered to be non-coding now known to be in fact active; and (2) the many potential RNAs and proteins that can often be produced from a single DNA sequence through alternative splicing and editing. Each of these phenomena come with their own complex regulation "rules". All this contributes to postulate such a complex cross-regulatory dynamic that it is a challenge to intelligibility, a fortiori when viewed from the GDP perspective.

Taken together, these two major phenomena imply that the hypothesis of the broad genetic homogeneity of clonal cells must be seriously relativized. This, in turn, raises formidable questions for the theoretical frameworks used to explain how organisms function.

## 6. Challenged, but Not in the Same Way

On the one hand, GDP can be seen as temporarily strengthened by this realization. Indeed, one of its major epistemological aporias is to presuppose difference in order to explain the appearance of difference. For example, when it is claimed that cells receive signals from other cells, which induce them into their own specialization or differentiation, this is based on the presupposition that there is a pre-existing asymmetry between the sending and the receiving cells. Of course, prior differentiation can explain this asymmetry ad hoc, but then the explanation is displaced without really being answered and, even more problematic, the existence of a difference between cells becomes both the *explanans* and the *explanandum*, creating a strong risk of *reductio ad infinitum*. In this context, awareness of the multiplicity of spontaneous sources of genetic differences between clonal cells within an organism can be seen as a providential windfall that relieves GDP of the responsibility of resolving its initial contradiction. Indeed, this is a Pyrrhic victory, for these phenomena in turn have far more severe consequences for its underlying logic. Clearly, they sweep away the idea that the coordinated collectivity that is supposed to be the sum total of somatic cells can be so as a strict consequence of Dawkinsian selfish gene dynamics. Strictly speaking, somatic cells can no longer be described as working together to maximize the organism's longevity and the probability of gamete transmission of copies of their shared gene pool, since this pool turns out to be heterogeneous. To put it another way, GDP initially rests on the idea that organisms function in a coordinated and precise manner because they are genetically homogeneous, and indeed measure the consequences of this when genetically different rogue sub-units appear (e.g. tumors). But this fundamental genetic homogeneity is increasingly being undermined. GDP is therefore unable to explain the rationality of the coordinated functioning of cells that are genetically heterogeneous, yet explaining this coordinated functioning is nothing less than its *raison d'être*.

Also the realization of the unsuspected extent of genetic heterogeneity within a clonal cell population of an eukaryotic organism challenges PAF, but it should be noted that this challenge is of a different order. The probabilistic explanation is largely based on the assumption that, *all other things being equal* at the cellular level, unpredictable cell behavior is observable. There is no need to presuppose any genetic differences between cells to explain their different behaviors. On the contrary, it is only on the basis of exploratory or even stochastic dynamics that cells can differentially use the same genomes to produce behavioral differences (i.e. differences in the way this common genome is exploited, mainly but not exclusively through the mechanism of stochastic gene expression). In concrete terms, an experimental demonstration of this functional power of stochastic expression potential is based on population observations of genetically identical cells placed in the most homogenizing conditions possible (same micro-environment, same cell cycle state). Cells are observed one by one as far as possible with adapted single-cell techniques, so that any observed behavioral variability (e.g. in transcription, translation, methylation) can then be attributed to a stochastic rather than programmed behavior, since these cells have the same gene content but behave differently. This clearly illustrates the challenge that genetic heterogeneity in clonal cell populations represents for this framework: the greater the heterogeneity, the more difficult it is to maintain the starting hypothesis of this experimental demonstration. Genetic heterogeneity within clonal populations acts here as a hidden variable, providing a possible "classical" explanation for differential behavior: it would not be based on random behavior with a constant genome but, much more classically, on

a long unsuspected difference in genetic composition, giving a potential selective advantage to some.

Both frameworks are therefore put to the test by the discovery of genetic heterogeneity in a clonal population, or at least by the extent of it. However, it should also be noted that the two frameworks are not affected in the same way. As wed have seen, GDP is affected at its very core. Designed to explain how variety can appear in the functioning of cells that are genetically identical and thus linked by a Dawkinsian-type community of destiny, GDP is criticized in its very starting presupposition: these cells do in fact differ genetically. According to consensus evolutionary principles, there is nothing to prevent them from exhibiting selfish or defective behavior, but this is not the case (apart from pathological situations). GDP may therefore be able to continue to explain the ways in which intercellular coordination is acquired, but it is unable to explain the evolutionary rationality of this maintenance, unless one enters into circular reasoning.

Likewise, PAF is challenged by the unsuspected extent of intercellular genetic variability in clonal populations, but this is a different kind of difficulty. Now, demonstrating that gene expression is stochastic, for example, requires an additional precaution, and perhaps a serious experimental headache. In fact, PAF needs to prove something previously taken for granted, i.e. that the clonal cells under study do not have an unsuspected genetic variability that would explain their differential behavior. This is a formidable experimental challenge, but it does not have the same epistemological status as the one faced by GDP. PAF is not weakened at its core; it just needs to be more cautious than it allows itself to be. Indeed, a probabilistic framework is based on the stochastic behavior of genetically identical cells, but there is nothing to prevent genetically different cells in a clonal population from also exhibiting stochastic behavior. In short, genetic variability here is rather added to gene expression variability in the generational sources of fate diversity between cells of the same clonal origin. The risk PAF may run is that it may fail to disentangle the causes (genetic or non-genetic) of stochastic cellular behavior, but not to minimize it, and genetic variability in clonal populations is not an observation likely to refute the intrinsic or extrinsic molecular causes of stochastic gene expression.

The fact that these recent approaches challenge both the frameworks shows how versatile is the role of single-cell studies in debates about genetic determinism. In fact, they can challenge determinism to explain so much stochasticity in supposedly precise and reproducible regulations, as well as they can rescue it by discovering countless unsuspected and providential sources of genetic variations in homogenous cells. This casts doubt on the possibility of using experimental approaches in molecular biology to compare and assess the relative validity of two competing theoretical models.

## Conclusion

In this schematic opposition between two frameworks, one might think that single-cell approaches could have played the role of justice of the peace: the more manifestations of stochastic gene expression were found, the more GDP would be challenged. Yet, after thirty years of development of single-cell techniques, GDP is resisting. The dominant discourse in molecular and cellular biology is admittedly more nuanced than it was half a century ago, but it remains deterministic at its core. This is stressed by the calls for projects that, from genes to genomes, most often continue to aim for their exhaustive description, with a view to eventually producing ever more sophisticated syntheses of all the cross-relationships between genes. Not to mention, of course, the economic context in which this research program is unfolding, where the atomization of organisms into stocks of genes to which a precise task can be assigned within a network of precise regulations, is compatible with patentability and therefore commercial appropriation. In addition, one also must not underestimate the power of the imaginary that emanates from GDP: it places biologists, or whoever controls biological processes, in the position of demiurges, able to modify living organisms by bioengineering in the same way that engineering can modify machines. It also contributes to create a reassuring narrative of our biological condition that leaves no room for the distressing dimension of chance in our daily functioning, or in our origins. All of this probably makes GDP much more than a dominant scientific framework. Therefore, the balance of power is not that of two theoretical frameworks of equal strength.

It follows that, with the benefit of a few years' hindsight, the contribution of single-cell studies to the clarification of theoretical biases is ambivalent. It is undeniable that interest in the cellular level has made

biologists more sensitive to cellular individuality, and to the delusional nature of the cellular homogeneity postulate. But it can also be said that this profusion of results, which one might have thought would allow for deciding between different theoretical frameworks, ultimately seems to alter the balance of forces very little. GDP remains largely dominant, as if it had metabolized the genetic heterogeneity of clonal populations and certain explicitly probabilistic dynamics, as special cases that do not call into question its founding principles. From this perspective, the little machines that cells are supposed to be are certainly less reliable than expected, but remain little machines nonetheless: at the very least, this does not do justice to the intense theoretical debates within biology over the machine conception of organisms with strong proponents (e.g. Bongard & Levin 2021) and opponents (e.g. Nicholson 2013). The issue at stake here is not to regret or support this persistence of GDP, but to engage in a collective discussion on how to question its relevancy, that seems overshadowed by its dominance. Such a paradigm can thrive because it is scientifically fruitful and productive, but also because, when dominant, it relies on its past successes (more than its own updated merits) to raise the bar high for any potential contestation. In other words, an important inertia can exist even if it is convincingly criticized. Further, this dominance comes with a powerful narrative, the above mentioned possibility to "engineer life" as we do for machines. Many researchers got acquainted with it, get advantages from it (in particular, a position of power), and thus hardly accept to let it go, even when a substantial number of observations challenge and even undermine it. Because of their aforementioned versatility, single-cell techniques have not been, so far, an efficient tool for this much needed falsifiability, even though they still have potential in this respect as well as in many areas of theoretical research in biology. The landscape we have described here has shown that the sources of genetic variability between cells within an organism are multiple. Moreover, they must be combined with even more radical sources of variability: the somatic cells of a multicellular organism, notably metazoans, cohabit with others, the cells of our microbiota, or even the cells of maternal origin that make each of us chimeric, mosaic individuals. In short, different ways of being different contribute to shaping individuals. Understanding the overall logic and functionality of all these sources of variability is a new frontier and a story yet to be told: do selective and instructional dynamics cohabit? Do different selective dynamics co-exist? Is there a competition of competitions between these different sources of genetic variability, and the cell populations that embody them? These open questions are crucial, and it would be desirable for single-cell approaches to tackle them head-on, rather than feeding the endless quest for details that molecular biology loves to accumulate.

## Acknowledgements

The author would like to thank the reviewers for their comments, which have helped to make the manuscript clearer and more coherent.

## References

Ansel, J, Bottin, H, Rodriguez-Beltran, C, Damon, C, Nagarajan, M, Fehrmann, S, *et al.* 2008 "Cell-to-cell stochastic variation in gene expression is a complex genetic trait", *PLoS Genetics*, vol. 4, no. 4, art. e1000049. https://doi.org/10.1371/journal.pgen.1000049

Bheda, P, & Schneider, R 2014 "Epigenetics reloaded: The single-cell revolution", *Trends in Cell Biology*, vol. 24, no. 11, pp. 712–723.

Bongard, J, & Levin, M 2021 "Living things are not (20th century) machines: Updating mechanism metaphors in light of the modern science of machine behavior", *Frontiers in Ecology and Evolution*, vol. 9, art. 650726.

Buganim, Y, Faddah, DA, Cheng, AW, Itskovich, E, Markoulaki, S, Ganz, K, Klemm, SL, van Oudenaarden, A, & Jaenisch, R 2012 "Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase", *Cell*, vol. 150, no. 6, pp. 1209–1222.

Buss, L 1987 *The evolution of individuality*, Princeton: Princeton University Press.

Darwin, C 1859 *On the origins of species*, London: Murray.

Di Carlo, D, & Lee, LP 2006 "Dynamic single-cell analysis for quantitative biology", *Analytical Chemistry*, vol. 78, no. 23, pp. 7918–7925.

Dobzhansky, T 1964 "Biology, molecular and organismic", *American Zoologist*, vol. 4, no. 4, pp. 443–452.

Dussiau, C, Boussaroque, A, Gaillard, M, Bravetti, C, Zaroili, L, Knosp, C, Friedrich, C, Asquier, P, Willems, L, Quint, L, Bouscary, D, Fontenay, M, Espinasse, T, Plesa, A, Sujobert, P, Gandrillon, O, & Kosmider, O 2022 "Hematopoietic differentiation is characterized by a transient peak of entropy at a single-cell level", *BMC Biology*, vol. 20, no. 1, p. 60.

de Beco, S, Ziosi, M, & Johnston, LA 2012 "New frontiers in cell competition", *Developmental Dynamics*, vol. 241, no. 5, pp. 831–841.

Elmentaite, R, Domínguez Conde, C, Yang, L, & Teichmann, SA 2022 "Single-cell atlases: Shared and tissue-specific cell types across human organs", *Nature Reviews Genetics*, vol. 23, pp. 395–410.

Ellis, RJ 2001 "Macromolecular crowding: Obvious but underappreciated", *Trends in Biochemical Sciences*, vol. 26, no. 10, pp. 597–604.

Elowitz, MB, Levine, AJ, Siggia, ED, & Swain, PS 2002 "Stochastic gene expression in a single cell", *Science*, vol. 297, 1183–1186.

Gawad C, Koh W, Quake, SR 2016 "Single-cell genome sequencing: Current state of the science", *Nature Reviews Genetics*, vol. 17, no. 3, pp. 175–188.

Guptasarma, P 1995 "Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli?", *Bioessays*, vol. 17, pp. 987–997.

Heams, T 2004 *Approche endodarwinienne de la variabilité intercellulaire de l'expression génétique*, PhD Dissertation, Institut National Agronomique Paris-Grignon, Paris.

Heams, T 2012 "Selection within organisms in the nineteenth century: Wilhelm Roux's complex legacy", *Progress in Biophysics and Molecular Biology*, vol. 110, no. 1, pp. 24–33.

Heams, T 2014 "Randomness in biology", *Mathematical Structures in Computer Science*, vol. 24, no. 3.

Kehr, J 2003 "Single cell technology" *Current Opinion in Plant Biology*, vol. 6, no. 6, pp. 617–621.

Kolodziejczyk, AA, Kim, JK, Svensson, V, Marioni, JC, & Teichmann, SA 2015 "The technology and biology of single-cell RNA sequencing", *Molecular Cell*, vol. 58, no. 6, pp 610–620.

Kumazaki, T, Hamada, K, & Mitsui, Y 1994 "Detection of mRNA expression in a single cell by direct RT-PCR", *Biotechniques*, vol. 16, no. 6, pp. 1017–1019.

Kupiec, JJ 1981 "Théorie probabiliste de la différenciation cellulaire", *XIIème rencontre de Méribel*, pp. 161–163.

Kupiec, JJ 1983 "A probabilist theory for cell differentiation, embryonic mortality and DNA C-value paradox", *Speculations in Science and Technology*, vol. 6, pp. 471–478.

Kupiec, JJ 2009 *The origin of individuals*, Singapore: World Scientific.

Lehner, B 2008 "Selection to minimise noise in living systems and its implications for the evolution of gene expression", *Molecular Systems Biology*, vol. 4, p. 170.

Lestas, I, Vinnicombe, G, & Paulsson, J 2010 "Fundamental limits on the suppression of molecular fluctuations", *Nature*, vol. 467, no. 7312, pp. 174–178.

Liang, SB, & Fu, LW 2017 "Application of single-cell technology in cancer research", *Biotechnology Advances*, vol. 35, no. 4, pp. 443–449.

Longo, D, & Hasty, J 2006 "Dynamics of single-cell gene expression", *Molecular Systems Biology*, vol. 2, p. 64.

McAdams, HH, & Arkin, A 1999 "It's a noisy business! Genetic regulation at the nanomolar scale", *Trends in Genetics*, vol. 15, no. 2, pp. 65–69.

Michod, RE 1996 "Cooperation and conflict in the evolution of individuality. II. Conflict mediation", *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 263, no. 1372, pp. 813–822.

Nicholson, D 2013 "Organisms ≠ Machines", *Studies in History and Philosophy of Biology and Biomedical Sciences*, vol. 44, no. 4, pt. B, pp. 669–678.

Noble, D 2006 *The music of life*, Oxford: Oxford University Press.

Ogawa, H, Horitani, K, Izumiya, Y, & Sano, S. 2022 "Somatic mosaicism in biology and disease", *Annual Review of Physiology*, vol. 84, pp. 113–133.

O'Huallachain, M, Karczewski, KJ, Weissman, SM, Urban, AE, & Snyder, MP 2012 "Extensive genetic variation in somatic human tissues", *Proceedings of the National Academy of Science of the United States of America*, vol. 109, no. 44, pp. 18018–23.

Parmentier, R, Racine, L, Moussy, A, Chantalat, S, Sudharshan, R, Papili Gao, N, Stockholm, D, Corre, G, Fourel, G, Deleuze, JF, Gunawan, R, & Paldi A 2022 "Global genome decompaction leads to stochastic activation of gene expression as a first step toward fate commitment in human hematopoietic cells", *PLoS Biology*, vol. 20, no. 10, art. e3001849.

Pearson, H 2008 "Cell biology: The cellular hullabaloo", *Nature*, vol. 453, pp. 150–153.

Raj, A, van Oudenaarden, A 2008 "Nature, nurture, or chance: Stochastic gene expression and its consequences", *Cell*, vol. 135, no. 2, pp. 216–226.

Roux, W 1881 *Der Kampf der Theile im Organismus* [The Struggle of the Parts in the Organism]. Leipzig: Engelmann.

Sood, V, & Misteli, T 2022 "The stochastic nature of genome organization and function", *Current Opinion in Genetics & Development*, vol. 72, pp. 45–52.

Spudich, JL, & Koshland, DE 1976 "Non-genetic individuality: Chance in the single cell", *Nature*, vol. 262, no. 5568, pp. 467–471.

Templer, RH, & Ces, O 2008 "New frontiers in single-cell analysis", *Journal of the Royal Society Interface*, vol. 5, suppl. 2, pp. S111-2.

Wang, D, & Bodovitz, S. 2010 "Single cell analysis: The new frontier in 'omics'", *Trends in Biotechnology*, vol. 28, no. 6, pp. 281–290.

Wernet, M, Mazzoni, O, Çelik, A, Duncan, D, Duncan, I, & Desplan, C 2006 "Stochastic spineless expression creates the retinal mosaic for colour vision", *Nature*, vol. 440, no. 7081, pp. 174–180.

Wijgerde, M, Grosveld, F, & Fraser, P 1995 "Transcription complex stability and chromatin dynamics in vivo", *Nature*, vol. 377, pp. 209–213.