Organisms

Journal of Biological Sciences

Single-cell Analysis: Epistemological Inquiries

Sapienza Università di Roma

Vol. 6, No. 2 (2023)

ISSN: 2532-5876

EDITORS

ORGANISMS is a biannual, open access, peer-reviewed, online journal devoted to basic, translational, theoretical and clinical research. Topics of interest include but are not limited to the following areas: complexity in biology; carcinogenesis; theory of biological organization; biophysics; mathematical and computational approaches in biology and medicine; developmental processes, and degenerative diseases.

ORGANISMS' Editorial Board is involved in research programs into many aspects of normal and pathological inflammation. The Board promises to conduct and facilitate unbiased and critical reviews of manuscripts.Every proposal submitted for publication is read at least by an editor for a preliminary review. If the paper matches the editorial policies and quality level, then it is sent to reviewers for in-blind evaluation via the OJS platform.

ORGANISMS provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. Authors of articles published in ORGANISMS retain the copyright of their works. Authors can archive pre-print, post-print, and publisher's versions.

The journal is committed to maintaining the highest level of integrity in the content published. All research must have been carried out within an appropriate ethical framework. If there is suspicion that work has not taken place within an appropriate ethical framework, Editors will follow the COPE guidelines and may reject the manuscript, and/or contact the author(s)' institution or ethics committee.

Research involving human subjects, human material, or human data, must have been performed in accordance with the Declaration of Helsinki, and must have been approved by an appropriate ethics committee. A statement detailing this, including the name of the ethics committee and the reference number where appropriate, must appear in all manuscripts reporting such research. If a study has been granted an exemption from requiring ethics approval, this should also be detailed in the manuscript (including the name of the ethics committee that granted the exemption). Further information and documentation to support this should be made available to Editors on request. Manuscripts may be rejected if the Editor considers that the research has not been carried out within an appropriate ethical framework. In rare cases, Editors may contact the ethics committee for further information.

When reporting experiments on animals, authors should indicate whether the institutional and national guide for the care and use of laboratory animals was followed. See International Committee of Medical Journal Editors ("Uniform Requirements for Manuscripts Submitted to Biomedical Journals") February 2006.

https://rosa.uniroma1.it/rosa04/organisms/index

CONTENTS

Special Issue, Single-cell Analysis: Epistemological Inquiries

Editorial

Special Issue

Perspectives and Hypotheses

Cover image: *Detail of a wooden inlay*. Photo: S. Serafini, 26 October 2023.

SAPIENZA UNIVERSITÀ DI ROMA

Editorial

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

DOI: 10.13133/2532-5876/18311

Special Issue, "Single-cell analysis: Epistemological inquiries"

Caroline Angleraux[a] *& Matteo Mossio*[b*]

[a] *UMR 1253, iBrain, Université de Tours, Inserm, Tours, France*
[b] *Institut d'Histoire et de Philosophie des Sciences et des Technique (CNRS/Paris1), Paris, France*

***Corresponding author:** Matteo Mossio, Email: matteo.mossio@univ-paris1.fr

Citation: Angleraux, C, & Mossio, M, "Special Issue, 'Single-cell analysis: Epistemological inquiries'" [editorial], *Organisms: Journal of Biological Sciences*, vol. 6, no. 2 (2023), pp. 5-9. DOI: 10.13133/2532-5876/18311

This special issue gathers four articles that aim at making explicit and discussing some epistemological issues arising from single-cell analysis.

What is single-cell analysis? The label does not refer to the general practice of observing and describing individual cells, which is something that biologists have been doing for a long time. Single-cell analysis consists more specifically in a set of recent -omics sequencing techniques that enable scientists to study the different so-called "-omes" of single cells, and thereby to describe their "molecular profile". In doing so, the aim is to have a better and a more comprehensive view about how each individual cell works within a biotic or abiotic environment.

Single-cell analysis involves different techniques for isolating cells. Depending on the abundance of cells within the sample, the cells' shape, the accuracy and the granularity of the requisite results or the available funds, single-cell isolation techniques encompass serial dilution, robotic micromanipulation, microfluid platforms, laser-capture microdissection and others. Then, following the kind of target -omes (whole genome, transcriptome, epigenome, etc.), there is also a variety of methods for amplifying them (degenerative-oligonucleotide-PCR [DOP-PCR]) and multiple-displacement-amplification [MDA] or Oligo dT-anchoring) as well as for sequencing the molecular material (like SMART-Seq or using unique molecular identifiers). A huge amount of data results from the sequencing procedures. This is usually categorized within barcoded libraries (Wang & Song 2017).

These overall steps (isolating, amplifying, sequencing, and categorizing) are usually common to any single-cell analysis, whatever the field involved—carcinogenesis, immunology, microbiology, neurobiology, etc. And yet, the variety of techniques and methods used at each step, together with a lack of standardized practices between laboratories and fields, affects collaborations between research infrastructure and communication of data (Lähnemann *et al.* 2020). In this respect, single-cell analysis may be an interesting object for sociology of sciences. In this special issue, we leave this range of questions aside and focus on epistemological issues.

Generally, there is a huge enthusiasm by biologists' communities that employ these techniques. The promise of a high degree of precision in the data collected, and the ultimate ambition of connecting different explanatory levels in order to achieve a broader understanding of living beings are the main reasons why single-cell analysis is so widespread in laboratories today. The questions addressed in this special issue concern the contribution of single-cell analysis to the advancement of biological knowledge. Is biologists' enthusiasm *vis-à-vis* single-cell analysis epistemologically justified? To what extent does single-

cell analysis contribute to provide a more adequate explanation of biological phenomena?

In the study of multicellular systems, sequencing a cell sample usually implies a global genetic knowledge of the tissue, which hides the intrinsic heterogeneity of molecular profiles within the sample. The rationale behind the use of single-cell analysis might be described as follows (Qian & Bao 2019):

- Understanding multicellular organization requires understanding the different functions exerted by different types of cells, assembled in tissues and organs;
- Cell types can be identified by a certain molecular profile. By hypothesis, all the cells belonging to the same type share the same molecular profile;
- Single-cell analysis allows biologists to detect different molecular profiles within a single tissue, and to distinguish between different cell types, beyond global "means" established over populations of cells.

In the case of unicellular systems, see for instance (Ku & Sebé-Pedros 2019), single-cell analysis is used within an evolutionary approach:

- Understanding how unicellular organisms interact with their environment requires understanding how they adapt or are more specified (for studying symbiosis, ecological or evolutionary processes);
- Types can be identified and clustered into phyla depending on the molecular profiles;
- Single-cell analysis allows biologists to track the specification within different unicellular phyla, by detecting different molecular patterns.

Overall, this rationale relies on a twofold background presupposition, according to which:

- Molecular characteristics are more reliable than other criteria (like metabolic behavior) to typify cells;
- The functional role of cells is subtended by a combination of molecular characteristics, i.e. their "molecular profile".

Such a presupposition is theoretically loaded, and raises an epistemological problem that might be broken down into three related issues.

(1) Single-cell analysis is claimed to provide a global and even holistic approach, insofar as it can combine datasets obtained through various -omics technologies, and referring to different biological objects such as mRNA, DNA, ribosomal RNA, etc. (Anam *et al.* 2019). Yet, the question is how single-cell analysis would realize a more comprehensive view of living beings. Indeed, the very idea according to which the molecular profile of a cell subtends its functional role is consistent with a reductionist approach to biological phenomena. And it might be argued that looking at the various molecular characteristics of a cell does not inform as such about its dynamic organization. Understanding how a cell works would imply a wider vision that does not focus on the molecular level alone. If so, then single-cell analysis provides data that are not sufficient to make sense of cell functions.

(2) The presupposition that the molecular profile is relevant (and sufficient) to categorize cells into types seems to overlook the processual dimension of biological phenomena. Cells are dynamic entities that undergo a life cycle, during which their molecular profile changes over ontogenetic time. Cells are plastic, some can dedifferentiate or transdifferentiate; ontogeny is quite reversible. In contrast, single-cell analysis provides stable data, that describe a biological mapping at a given moment (a "snapshot"), with a given set of spatial interactions. Single-cell analysis (for now) is only able to take snapshots of cells' life cycles, which means that it cannot produce any description of individual trajectories and it cannot determine whether a certain snapshot is representative of a certain cell type (Trapnell *et al.* 2014). This snapshot has to be put into perspective and compared either with other snapshots of the same biological process at a different time, or stated as the representative of a given cell type, based on previous knowledge. For now, the description of developmental or transitional dynamics in cells relies on a pseudo-time derived from a comparison of quantitative measurements between proximate moments in different cells in order to infer the states that precede or follow each other.

These limitations raise several questions: (a) To what extent does the pseudo-time account for the developmental time of a living organism? (b) By relying on molecular patterns only, how distinguishing between two different cell types, on the one hand, and the same cell at two different moments, on the other hand? (c) How determining which snapshot better characterizes a cell type? As a consequence of the dynamic nature of living processes, a (theoretical) choice should be made regarding what "moment" in the cell's life cycle is the relevant one to determine its type.

(3) More generally, single-cell analysis claims to be data-driven only. In fact, it seems to rely rather on epistemological choices, which are not always explicit (Leonelli 2019). There is nowadays an increasing acknowledgement of the "heterogeneity" of data across individual cells. Cells that supposedly belong to the same type and perform the same function (because they are in the same tissue for instance) do not exhibit the same molecular profile. In general, molecular data exhibit a continuum among the various categories of cell rather than sharp discontinuities. Instead of finding similarities, single-cell analysis finds differences. Such heterogeneity pushes biologists to multiply categories ("rare" cell types), which risks to be a useless process that would result in obtaining as many types as individual cells.

Therefore, single-cell analysis requires making epistemological choices while structuring categories. It is what happens when bio-analysts define clusters by selecting a set of features, and measure to what extent each cell possess such features. As Gross (this issue) puts it, "membership is based on overall similarity, that is, the degree to which objects share a set of properties". Clustering requires to make a choice about what features are considered as relevant, and how the "distance" or "score" is measured. Depending on these choices, different categories emerge. Single-cell analysis is not (and cannot be) entirely data-driven: categorization does not emerge from data themselves. Categories depend on choices, which in turn depend on previous knowledge about cell types, background theories and assumptions.

The issue of identifying and making explicit epistemological choices also raises a number of questions: (a) What criteria should determine the features by which categories and clusters are elaborated? And how do these criteria promote different research directions? (b) How does background (structural, functional and genealogical) knowledge affect the elaboration of cells categories? (c) How do background epistemological choices impact information sharing and communication? A research team produces databases that can be hard to understand by a different team. The way to classify and identify nomenclatures (which exacerbates the variable number and the tendency towards multiplications of categories, as we mentioned before) may complexify the adequation between different databases. It also questions, in another way, the ability to reproduce results. In this sense, making explicit epistemological choices is an absolute necessity for securing disclosable data.

In a word, the focus of single-cell analysis on cells molecular profile raises a number of questions about reductionism, cell dynamics and implicit epistemological choices. More generally, single-cell analysis can be critically examined in terms of the characterization of cells and different biological processes that it puts forward, as well as the criteria for biological identity that it adopts. During its history, biology has oscillated between structural, functional and genealogical criteria, and the debate about their relation is a never-ending one. The epistemological enquiries about single-cell analysis should also be located within this larger and fascinating debate.

The four contributions to this special issue, authored by biologists and philosophers alike, examine the above questions from different, and yet complementary perspectives.

In his contribution, Fridolin Gross examines how single-cell analysis impacts the very concept of cell type. He emphasizes the tension existing between the idea of using single-cell analysis to elaborate more solid cell types and the recognition of huge spatial and temporal cellular heterogeneity. He describes what might possibly be labelled a "molecular pheneticist account" to cell types, and focuses on (and questions) the claim that such account might be theory-free. Gross shows that fundamental steps in single-cell analysis (as dimensionality reduction and clustering methods) do require to make choices (based on theories or at least on background knowledge) about the number of dimensions and the parameter values. Above all, clustering methods are "importantly driven by the concern to reproduce previously accepted cell type classifications".

Gross concludes by claiming "it seems inappropriate to refer to them as 'theory-free' or purely data-driven as this would ignore the clearly theory-guided process of method selection". Gross generalizes his argument. According to him, thinking that, in principle, the more data are added, the more they would converge in creating stable categories in a theory-free manner is delusional. Even more generally, Gross mentions the fact that focusing on "structures" is not a straightforward choice, because of the everlasting tension with functional and genealogical criteria. So Gross asks: "Why then should biologists focus

so much on a theory-free classification approach if that approach misses the central goal that cell type classifications are meant to achieve?".

Racine and Paldi also underscore that single-cell analysis is supposed to contribute to our understanding of cells identity and differentiation. Classifications based on origins vs. similarities have always co-existed, but the more recent idea is that "cells of the same type must express the same genes and can be identified on the basis of the transcriptional regulator (transcription factors) they express".

As Gross, they emphasize the strong heterogeneity in gene expression, and in the resulting molecular profile. So, while cells belonging to different tissues or organs tend to exhibit distinguishable gene expression patterns, the same is not true for supposedly different cells belonging to the same tissue, for instance. According to them, the continuous nature of gene expression in cell population makes that "If one picks up randomly a cell from the population, there are good chances that it is impossible to say on the basis of its gene expression pattern to which type it belongs". Again, clustering methods, no matter how powerful they are, do not produce types by themselves, but depend on several background choices made by the biologist (about p-values, thresholds, filters, the presumed number of clusters one expects, etc.).

Racine and Paldi also suggest that the more our analysis is fine-grained, the more we find molecular differences. We find "rare cell types", and "cell states" which, according to the authors, do not change anything to the initial problem. The identity of cells is contextual and dynamic, and any search by single-cell analysis should rely on a definition elaborated beforehand. This means in particular deciding what counts as relevant stability in cell life cycles, which are profoundly variable and dynamic. We should not forget that "stability" itself is a scale-dependent notion, with respect to which a decision also has to be taken. The Authors conclude by calling for "a new interpretation framework based on solid theoretical ground", possibly centered on the organicist tradition.

Heams' article agrees with the previous ones about the fact that the questions raised by single-cell analysis are not just technical, but profoundly epistemological and theoretical. Heams underscores that single-cell analysis is mainly used within the Genetic Determinism Paradigm, according to which (among other things) similarities and differences among cells should unambiguously correspond to similarity and differences in their gene expression and in their overall molecular profile. Yet, single-cell analysis has shown for 20 years now that even a population of clonal cell shows very heterogeneous gene expression, to the point that the idea of stochastic gene expression was proposed, and it constitutes now a solid hypothesis in molecular biology. Single-cell analysis, in this sense, contributed not to find stability and categories, but to shake the very foundations of GDP. Heams discusses the various ways of interpreting unpredictable variability, ranging from the more conservative and GDP-related to the more original and alternative one, which Heams calls the "probabilistic alternative framework (PAF)". PAF claims that gene expression is fundamentally stochastic and incompatible with GDP. Heams discusses the strength and weaknesses of PAF, and in particular the extent to which it is at odds with some of the theoretical pillar of evolutionary theory, i.e. the necessity of cooperative behavior. On this crucial point, Heams argues that PAF does not exclude cooperative behavior, but that this very notion should be reconceptualized within a probabilistic framework. Heams goes farther in discussing how both GDP and PAF are challenged by the discovery that cells constituting a multicellular system are not genetically homogeneous, although he argues that the challenge is not the same. GDP "is affected at its very core", while "there is nothing to prevent genetically different cells in a clonal population from also exhibiting stochastic behaviour".

The upshot of his analysis is that while single-cell analysis has shaken the GDP at its foundations, the mainstream paradigm is still... mainstream. This raises the question of how experimental results (in this case, obtained by single-cell analysis) can actually falsify a paradigm, and open the way to innovative research directions.

Angleraux's article (which will be published in the following issue, because of editorial reasons) follows the same line of Heams, Racine and Paldi regarding the need to clarify the theoretical ground of single-cell analysis. She questions the type of biological explanation underlying single-cell sequencing, and she applies general frameworks in philosophy of biology (especially new mechanism and systems biology) to specify how these techniques explain biological phenomena. She comes to

the conclusion of a gap between the scientific narrative and what single-cell sequencing de facto produces.

Indeed, by combining different databases, single-cell analysis (as a kind of -omics sequencing techniques) aims to achieve a comprehensive and a more integrative explanation of biological phenomena, which matches with the *zeitgeist* against reductionism of current theoretical and philosophical perspectives in biology. New mechanism, on the one hand, claims to be non-reductionist because it takes into account emergent properties of organisms and explains them by integrating elements at different levels of description. Systems biology, on the other hand, also alleges a holistic view of life by combining biological subsystems. Angleraux examines to what extent single-analysis embraces new mechanism's and systems biology's *zeitgeist*. However, this is the case also because single-cell analysis shares the same theoretical and philosophical limits with these perspectives. In particular, both mechanism and system biology keep favoring bottom-up, rather than top-down explanations of living phenomena. As a consequence, Angleraux underscores the hiatus between the scientific narrative—what single-cell analysis declares to accomplish—and what it actually does for now.

We would like to thank the "Who am I? Single Cell Initiative" (WISCI), led by Pierre-Antoine Defossez at Paris Cité, and funded by the Labex "Who am I?", which provided financial support for the workshop "Enjeux épistémologiques de l'approche «single cell»" held in Paris in May 31-June 1, 2022. The contributions to this special issue were initially presented at the workshop. We thank in particular Pierre-Antoine Defossez and Nikos Konstantinides who also attended the workshop and hosted Caroline Angleraux in their team as a postdoc. Their work provided a relevant case study to work on and think about single-cell analysis. Our warmest thanks go to the editors of the journal Organisms for accepting to publish this special issue, and to the reviewers of the articles for their rigorous and constructive critical comments.

References

Ku, C, & Sebé-Pedrós, A 2019, "Using single-cell transcriptomics to understand functional states and interactions in microbial eukaryotes", *Philosophical Transactions of the Royal Society B: Biological Sciences,* vol. 374, art. 20190098.

Leonelli, S. 2019, "The challenges of big data biology", *eLife,* vol. 8, art. e47381.

Naz, A, Obaid, A, Ikram, A, Awan, FM, Nisar, M, Paracha, RZ, & Ali, A 2019, "Systems-level understanding of single-cell omics", *Single-Cell Omics*, vol. 1, pp. 433-456.

Qian, Z, & Bao, L 2019, "Cellular heterogeneity and single-cell omics" in D Barh & V. Azevedo (eds.), *Single-cell omics*, vol. 1, pp. 35-44, Academic Press.

Trapnell, C, Cacchiarelli, D, Grimsby, J, Pokharel, P, Li, S, Morse, M, Lennon, NJ, Livak, KJ, Mikkelsen, TS, & Rinn, JL 2014, "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells", *Nature Biotechnology*, vol. 32, pp. 381-386.

Wang, J, & Sang, Y 2017, "Single cell sequencing: A distinct new field", *Clinical and Translational Medicine*, vol. 6, no. 1, p. 10.

Special Issue, Single-cell Analysis: Epistemological Inquiries

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

DOI: 10.13133/2532-5876/18131

Do Single-cell Experiments Challenge the Concept of Cell Type?

Fridolin Gross[a*]

[a] *CNRS UMR5164 ImmunoConcEpT, University of Bordeaux, 146 rue Leo Saignat, Bordeaux, 33076 France*

***Corresponding author:** Fridolin Gross, Email: fridolin.gross@u-bordeaux.fr

Abstract

Recent debates among biologists have highlighted problems with the traditional concept of cell type, which is considered vague and subjective. Single-cell technologies reveal the limitations of the current concept by exposing a high degree of heterogeneity in cell populations. At the same time, some biologists believe that these technologies provide the basis for a more objective and precise concept of cell type that is not dependent on prior theoretical assumptions. In this paper, I explore the impact that single-cell experiments and analyses will have on the concept of cell type. Drawing on the practices of biologists using these methods, but also on more principled arguments, I argue that the idea of a purely theory-free classification is unlikely to be realized. However, single-cell technology may affect the concept of cell type in more subtle ways.

Keywords: ontology, cell type, theory-free classification; pheneticism, single-cell technology

Citation: Gross, F 2023, "Do Single-cell Experiments Challenge the Concept of Cell Type?", *Organisms. Journal of Biological Sciences*, vol. 6, no. 2, pp. 11–23. DOI: 10.13133/2532-5876/18131

Introduction

Biologists use many concepts without worrying too much about clear definitions. Some of these are quite fundamental, such as the concept of gene or even of life itself. This is not necessarily problematic, and it has even been argued that precisely ambiguity and indeterminacy contribute to the fruitfulness of some scientific concepts (Neto 2020). Sometimes, however, a discipline may undergo certain developments that force scholars to think more deeply about a particular concept and clarify it by making some of their tacit assumptions explicit. The concept of cell type is an interesting illustration of this pattern. For more than a century, biological and medical practice has identified and distinguished cell types according to various and often ill-defined sets of criteria related, for example, to morphology, function, location, or developmental origin, without ever converging on an explicit and general account. However, recently introduced experimental techniques, along with computational methods of data analysis, seem to be forcing biologists to clarify their ideas about what they mean when they talk about cell types. In particular, single-cell sequencing experiments provide much more detailed insight into the diversity and heterogeneity of cell populations. This has led some people to argue that biology needs a more principled and possibly more fine-grained classification of cells into types, sub-types, or states. At the same time, many biologists think that the new experimental techniques offer the possibility of achieving a delineation of cell types that, because

purely data-driven, is more objective and precise and thus superior to the subjective, vague, and potentially biased classifications based on the traditional concept of cell type.

In this paper, I investigate what impact such technological advances can be expected to have on the concept of cell type. In particular, I will address the claim that such an approach to classification can be based solely on data-driven methods. Plausibly, such a "theory-free" account may be desirable for a variety of reasons, but it is unclear to what extent scientific concepts and classifications could be based on such foundations alone. Interestingly, philosophers have been discussing very similar questions about classificatory concepts in different contexts, notably with regard to the classification of organisms into species and higher taxa. Therefore, the debate about cell types might benefit from an awareness of some of the problems and arguments that were debated elsewhere in the past.

The paper is structured as follows. In Section 1, I provide an overview of the current debates around cell types, and explain why biologists feel the need to revise or clarify the concept. Section 2 offers some philosophical background on classification and shows how biologists, over time, have endorsed different approaches to the classification of cells that may correspond to different philosophical positions. In particular, the new approaches based on single-cell technologies fit a pheneticist or clustering approach to classification that is familiar in the biological taxonomy debate. In Section 3, I argue that a purely data-driven version of such a pheneticist approach is unlikely to be successful, before wrapping up the matter with some remarks in the Conclusion.

1. Cell types and single-cell experiments

Cell theory, dating back to the 19th century, established the idea that the tissues of animals and plants are made up of basic building blocks, all of which originate from the same fertilized egg cell (Duchesneau 1987; Canguilhem 1995). Although all cells in a multicellular organism contain much the same genetic material, they can differ radically in size, morphology, and the role they play in the context of the organism. Based on early microscopy and staining techniques, cell types were at first distinguished using phenotypic criteria, for example in terms of the functions they carry

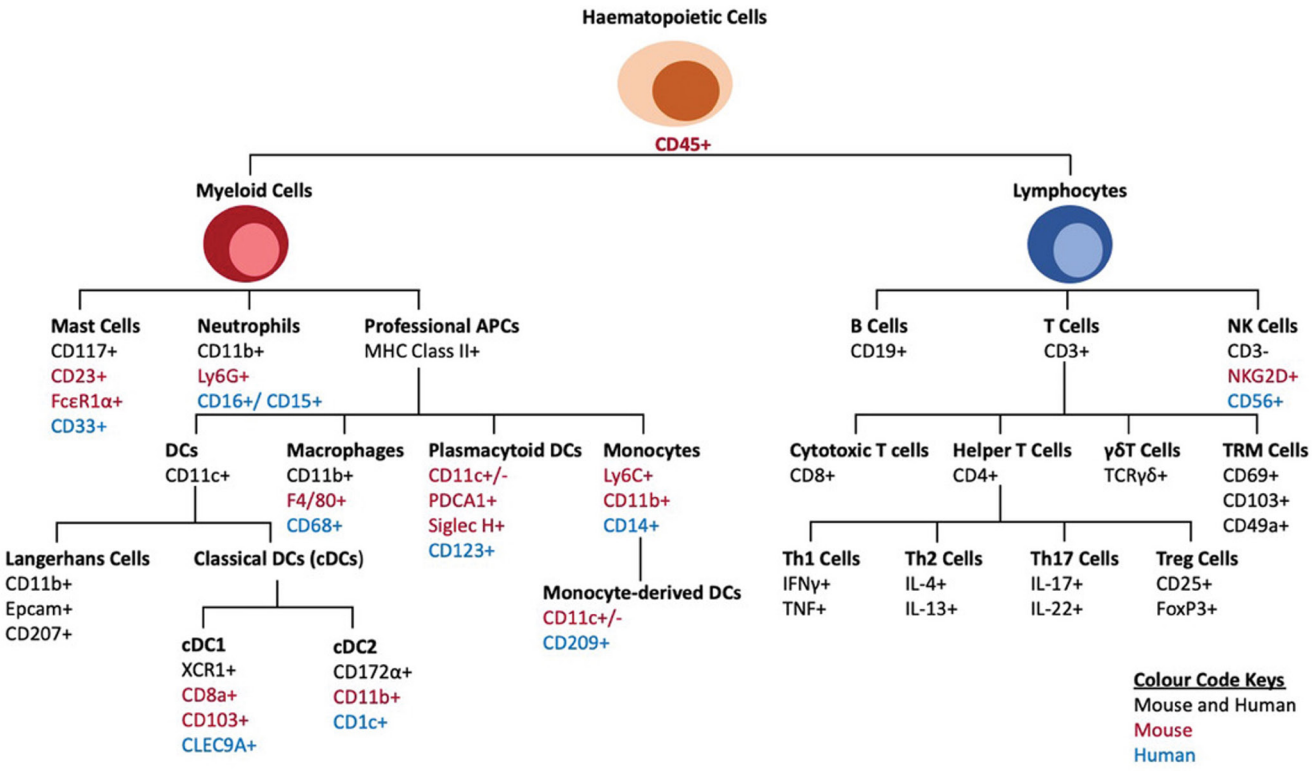

Figure 1: Lineage tree of human hematopoietic cells. Adapted from Murphy et al. (2022) (CC BY-NC-ND 4.0).

out within the organism or according to morphological features such as size or shape.

Ramon y Cajal and Camillo Golgi for the study of the fine structure of the nervous system (Jones 1999) and Alexander Maximov for the discrimination of different types of blood cells (Novik *et al.* 2009) are among the pioneers in applying these methods. Over the years these techniques were refined, often by linking cell types to specific "marker" molecules (Baskin 2015). Notably in the context of immunology, a sophisticated system of such markers, known as "cluster of differentiation" (CD), was developed. It allows biologist to detect and distinguish different cell types using techniques such as immunohistochemistry and flow cytometry that are based on the detection of specific molecular patterns on the cell surface (Chan *et al.* 1988). As an example, Figure 1 shows the lineage tree of hematopoietic cells (i.e., white and red blood cells) along with common markers.

Overall, however, these approaches have remained largely qualitative and context-dependent. And while they have given rise to very precise methods of detection, and thus to operational definitions of specific cell types, they have not led to a general agreed-upon conceptual definition that can be straightforwardly applied across different biological sub-disciplines (Clevers *et al.* 2017). In line with this, one does not find much explicit discussion of the concept of cell type by biologists up until very recently. For instance, a standard textbook, such as (Alberts *et al.* 2015) does not contain any clear definition, nor does it its historical equivalent from 1924 (Cowdry 1924).

Despite this lack of a clear definition, there seems to be a shared intuitive understanding. Among the main criteria commonly alluded to in the discussions around cell types, we find *structure*, *function*, and *lineage* (Clevers *et al.* 2017). Structural criteria classify cells according to differences in the arrangement of and the relations between their parts. This includes broad features, such as shape, size and morphology, but also comprises details that are more specific, such as distinctive expressed molecules, or the presence of particular cellular substructures. Functional criteria, by contrast, classify cells according to the role that they carry out in the context of the organism. For example, fibroblasts are sometimes defined as cells that contribute to the formation of connective tissue by secreting collagen proteins (National Human Genome Research Institute 2022). Finally, lineage-based criteria classify cells according to their developmental ancestry. This means that we identify a type of cell in terms of its position in a lineage tree such as the one shown in Figure 1. Much of biological research seems to be based on the tacit assumption that these criteria neatly coincide and yield one objective classification scheme. However, it is by no means obvious that this is the case, and the assumption that different perspectives on a system lead to matching ways of decomposing it into parts may reflect a serious underestimation of its actual complexity (Wimsatt 2007).

Such complexity is revealed by recent research and advances in experimental methods. On the one hand, observations of cellular plasticity, dedifferentiation, transdifferentiation, and especially the "reprogramming" of terminally differentiated cells to a pluripotent state have led to a fundamental rethinking of some of the basic assumptions of the field (Andrews 2002; Sánchez Alvarado and Yamanaka 2014; Laplane and Solary 2019). The idea of cell types as clearly demarcated and irreversibly committed end points of differentiation has been put into question and given way to a more fluid picture.

In parallel, advances in genomics have led to the realization that the activities of cells are based on a complex and dynamic orchestration of genetic and epigenetic factors that influence their development as well as their morphology and functional properties. Biologists have revisited earlier ideas from Conrad Waddington who in the 1950s coined the metaphor of the epigenetic landscape, which compares the differentiation of cells and tissues to a marble rolling down an inclined surface (Waddington 1957). The particular shape of the surface, with hills and valleys, creates preferred paths and branching points for the marble, corresponding to developmental trajectories and decision points that eventually lead the developing system towards one of several possible ends or 'fates.' Drawing in particular on the work by (Kauffman 1974), it has been proposed that cell types should be understood as different "attractor states" of the complex dynamical system constituted by the gene regulatory network that is shared by all cells of an organism (Kauffman 2004; Huang 2009).

Finally, the advent of next-generation sequencing techniques, particularly single-cell sequencing has enabled biologists to measure the diversity of cell populations with an unprecedented level of detail.

For example, single-cell RNA sequencing provides a snapshot of the simultaneous expression of thousands of genes in individual cells, while single-cell ATAC sequencing captures the accessibility and therefore the regulatory state of the genome at the single-cell level (Van den Berge *et al.* 2019; Yan *et al.* 2020). Sophisticated computational techniques are required to convert the resulting high-dimensional data sets into representations that can be meaningfully analyzed, notably using various clustering algorithms along with methods to annotate these clusters to known cell types. The picture that emerges from this type of analysis suggests that cell populations previously considered to be of the same type are often much more heterogeneous than expected and appear to contain different sub-populations or cell states (Trapnell 2015).

Interestingly, there have recently been a number of articles by biologists that explicitly raise the question of how the cell type concept should be defined given the recent scientific and technological developments (Arendt *et al.* 2016; Clevers *et al.* 2017; Fishell & Heintz 2013; Morris 2019; Zeng 2022). Some of them suggest that the new techniques will allow a more formal and objective classification of cells into types and states analogous to the classification of chemical elements and their isotopes in the periodic table (Xia & Yanai 2019), while others are concerned that the observed degree of plasticity and heterogeneity will render any attempt to classify them into discrete types entirely subjective (Clevers *et al.* 2017). Overall, the impression is that intuitions about what constitutes a cell type vary considerably among biologists and seem to include different ways of capturing the relationships between the three criteria of structure, function, and lineage mentioned above. Corresponding to this is a lack of standardization in the field and a variety of often conflicting methods by which cell classifications are performed in practice based on the new experimental techniques, a problem well summarized in the following quote from a recent article in Cell, one of the leading biology journals:

> "Single-cell biology is facing a crisis of sorts. Vast numbers of single-cell molecular profiles are being generated, clustered and annotated. However, this is overwhelmingly ad hoc, and we continue to lack a principled, unified, and well-moored system for defining, naming, and organizing cell types" (Domcke & Shendure 2023, p. 1103).

The crisis described motivates the guiding questions of this paper: Are single-cell methods by themselves sufficient to enable a more coherent classification of cells? If not, what is their role in improving the traditional concept, which is considered deficient in important aspects? Before addressing these questions directly, I will provide some philosophical background on classification that will help illuminate some of the conceptual issues involved.

1. Cell types and the philosophy of classification

Creating a scheme according to which cells are assigned to specific types means creating a classification. Philosophers have been thinking about classification and related practices for millennia, and so it might be useful to look at some of this work to see if some insights might illuminate the search for the right cell type concept. In particular, the debates among biologists and philosophers of biology about the classification of organisms into a system of taxonomy have interesting parallels with the case of cell types.

A first distinction can be made between classifications that are arbitrary or simply based on human interests and classifications that in some way reflect actual patterns in the world. A common traditional way of thinking about such "natural" classifications is that objects in the world belong to the same class if they share certain basic properties or *essences*. For example, all water molecules share a common molecular structure described by the chemical formula H_2O. In general, the real essence of a class may not be known, just as the molecular structure of water was not known until relatively recently. John Locke thought that most actual classifications used by humans are based on nominal essences, by which he meant that they are based on observable macroscopic properties that do not necessarily coincide with the underlying and unknown real essences. While essentialism may be a defensible position with respect to chemical elements and molecules, it has been largely discarded in biological debates about the classification of organisms into species and higher taxa. The insights of evolutionary biology have shown that species are not static entities but are subject to change over time, and that the organisms within a species exhibit significant variation at any point in time. It has been doubted, therefore, whether in general any fixed set of

properties can be found that is shared by all and only the members of a given taxon (Hull 1965).

Following Ereshefsky (2000), two main approaches to classification have been proposed as an alternative to essentialism: cluster approaches and historical approaches. Cluster approaches are similar to essentialism in that they are also based on the properties of the objects being classified, but they are less rigid in that they do not require properties to be shared by all members of a class. Instead, membership is based on overall similarity, that is, the degree to which objects share a set of properties. A prominent example of such an approach in the context of biological taxonomy is pheneticism. Pheneticists hold that biologists should record as many properties of individual organisms as possible, usually in morphology or other observable traits, and then classify them according to a measure of distance in the "phenotypical space" constituted by these properties. Phylogeny or other evolutionary relationships are deliberately ignored. Historical approaches, by contrast, are not based on the shared properties of objects at all, the criterion instead is whether they share a causal history. In the context of biological taxa this causal history is provided by common evolutionary descent. David Hull's account of species as individuals is an example of a historical approach (Hull 1976).

It is interesting to see how different stages in the history of cell type classification can be reconciled with different philosophical approaches to classification, without this necessarily being in the minds of the biologists involved. I will admit right away that such an assignment is based on a sketchy and probably somewhat caricatured representation of the actual history. It should also be noted that this account is not intended to represent a purely conceptual development but is clearly determined to a considerable extent by the experimental technologies available for the study of cells. Nevertheless, I think this perspective illuminates some of the conceptual issues surrounding the problem of cell type classification.

The early investigations based on light microscopy and staining techniques can be interpreted as classifications based on nominal essences, in Locke's sense. Cells were identified and distinguished based on readily observable, macroscopic features, such as morphology, size, or color after staining. Already in the late 19th century, biologists suspected that chromatin, a stainable nuclear substance, was involved in cellular differentiation, assuming that stem cells preserve and pass on the complete chromatin of the fertilized egg, while differentiated somatic cells preserve only specific parts of it (Maehle 2011). However, these ideas could at the time not be linked to specific experimental measurements, and therefore the presumed "real essences" of cell types were out of reach.

At the same time, comparative embryologist studies revealed the developmental relationships of differentiating cells. Studying the formation of blood cells, researchers such as Artur Pappenheim and Alexander Maximow revealed complex 'stem trees' that displayed the genealogical relationships between different types of blood cells. These studies suggest an alternative criterion for the classification of cells into types according to their developmental ancestry, which is in analogy to the historical approach to classifying organisms according to phylogenetic relationships (Lancaster 2017).

The first half of the 20th century saw tremendous advances in how the genetic material affects the properties of cells and organisms, especially with the transition from classical genetics to molecular genetics. According to the central dogma of molecular biology, which can be considered as the culmination of these developments, the information-bearing part of chromatin is DNA, and this information is transferred from nucleic acids to proteins, determining phenotypic characteristics by specifying functionally active molecules. Consistent with this picture, cell types were conceptualized as endpoints of unidirectional and irreversible differentiation pathways, during which cells acquire the ability to produce specific types of proteins that enable them to carry out their respective functions in the organism. Historian of biology, Richard Burian summarizes this view as follows:

> "The underlying hypothesis was that differentiation is an irreversible commitment of a cell lineage to the manufacture of a coordinated set of "luxury" proteins—i.e., specialized proteins not needed to maintain the life of the cell. Thus, the primary differences among nerve, kidney, skin, and blood cells were thought to depend on the specialized sets of proteins that they make, which, in turn, affect their morphologies, interactions with other cells, and responses to biological signals and stimuli" (Burian 1993, p. 391).

We may interpret this view as the confirmation of an essentialist position, in which real essences are identified with the unique set of molecules that characterize the phenotype and function of a differentiated cell. In particular, in the context of immunology, it was assumed that cell types could be characterized in terms of their "surface phenotype" of specific proteins expressed on the cell surface and measured by flow cytometry (e.g., Lanier *et al.* 1983).

As mentioned above, more recent research has undermined this view by revealing, on the one hand, that cell fates are much less static and irreversible than previously thought. Dynamic transitions between cell fates can be induced experimentally and may occur also under physiological conditions, notably the "reprogramming" of terminally differentiated cells into a pluripotent state. On the other hand, genome-wide single cell sequencing methods have enabled a much more detailed exploration of the heterogeneity of cellular population and notably have provided a refined idea of how gene expression relates to cellular phenotypes and functions. The view that well-delineated sets of genes are switched either on or off to determine the characteristics of cell types now appears simplistic. Instead, the idea of basing cell type classifications on exhaustive molecular measurements seems much more solid. In addition, it seems that computational methods of identifying clusters in these high-dimensional datasets are nowadays available to make classifications easily available. This development can be understood as a move towards a pheneticist conception of cell types, and many of the arguments put forward in favor of such an approach resemble the arguments that pheneticists have formulated against alternative views on the taxonomy of organisms. One common argument, in particular, is that a pheneticist account is desirable because it would make classifications independent of any theoretical assumptions. In the following section we will take a closer look at the prospects of such a view in the context of cell type classification.

2. A theory-free account of cell types?

When looking at the recent discussions among biologists concerning the definition of cell types, a recurring motif is the idea that such a definition should be independent of theoretical assumptions, and that single-cell techniques can provide the basis for such a definition. Developmental biologist Samantha Morris, for instance, points out:

> "These methods enable the capture of many thousands of features, without the requirement for experimental cell enrichment, thus generating a rigorous and unbiased picture of the range of cell phenotypes that exists within any given tissue" (Morris 2019, p. 2).

In the context of neuroscience, Hongkui Zeng makes a similar case for data-driven classification:

> "To untangle this complexity, it is necessary to adopt approaches that provide comprehensive, unbiased, quantitative, and standardizable measurements and are scalable to densely sample a sufficient number of cells within a brain region or tissue organ as well as across the entire brain and body to eventually reach completeness, and then perform data-driven computational clustering and analysis to obtain cell type classification" (Zeng 2022, pp. 2739–2740).

Similarly, the neuroscientist Ed Lein emphasizes the superiority of those approaches to traditional ways of classifying cells:

> "...traditional approaches to neuronal classification rely on single-cell anatomy and physiology, which are typically qualitative and under-sampled. Transcriptomics has recently offered an unbiased, quantitative, and high-throughput alternative" (Clevers *et al.* 2017, p. 256).

And the authors of the article observing a "crisis" of single-cell biology, already quoted above, explicitly mention this as one of the desiderata for a successful cell type classification:

> "In our opinion, we should be pushing for a cell type nomenclature that meets some of the same key criteria as Linnaean taxonomy, as well as additional ones, including: (1) accommodating all cells arising during the life cycle of a given organism; (2) accommodating inter-individual variation, both normal and disease-related; (3) relating cell types to one another in a biologically meaningful way; (4) being stable to the incorporation of new data or new data types; and (5) *being constructed in a largely, if not entirely, data-driven manner* (Domcke & Shendure 2023, p. 1104, emphasis added).

Thus, there seems to be a common understanding that a purely data-driven classification of cells is both

desirable and feasible. In the remainder of this section, I will challenge this common understanding, drawing in particular on lessons learned from the debate on taxonomy. I start by providing some necessary background on single-cell experiments and analyses. This is followed by two lines of argument. First, I argue that practice shows that biologists do not believe that these types of experiments provide sufficient evidence to refute or justify any typology classification claim. Instead, such claims are always validated by more conventional and "theory-based" methods. Second, I provide more principled reasons for why a purely data-driven account of classification is destined to fail. These are analogous to some of the arguments that have been put forward against pheneticism in the context of taxonomy.

To avoid possible misunderstandings, I would like to point out at the outset that when I speak of "theory" in this context, I do not mean the traditional narrow sense of an axiomatic system based on laws of nature. Rather, the term "theory" here refers to any prior assumptions about underlying biological processes and mechanisms. Thus, a theory-free classification is one whose criteria depends solely on regularities in the observed data and do not presuppose any domain-specific knowledge. This corresponds to the use of "theory" and "theory-free" in the debates about the classification of organisms (see Ereshefsky 2000) and seems to capture the sense that contemporary biologists such as those cited above have in mind when they speak of "data-driven" or "unbiased" approaches.

3.1. A primer on single-cell experimentation and analysis

Progress in sequencing technology in the past two decades has enabled the quantification of gene expression on a genome-wide scale. To determine gene expression based on sequencing, the RNA isolated from a tissue is fragmented into small pieces that are afterwards sequenced in parallel. Counting the number of fragments that can be aligned to a particular gene sequence provides a quantitative proxy for the expression of that gene. Traditional RNA-sequencing (or "bulk" sequencing) experiments are based on mixed samples of thousands of cells and therefore provide an idea of the average expression of a gene in the sample. However, they do not provide information about the composition of the sample and about differences between individual cells. Single-cell sequencing technologies circumvent this problem by isolating single-cells in tiny droplets in a microfluidic device and adding a unique "barcode" sequence to each of them that allows the assignment of each RNA molecule to its cell of origin. Single-cell experiments thus provide a much higher-resolution image of gene expression in a population of cells. The result of a single-cell experiment is typically represented in the form of a large count matrix in which columns correspond to the individual cells and rows correspond to genes. Thus, each entry in this matrix indicates the number of reads (i.e., sequence fragments) of a particular gene in a particular cell. However, due to the small amounts of starting material, the resulting data are extremely noisy and sparse, which is to say that for any given cell in the sample a large fraction of genes will not be detected and appear as zeros in the count matrix. Therefore, perhaps paradoxically, single cell experiments cannot generally be used to obtain meaningful information about individual cells. However, they do provide information about the detailed structure of a cell population, which can be used to answer a range of biological questions.

The data analysis necessary to identify cell types based on single cell experiments consists of several steps. For the sake of brevity, I will focus on only two of them: dimensionality reduction and clustering. Other steps such as quality control, imputation, or normalization are part of the overall pipeline, but they are less directly related to the conceptual question at stake in this paper. In principle, each cell can be thought of as a data point in gene expression space, with each gene corresponding to one dimension of the space. Importantly, data analysis typically includes a step of dimensionality reduction. This means that the data are not analyzed and represented directly in the full gene expression space, but in a lower-dimensional space whose dimensions correspond to appropriate combinations of genes that capture important structural information in the given data set. Dimensionality reduction mitigates both the problem of noise and sparseness of data and the more fundamental "curse of dimensionality", which refers to the fact that as the number of dimensions increases, the distances between data points become more similar and thus less informative (Kiselev *et al.* 2019). Finally, dimensionality reduction makes subsequent analyses computationally more tractable. It corresponds to a

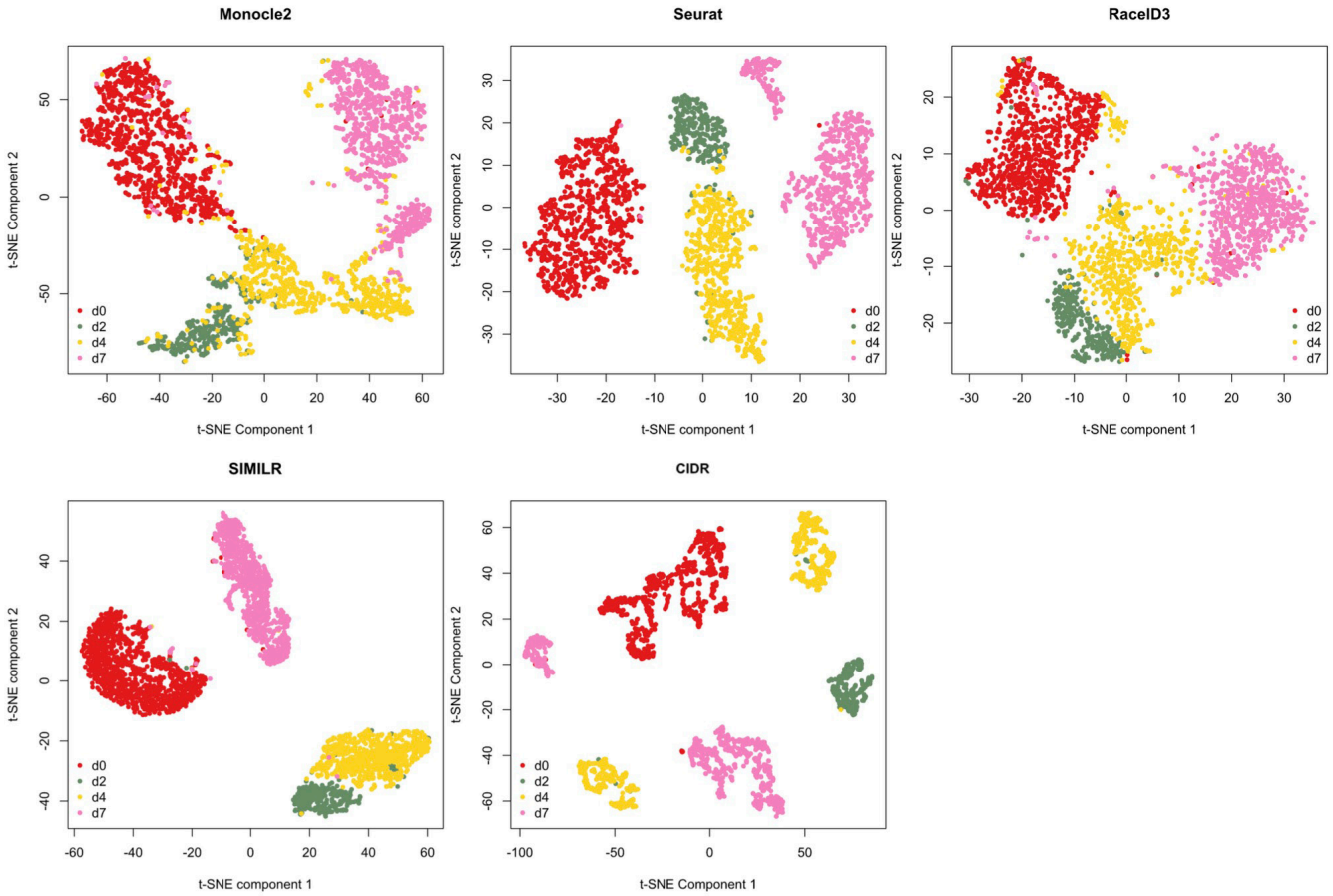

Figure 2: Comparison of different analysis pipelines for the identification of cell types based on a test dataset. Colors correspond to the "ground truth" annotations. Adapted from Zhang *et al.* (2023) (CC BY-NC 4.0)..

complex mathematical transformation of the data, which can be performed by various algorithms that may differ considerably in their results (Sun *et al.* 2019).

After dimensionality reduction, cluster analysis is used to identify cell types or other subsets of cells. In general, clustering methods are based on a measure of similarity between the objects to be clustered and then group the objects so that the objects in the same cluster are more similar to each other than the objects in other clusters. As with dimension reduction, there are a variety of different methods for performing this task, which may produce different results. Examples of widely used methods are k-means clustering and hierarchical clustering.

3.2. Are classifications based on single-cell methods really theory-free?

If one considers the current practice of biologists in establishing and using single-cell experiments to identify cell types, it quickly becomes apparent that they generally do not consider these experiments to provide a definition of cell types or to form a sufficient basis for classification. Instead, these methods are evaluated and calibrated based on previous biological knowledge. Thus, the choice of the appropriate clustering method and specific parameters is not imposed on scientists by the properties of the data alone. This means that the classifications resulting from these methods are not—strictly speaking—theory-free or unbiased, even though they are based on comprehensive and unsupervised methods of data analysis. The following statement from a recent review summarizes this current state of affairs:

> "Although considerable progress has been made in terms of clustering algorithms over the past few years, a number of questions remain unanswered. In particular, there is no strong consensus about what is the best approach or how cell types can be defined based on scRNA seq data" (Kiselev *et al.* 2019, p. 273).

One problem is that such methods usually base classifications on a specific class of molecular constituents, typically RNA, and neglect other potentially relevant classes (e.g. proteins or metabolites). However, the idea that a data set constrained in this way can lead to successful classifications amounts to an important theoretical assumption in itself that is not necessarily justified.

Another problem is due to the variety of methods that are available for important parts of the computational analysis, such as dimensionality reduction and cluster analysis. A theory-free account could be salvaged by assuming that all these methods lead to essentially the same classification. However, this does not seem to be the case. Figure 2 shows results from a study that compared different data analysis pipelines on the same data set. While there is clearly some agreement between the methods, the differences between results are perhaps even more striking. In particular, one can observe that cell groups, which are clearly separated by one method end up mixed or overlapping when another method is used.

Furthermore, even when focusing on one particular method alone, biologists are confronted with various choices. Dimensionality reduction obviously requires a decision on the dimension of the reduced space, which in turn affects the results of subsequent cluster analysis (Sun *et al.* 2019). Both too many and too few dimensions will lead to unsatisfactory results. An additional problem for some of these methods is that they are non-deterministic. For example, the widely used t-distributed stochastic neighbor embedding (tSNE) method is based on a non-deterministic algorithm, which means that different runs on the same dataset and with the same settings will lead to different lower-dimensional representations of the data (Zhang *et al.* 2023).

While clustering methods are usually considered as unsupervised methods, i.e., they identify features or structure in data without directly relying on prior information, they do rely on the choice of important parameter values. For example, k-means clustering requires the number of desired clusters (k) to be specified in advance, and most clustering methods rely on a distance measure between cells (when represented as data points in the reduced space), for which there are various possible choices. An obvious way out is to make the choice of these parameters automatic and data-driven as well, but then one has to choose the corresponding property to be optimized, for which again there are several possibilities.

A further issue is that clustering methods cannot be considered as completely devoid of biologically relevant assumptions. For example, k-means clustering relies on the assumption that there are discrete groups of cells in the first place, an assumption that is of course difficult to assess if one has no prior idea of the structure of the underlying cell population. Moreover, it tends to identify spherical clusters, which amounts to a strong assumption about the way in which cells of one type differ in their gene expression patterns. These assumptions can either lead to the failure to detect biologically relevant subpopulations (e.g., rare cell types) or, conversely, to the detection of spurious clusters.

The most important point, however, is that clustering methods in practice are evaluated based on a "ground truth", which consists in pre-labeled data sets (Zhang *et al.* 2023). As highlighted in the review cited above:

> "Perhaps the most challenging aspect of scRNA seq analysis (and this is not restricted to clustering) is how to validate a computational analysis method. The best strategy currently available is to have a setup where the cell types are known through other means, for example, by selecting cells from distinct cell lines, using tissues that are very well studied and understood (...), or considering cells taken from the earliest stages of embryonic development" (Kiselev *et al.* 2019, p. 278).

Thus, the choice of method and the specific settings are not determined based on "theory-free" considerations alone. Instead, it is importantly driven by the concern to reproduce previously accepted cell type classifications. If data-driven methods were indeed considered constitutive of the cell type concept, then the idea of an assessment based on a previously established baseline data set would not make sense, and other non-theoretical considerations would have to determine which method and settings should be used to identify and classify cells.

While it is possible that accepted single-cell based methods may subsequently be used to discover new cell types or even to correct and refine previous annotations, it seems inappropriate to refer to them as "theory-free" or purely data-driven as this would ignore the clearly theory-guided process of method selection.

It is also telling in this respect, that biologists usually do not accept the discovery of a new cell type based on single-cell experiments alone:

> "...for a new cell type to be accepted, it is necessary to go beyond characterization of the transcriptome. Researchers must demonstrate that the newly identified cluster is also functionally distinct. There are no universally applicable rules that can be applied here, and which assay is appropriate depends on the biological context" (Kiselev *et al.* 2019, p. 280).

This quote shows at the same time that some biologists seem to think that functional considerations that cannot be captured by gene expression data alone are relevant for cell type classifications, a point to which I will return later.

For the purpose of this paper, I can only hint at the full complexity of single-cell analysis, and I have neglected many aspects that may be considered equally relevant to the problem of identifying cell types. However, I think it has become clear that current scientific practices do not easily support the idea of a theory-free account of cell types.

3.3. General problems of a pheneticist approach

My previous arguments do not preclude the development in the future of methods based on single-cell experiments that can be considered theory-free in the relevant sense and accepted by biologists as truly constitutive of cell type classifications. In particular, an objection to the line of argument put forward in the previous section might be that it relies on the contingent imperfections of current single-cell technologies. Perhaps, an ideal single-cell experiment, unaffected by the noise and incompleteness of existing methods, could serve to build a satisfactory account of cell types. Consistent with this idea, some biologists have argued that the desired classification must be based on the integration of many different data modalities beyond gene expression, such as proteomic analysis and genome accessibility (e.g., Zeng 2022; Domcke & Shendure 2023). The underlying thought is that the more comprehensive data become, the more data-driven methods will approach the "natural" classification of cells into types.

In this section I will therefore move to some more principled reasons for doubting that this will be possible in a straightforward way. In particular, I will discuss some arguments that can be put forward against theory-free approaches in general, and in particular to the pheneticist approach to taxonomy. Further points take into account some specific features of the particular context of cell type classification.

One common argument against pheneticism is that the idea of "overall similarity" between the objects to be classified is not well-defined. Similarity is usually understood in terms of shared properties, but there is potentially an infinite number of properties that may be used for this assessment, and depending on the properties one chooses and how one weights their relative importance, one may arrive at very different and even diametrically opposed outcomes (Goodman 1972). The idea that this problem can be solved simply by measuring as many properties as possible rests on the tenuous "asymptote hypothesis". It states that, as the number of measured properties increases, the similarity converges to a constant value (Sneath 1995). The discussed "curse of dimensionality" illustrates the difficulties with this hypothesis. In defense of pheneticism, one might argue that the threat to a coherent notion of overall similarity is based on the mistaken idea that there is no restriction for allowed candidate properties, and that there is instead a set of "natural properties" on which a measure of similarity can be based (Lewens 2012). This latter move, however, presupposes prior ideas about which properties are biologically relevant; and while it might lead to a respectable version of pheneticism, clearly it would not be theory-free.

Another objection against pheneticism is that it is mistaken about the goals of taxonomy. The idea is that phenetic criteria of clustering organisms according to overall similarity will not pick out the evolutionarily salient actors. For instance, Ereshefsky (2000) points out that a pheneticist account would assign different developmental stages of the same organism or males and females of the same species to different groups. Similar considerations can be made for the case of cell types. It is conceivable that small differences in the expression of only a few genes can cause large phenotypic differences. On the other hand, there might be considerable differences in the transcriptomes of closely related cells due to stochastic variations or to transient differences (e.g. cell cycle stages). In such cases, it would be quite misleading to rely on a measure

of overall similarity which weights every feature equally. In response to such arguments, Lewens (2012) thinks that pheneticism should not be construed as a proposal that replaces other approaches to taxonomy that pursue specific goals. Rather, pheneticism provides a general-purpose taxonomy that allows for the investigation of more specific hypotheses regarding a variety of scientific problems. If general-purpose taxonomy clashes with groupings established by different means, this can be taken as a reason for refining the former. In the context of cell types, this might be an attractive option, notably in light of the fact that many biologists explicitly strive for an account of cell types that is universally applicable across all biological contexts (as manifested in the attempts to build comprehensive reference classifications, such as the human cell atlas). However, it should be clear that a classification obtained as a result of such a process of iterative refinement will not itself be theory-free. In addition, one may ask whether in the context of cell types there is a similar plurality of purposes as in taxonomy. One recurring idea in recent literature is that cell types should ultimately be defined in terms of their function (Clevers *et al.* 2017). Thus, if there is indeed overwhelming consensus that cell type classifications should track functional differences, then the argument of mistaken goals regains at least some of its bite. While it is plausible that divisions based on functional differences will roughly coincide with structurally defined differences, conflict between the two approaches is not at all excluded. Why then should biologists focus so much on a theory-free classification approach if that approach misses the central goal that cell type classifications are meant to achieve?

Finally, it should be noted that there are important differences between the questions faced by evolutionary biologists and those faced by biologists interested in classifying cell types. For example, many of the debates between different approaches to the taxonomy of organisms reflect the difficulty of inferring phylogenetic relationships because of incomplete evidence about past evolutionary events. Therefore, a pheneticist approach is attractive because it does not make any assumptions about unobservable events and processes. This problem is less severe in the case of ontogenetic relationships between cells because it is possible, at least in principle, to directly study the events involved in cellular differentiation and organismal development. The concern about independence from "theory", therefore, has a different urgency in evolutionary contexts because such theory usually involves weak hypotheses that likely will be overturned by new evidence.

All these considerations lead to think that even in the long run, single-cell technologies will not be able to provide a purely theory-free classification of cell types.

Conclusion

In this contribution, I considered the question whether and to what extent recent single-cell technologies challenge the notion of cell type. There is an intuitive concept of cell type that is based in some way on a combination of structural, functional, and developmental criteria. I have suggested that the cell type concept at different historical stages can be aligned with different approaches to classification. In particular, the idea of grounding cell classifications in the unbiased and theory-free clustering of single-cell data can be understood as the application of pheneticism to the context of cells. I have provided arguments to question that such a theory-free account can be achieved, both based on current scientific practice and on more principled grounds. It is interesting to see that concrete proposals of how cell types should be classified based on single-cell experiments are clearly theory-based in important ways. For example, the "periodic table" of cell types presented by Xia and Yanai (2019) does not use comprehensive gene expression, but relies on the idea of "core regulatory complexes" to provide the subsets of genes that are relevant for comparison. Similarly, Domcke and Shendure (2023) argue that a satisfactory description of cell identity must go beyond static molecular profiles and include information about ontogeny, i.e., the lineage tree of cells that corresponds to the development of the organism. Bioinformaticians are working on techniques to estimate phylogenetic relationships based solely on single-cell data (e.g., Farrell *et al.* 2018), but I strongly suspect that upon closer inspection these methods will not prove to be theory-free in the sense discussed in this paper either. I will save a more detailed discussion of this topic for a later occasion.

Does this mean that single-cell experiments do not affect cell type classifications at all? This does not seem plausible. However, overemphasizing the idea that a respectable approach to cell classification must be theory-free is wrong. One way for single-cell

experiments to affect cell classifications is by correcting particular assignments of cells to certain types. They simply provide additional information that may lead biologist to reconsider assignments they have made based on an incomplete evidence. The more interesting question, however, is whether single-cell experiments will affect classification criteria. This is less clear, but could be envisaged if one drops the requirement that classification should be theory-free. In particular, we could think of single-cell experiments as a way to iteratively refine the traditional concept, rather than replace it. Once an analysis pipeline has been validated based on test data of prior classifications, it can be used to make predictions on unseen data. While in case of mismatch previous classifications or biologists' intuitions might initially be given more weight, in the long run one may end up with a "reflective equilibrium" that represents the best compromise between fit and certain theoretical desiderata. The analysis method would then effectively be a theory that embodies, extends, and systematizes biologists' prior intuitions about what a cell type is.

References

Andrews, PW 2002, "From teratocarcinomas to embryonic stem cells", *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, vol. 357, no. 1420, pp. 405–417. https://doi.org/10.1098/rstb.2002.1058.

Arendt, D, Musser, JM, Baker, CVH, Bergman, A, Cepko, C, Erwin, DH, Pavlicev, M *et al.* 2016, "The origin and evolution of cell types", *Nature Reviews Genetics*, vol. 17, no. 12, pp. 744–757. doi: 10.1038/nrg.2016.127.

Baskin, DG 2015, "A historical perspective on the identification of cell types in pancreatic islets of Langerhans by staining and histochemical techniques", *Journal of Histochemistry & Cytochemistry*, vol. 63, no. 8, pp. 543–558. https://doi.org/10.1369/0022155415589119.

Bruce, A, Johnson, A, Lewis, J, Morgan, D, Raff, M, Roberts, K, & Walter, P 2015, *Molecular biology of the cell*, 5th ed. New York: Garland Science.

Burian, RM 1993, "Technique, task definition, and the transition from genetics to molecular genetics: Aspects of the work on protein synthesis in the laboratories of J. Monod and P. Zamecnik", *Journal of the History of Biology*, vol. 26, no. 3, pp. 387–407. https://doi.org/10.1007/BF01062055.

Canguilhem, G 1995, *Études d'histoire et de philosophie des sciences*, 5th ed. Paris: Vrin.

Chan, JKC, Ng, CS, & Hui, PK 1988, "A simple guide to the terminology and application of leucocyte monoclonal antibodies", *Histopathology*, vol. 12, no. 5, pp. 461–480. https://doi.org/10.1111/j.1365-2559.1988.tb01967.x.

Clevers, H, Rafelski, S, Elowitz, M & Lein, E 2017, "What is your conceptual definition of 'cell type' in the context of a mature organism?", *Cell Systems*, vol. 4, no. 3, pp. 255–259. https://doi.org/10.1016/j.cels.2017.03.006.

Cowdry, EV, ed. 1924. *General cytology: A textbook of cellular structure and function for students of biology and medicine*, Chicago: University of Chicago Press.

Domcke, S, & Shendure, J 2023, "A reference cell tree will serve science better than a reference cell atlas", *Cell*, vol. 186, no. 6, pp. 1103–1114. https://doi.org/10.1016/j.cell.2023.02.016.

Duchesneau, François. 1987. *Genèse de la théorie cellulaire*. Paris: Vrin.

Ereshefsky, M 2000, *The poverty of the Linnaean hierarchy: A philosophical study of biological taxonomy*, Cambridge: Cambridge University Press.

Farrell, JA, Wang, Y, Riesenfeld, SJ, Shekhar, K, Regev, A, & Schier, AF, 2018, "Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis", *Science*, vol. 360, no. 6392, pp. eaar3131. https://doi.org/10.1126/science.aar3131.

Fishell, G, & Heintz, N 2013, "The neuron identity problem: Form meets function", *Neuron*, vol. 80, no. 3, pp. 602–612. https://doi.org/10.1016/j.neuron.2013.10.035.

Goodman, N 1972, *Problems and projects*, Indianapolis: Bobbs-Merrill.

Huang, S 2009. "Reprogramming cell fates: Reconciling rarity with robustness", *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, vol. 31, no. 5, pp. 546–560. https://doi.org/10.1002/bies.200800189.

Hull, DL 1965, "The effect of essentialism on taxonomy—Two thousand years of stasis", *The British Journal for the Philosophy of Science*, vol. 15, no. 60, pp. 314–326. https://doi.org/10.1093/bjps/XV.60.314.

Hull, DL 1976, "Are species really individuals?", *Systematic Biology*, vol. 25, no. 2, pp. 174–191. https://doi.org/10.2307/2412744.

Jones, EG 1999, "Golgi, Cajal and the neuron doctrine", *Journal of the History of the Neurosciences*, vol. 8, no. 2, pp. 170–178. https://doi.org/10.1076/jhin.8.2.170.1838.

Kauffman, SA 1974, "The large scale structure and dynamics of gene control circuits: An ensemble approach", *Journal of Theoretical Biology*, vol. 44, no. 1, pp. 167–190. https://doi.org/10.1016/S0022-5193(74)80037-8.

Kauffman, SA 2004, "A proposal for using the ensemble approach to understand genetic regulatory networks", *Journal of Theoretical Biology, Special Issue in Honour of Arthur T. Winfree*, vol. 230. No. 4, pp. 581–590. https://doi.org/10.1016/j.jtbi.2003.12.017.

Kiselev, VY, Andrews, TS, & Hemberg, M 2019, "Challenges in unsupervised clustering of single-cell RNA-seq data", *Nature Reviews Genetics*, vol. 20, no. 5, pp. 273–282. https://doi.org/10.1038/s41576-018-0088-9.

Lancaster, C 2017, *A history of embryonic stem cell research: Concepts, laboratory work, and contexts*, Doctoral thesis, Durham University.

Lanier, LL, Engleman, EG, Gatenby, P, Babcock, GF, Warner, NL & Herzenberg, LA, 1983 "Correlation of functional properties of human lymphoid cell subsets and surface marker phenotypes using multiparameter analysis and flow cytometry", *Immunological Reviews*, vol. 74, no. 143, pp. 143-160.

Laplane, L, & Solary, E 2019, "Towards a classification of stem cells", *eLife*, vol. 8, article e46563. https://doi.org/10.7554/eLife.46563.

Lewens, T 2012, "Pheneticism reconsidered", *Biology & Philosophy*, vol. 27, no. 2, pp. 159–177. https://doi.org/10.1007/s10539-011-9302-2.

Maehle, AH 2011, "Ambiguous cells: The emergence of the stem cell concept in the Nineteenth and Twentieth centuries", *Notes and Records of the Royal Society of London*, vol. 65, no. 4, pp. 359–378.

Morris, SA 2019, "The evolving concept of cell identity in the single cell era", *Development*, vol. 146, no. 12, article dev169748. https://doi.org/10.1242/dev.169748.

Murphy, PR, Narayanan, D, & Kumari, S 2022, "Methods to identify immune cells in tissues with a focus on skin as a model", *Current Protocols*, vol. 2, no. 7, article e485. https://doi.org/10.1002/cpz1.485.

Neto, C 2020. "When imprecision is a good thing, or how imprecise concepts facilitate integration in biology", *Biology & Philosophy*, vol. 35, no. 6, p. 58. https://doi.org/ 10.1007/s10539-020-09774-y.

National Human Genome Research Institute 2022, "Fibroblast". Available from: www.genome. gov/genetics-glossary/Fibroblast.

Novik, AA, Ionova, TI, Gorodokin, G, Smolyaninov, A, & Afanasyev, BV 2009. "The Maxow 1909 centenary: A reappraisal", *Cellular Therapy and Transplantation*, vol. 1, no. 3, pp. 31–34.

Sánchez Alvarado, A, & Yamanaka, S 2014, "Rethinking differentiation: Stem cells, regeneration, and plasticity", *Cell*, vol. 157, no. 1, pp. 110–119. https://doi.org/10.1016/j.cell.2014.02.041.

Sneath, PHA 1995, "Thirty years of numerical taxonomy", *Systematic Biology*, vol. 44, no. 3, pp. 281–298. https://doi.org/10.1093/sysbio/44.3.281.

Sun, S, Zhu, J, Ma, Y, & Zhou, X 2019, "Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis", *Genome Biology*, vol. 20, no. 1, p. 269. https://doi.org/10.1186/s13059-019-1898-6.

Trapnell, C 2015, "Defining cell types and states with single-cell genomics", *Genome Research*, vol. 25, no. 10, pp. 1491–1498. https://doi.org/10.1101/gr.190595.115.

Van den Berge, K, Hembach, KM, Soneson, C, Tiberi, S, Clement, L, Love, MI, Patro, R, & Robinson, MD 2019, "RNA sequencing data: Hitchhiker's guide to expression analysis", *Annual Review of Biomedical Data Science*, vol. 2, no. 1, pp. 139–173. doi:10.1146/annurev-biodatasci-072018-021255.

Waddington, CH 1957, *The strategy of the genes: A discussion of some aspects of theoretical biology*. London: Allen & Unwin.

Wimsatt, WC 2007, *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Xia, B, & Yanai, I 2019, "A periodic table of cell types", *Development*, vol. 146, no. 12, article dev169854. https://doi.org/10.1242/dev.169854.

Yan, F, Powell, DR, Curtis, DJ, & Wong, NC 2020, "From reads to insight: A hitchhiker's guide to ATAC-seq data analysis", *Genome Biology*, vol. 21, no. 1, p. 22. https://doi.org/10.1186/s13059-020-1929-3.

Zeng, H 2022, "What is a cell type and how to define it?", *Cell*, vol. 185, no. 15, pp. 2739– 2755. https://doi.org/10.1016/j.cell.2022.06.031.

Zhang, S, Li, X, Lin, J, Lin, Q, & Wong, KC 2023, "Review of single-cell RNA-seq data clustering for cell-type identification and characterization", *RNA*, vol. 29, no. 5, pp. 517–530. https://doi.org/10.1261/rna.078965.121..

Special Issue, Single-cell Analysis: Epistemological Inquiries

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

DOI: 10.13133/2532-5876/18129

Single-cell Molecular Analysis: When an Experimental Technique Reveals Conceptual Controversies

András Páldi[ab*] *& Laëtitia Racine*[ab]

[a] *Ecole Pratique des Hautes Etudes - PSL, Paris, France*
[b] *Centre de Recherche Saint-Antoine, Paris, France*

***Corresponding author:** András Páldi, Email: Andras.Paldi@ephe.psl.eu

Abstract

The last decade has witnessed a rapid evolution of highly sensitive single-cell molecular analysis techniques. These techniques allow the simultaneous detection and quantification of mRNA and protein molecules in a large number of individual cells. Some of these methods are already commercialized, making them readily available to any interested lab. While the pitfalls concerning the experimental extraction of biocomponents (mRNA and protein) and analytical bioinformatic methods are widely discussed in the literature, little is known regarding the conceptual difficulties raised by single-cell methodologies. Considered and treated as pure technical difficulties, these issues are rarely discussed explicitly. This is a problem as conceptual difficulties precede technical ones and contribute, to a large extent, to the failure of techniques. Consequently, a new theoretical framework is urgently needed to make sense of the ever-increasing amount of data.

Keywords: ontology, cell type, cell classification, single-cell technology

Citation: Páldi, A, & Racine, L 2023, "Single-cell Molecular Analysis: When an Experimental Technique Reveals Conceptual Controversies", *Organisms. Journal of Biological Sciences*, vol. 6, no. 2, pp. 25–30. DOI: 10.13133/2532-5876/18129

While central to biology, the process of cell differentiation is still not understood. The traditional molecular biology approach to differentiation uses a detailed description of gene expression changes and their correlation to the cell's morphological and physiological characteristics (phenotype). Study of individual cells has become a standard procedure for the investigation of a number of biological questions including differentiation. Single-cell techniques are considered as the best way to discover new and rare cell types, identify their differentiation pathways and the clonal structure of these cell populations (Mincarelli *et al.* 2018).

Multicellular organisms are composed of a large number of phenotypically different cells usually sorted in distinct categories called "cell types". The classification of living organisms and their parts is at the basis of biology as a science. The first classification of biological species was proposed by Carl von Linné in the 18th century in his work *Systema Naturae*. The system was based on hierarchical ranking of the living organisms in classes, orders, genera, species, and varieties. Linné's system, based on the similarity between the entities at each level of the hierarchy, is a perfect application of the essentialist ontology originally proposed by Aristotle and dominant in Western thinking since antiquity. Although Linné's binomial nomenclature is still in use nowadays, the system of classification based on similarity has been questioned by the Darwinian theory of evolution. Darwin proposed a new way of

classification based on descent rather than similarities. In this classification different entities belong to the same category if they are derived from the same ancestor. The Darwinian view emphasizes the importance of individuals instead of categories defined on the basis of a set of properties shared by all individuals. Species and higher taxa are reduced to a pragmatic and artificial category made for convenience. As put by Darwin in the Chapter 14 of the *On the Origin of Species*: "In short, we shall have to treat species in the same manner as those naturalists treat genera, who admit that genera are merely artificial combinations made for convenience. This may not be a cheering prospect, but we shall at least be freed from the vain search for the undiscovered and undiscoverable essence of the term species" (Darwin 1859, p. 485). The individuals that are usually classified within the same taxonomic category called species are better characterized by their genealogical proximity rather than their resemblance. As a result, the boundaries between species became blurred. How many generations separate two individuals of two different species? The answer to this question is a matter of convenience, there is no universal rule. We can consider any morphological, functional or genetic characteristics—the result is always circumstantial (Mallet 1995; Mayr 1996).

It is difficult not to notice the analogy between the concept of species and that of cell types. The fact that a multicellular organism always develops from a single initial cell leaves no doubt about the common origin of all cells of the body. Early studies of the embryo development first identified the three germ layers, ectoderm, mesoderm and endoderm, then the specific structures – the organs – derived from them. From the 19th century until recently, embryologists investigated the origin of the organs, tissues and cell lineages during development. As a result, the classification of the tissues and their cell types was naturally based on origin, rather than on similarity. Embryology textbook illustrations represent germ layers, organs, tissues and cells in a hierarchical graph reminiscent of a genealogical tree. This is a Darwinian way of considering cell types. In parallel, cell biologists, anatomists and physiologists used the well-known classification method based on morphological and functional features. The two visions co-existed and complemented each other until the last decades of the 20th century. When molecular biology became dominant in life sciences, then the situation changed. According to the molecular genetic vision, cells are controlled by a program that is "hard wired" in the genes, and differentiation is a process of this program. Hence, same-type cells must express the same genes and can be identified on the basis of the transcriptional regulator (transcription factors) that they express (Davidson & Erwin 2006).

When flow cytometry, the first single-cell analysis method, was introduced, it was generally admitted that a variation of the mRNA or of the protein expression in same type cells was a simple stochastic fluctuation and a cell type was represented by the average of these parameters (Levsky & Singer 2003). A flow cytometer provides rapid analysis of multiple parameters with physical and chemical characteristics on single cells such as size, granularity and surface protein profile. Usually, it measures the fluorescence intensity emitted by specific surface proteins labelled with a fluorescent tag, generally an antibody. The fluorescence intensity is proportional to the number of molecules on the cells' membrane. This approach allows to label and measure several proteins in a single run, thus obtaining single-cell information from a large number of individual cells. The analysis of the results is typically performed using graphical plots. The most striking systematic observation brought by this technique is the large variation between single-cell values. This means that the amount of any expressed protein varies systematically on an unexpectedly large scale even between cells belonging to the same clonal population. However, it is common to convert data to a logarithmic scale to simplify data representation, which inherently reduces the apparent variation rendering it irrelevant. In fact, most of the actors in the field used to consider (and many still consider) same type cells and same clonal population to be essentially identical. In their opinion, any observed variation comes from measurement noise or size differences due to cell cycle. Groups of cells are defined on a graph using a procedure called "gating". This is mostly guided by the subjective appreciation of the fluorescence intensity of the cells. Even though some procedures based on multi-parametric algorithms exist, these are not widespread, and most experts are still using software like Kaluza or Flowjo in which the gate definition is done by hand in a subjective manner.

These groups are considered as different cell types or subtypes and are subject of further investigation to determine their biological properties. Their analysis

usually focuses on the average value of the selected population of cells, so the individual cell-specific information gets lost. Average is considered as a kind of "essence" or "norm" of the cell type that shows how each cell would be on its own if the biological noise was irrelevant to individual variations. Genealogical relationships for the classification of cell types are usually not considered. Finally, the first technique aimed at studying single cell characteristics is used to provide population average, and the cell types defined in this way are approximate categories based on the subjective assessment of similarities between cells. The problem of information loss by the use of averages has already been recognized in biology long time along (Benzer 1953). A detailed discussion of the mathematical inadequacy of using average in biology can be found in (Rauch, Wattis & Bray 2023).

Thanks to the ability to amplify individual nucleic acid molecules by polymerase chain reaction, the resolution of the usual molecular detection techniques has increased more recently. Moreover, numerous new methods emerged and made possible the simultaneous detection and quantification of the whole sets of mRNA molecules, chromatin structural profiles, proteins etc. in a vast number of individual cells. Many authors consider that this technological advance represents an opportunity to redefine and systematically detect cell types (Wagner, Regev, & Yosef 2016; Morris 2019). The amount of single-cell resolution data generated by these experimental techniques is much higher than what is provided by flow cytometry, because they detect more features in a single cell. While flow cytometry can detect the abundance of a limited number of proteins in a single cell, the new technology can detect the approximate number of RNA transcripts of each individual gene in each individual cell simultaneously in a large number of cells. Each cell is described then by as many features as the number of genes, and the resulting data set may contain several hundred million of data points. As it is impossible to analyze huge amount of data by simple visual inspection on a graphical display, sophisticated computational analysis methods are required. However, those modern techniques did not immediately resolve a fundamental question: how to differentiate different cell types? As indicated above, this question is an adaptation to cell biology of a fundamental question of philosophical ontology about entities and identities. Over the last few years the question of cell types has become a subject of intense discussion among biologists (Mincarelli *et al.* 2018; Wagner, Regev, & Yosef 2016; Morris 2019; Han *et al.* 2020; Xia & Yanai I 2019). Surprisingly, however, the nature of the difficulties in answering the question about cell types in most cases is considered technical. For example, some authors explicitly declare that "classification" of cells into discrete types from single-cell profiles is a problem of "unsupervised clustering in high dimensions" (Wagner, Regev, & Yosef 2016). The pre-Darwinian essentialist way of conceptualizing cell types is never questioned. Instead, it is admitted that each individual cell in the organism can be assigned to a well-defined class. This classification is considered as one of the primary objectives of single-cell technologies (Mincarelli *et al.* 2018). A significant effort is made to establish cell catalogues (cell atlases) of various multicellular organisms (for example: www.humancellatlas.org). The cells are grouped on the basis of the similarity of their gene expression patterns, a unique "ID card" for each cell type. In other words, cells belonging to the same type are supposed to share a minimal set of expressed genes. The overall difference between the gene expression patterns of the cells isolated from different organs or tissues of the developing embryo or adult organism is easily distinguishable. However, distinguishing groups of cells with clearly different gene expression patterns from a mixture of cells isolated from the same tissue is far more difficult. Perhaps, the best illustration comes from the study of the human hematopoietic stem cells lineage, that demonstrated the highly variable and continuous nature of mRNA profiles between cells considered as different cell types on the basis of their functional characteristics (Velten *et al.* 2017). A very high number of cells have intermediate gene expression patterns, that is, no minimal set of genes is expressed only in a well-defined group or cluster. Highly likely, a gene expression pattern does not allow identifying the type of a cell randomly picked up from a population. Whatever the mathematical method to cluster the data, some subjective decision is always required and the final result depends on the choice of some key parameters used by the algorithm (p-values, thresholds, filters, the presumed number of clusters one expects, etc.) (Luecken & Theis 2019; Breda, Zavolan, & van Nimwegen 2021). As a result, the number of identifiable cell clusters depends as much on those biased parameters as on data. This procedural subjective component rarely

emerges during discussions about cell type analysis methods. In the light of this, it is not surprising that more and more studies report on new rare cell types identified based on single-cell data analysis. There is a high risk that those discoveries are in fact the results of an over interpretation of single-cell data. This problem illustrates the ambiguities of the "cell type" concept defined solely by single cell mRNA profiling. It also shows how the misuse of a concept imposes limitations to our thinking by canalizing the discussions on the technical aspects, leading to the conclusion that collecting more data will solve the difficulties.

To circumvent the problem of rare cell types, one of the most popular *ad hoc* explanations proposed is to further divide the cell type into smaller categories named "cell states", etc. The idea is that single-cell data represent a snapshot of the studied population and rare cell profiles may represent a short-lived transitory cell state. For example, Wagner and colleagues "refer to the more permanent aspects in a cell's identity as its type (e.g., a hepatocyte typically cannot turn into a neuron) and to the more transient elements as its state. Cell types are often organized in a hierarchical taxonomy, where types may be further divided into finer subtypes; such taxonomies are often related to a cell fate map, reflecting key steps in differentiation" (Wagner, Regev, & Yosef 2016). Unfortunately, further dividing a population of cells into "types" or "states" changes nothing to the initial problem since it does not provide any better solution for the classification. This way of categorizing cell types and states into hierarchies merely changes the name of the class from "type" to "state" and suffers from the same conceptual shortcomings as exposed above. As the species concept in population biology, this vision of cell types has operational utility when applied to whole cell populations or whole organisms, but fails when individual cells are considered. Paradoxically, single-cell technologies revealed that the "cell type", contrary to what its name suggests, is a concept that describes the features of a cell population and not that of an individual cell. For purely pragmatic reasons, cell populations but no single cells can be grouped on the basis of their gene expression profiles. The concept of cell type as mentioned above is the closest biological analogy of the concept of "species". Cell "type" can fit a group of cells based on their biological function. It can emerge from the characteristics of individual cells but it is inapplicable to them in the same way as the concept of "pressure" describes a gas but cannot be applied to individual gas molecules. The way the cell type is inferred from the single cell mRNA data captures in fact some kind of "average", a rather statistical reflection of a more or less arbitrary chosen population of cells. Contrary to what is asserted by many authors, the average does not describe the intrinsic, functional and context-independent biological features of individual cells. Claims such as: "it is possible to practically define cell types according to their expressed transcription factors (TFs)" (Xia & Yanai 2019) are simply not supported by observation (Weinreb, Rodriguez-Fraticelli, Camargo, & Klein 2020). We do not know yet how individual cells behave and to what extent they can change their function and morphology (what we call "phenotype"). Single-cell mRNA profiling is a simple "cross section" of a temporal process at a given time-point; alone it cannot provide the information many experts expect without taking into account the temporal character of the cell, her lineage history and environment.

Therefore, calling for the revision of the "cell type" concept is one of the unrecognized but important contributions of single-cell technologies. Such a revision is, however, impossible without rethinking another key concept, i.e. "cell identity". The identity of an individual cell—its phenotype—is not simply an intrinsic property of the cell that can be deduced from its molecular composition. The cell is continuously interacting with the biological (the other cells), physical (intracellular matrix) and chemical (available nutrients, oxygen, pH, etc.) micro- and macro-environments. These interactions act as extrinsic constraints; their changes promote and canalize the phenotypic change of the cells. In turn, the cell also modifies its environment, forming in this way a complex interacting system. On the other hand, the phenotype is also constrained by the cell's own life history and genealogy, conveyed by what is usually called cellular or epigenetic memory. Cellular memory represents an intrinsic limitation to the change by restraining the repertoire of genes that can be easily expressed (Páldi 2020). As a result, at any moment, the cell phenotype is determined by the outcome of the interplay between intrinsic and extrinsic constraints and reflects a dynamic equilibrium of rapid change-promoting and inhibiting processes. The phenotype encompasses the whole life cycle of the cell and it is impossible to specify the exact moment when the "true identity" appears. Therefore, the phenotype

or "identity" of a cell is better described as a dynamic equilibrium of many different and frequently opposing processes than as a static state (Dupré & Nicholson 2018). Recent observations suggest that the transition of the cell toward a new phenotype, usually called fate choice or differentiation, is indeed highly dynamic and not a simple switch as previously thought (Moussy *et al.* 2017; Parmentier *et al.* 2022). It is more like a trial-and-error process based on the permanent dynamic exchange between the cell and the micro-environment. From the point of view of the "process", the capacity to change requires no specific explanation as variation is its true nature. What requires explanation however, is the lack of change, i.e., stability, the equilibrium of the antagonistic processes. Only if the equilibrium is maintained for an extended period, then the cell morphology and function appear stable. It may be tempting to consider the cell's appearance during this period as the "true" phenotype. Nevertheless, a simple snapshot is unsuitable to determine the stability of a cell phenotype. This is only possible based on continuous observation over a period of time or using a time series of snapshots of the same cells. It is worth remembering however, that "stable" or "transient" depends entirely on the time scale of these observations. There is no privileged time scale. If the frequency of the snapshots is lower and the period of observation is longer, the rapid changes are not detected and the proportion of different morphologies or gene expression patterns will appear constant in a cell population (Brock, Chang, & Huang 2009). Current single-cell mRNA detection technologies provide only a single snapshot for an individual cell because they are invasive to the point of destroying the cells during the analysis. Although there are promising attempts to overcome this limitation (Chen *et al.* 2022; Boersma *et al.* 2019; Lyon, Aguilera, Morisaki, Munsky & Stasevich 2019), it is currently impossible to repeat the same measurement on the same cell or repeat the analysis of the same cell at a later point. Taken together, these considerations suggest that single-cell molecular approaches, as they stand today, can only be used to follow the general trend of changes if applied to a time series of pre-defined groups or cell populations. These general trends tell us little about the trajectory of individual cells; they only allow for conjectures.

Over the past decade, single-cell molecular technologies have produced a huge amount of data. Although this gives us the illusion of knowledge, only a small fraction of such information is really exploited to improve our understanding of the process of cell differentiation. What we really need now is a new interpretation framework based on solid theoretical ground to develop analytic methods and go beyond the calculation of gene expression profile and resemblance between groups of cells. Such a method should establish a true association between the single-cell gene expression pattern and the individual cell's phenotype that can be used for functional studies. A promising way to build a new paradigm is to capitalize on the organicist tradition of the pre-molecular biology period, as suggested by several authors (Dupré & Nicholson 2018). As Paul Weiss put it: "Life is a dynamic process. Logically, the elements of a process can be only elementary processes, and not elementary particles or any other static units"... Life "can never be defined in terms of a static inventory of compounds, however detailed, but only in terms of their interactions" (Allen 1962).

References

Allen, JM (ed.) 1962, *The Molecular Control of Cellular Activity*. New York - Toronto - London: McGraw Hill. Available from: https://archive.org/details/molecularcontrol0000alle/mode/2up.

Benzer, S 1953, "Induced synthesis of enzymes in bacteria analyzed at the cellular level", *Biochimica et Biophysica Acta*, vol. 11, no. 3, pp. 383–395. Available from: https://doi.org/10.1016/0006-3002(53)90057-2.

Boersma, S, *et al.* 2019, "Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding", *Cell*, vol. 178, no. 2, pp. 458–472.e19. Available from: https://doi.org/10.1016/j.cell.2019.05.001.

Breda, J, Zavolan, M, & van Nimwegen, E 2021, "Bayesian inference of gene expression states from single-cell RNA-seq data", *Nature Biotechnology*, vol. 39, no. 8, pp. 1008–1016. Available from: https://doi.org/10.1038/s41587-021-00875-x.

Brock, A, Chang, H, & Huang, S 2009, "Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours", *Nature Reviews. Genetics*, vol. 10, no. 5, pp. 336–342. Available from: https://doi.org/10.1038/nrg2556.

Chen, W, *et al.* 2022, "Live-seq enables temporal transcriptomic recording of single cells", *Nature*, vol. 608, no. 7924, art. 7924. Available from: https://doi.org/10.1038/s41586-022-05046-9.

Darwin, C 1859, *On the Origin of Species*, London: John Murray.

Davidson, EH, & Erwin, DH 2006, "Gene regulatory networks and the evolution of animal body plans", *Science*, vol. 311, no. 5762, pp. 796–800. Available from: https://doi.org/10.1126/science.1113832.

Dupré, J & Nicholson, DJ 2018, "A manifesto for a processual philosophy of biology", in: Nicholson, DJ, & Dupré, J (eds.) *Everything Flows: Towards a Processual Philosophy of Biology*, Oxford - New York: Oxford University Press, pp 3–46. Available from: https://doi.org/10.1093/oso/9780198779636.003.0001.

Han, X, *et al.* 2020, "Construction of a human cell landscape at single-cell level", *Nature*, vol. 581, no. 7808, pp. 303–309. Available from: https://doi.org/10.1038/s41586-020-2157-4.

Levsky, JM, & Singer, RH 2003, "Gene expression and the myth of the average cell", *Trends in Cell Biology*, vol. 13, no. 1, pp. 4–6. Available from: https://doi.org/10.1016/s0962-8924(02)00002-8.

Luecken, MD, & Theis, FJ 2019, "Current best practices in single-cell RNA-seq analysis: A tutorial", *Molecular Systems Biology*, vol. 15, no. 6. Available from: https://doi.org/10.15252/msb.20188746.

Lyon, K, Aguilera, LU, Morisaki, T, Munsky, B,& Stasevich, TJ 2019, "Live-cell single RNA imaging reveals bursts of translational frameshifting", *Molecular Cell*, vol. 75, no. 1, pp. 172–183.e9. Available from: https://doi.org/10.1016/j.molcel.2019.05.002.

Mallet, J 1995, "A species definition for the modern synthesis", Trends in Ecology & Evolution, vol. 10, no. 7, pp. 294–299. Available from: https://doi.org/10.1016/0169-5347(95)90031-4.

Mayr, E 1996, "What is a species, and what is not?", *Philosophy of Science*, vol. 63, no. 2, pp. 262–277. Available from: https://doi.org/10.1086/289912.

Mincarelli, L, Lister, A, Lipscombe, J, & Macaulay, IC 2018, "Defining cell identity with single-cell omics", *Proteomics*, vol. 18, no. 18, p. 1700312. Available from: https://doi.org/10.1002/pmic.201700312.

Morris, SA 2019, "The evolving concept of cell identity in the single cell era", *Development*, vol. 146, no. 12, art. dev169748. Available from: https://doi.org/10.1242/dev.169748.

Moussy, A, *et al.* 2017, "Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment", *PLOS Biology*, vol. 15, no. 7, art. e2001867. Available from: https://doi.org/10.1371/journal.pbio.2001867.

Páldi, A 2020, "Random walk across the epigenetic landscape", in: Levine, H, Jolly, MK, Kulkarni, P, & Nanjundiah, V (eds.) *Phenotypic Switching: Implications in Biology and Medicine*, London - San Diego - Cambridge - Kidlington: Academic Press - Elsevier, pp. 53–76. doi: 10.1016/B978-0-12-817996-3.00008-6.

Parmentier, R, *et al.* 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. Available from: https://doi.org/10.1371/journal.pbio.3001849.

Rauch, C, Wattis, J, & Bray S 2023, "On the meaning of averages in genome-wide association studies: What should come next?", *Organisms*, vol. 6, no. 1, pp. 7–22. Available from: https://doi.org/10.13133/2532-5876/17811.

Velten, L, *et al.* 2017, "Human haematopoietic stem cell lineage commitment is a continuous process", *Nature Cell Biology*, vol. 19, no. 4, pp. 271–281. Available from: https://doi.org/10.1038/ncb3493.

Wagner, A. Regev, A, & Yosef, N 2016, "Revealing the vectors of cellular identity with single-cell genomics", *Nature Biotechnology*, vol. 34, no. 11, pp. 1145–1160. Available from: https://doi.org/10.1038/nbt.3711.

Weinreb, C, Rodriguez-Fraticelli, A, Camargo, FD, & Klein, AM 2020, "Lineage tracing on transcriptional landscapes links state to fate during differentiation", *Science*, vol. 367, no. 6479, art. eaaw3381. Available from: https://doi.org/10.1126/science.aaw3381.

Xia, B, & Yanai, I 2019, "A periodic table of cell types", *Development*, vol. 146, no. 12, art. dev169854. Available from: https://doi.org/10.1242/dev.169854.

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.

Perspectives and Hypotheses

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

DOI: 10.13133/2532-5876/18130

The Common Origin of Multicellularity and Cancer: Lessons from the Fossil Record

Daniela Montagna[a] *& Raúl Ruggiero*[a**]

[a] *Instituto de Medicina Experimental (IMEX-CONICET), Academia Nacional de Medicina de Buenos Aires. Pacheco de Melo 3081, PC 1425, Buenos Aires, Argentina*

***Corresponding author:** Raúl Ruggiero, Email: *ruloruggiero@yahoo.com.ar*

Abstract

Despite the methodological limitations in the study of fossil record and some confusion in the literature about the diagnostic distinction between real neoplasia and other types of proliferation or even malformations in species very distant from mammals, paleopathological studies have revealed many cases of *bona fide* benign as well as malignant neoplasms in animals and land plants since Paleozoic Era. Further, almost all types of modern neoplastic diseases have been documented in ancient *Homo sapiens* bone remains. It is worth to note that, despite the major changes in the structure of animal populations, the prevalence of malignant as well benign neoplasms has remained relatively constant (and in some cases it has even increased) among the different taxa of animals for hundred million years. This suggests that malignancies as well as benign neoplasms are rooted quite deeply in the evolutionary life of organisms. This seemingly unremarkably fact represents a remarkable riddle for evolutionary biologists. If natural selection, working on living organisms has been powerful enough to produce complex adaptations, from the eye to the immune system, why has it been unable to eliminate or even reduce the incidence of cancer, even though many apparently less harmful traits have been eliminated during species evolution? Based on the fact that, both today and in the fossil record, cancer seems to occur in organs that have experienced a decline or loss of their regenerative ability we suggested that cancer may be an ultimate, even futile, reparative attempt. Therefore, the permanence of cancer by hundred million years might be understood as if its existence is coupled to the normal regenerative mechanisms of the organisms without which no pluricellular organism could survive. This interpretation, encoded in the so-called hypothesis of the biological sense of cancer, was built within the broad framework of tissue organization field theory (TOFT) by assuming that cancer is primarily a disease of higher levels of organization, that is, an organismic, organ- or tissue-based disease rather than a cellular one.

Keywords: multicellularity, cancer, fossil record, paleopathology, TOFT

Citation: Montagna, D & Ruggiero, R 2023, "The Common Origin of Multicellularity and Cancer: Lessons from the Fossil Record", *Organisms: Journal of Biological Sciences*, vol. 6, no. 2, pp. 43–69. DOI: 10.13133/2532-5876/18130

1. The Many Routes to Multicellularity and Cancer

In the ancient seas of the Earth, about a billion (10^9) years ago, multicellular organisms began to evolve from eukaryotic unicellular ancestors. There is convincing evidence that this process was not unique, and that multicellularity has evolved independently many times in the history of life: once in animals (*metazoa*), once in land plants (*embryophytes*); once in each sac (*ascomycetes*) and club (*basiodiomycetes*) fungi; once in green algae (*chlorophytes*) and at least one in both red algae (*rhodophytes*) and a complex group that includes brown algae called heterokontophytes (Aktipis *et al.* 2015).

Remarkably, however, regardless of the different evolutionary lines of multicellular organisms, all of these processes share two central features as a precondition to preserve and restore the organismic homeostasis. First, the capacity for cooperation among their cells—that includes different types and levels of differentiation associated with division of labors, resource transport and creation and maintenance of the extracellular environment. And second, the development of mechanisms designed to regulate cell death and to control normal cell proliferation, although the intimate nature of these mechanisms remains controversial (Sonnenschein & Soto 2021; Sanchez Alvarado & Yamanaka 2014).

A dis-regulation of the latter mechanisms is assumed to be the basis of the *normal architecture-modifying proliferative growth of tissue*, which is collectively called "neoplasia" (etymologically, "new growth"), in which tissue organization and function are altered.

If this new growth has relatively little effects on the organismic homeostasis, it can be defined as benign neoplasia. On the other hand, if it affects the organismic homeostasis in ways that may have profound effects for organism fitness and survival, it can be defined as malignant neoplasia or cancer.

Herein, we define these terms in a broader sense than that clinically defined for humans and, for extension, for mammals. In effect, in clinical settings, the term "benign" is reserved for slow-growing, relatively well differentiated neoplasms that remain localized in the tissue of origin. In contrast, the term "malignant" is used to denote fast-growing neoplasms that invade and destroy adjacent or distant (metastases) tissues and display many additional features such as less cellular differentiation or anaplasia, acceleration of cell cycle and high number of mitotic figures, genomic alterations, increase cell mobility, chemotaxis, changes in the cellular surface, secretion of lytic factors, etc. (Robert 2010). We do not use the classical definition of Ewing (1940) in which "a neoplasm (either benign or malignant) is an autonomous, or relatively autonomous, growth of tissue" (autonomous meaning the ability to disobey the rules that control normal cell proliferation) because this statement, that has guided cancer research for more than 80 years, is actually a postulation rather than a true definition. In effect, pathologists do not use it as an operational tool to diagnose the presence of a tumor. In fact, the means to diagnose cancer have not changed that much since the 19th century, when pathologists began describing the histological pattern of tumors using the light microscope (Sonnenschein & Soto 1999; Mayo Clinic 2023). In addition, if the mechanisms that control normal cell proliferation are still unknown, how can anyone be assured that cancer cells are disobeying those mechanisms?

It is possible that not all features of human cancer are present in species or lineages very distant from mammals such as invertebrates or land plants. In consequence, it might be said that those species or lineages do not get cancer. However, if we do not focus exclusively on human cancer and we adopt the more general definitions stated above, "cancer" or "cancer-like phenomena" (as proposed by some authors) might be present in a much larger collection of multicellular organisms than originally thought (Aktipis *et al.* 2015; Dujon *et al.* 2022).

2. Limitations of the Fossil Record

The broad definition of neoplasia given above distinguishes it from other common proliferations or phenomena such as malformations, hyperplasia, regenerative growths, inflammation, etc. although the boundaries among them are not always very clear. This statement is particularly true for extinct organisms. In fact, the interpretation of neoplasms in the fossil record is one of the more challenging aspects of paleopathology.

In the first place, the material available to paleopathologists consists, in most cases, of osseous remains, apparently limiting the detection of neoplasms to bone tumors of ancient vertebrates and leaving behind the most ancient pluricellular organisms.

However, upon certain circumstances, structures of invertebrates such as exoskeletons and shells could be preserved. In effect, the precipitation or growth of mineralized exoskeletons and shells is widely distributed in invertebrate taxa within the phylum Mollusca as well as the subphyla Crustacea. The composition of these structures varies throughout invertebrates, and, similar to vertebrate bones and teeth, consists of mineral and organic components that can be fossilized depending on special biological, physical, and chemical conditions. Furthermore, upon very rare circumstances, soft tissues can also be preserved. One mechanism that facilitates soft tissue preservation is phosphatization, where the tissue is replaced by calcium-phosphate minerals. However, in these cases, the process does not preserve the physical structure of the organs. Exceptionally rare, intact or almost intact soft-tissue fossils have been found in some rocks. This process is known as Burgess Shale-type (BST) preservation. Burgess Shale is a fossil-bearing deposit exposed in the Canadian Rockies of British Columbia, Canada, famous for the exceptional conservation of fairly tough tissues such as cuticle as thin films, and soft tissues as solid shapes, even those pertaining to organisms of extreme antiquity. Consequently, soft normal and neoplastic tissues might have also been eventually preserved associated with both ancient vertebrates and invertebrates. The BST preservation is not yet completely understood although latest investigations suggest that soft tissue fossil-bearing rocks apparently contain some minerals that inhibit bacteria, preventing the process of decomposition after death. On this basis, scientists hope to elucidate the underlying mechanisms of this process to find more soft-tissue fossils (Keenan 2021).

In the second place, the diagenetic process that affects the fossil remains may produce post-mortem alterations that either simulate or overshadow cancer-linked lesions that occurred during life. In fact, focal and multiple alterations induced in bone by different physical, chemical and/or biological factors may produce erosions that can mimic lesions associated with primary or metastatic neoplasms. On the other hand, the diagenetic process may also superimpose alterations (for example, incrustations) that may hide cancer lesions or modify their original appearance (Capasso 2005). Today, however, methodological progress, especially in the field of archeometry, has improved our capacity to distinguish a variety of lifetime parameters, including cancer images and lesions, from alterations produced after the death of the organisms (Grupe & Harbeck 2014).

In the third place, the difficulty to find neoplasms and especially cancer in the fossil record is associated with two characteristics of the wild life: first, most individuals tend to die at a young age due to starvation, infections or predation, at a time when the incidence of cancer is very low; second, when the age to get cancer is reached, tumor-bearing organisms could be more susceptible to predation than healthy individuals, limiting the possibility to appear in the fossil record. Moreover, predators are thought to prey on individuals that are in poor physical condition. This can explain why benign tumors or early but not metastasized cancer are more commonly detected in organisms in the wild (Perret *et al.* 2020).

In the fourth place, a fossil record represents an instantaneous picture and not a moving process. Consequentially, the chance of determining whether, during the life of an extinct organism, a neoplasm could have affected—and to what extent—its fitness and survival, is an inference based on the peculiar traits of the neoplasm but not a direct observation. In

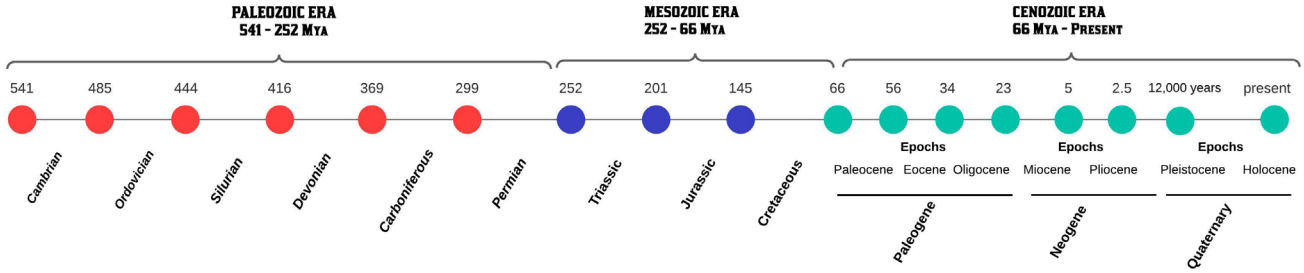

Figure 1. Chronology of the different geological eras, periods, and epochs.

Table 1: Cases of benign and malignant tumors in the fossil record of animals and land plants.

Geological eras		***Species***	***Tumor type and location***	***References***
Paleozoic Era	Invertebrates (trilobites)	*Centropleura loveni* (Cambrian)	Nearly circular prominent bubble-shaped structure. Anterior pleura ridge of thoracic segment	*De Baets et al. 2021; Babcock 1993*
		Conomicmacca hyperion (Cambrian)	Large neoplasia. Posterior pleura region adjacent to the furrow that separates the pygidial axis from the pleuron	*Elicki & Geyer 2013*
		Toxochasmops (Ordovician)	Metaplasia	*Nielsen & Nielsen 2017*
		Bohemoharpes ungula (Silurian)	Neoplasm accompanied by radiating circulatory canals	*Owen 1983*
	Vertebrates	*Dinichthys* (Devonian)	Bone resorption due to a malignant tumor of the soft tissues of the mouth floor	*Capasso 2005; Scheele 1954*
		Phanerosteon mirabile (Carboniferous)	Osteoma including a bone focal hyperostosis	*Capasso 2005; Moodey 1927*
		"Mammalian" forebear (Permian)	Compound odontoma (a benign neoplasia of calcified dental tissue)	*Whitney, Mose, & Sidor 2017*
	Plants	*Odontoperis* (Early Permian) *Pteridiorichnos stipitopteri, Walchia piniformis* (Carboniferous)	Abnormal outgrowths of plant tissues denominated galls. Galls are produced by host plant cells in response to infection by fungi, bacteria, nematodes, insects, mites or other agents.	*Labandeira 2021; Scott, Stephenson, & Collinson 1994; Schachat & Labandeira 2015; Impson, Post, & Hoffmann 2013; Schread 1971*
Mesozoic Era	Vertebrates	*Triassic capitosaurid,* amphibian (Early Triassic)	Parostotic osteosarcoma in a cranial bone	*Gubin et al. 2001*
		Shell-less stem-turtle *Pappochelys rosinae* (Middle Triassic)	Osteosarcoma on the femur	*Haridy 2019*
		Metoposaurus, krasiejowensis (Late Triassic)	Osteosarcoma	*Surmik et al. 2022*
		Comanchean dinosaur (Early Cretaceous)	Hemangioma between two caudal vertebrae	*Moodie 1921*
		Mosasaurus, Pachyrhinosaurus, Vagaceratops irvinenesis, Titanosaurus, Hadrosaurs (Cretaceous)	Osteomas	*Moodie 1921; Rothschild et al. 2003; Rothschild & Martin 1993; Rega, Holmes, & Tirabasso 2010; Souza Barbosa et al. 2016; Norman & Milner 1989*
		Edmontosaurus (Cretaceous)	Metastatic cancer	*Rothschild et al. 2003*
		Apatosaurus, Allosaurus, Vagaceratops irvinenesis (Jurassic)	Osteochondroma	*Capasso 2005; Rega, Holmes, & Tirabasso 2010;Foth et al. 2015*

	Plants	*Viaznikopteris rigida, Dicroidim odontopteroides* (Early Triassic) *Ginkgoites sp., Desmiophyllum sp.* (Cretaceous)	Galls	*Labandeira 2021; Krassilov & Karasev 2008; McLoughlin 2011; Vasilenko 2005*
Cenozoic Era	Vertebrates	*Fishes* (Tertiary and Quaternary) *Sirenian mammals* (Oligocene) *Elephants* (Quaternary)	Osteomas	*Capasso 2005; Przyklady 1965*
		Hesperocyon (Oligocene)	Osteochondroma	*Wang & Rotschild 1992*
		Mammoth (Late Oligocene)	Osteoblastoma	*Krzeminska 2008*
		Ungulates from Argentina, horses, European mammoths, Japanese elephants and in a walrus from Alaska.	Neoplasms of the dental tissues	*Capasso 2005; Cabrera 1934; Patte 1937; Hunter & Langston 1964; Kobayashi 1937*
		Bovidae, Canidae, Nothroterium Ursusus spelaeus	Benign tumors	*Miralles & Crusafont Pairo 1952; Pales & Wernert 1953; Moodie 1929; Scott 1898; Pales 1959*
		Buffalo, Capra, Nothrotherium maquinense	Osteosarcoma	*Conkling 1990; Capasso & Di Tota 1996; Souza Barbosa et al. 2021; Baker & Brothwell 1980*
	Plants	*Taxodium dubium, Alnus julianiformis* (Paleogene)	Galls	*Chen & Appleby 1984; Jiang et al. 2021*
	Pre-human and ancient human populations	*Australopithecus sediba*	Osteoid osteoma	*Quinney et al. 2016*
		Homo ergaster	Osteosarcoma	*Odes et al. 2016*
		Homo erectus	A possible Burkitt lymphoma or an ossifying sarcoma	*Capasso 2005*
		Homo steinheimensis	Meningiomas	*Czametzki, Schwaderer, & Pusch 2003; Czametzki 1980*
		Homo neanderthalensis	Meningiomas, intradiploic epidermal cyst	*Hublin et al. 2009*
		Homo sapiens	Meningiomas, hemangiomas, osteoclastomas, histiocytomas, osteomas, osteocondromas, osteosarcomas, condrosarcomas, hemangiosarcomas, Ewing's sarcoma. Bone metastases of nasopharyngeal, breast and prostatic carcinoma and lytic lesions due to multiple myeloma and melanoma.	*Capasso 2005; Czametzki, Schwaderer, & Pusch 2003; Shimkin 1977; Strouhal 2001; Pahl 1986; Luna et al. 2008; Luna et al. 2015; Arrieta, Mendonca, & Bordach 2018*

fact, strictly speaking, the very existence of a neoplasm (new growth) in a fossil is also an inference, because no "growth" during life can be directly demonstrated in a dead body.

3. The Fossil Record of Cancer

Despite the limitations mentioned in the precedent paragraph and despite some confusion in the literature about the diagnostic distinction between real neoplasia and other types of proliferation or even malformations in species very distant from mammals , the fossil record has revealed many cases of *bona fide* benign as well as malignant neoplasms since Paleozoic Era. Figure 1 shows the chronology of the different geological eras, periods, and epochs. Table 1 summarizes the many cases of benign as well as malignant neoplasms observed in fossils of both animals and plants that will be described below.

3.1. Non-human Organisms: Paleozoic Era (541-252 Million Years Ago)

The most ancient reported neoplastic cases may be traced to the Paleozoic Era. In effect, 23 neoplasia have been detected in fossils of trilobites, an extinct class of marine arthropods with over 20,000 species having been described, that lived for almost 270 million years, from the early Cambrian [Cambrian period lasted 56 million years between 541 to 485 million years ago (Mya) and it was the time when practically all major animal phyla first appeared in the fossil record] up to the late Permian (299-252 Mya). Because trilobites had wide diversity and an easily fossilized exoskeleton, they have left an extensive fossil record. Some of these neoplasms have been attributed to parasitism and/or traumatic injuries while the origin of others remains uncertain. Some examples show simple bulbous swellings with a central crater-like depression that could be produced in slow-healing ulcers induced by infections of pre-existent injuries. In other cases, however, the growth seems to invade and damage adjacent structures resembling the invasive neoplasia observed in human beings and other mammals. Probably the best example has been detected in an incomplete carapace of a specimen of *Centropleura loveni* from the Cambrian (more than 500 Mya): the neoplasia was a nearly circular prominent bubble-shaped structure developed from the anterior pleura ridge of thoracic segment 6. The neoplasia affected not only the part of the pleura from which it originates but also the posterior part of the anteriorly neighboring pleura, the posterior margin of which is indented (De Baets *et al.* 2021; Babcock 1993). A large neoplasia was also observed in a specimen of the bathynotid trilobite *Conomicmacca hyperion* from the Cambrian. The neoplasia was located in the posterior pleura region adjacent to the furrow that separates the pygidial axis from the pleuron and it could have affected the pleural area immediately adjacent to the neoplasm as well as the pygidium growth (Elicki & Geyer 2013).

In addition, in a specimen of *Bohemoharpes ungula* from the Silurian (between 444 and 416 Mya), the neoplasm is accompanied by radiating circulatory canals, similar to the way that tumors often attract blood vessel development (Owen 1983). Further, in a *Toxochasmops* trilobite from the Ordovician (485-444 Mya), putative images of metaplasia were observed in an anomalous growth of tissue although it is not easy to distinguish herein a true neoplasm from a diagenetic process or a regeneration after an injury (Nielsen & Nielsen 2017).

In vertebrates, the earliest known possible case of neoplasm was found in a fossil of an armored large fish from the extinct genus *Dinichthys*, which lived in the late Devonian, about 360 Mya. The case consists of a profound depression on the internal surface of the lower jawbone. The lesion, which certainly occurred during the life of the fish (that is, it was not produced by diagenetic processes) could have been caused by trauma linked to intra-specific aggression among these combative animals; however, the paleopathologists better interpreted it as the result of bone resorption due to a malignant tumor of the soft tissues of the mouth floor (Capasso 2015; Scheele 1954). The armored fishes known as placoderms are considered to be the earliest branch of the jawed fishes and in consequence, they are one of the first groups of vertebrates to appear on the Earth after the jawless fishes.

More direct evidence of neoplasia was obtained from a fossil of the extinct bony fish *Phanerosteon mirabile* that lived in the lower Carboniferous, about 300 Mya. This neoplasia is a classic osteoma including a bone focal hyperostosis (excessive bone growth) similar to that observed in bony fishes living today (Capasso 2015, Moodie 1927).

In terrestrial vertebrates, the oldest case reported up to date is a lesion characterized as a compound odontoma (a benign neoplasia of calcified dental tissue) in a specimen of a "mammalian" forebear, a premammalian synapsid that lived in the late Permian, about 255 Mya. Odontomas are the most common odontogenic tumors and its recognition in such a distant specimen suggests that this condition is unlikely related to characteristics of mammalian dentition but rather evolved much earlier in vertebrate evolution (Whitney, Mose, & Sidor 2017).

The fossil record of tumors in the Paleozoic is not restricted to animals.

Many land plants exhibit abnormal outgrowths of plant tissues (tumors) that are denominated galls. Galls are produced by host plant cells in response to infection by fungi, bacteria, nematodes, insects, mites or other agents. In most cases, galls do not seriously harm the host plant and could be considered benign neoplasia. In the more highly developed galls, these self-limiting neoplastic growths are almost comparable, in the determinate growth of their structures, to a leaf or a fruit (Bayer, Kaiser, & Micozzi 1994). A few of them, however, may be highly deleterious for their hosts. The best—but not the only—example of the latter are the crown galls caused in many plants (such as nut trees, perennial fruit trees, vines and roses) by the soil-inhabiting bacterium *Agrobacterium tumefaciens* that, in some cases, especially when the gall completely encircles the main stem, can severely harm and kill the hosts (Armstrong 1995; Chen *et al.* 2016; Kluepfel *et al.* 2017; Gohlke & Deeken 2014; Zhu *et al.* 2020; Grabowski & Koetter 2019). Figure 2 shows an example of crown galls induced by *Agrobacterium tumefaciens* in a present-day tree.

In the narrow framework of human pathology, crown galls as well as other galls that may affect the survival of the host plant, cannot be considered cancer because they do not invade or metastasize. In effect, the rigid wall of the plant cells as well as the absence of a vascular system able to transport cells, prevents the galls to invade or metastasize in the true sense of these words. However, in the broader sense of malignant neoplasia stated in this work, the galls that are harmful for their hosts may be considered genuinely cancer plants. Recognition of crown galls as true cancer may be traced up to the onset of the 20th century (Smith 1916). Its malignant behavior is also exemplified by the fact that, as with many animal tumors, unless caught very early in tumorigenesis, surgical excision of crown gall tumors from the infected plants is ineffective in controlling the disease (Lacroix & Citovsky 2001).

The physical features of most galls (hardened, three-dimensional and resistant to flattening) allow their

Figure 2. An example of crown galls induced by *Agrobacterium tumefaciens* in a present-day tree (Picture by the Authors).

preservation in the fossil record and provide a basis for evaluating their external and eventually internal structure (Labandeira 2021). In fact, galls on land plants have a spotty but periodically rich and abundant fossil record. Many galls in the fossil record may have been induced by arthropods although in some cases, the causes remain undetermined.

The earliest gall registered is a putative insect or mite-induced gall on a liverwort (an early cryptogam) of the Middle Devonian Period at 385 Mya. Afterwards, gall activity was registered about 315 Mya, during the Carboniferous, on vegetative and reproductive axial organs of horsetails, ferns and probably conifers but not on foliage (Labandeira 2021). The earliest galls in leaves were detected on *Odontoperis* from the early Permian (Scott, Stephenson, & Collinson 1994), followed by the expansion of foliar galling through the late Permian as supported by plant damage, an extensive diversification of small, early *hemipteroid galler* lineages on seed-plant foliage and the paleoclimate record (Schachat & Labandeira 2015).

The nature (whether benign or malignant) of these very ancient neoplasms is unknown but it might be inferred from the behavior of analogous present galls. For example, the anatomical and three-dimensionally preserved rachis gall of the extinct tree-fern *Pteridiorichnos stipitopteri* from the Carboniferous is similar to some modern fern rachis galls caused by gall midges, which do not negatively affect the growth of its host. On the other hand, a beaked gall described on the axes of the extinct conifer *Walchia piniformis* from the Carboniferous is similar to the aphid-induced gall on Norway spruces that, in severe cases, may cause disfigurement, stunting and eventually death of the affected trees (Labandeira 2021; Impson, Post, & Hoffmann 2013; Schread 1971).

3.2. Mesozoic Era (252-66 Million Years Ago)

The first period of the Mesozoic Era was the Triassic (252-201 Mya) that began after Earth's worst-ever life devastation, the Permian-Triassic extinction, also known as the Great Dying, when an unidentified event killed some 90 percent of the planet's species.

Among animals, the earliest cases of neoplasia in the Mesozoic Era were a parostotic osteosarcoma reported in a cranial bone of an early Triassic capitosaurid amphibian that lived between 252 and 247 Mya (Gubin *et al.* 2001) and an osteosarcoma on the femur of a specimen of the extinct shell-less stem-turtle *Pappochelys rosinae* that lived 240 Mya in the middle Triassic (Haridy *et al.* 2019). The appearance of the latter tumor conforms with present-day periosteal osteosarcoma in humans and represents the oldest instance of bone cancer in an amniote. Another case of osteosarcoma has recently been reported in the vertebral intercentrum of a temnospondyl amphibian, *Metoposaurus krasiejowensis*, that lived between 237 and 201 Mya ago in the late Triassic (Surmik *et al.* 2022).

Afterwards, there are many documented cases of benign as well as malignant neoplasms in fossils from extinct animals that lived during the Jurassic (201-145 Mya) and Cretaceous (145-66 Mya). Among benign tumors, the most represented cases were hemangioma, osteomas, and osteochrondoma. The earliest case of hemangioma was reported in a fossil fragment—presumably a vertebral centra—of an unidentified dinosaur that lived between 165 and 145 Mya in the late Jurassic (Rothschild *et al.* 1998). Afterwards, hemangioma have been found between two caudal vertebrae of a not identified Comanchean dinosaur that lived in the early Cretaceous between 145 and 100 Mya (Moodie 1921) and in some late *Cretaceous hadrosaurus* or duck-billed herbivorous dinosaurs which lived about 80 Mya (Rothschild *et al.* 2003). Similarly, osteomas have been described in two Cretaceous specimens belonging to mosasaurus family—an extinct group of large marine reptiles that were positioned at the top of the food chain in the late Cretaceous oceans (Moodie 1921; Rothschild *et al.* 2003), in the left scapula of a specimen of *Pachyrhinosaurus*—a ceratopsid dinosaur of the late Cretaceous—in the right foot of a *Vagaceratops irvinenesis*—an herbivorous ceratopsian dinosaur which lived during the late Cretaceous about 75 Mya (Rega, Holmes, & Tirabasso 2010)—in a bone tail of a titanosaurus, a gigantic long-necked and long-tailed sauropod dinosaur from the late Cretaceous (Souza Barbosa *et al.* 2010) and in some specimens of Cretaceous hadrosaurs (Rothschild *et al.* 2003; Norman & Milner 1989). In the same way, the earliest case of osteochondroma was found in a rib of a specimen of the extinct genus of Apatosaurus, a giant herbivorous sauropod that lived in the late Jurassic between 156 and 150 Mya (Capasso 2005). At least two additional

cases of osteochondroma have been informed: one in a specimen of *Allosaurus*, a large predatory theropod dinosaur which lived in the Upper Jurassic, between 155 and 145 Mya (Foth *et al.* 2015) and another, in a specimen of *Vagaceratops irvinenesis* (Rega, Holmes, & Tirabasso 2010) Other less frequent benign neoplasms or neoplasm-like bone lesion have also been described such as osteoblastoma, desmoplastic fibromas and Langerhans Cell Histiocytosis in some specimens of hadrosaurs (Rothschild, Tanke, & Helbling 2003; Rothschild *et al.* 2020), an ameloblastoma in the lower jaw of a specimen of a dinosaur *Telmatosaurus transsylvanicus* from the late Cretaceous (70-66 Mya) (Dumbravá *et al.* 2016) and an osteoblastic tumor—identified by the presence of a large outgrowth of ovoid appearance with a spiculated microstructural pattern—in the femur of a specimen of the sauropod *Bonitosaura salgadoi* that lived approximately 84 Mya near the end of Cretaceous period (González, Gallina, & Cerda 2017).

Different malignant neoplasms have also been reported—especially in the last few years—in fossil remains from the Jurassic and Cretaceous periods although, in certain cases, diagnosis needs to be further confirmed. Stadtman (Stadtman 1992) reported a probable chondrosarcoma invading the surrounding normal bone in the humerus of a theropod dinosaur (*Allosaurus fragilis*) from the late Jurassic. Another case of chondrosarcoma was reported in a specimen of Vagaceratops irvinenesis from the Late Cretaceous (Rega, Holmes, & Tirabasso 2010).

In the same way, at least two cases of osteosarcoma have been informed: one, in a specimen of *Dilophosaurus wetherilli*, a theropod dinosaur that lived in the early Jurassic, between 200 and 190 Mya (Senter & Juengst 2010) and another, in a specimen of *Centrosaurus apertus*, a herbivorous ceratopsian (horned) dinosaur, which dates from approximately 77·75 Mya (Ekhtiari *et al.* 2020). In addition, a putative thumb-sized brain tumor, was found in a skull fossil of *Gorgosaurus*, a 7.5 m long meat-eater giant closely related to *Tyrannosaurus rex*, that lived 72 Mya (Pickrell 2003). The tumor, possibly an unusual type of bone-forming cancer called an extraskeletal osteosarcoma, filled nearly the entire area formerly occupied by the cerebellum and brainstem and probably impaired the cerebrum, the part of the brain that controls thought and memory.

In addition, two putative cases of multiple myeloma have been described in the cranial bones of both, a specimen of *Torosaurus latus*—a herbivorous horned dinosaur that lived in the Late Cretaceous between 68 and 66 Mya—and an *ornithischian* dinosaur (Capasso 2005). Metastatic cancer has been reported at least twice in fossil remains. The first case was described in a fossil sawed bone section from an unspecified large-sized terrestrial dinosaur that lived between 156 and 148 Mya, in the Upper Jurassic. The permineralized bone contains an ovoid agate filling occupying a large hole whose appearance is that of a lytic zone that is penetrated by irregular trabeculae and that seems to have originally contained a mass of soft tissue. This image together with both the existence of a transition zone between normal bone and the tumorous space characterized by a pattern of bone destruction, and a radiographically detected cortical bone invasion with residual cortical shell, strongly suggest the existence of metastatic cancer (Rothschild *et al.* 1999). Another metastatic cancer was reported in a specimen of *Edmontosaurus* belonging to the family of hadrosaurs, that lived about 70 Mya, in the Late Cretaceous (Rothschild *et al.* 2003). In both cases, the primary origin of metastatic cancer is unknown.

As for land plants, relatively few galls were reported, in the Mesozoic pre-Cretaceous, since the Great Dying at the Permian End extinguished most gall lineages. In the early Triassic only two cases were reported: a gall apparently induced by a leaf mining fly in the leaf of a specimen of the extinct *Viaznikopteris rigida*, a rare plant belonging to the group of *Pteridospermatophyta* or seed ferns (Krassilov & Karasev 2011) and a gall that occurred on the pinnate leaf of a specimen of *Dicroidim odontopteroides*, belonging to the extinct group of *corystospermales* (McLoughlin 2011).

Recovery of the moderate level of plant-insect interactions, including gall associations, that was present during the late Permian, was not matched until the middle of Triassic, 237 Mya. During the late Triassic and Jurassic periods, new groups of galling insects began to colonize Ginkgoales, Bennetitales, Conifers and other gymnosperms (plants without flowers) but cases found in the fossil record are rather sparse. In fact, only two groups (both Coleoptera) of the major modern gall-inducing insects have a pre-Cretaceous fossil record (Labandeira 2021; Alvin *et al.* 1967).

A great expansion of both plant-insect interactions and galls occurred during the 35-million-year-long interval from 125 to 90 Mya of the mid-Cretaceous, largely associated with the initial expansion of

angiosperms or flowering plants. During this period, there was a major transformation of flora from gymnosperms dominance to angiosperm dominance when the latter expanded in a wide variety of ecosystems becoming the largest and most diverse group within the kingdom Plantae. Comparison with the modern material suggests that these numerous Cretaceous galls were mainly produced in response to mites, aphides, midges and wasps.

As occurred in the Paleozoic Era, benign as well as malignant behaviors seem to have been associated with Mesozoic galls. For example, in the locality Chernovskie Kopi of Transbaikalia, Russia, two markedly different types of insect-induced galls were found in two different specimen fossils of gymnosperms of the latest Jurassic to earliest Cretaceous boundary interval, about 145 Mya (Vasilenko 2005). The first type of gall, that was described on specimens of the ginkgolean host *Ginkgoites sp.*, was small-sized, hemispheroidal with smooth surfaces and it does not appear to have been significantly harmful for its host. In contrast, the second type of gall, that was described on specimens of the pinalean host *Desmiophyllum sp.*, behaved as a canker-like lesion, producing disruptions of leaf tissue with considerable internally disrupted tissue and thin, unhardened gall walls. This gall seems to have had the power to break branches and to structurally weaken and even kill its host plant (Labandeira 2021).

3.3. Cenozoic Era (66 Million Years Ago to Present)

The end of the Mesozoic Era, associated with a massive extinction of millions of animal species, included all the dinosaurs, marks the beginning of the Cenozoic Era. This extinction—known as the Cretaceous-Paleogene (K-Pg) event—was second only to the Permian-Triassic one as the most perilous period to affect life on Earth during the past 450 million years. Cenozoic is divided in three periods, namely Paleogene [that includes Paleocene (66-56 Mya), Eocene (56-34 Mya) and Oligocene (34-23 Mya) epochs], Neogene [that includes Miocene (23-5 Mya) and Pliocene (5-2.5 Mya) epochs] and Quaternary (that includes Pleistocene, 2.5 Mya -12000 years ago) and Holocene (12,000 years ago-present). Tertiary was the old denomination for Paleocene and Neogene together and it is sometimes still used in the literature. Cenozoic Era is associated with the great diversification and spread of mammals (and birds to a lesser extent) by which this period is also known as the Age of Mammals. In addition, the continents moved into the current positions during this Era. Among animals, many cases of benign as well malignant neoplasms have been reported in the different epochs and periods of this Era.

Among benign tumors, osteomas were very frequent among different fishes of the Tertiary and Quaternary and in sirenian mammals (order *Sirenia*) from Oligocene (Capasso 2005). Osteomas have also been detected in Quaternary fossil elephants from Poland (Przyklady 1965) and multiple hereditary osteochondroma have been observed in 61% (19 of 31) of fossil remains of the North American Oligocene *Canidae Hesperocyon* (Wang & Rotschild 1992). In the same way, an osteoblastoma has been reported in a mammoth that lived about 24,000 years ago at the Late Oligocene in a locality of the actual Poland, which is known for the presence of a substantial assemblage of mammoth bones accompanied by human artifacts from the Gravettian technocomplex (Krzeminska 2008). Neoplasms of the dental tissues have also been demonstrated in many extinct Cenozoic animals such as Tertiary ungulates from Argentina, fossil horses, European mammoths, Japanese fossil elephants and in a specimen of a Holocene fossil walrus from Alaska (Capasso 2005; Cabrera 1934; Patte 1937; Hunter & Langston 1964; Kobayashi 1937). Other benign tumors were reported in Tertiary *Bovidae*, in some Tertiary and Quaternary *Canidae*, in a specimen of *Nothroterium* (an extinct ground sloth from South America) and in a specimen of *Ursusus spelaeus* or cave bear (an extinct species of bear that lived in Europe and Asia at the late Pleistocene) (Miralles & Crusafont Pairo 1952; Pales & Wernert 1953; Moodie 1929; Scott 1898; Pales 1959).

Malignant tumors have also been reported in Cenozoic Era. In effect, osteosarcoma have been demonstrated in a specimen of a Pleistocene buffalo (Conkling 1990), in a *Holocene Capra* (Capasso & Di Tota 1996) and in a right femur assigned to a specimen of the Quaternary ground sloth *Nothrotherium maquinense*, that lived about 12,000 years ago in the actual Brazil (Souza Barbosa *et al.* 2021). In addition, chondrosarcoma was reported in some species of fossil *Canidae* (Baker & Brothwell 1980, pp.110–114).

As for land plants, after the K-Pg crisis, recovery of gall and other associations, started at the middle of Paleocene, about 60 Mya. Afterwards, between 49 and 40 Mya, distinctive new gall associations, similar to extant plant-gall interactions, make their earliest appearance as fossils. During the Neogene, the expansion of galls involved a broad diversity of plant hosts and gall-inducers, especially arthropods. For example, the early Neogene (20 Mya) flora whose remains were found in a region of the actual Czech Republic, provides 16 excellently preserved gall types, some with remarkable resemblance to their modern analogues attributable to extant families or genera, suggesting prolonged evolutionary stasis (Labandeira 2021). As occurred in Paleozoic and Mesozoic Eras, most Cenozoic galls seem to exhibit benign behaviors although in some cases, a rather malignant behavior may be suspected. Both benign as well malignant behaviors may be inferred quite accurately from the study of galls induced by modern gall-inducing species showing the closest resemblance to those found in the fossil record. For example, the extinct *Taxodium dubium* (belonging to the cypress family) found in that Neogene flora, exhibits galls very similar to those induced in the present days by the midge *Taxodiomya* in cypresses. These oval shaped galls are formed on the terminal portion of the branchlets and when mature, they resemble small pineapples which do not appreciably harm the tree health (Chen & Appleby 1984) On the other hand, the extinct *Alnus julianiformis* (belonging to the family of *Betulaceae*) found in that very Neogene flora, exhibits galls almost identical to those induced today by the mite *Eriophyis inangulis* in modern alnus. These galls develop as sub-spherical distortions rising up to the upper surfaces of the leaves and may vary in color from pale yellow-green to deep red. Although few galls may be not harmful for their host, many of them may cause strong decrease of both photosynthetic activity and stomatal conductance which may affect the survival of the affected tree (Jiang *et al.* 2021).

For the last 3 million years of late Pliocene and Pleistocene, however, the fossil record of galls is relatively scarce, probably because the several cycles of glaciation and deglaciation characteristics of this period, have eroded or otherwise prevented the formation of many persistent deposits.

Lastly, the Holocene marks the beginning of the actual large collection of galls in land plants.

3.4. Pre-human and Ancient Human Populations

The earliest evidence for neoplastic disease in the hominin lineage was reported in a specimen of the extinct *Australopithecus sediba* (belonging to the family of *Hominidae*) from the fossil-bearing cave of Malapa, located about 45 km north-northwest of Johannesburg, South Africa, dated to 1.98 Mya. The affected individual was male and developmentally equivalent to a human child of 12 to 13 years of age. The specimen exhibited a penetrating lytic lesion that affected the sixth thoracic vertebra, which was diagnosed as an osteoid osteoma, a benign osteoid and bone-forming tumor (Quinney *et al.* 2016). In the genus *Homo*, the two earliest known examples are an osteosarcoma present in a metatarsal specimen probably belonging to a *Homo ergaster* who lived in South Africa 1.6-1.8 Mya (Odes *et al.* 2016) and a possible Burkitt lymphoma or an ossifying sarcoma observed in a fragment of mandibular ramous attributable to *Homo erectus* who lived in Kenya about 1.5 Mya (Capasso 2005). As for the last case, however, some researchers have suggested that, alternatively, it might have been an overabundant bone callus associated with a healed fracture, which, incidentally, would reveal the similarity between cancer and a regenerative process, just as several analogous cases have been reported since the Paleozoic.

Many years later, meningiomas were reported in the fossil bones pertaining to a *Homo steinheimensis* and to a *Homo neanderthalensis* that lived in Germany 365,000 and 35,000 years ago, respectively (Czametzki, Schwaderer, & Pusch 2003; Czametzki 1980). In addition, a fibrous dysplastic neoplasm was described in a Neanderthal rib from a specimen that lived in present-day Croatia about 120,000 years ago (Monge *et al.* 2013). In this case, the incomplete nature of the rib and the lack of associated skeletal elements, prevented the authors to speculate on the health effects the tumor had on the individual. A benign tumor called intradiploic epidermal cyst, was also described in the frontal bone of another *Homo neanderthalensis* that lived between 50,000 and 70,000 years ago in Doggerland, the prehistoric landscape now under the sea off the Dutch coast (Hublin *et al.* 2009).

Later, almost all types of modern neoplastic diseases have been documented in ancient Homo sapiens bone remains. For example, meningiomas have been

reported in skeletons from Ancient Egypt since the time of the Fifth Dynasty between 2500 and 2350 years ago, and pre-historic America. Other benign tumors such as hemangiomas, osteoclastomas, histiocytomas, osteomas, osteochondromas as well as neoplasms in other organs that affect bones—such as pituitary adenoma and fibroleiomyomas of the uterus—have been documented since prehistory in Europa, North Africa and South and North America. Malignant primary bone tumors such as osteosarcomas, chondrosarcomas, hemangiosarcomas and Ewing's sarcoma have been reported in ancient populations of Europa and Egypt and prehistoric populations of Peru, Chile and Hawaii. Paleopathological studies have also revealed the existence of bone metastases of nasopharyngeal, breast and prostatic carcinoma and lytic lesions due to multiple myeloma and melanoma in the skeleton of individuals of prehistoric populations of Europa, Iran, Egypt and Pre-Columbian America, including pre-historic sites at Peru, California, St. Lawrence island (Alaska) and the western Pampean region and northwest Argentina. Furthermore, investigations of naturally or artificially mummified human bodies, excavated in Egypt, Nubia, Peru, Chile, Alaska, China and Europa have revealed the existence of some malignant primary tumors of soft tissues including carcinomas of the prostate and rectum, naso-orbital cancer, rhabdomyosarcomas, nasopharyngeal carcinoma, melanoma and multiple myeloma among others (Capasso 2005, Odes *et al.* 2016; Shimkin 1977; Strouhal 2001; Pahl 1986; Luna *et al.* 2008; Luna *et al.* 2015; Arrieta, Mendonca, & Bordach 2018). For a more comprehensive and detailed description of the up to date 154 paleopathological studies documenting 272 archeologically recovered individuals exhibiting skeletal or soft tissue evidence of cancer (that is, including only malignant neoplasms) between 1.8 Mya and 1900 CE see (Hunt, Roberts, & Kirkpatrick 2018) and the Cancer Research in Ancient Bodies (CRAB) Database (Hunt *et al.* 2017).

4. The Incidence of Cancer over Time and Geological Ages

A conclusion derived from the record fossil from Paleozoic Era onwards, as well as from the study of skeletons and mummies from pre-human and ancient human populations, suggests that cancer or cancer-like phenomena as well as benign neoplasia are very old diseases, which have afflicted animals and land plants since long before man appeared on Earth and human beings since prehistoric times.

Until relatively recently, it was assumed that the prevalence of cancer in the remote past was quite rare in animals on the basis of the apparently very low ratio between the number of reported cases of metastatic cancer in fossil bones (it must be remember that about 95% of malignant neoplastic lesions in bones are associated with metastases of soft tissues and the remaining 5% is related to multiple myeloma and primary bone cancer) and the vast number of fossil bones that have been excavated and examined by specialists. However, this ratio may strongly underestimate the cases of cancer if the remains are represented by minimal fragments of the whole body, as it occurred in many extremely ancient fossil deposits. In effect, the probability to find a cancer in a solitary bone from a specimen with many bones is many times lower than finding cancer in a complete specimen. Therefore, for the sake of comparing properly that metastatic cancer incidence between extinct and modern animals, it is necessary that both collections contain a similar number of bones by each specimen. In addition, it is worth noting that metastatic cancer in bones may be useful for comparative purposes but not as an absolute measure of cancer incidence because there are many cancers that may not produce bone metastases.

The largest epidemiological study of tumors in dinosaurs to date, undertook by Rothschild and colleagues using computed tomography for fluoroscopically screening dinosaur vertebrae, showed that out of a total of 10,312 vertebrae from 708 individual dinosaurs of varying families, only one malignant metastatic tumor was found. This ratio 1/708 = 0,141 % is significantly lower than that obtained in humans using the Hamann-Todd Collection, that is one of the largest and best-preserved compilation of modern human skeletons for which a background demographic is known (Rothschild *et al.* 1993; Rothschild & Woods 1991). In this collection, from a total of 2906 defleshed skeletons, 33 cases of metastatic disease were identified fluoroscopically, yielding a probability of 1.136 % ($p < 0.05$ versus the ratio 0.141% observed in dinosaurs, X2 test). However, if the comparison is made with modern reptiles, based on necropsy results of captive wild animals (Effron, Griner, & Benirschke 1977; Kitsoulis, Baxevanis, & Abatzopoulos 2020), the ratio of cancer in

them is approximately 0.142 %, that is almost identical to that observed in dinosaurs (Natarajan *et al.* 2007). Further, discovery and study of tumors in dinosaurs has revealed that they are indistinguishable from tumors from modern reptiles and humans suggesting that this global disease has barely changed over 100 million years.

Similar conclusion may be achieved when comparison is made between extinct and modern avian and mammals. The relatively constant incidence of cancer within each major taxa of animals over long periods of time, is also supported for the similar cancer incidence observed among modern animals belonging to a same order, even though substantial differences in cancer incidence and mortality across major animal orders occur. For example, among mammals, all or almost all members of the order *Carnivora* (lion, tiger, hyena, bear, wolf, dog, cat, etc.) display an elevated risk to get cancer throughout their lives while, on the other hand, all or almost all members of the order *Artiodactyla* (camel, pig, cow and bull, sheep, deer, giraffe, hippo, etc.) exhibit a significant lower risk (Vincze *et al.* 2022). In the same way, among arthropods, all or almost all members of the order *Diptera* (fly, midge, jig, horsefly, etc.) exhibit a rather frequent occurrence of tumors while in contrast, all or almost all members of the order *Decapoda* (crab, lobster, crayfish, shrimp, pawn, etc.) display a low incidence (Vogts 2008). This means that the extinct ancestors of *Carnivora* and *Diptera* probably had a similar high risk of having cancer as their modern descendants. On the other hand, the extinct ancestors of *Artiodactyla* and *Decapoda* probably displayed a relatively lower one.

5. Exceptions to the rule of constancy of tumor incidence over time: examples of increased tumor incidence over time

Human beings followed the same rule of constancy up to the turn of the 20th century, after which the trend changed. In effect, comparison between ancient and modern human populations suggests that incidence of cancer remained relatively constant for many years but it started to increase progressively from 1900 onwards. In fact, Nerlich *et al.* (2006) searched for malignant growth affecting the skeleton in both, a collection of 905 individuals that have been excavated from the necropolis of Thebes-West and Abydos, Upper Egypt, covering the time period between 3200 and 500 BC, and a collection of 2547 individuals that have been buried in a Southern German ossuary dating from between AD 1400 and 1800. According to the authors, the skeletal tissue preservation of both the Egyptian and Southern German material was excellent. All available specimens were subjected to a very careful macroscopic examination and isolated findings were also radiologically investigated. In parallel, anthropological data, such as gender and age at death, were recorded. The study identified 5 cases of malignant tumors affecting the skeleton in the Egyptian material (ratio: 5/905 = 0.552 %) and 13 cases affecting the skeletal material from Southern Germany (ratio: 13/2547 = 0.510 %, p: NS). In most instances, multiple osteolytic lesions with slight osteoblastic reaction, were strongly suggestive for metastatic carcinoma. The ratios were very similar indicating that malignant tumors were present in spatially and temporarily different populations over the last 5000 years with an age-and gender-adjusted frequency not different from Western industrial populations before 1900. Afterwards, cancer incidence began to increase significantly. In effect, in the Hamann-Todd Collection that contains human skeletons from persons that passed away between 1912 and 1938, the ratio of metastatic cancer in bones had increased (33/2906 = 1.136 %) over basal values before 1900 ($p < 0.01$). Later, in the William M. Bass Forensic Skeletal Collection of the University of Tennesse, USA, which contains 868 skeletons from persons that passed away between 1970 and present time, 19 metastatic cancer in bones were reported (Fatula 2020), which represents a ratio of 19/868 = 2.19 % ($p < 0.001$ versus human populations before 1900) that is even higher ($p < 0.05$) than that reported in the Hamann-Todd collection. The increase of metastatic cancer in bone remains during the 20th century is correlated to the increased incidence of cancer reported clinically. In effect, in developed countries, mortality for cancer was only 5% in 1900 and it had climbed to 20% in 1970 and to 33% in 2018 (Capasso 2005; Khatami 2018).

Two main causes have been invoked to explain the great increase of human cancer over the last century. The first is linked to the aging of modern populations since cancer is an age-associated disease whose prevalence ranges from about 1.8 % for those with <39 years old to 27.2 % among those with 60-79 years

old. Taking into account that life expectancy increased from about 30-40 years to 70-80 years during the 20th century, age alone could be expected to reduce the incidence of malignancy in past centuries by about 90% with respect to the modern rate. The second cause is associated with the fact that ancient humans were not exposed to both chemical agents responsible for the modern environmental pollution and physical factors such as radioactivity due to nuclear assays that only began in the 1950s. In summary, in humans, the incidence of cancer remained constant for many years but the longest life expectancy and the chemical and physical contaminants associated with urban modern civilization seems to have increased the incidence of tumors.

Increased tumor incidence over time might also be associated with the fact that some organisms adopted some kind of tumors as a biologic strategy to increase their adaptability to difficult environmental conditions. For example, fossil fishes of the genus *Pachylebias* (now referred to as *Aphanius crassicaudus*) that lived about 8-5 Mya, adopted pachyostosis to facilitate immersion in the hypersaline waters of the Mediterrean Sea at the time of the Miocene desiccation period. This condition, characterized by an extraordinarily thick skeleton that occupied almost the entire body did not differ from a benign tumor originating from bone tissue. An almost identical condition occurred in the larger cyprinid fish *Hsianwenia wui* that lived in the Pliocene period (5-2.5 Mya) in the hypersaline lakes of the Qaidam Basin on the northern Tibetan Plateau. Both these unusually thick-bone fishes represented an adaptive mode to the extreme conditions resulting from continuing aridification in the two areas (Capasso 2005; Chang *et al.* 2008). In the same way, mammals of the Sirenian group that lived about 30 Mya ago during the Oligocene acquired tumor-like forms in their axial skeletons to consent browsing on the bottom in shallow waters (Capasso 2005). A similar adaptative strategy had been developed many millions of years before, by the plesiosaur *Tatenectes laramiensis* that lived in shallow marine waters during the Upper Jurassic between 164 and 157 Mya (Street & O'Keefe 2010).

In land plants, the insect-induced gall tumors have experienced a large increase in both incidence and diversification over time, since a few cases reported in Paleozoic Era up to the huge number of about 130,000 plant species that harbor 130,000 different types of insect-induced galls, in present times (Labandeira 2021; Espirito-Santo & Wilson Fernandes 2007). Galls induced by other organisms have also experienced a significant expansion although less important than the former. Sometimes, gall-inducers are plant parasites and, in such cases, they are the only ones who benefit from their inter-specific association with host plants. In consequence, the expansion of these galls over time is exclusively associated with the evolutionary expansion of their gall-inducers. In other cases, however, galls are also beneficial to plants, as in brood-site pollination mutualism where plants trade insect development sites against seed production. In these cases, the expansion of these galls over time could also be related to the fact that these plants adopted these gall tumors as a biological strategy to increase their descendants. A typical example of this mutualism is the ancient interaction between figs (*Ficus*, *Moraceae*) and their pollinating fig wasps (Borges 2021).

6. The two main evolutionary riddles of Cancer

6.1. First Riddle

Despite the major changes in the structure of animal populations, the prevalence of malignant as well benign neoplasms has remained relatively constant (and in some cases it has even increased), among the different taxa of animals for hundred million years (Capasso 2005) suggesting that malignancies as well as benign neoplasms are rooted quite deeply in the evolutionary life of organisms.

However, this seemingly unremarkably fact represents a remarkable riddle for evolutionary biologists. If natural selection, working on living organisms has been powerful enough to produce complex adaptations, from the eye to the immune system, why has it been unable to eliminate or even reduce the incidence of cancer, even though many apparently less harmful traits have been eliminated during species evolution.

For some researchers this question is neither original for cancer nor enigmatic.

They claim that it would not be original because a similar case may be stated for ancient infectious diseases and for rare genetic disorders. For example, malaria produced by Plasmodium falciparum has affected

human beings for 50,000 - 100,000 years and it may be a pathogen that accompanied the whole history of our specie; this hypothesis is further supported by the observation that close relatives of the human malaria parasites remain common in chimpanzees (Joy *et al.* 2003). However, in this as well as in other similar cases, throughout the years, foreign infectious agents may have acquired, sophisticated evolutionary strategies for attacking our bodies that may have overcome our evolutionary renewed strategies of immune defenses against them. In contrast, the tumors that affect our bodies are not foreign agents; they are made of our own cells that have evolved over millions of years to preserve the homeostasis of the organism, not to attack it. In the same way, it is true that natural selection cannot drive the spread of new defenses against rare genetic diseases because the acquisition of such new defenses would make very little difference to the average reproductive success of a population. In contrast, cancer is not some bizarre rarity: in developed countries a person has 30-40 per cent chance or more of being diagnosed with some type of cancer in its lifetime.

On the other hand, the argument that cancer should have been removed or reduced by natural selection has often been challenged on the ground of two main objections. In the first place, it is invoked that most tumors develop after reproduction has ceased and, in consequence, negative selection could not have been operative against them. In the second place, it is argued that natural selection works with hereditable traits and cancer is not, in general, a hereditable disease, taking into account that only about 5 % of cancers—the so called "familial cancers"—are transmitted by the germinal route (Ewing 1940).

These objections are highly questionable. In effect, although many species show a decline of reproductive function with age, the female menopause occurs only in humans, a few other primates and in the killer and pilot whales (Woodruffs 1982; McAuliffe & Whitehead 2005; Brent *et al.* 2015). The latter means that in most living and extinct animals over three geological eras, cancer has not been a post reproductive disease. In the second place, it is known that the "familial cancers" are associated with the hereditary transmission of a mutated allele of some genes such as RB1 and BCRA1 which confers high susceptibility to the development of retinoblastoma and breast cancer. However, even assuming that mutated RB1 and BCRA1 are the main etiological factors related to those cancers, the trait "susceptibility to cancer" could also be transmitted through deficient immunologically or biochemically-mediated anti-tumor mechanisms that would prevent the host to limit the development of the tumors much the way that, in infection diseases produced by foreign pathogens, the trait "susceptibility to the disease" may be transmitted not by vertical transmission of the main etiological agent (the foreign pathogen) but by vertical transmission of a deficient anti-pathogen host defense.

In order to explain why natural selection did not eradicate or at least ameliorate cancer from species over three geological eras, some authors have advanced the idea that cancer may play a real role within the organism (Zajicek 1996; Muller 2017), or it may be coupled to essential physiological functions that would prevent it from being removed by natural selection (Zimmer 2007). However, up to date the nature of these putative tumor roles or the normal essential functions with which it would be coupled, remain obscure.

6.2. Second Riddle

The constancy of cancer incidence over hundred million years also demands an extremely remarkable mechanistic explanation of carcinogenesis. In effect, comparison between the record fossil and present evidence of cancer reveals that, within a taxon (for example carnivorous mammals), cancer incidence is very similar regardless of the tumor host's body size and life length. Further, even considering mammals as a whole, it is clear that a higher risk of cancer does not correlate with increased body mass and lifespan (Abegglen *et al.* 1996) The prevalent somatic mutation theory (SMT) of cancer posits that the malignant cell is the physiological and anatomical unit of cancer disease (Boveri 1929; Hanahan & Weinberg 2000; Bignold 2002; Tomasetti *et al.* 2015). Implicit in this contention is the assumption that the probability of origin of an aberrant, neoplastic cell lineage may be the same per unit of both cell population and time, regardless of species or cell type concerned. However, this assumption evokes one of the most intriguing enigmas in cancer research, which remains unsolved. The riddle, currently called Peto's Paradox, asks (Dawe 1969; Peto *et al.* 1975; Peto 2015): Why do not extremely large animals with a long lifespan develop neoplasms with a much higher incidence than very small ones displaying

a short lifespan since both the cell population at risk and the exposure time at putative carcinogens are greater by several orders of magnitude? Let us consider the blue whale, the human and the mouse. If one takes the weight of a mouse as 20 g, that of a human as 60 kg and that of a blue whale as 200 ton, a blue whale and a human are equivalent to 10,000,000 and 3,000 mice, respectively. Therefore, we should expect the blue whale and the human being to develop cancer, respectively, 10,000,000 and 3,000 times more often than a mouse by unit of time (Figure 3). Furthermore, since the lifespan is 2.5 years for the mouse and about 80 years for blue whales and humans, the relative risk of cancer should be also increased in function of the ratio between both lifespans. In fact, according to the hypothesis of the multistage carcinogenesis, this increase should not be linear but exponential with the sixth power of age (Nordling 1953; Weiss 2004; Prejean *et al.* 1973; Pugh *et al.* 1999).

Some *ad hoc* hypotheses have been invoked to explain Peto's Paradox. For example, the animal fat depots might sequester fat-soluble carcinogens with an efficiency proportional to animal's size and thereby proportionately diminish the exposure of other tissues. Other putative explanations hold that faster metabolism of small animals generate more putative cancer inducing-free radicals, or that the efficiency of defenses against neoplasia, such as mechanisms of DNA repair, cellular resistance to metabolism and mutagenic activation of putative carcinogens, number of copies of the tumor suppressor gene TP53, immunological surveillance, etc. could be proportional to animal size (Wheatley & Clegg 1994; Dung 2014; Vineis *et al.* 2009; Dunn, Koebel, & Schreiber 2006; Downs *et al.* 2020; Nunney 2020). However, these invoked mechanisms remain largely unproven as general rules and in fact there is evidence that argues against them. For example, even though the African savannah elephant (*Loxodonta Africana*) genome contains 20 copies (40 alleles) of TP53 and the human genome contains only one copy; on the other hand, the mouse genome has 2 copies and whales neither exhibit extra copies of TP53 nor of any other known tumor suppressor gene (Tollis, Boddy, & Maley 2017).

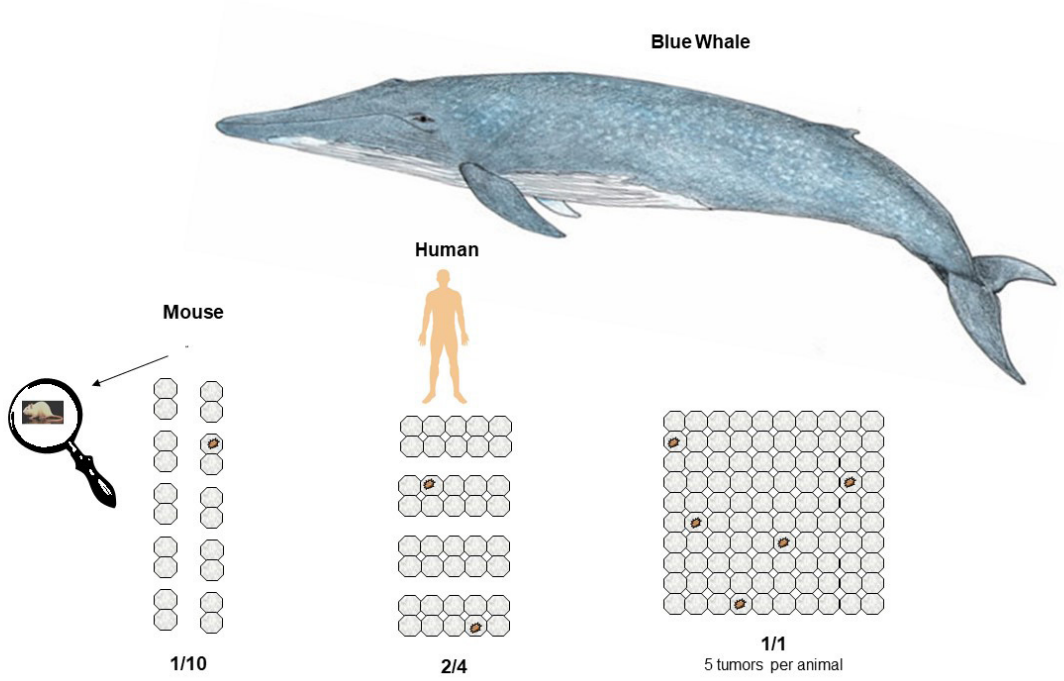

Figure 3. An illustration of Peto's Paradox: Theoretical influence that size of whole body would have on tumor incidence per unit time on the assumption that the individual cell is the unit at risk of carcinogenesis. We have arbitrarily assumed that a carcinogenic mutation occurs at a rate of 1 per 20 cell units per unit time (tumor cells are identified by an internal mark). As a consequence, animals with 2, 10 and 100 cells should develop, respectively, 1 neoplasm in every 10 animals, 2 neoplasms in every 4 animals and 5 neoplasms per animal, per unit time. The correspondence between organisms [mouse (20 g); human (60 Kg) and blue whale (200 Ton)] and number of cells (2, 10 and 100, respectively) is only illustrative. These theoretical expectations do not match reality: long-lived and large-sized animals do not have more cancer than short-lived and small-sized animals (image by the Authors).

Using a mathematical model of carcinogenesis, Nunney (2020) proposed that neither intrinsic changes in metabolic rates nor different mutation rates nor changes in immune surveillance, may resolve by themselves Peto's Paradox. Instead, he proposed that in order to compensate the sharp intrinsic increase of cancer risk associated with increased body size and longevity, large-sized and long-lived organisms (such as human beings and blue whales) may have acquired much more genetic controls of cancer (the sum of proto-oncogenes and tumor suppressor genes, that is, not only suppressor genes such as TP53) than small-sized and short-lived ones. However, although this proposal is attractive and it might theoretically explain the similar accumulated cancer incidence observed in all animals at the end of their lives, it would not explain the fact that, for all tested animals, not only the final incidence but also the shape of the curves of cancer incidence throughout their lives, are also very similar (Nordling 1953; Weiss 2004;Prejean *et al.* 1973; Pugh *et al.* 1999) (as seen above, roughly proportional to the 6^{th} power of age). This suggests that the number of oncogenic steps (each one assumed as oncogenic genetic mutations by the hypothesis of multistage carcinogenesis) are also very similar among all animals.

A recent paper aimed to study the landscape of somatic mutation across 16 mammalian species displaying 30-fold variation in lifespan and 40,000-fold variation in body mass, has demonstrated that somatic mutation rate per year varied greatly across species and exhibited a strong inverse relationship with species lifespan (Cagan *et al.* 2023). These results suggested that the somatic—and presumably the oncogenic—mutation burden by unit of mass at the end of lifespan was roughly similar among long- and short-lived animals. However, since the mutation rate did not exhibit significant variation with tumor mass (Cagan *et al.* 2023), even if individual end-of-life cells across species have a fairly similar mutation burden, overall cancer risk should still be expected to scale with the number of cells in an organism, which we know it does not happen.

7. A Light in the Dark

Most attempts to explain the evolutionary riddles of cancer were based explicitly or tacitly on the SMT, that is the hegemonic paradigm in cancer research. The theory states that cancer is the outcome of the constitutive activation or mutation of some genes (protooncogenes) or the inactivation of others (tumor suppressor genes) allowing the cell to evade the mechanisms controlling cell proliferation. These genetic changes would define the attributes of the malignant cell, which, in turn, should be the target of specific therapies against cancer. This theory has the merit of unifying, through an immediate common cause, the numerous different mediate causes of cancer such as chemicals, radiation, viruses, etc. However, it has some theoretical difficulties that have been addressed by some authors (Sonnenschein & Soto 2021; Peto 2015; Prehn 2005) which have also emphasized that—apart from some particular advances in targeted molecular therapies against certain neoplasia (Danthala 2017)—cancer remains a major cause of morbidity and mortality, despite the explosive development of our knowledge about the molecular mechanisms associated with the control of cell cycle and survival (Bailar & Gornik 1997; Sung *et al.* 2021). Of course, these theoretical difficulties and the failure to treat malignant diseases, especially disseminated cancer, do not necessarily imply that the SMT is incorrect, but they encourage us to explore other approaches. SMT or some of its variants posit that the origin of cancer must be placed at the cellular or subcellular level of biological organization. On the other hand, some authors have raised the idea that cancer is primarily a disease of higher levels of organization, that is, an organismic, organ or tissue-based disease rather than a cellular one. This possibility has been advocated by Waddington, Smithers and others many years ago (Waddington 1935; Smithers 1962), and more recently by the group of Sonnenschein and Soto in their tissue organization field theory (TOFT) of cancer (Sonnenschein & Soto 2000; 2016; 2020). TOFT states that carcinogenesis occurs when some factors called carcinogens disrupt the flow of information between the stroma and the adjacent epithelium and unlock the constitutive proliferative capacity of the epithelial cells (Maffini *et al.* 2004). This is not to say that tumor cells do not harbor mutations, but they would not have the pivotal carcinogenic role that SMT attributes to them.

Several lines of evidence from both the record fossil and comparative oncology, seem to favor the latter interpretation.

For example, an osteosarcoma was recently diagnosed in the vertebral intercentrum of a

temnospondyl Mesozoic amphibian that lived more than 200 Mya in the current locality of Krasiejów, southern Poland (Surmik *et al.* 2022). The authors claim that the growth dynamics and development of the tumor are consistent with the postulates of TOFT which locates the cause of cancer in disorders of tissue architecture. This consistency is expressed in different ways: a) the fast growing characteristics of the newly formed bone, which mixes a slowly deposited matrix type with spatial distribution typical for rapidly growing bone; b) both the affected intercentrum and the overgrowth being subject to physiological remodeling processes, as evidenced by the numerous areas of bone tissue destruction within the tumor and the vertebra itself suggesting that the physiological processes occur in the neoplasm and the original bone alike; c) the difficulty to explain why the border between the physiological bone and the overgrowth is ordered and clearly marked taking into account that multiple lesions and a chaotic organization could be expected from an invasion of a collection of autonomous and independent mutated neoplastic cells. The existence of common physiological processes in both the normal remodeling bone and the neoplasm in this ancient fossil remain highlights the resemblance between cancer and regenerative processes and it is paralleled with the recent findings of oncogenic mutations in many normal aging tissues (Martincorena *et al.* 2018; Kakiuchi & Ogawa 2021) which, in turn, challenges the causal direct role of these mutations in the genesis of cancer.

In addition, although cancer or cancer-like phenomena have been observed in many of the largest groups of pluricellular organisms, including not only animals and land plants but also fungi and red and green algae (Aktipis *et al.* 2015), not all taxa exhibit it. Considering only the animal kingdom, cancer is rarely (if ever) produced in animals or body regions displaying regenerative abilities that remain efficient throughout life. These regenerative abilities are generally "strong" (strong meaning the capacity to regenerate complex structures such as a whole limb); and the regions that exhibit such abilities can encompass the whole body, as in sponges, ctenophores, cnidarians, echinoderms, annelids, etc. (Aktipis *et al.* 2015; Wellings 1969; Sparks 1969; Tascedda & Ottaviani 2014; Edgar *et al.* 2021) or parts of the body, as in the upper body regions of Planaria, phylum *Platyhelminthes* (Saló 2006) and limbs, tails and some other tissues of urodele amphibians (Prehn 2007; Stocum 2017). In contrast, cancer is relatively frequent in animals that display regenerative abilities that are efficient mainly during youth and wane progressively as the animals

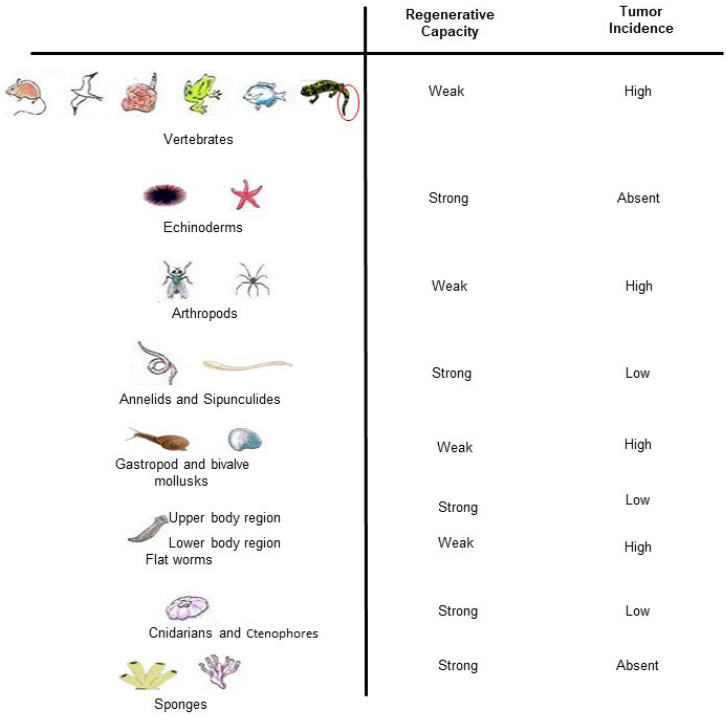

Figure 4. Comparing regenerative capacity and tumor incidence among different phyla.

age (Sharpless & DePinho 2004). These regenerative abilities are generally "weak" (by weak we mean having the capacity to repair o regenerate relatively simple structures only, as in compensatory hyperplasia of the liver, skin regeneration, etc.) and can encompass the whole body such as seen in most vertebrates others than urodele amphibians, nematodes, arachnids, insects, gastropods and bivalve mollusks (Aktipis *et al.* 2015; Kitsoulis *et al.* 2020; Gubin *et al.* 2001; Tascedda & Ottaviani 2014; Prehn 1997; Robert 2010; Ostrander *et al.* 2004; Caussinus & Gonzalez 2005; Kiriakakis, Markaki, & Tavemarakis 2015). A similar relatively high frequency of tumors has been observed in the body regions of urodele amphibians that lack a strong reparative capacity (Prehn 1997) and in the lowest body region of Planaria where the regenerative ability gradient is minimal (Hall, Morita, & Best 1986) (Figure 4). In animals in which cancer is relatively frequent, cancer incidence rises exponentially with age (Nordling 1953; Weiss 2004; Kiriakakis, Markaki, & Tavemarakis 2015; Hall, Morita, & Best 1986; Campisi 2013; Rozhok & DeGregori 2016) coincident with decreased reparative capacity. In addition, when cancer develops in young animals, it is usually associated with injured organs and tissues such as cirrhotic liver, gastric tissues exhibiting chronic atrophic gastritis, radiation-damaged skin, colon displaying ulcerative colitis, breasts of nulliparous women, non-secreting prostate alveoli, pulmonary fibrosis, etc., which may have a significant decrease of their regenerative abilities (Edgar *et al.* 2021; Karin, Lawrence, & Nizet 2006; Bustuoabad *et al.* 2021). The fossil record might also support this contention: although neoplasms have been described in fossils of many vertebrates and invertebrates groups (as trilobites) no neoplasms have been described in the abundant record of echinoderm (mainly crinoids) fossils, animals in which strong regenerative abilities—such as present in living echinoderms—have been extensively documented (Gahn & Baumiller 2010).

Strictly speaking, even animals that exhibit a strong reparative capacity, such as cnidarians, can exhibit tumors termed "calicoblastic epitheliomas", upon the action of exceptional environmental stressors that are strong enough to injure seriously their organisms and to impair their reparative capacity. This seems to have occurred, especially, but not exclusively, in some coral reefs of the genus *Acropora* in some locations of Caribbean where, during the last 40 years, water pollution and other diseases have produced rates of coral mortality without precedent in the late Holocene (Ruggiero *et al.* 2008).

In summary, throughout the animal kingdom, cancer seems to occur in organs and tissues that have experienced a decline or loss of their regenerative ability. In these organs and tissues, any injury causing loss of cells or cellular function could not be adequately compensated by cellular division or increased cellular size (Mitchell & Valk 1962; Castle & McDougal 1984; Fankhauser 1945), and in consequence the original size and function could not be restored. We suggest that this situation would induce a crisis, which might promote some degree of variability in the remaining cells of the organ bearing low ability to regenerate. The outcome of this situation would be the emergence, by chance, of a cell variant bearing mitotic ability to respond to the reparative signal. If this new variant were still functionally active, normal function might be restored and this restored organ might reproduce the regulatory fields associated with the intact functional organ, after which further mitosis would be halted. However, if the injury were persistent or more profound, later or sooner, a poorly or non-functional variant bearing mitotic ability might finally arise. This new variant would begin to divide and the organ would be numerically but not functionally restored. In consequence, it would not score the regeneration as effective and it would continue to send mitotic signals. As a result of this, the new variant would grow over and over and the outcome would be a tumor.

According to this interpretation, cancer would not be autonomous and have a profound biological sense: it would eventually be the ultimate attempt to restore organ functions and structures that have been lost or altered by aging or noxious environmental agents. However, unlike normal structures, cancer would have no physiological value, because the usually poor-functional nature of its cells would make their reparative task unattainable. The fact that animals that are resistant to cancer do not exhibit neither decline of regenerative ability nor aging (Petralia, Mattson, & Yao 2014) reinforces the proposal raised here of cancer as an attempt (even futile) to restore the regenerative ability of the affected organ and to evade the process of aging. Naturally, someone could ask why individuals with less efficient regenerative abilities have evaded natural selection. We have not a definitive answer to

this question. We only may suggest that, especially for highly complex organisms in this precise moment of their evolutionary history, the maintenance of their regenerative abilities fairly efficient throughout their lives—that would eventually prevent tumor formation—might be achieved only at the cost of reducing the efficiency of growth during youth when reproduction is more probably to occur, which would be, as a whole, selectively unfavorable. The existence of undesired traits coupled with more beneficial ones that globally represent phenotypes which have been selected during evolution is highlighted in a recent paper concerning the price of human evolution (Erenpreisa *et al.* 2023).

The interpretation stated above, encoded in the so-called hypothesis of the biological sense of cancer (Ruggiero *et al.* 2008; Ruggiero & Bustuoabad 2006; Bustuoabad & Ruggiero 2017) was built within the broad framework of TOFT by assuming that cancer is basically a problem of tissue organization. In fact, according with this hypothesis, there would not have such thing as a cancer cell if it means a cell endowed with genetic anomalies that allow it to escape from the inhibitory signals of normal cell proliferation. Instead, the problem would be the reduction or absence of such tissue signals. In this context, this hypothesis could offer a relatively easy solution of the Peto's Paradox by assuming that the true basic unit at risk of carcinogenesis is the tissue or organ as a whole rather than the individual cell. In effect, according to the hypothesis, cancer originates in organs or tissues that display a significant decline of their regenerative capacities, and this would occur when a critical proportion of their cells have partially or wholly lost that capacity. In such a case, if an organ were x times larger than another one, the probability that its regenerative capacity is critically diminished would be x times lower, because an x times greater number of cells would have to be affected to depress that capacity. This lower probability would balance the proportionally higher number of their cells that could be transformed. As a result, if the unit at risk is, for example, one liver rather than 10^9 (mouse) as opposed to 3×10^{12} (human) or 1×10^{16} (blue whale) liver cells, then the human or the whale will be at no greater risk of developing liver cancer than the mouse, or any other animal with an equally efficient defense mechanism against neoplasia.

The hypothesis advanced in this paper that the tissue or organ as a whole rather than the individual cell is the basic unit of carcinogenesis might be questioned by the fact that cancer can be transplanted into healthy individuals. In effect, this universal laboratory practice—initiated by Novinsky in 1877 (Shimkin 1955)—demonstrates that only a small fragment of a tumor or a relatively small number of tumor cells [in the limit, only one euploid or polyploid cell (Weihua *et al.* 2011; Moein *et al.* 2020)] dispersed in physiological saline will suffice to transplant that tumor from a tumor-bearing donor to a normal recipient host. This would mean that the growth of a tumor does not need to be supported by any tissue, organ or organismic pathological condition but only by the nature of the tumor cells themselves. In other words, the basic unit of carcinogenesis would be the tumor cell that, in consequence, could be deemed as "autonomous". However, the whole of this apparently fatal objection pivots on the ambiguity of the word "autonomy". We can accept that tumor cells are deemed as "autonomous" if their inoculation into an appropriate recipient host is enough to induce a new tumor growth (the first meaning of autonomy). But this does not mean that the new growth has to be accomplished by evading the rules controlling normal cell proliferation (the second meaning of autonomy). In effect, tumor cells transplanted might need to injure the recipient organ and to reduce its regenerative ability as a precondition for regenerative signals produced by the injured organ to promote tumor growth. This last possibility concerning the mechanisms underlying both tumor transplantation among different individuals and strategies used by a tumor to invade adjacent or distant organs within the same individual, has significant experimental support: a) Benign tumors, which are not invasive and commonly produce little damaged to host tissues, seldom grow when transplanted in another host (Shimosato *et al.* 1976). b) In chickens, tumors induced by Rous sarcoma virus (RSV) typically form at the viral injection site but not at distant sites; the wound associated with the injection seems to be required for local tumor growth, because additional tumors can be induced at distant sites simply by wounded the infected birds (Kennt & Bissell 2003). c) The liver of a young rat, but not of an aged rat in which regenerative ability is diminished, can normalize the morphology and growth capacity of transplanted hepatocarcinoma cells. The most successful normalization occurred when cells were transplanted into the spleen and filtered as solitary cells into the liver without disrupting

normal liver architecture. On the other hand, when this architecture was disrupted by transplanting a greater number of malignant cells directly into the liver, normalization was less likely to occur (Rubin 2006). d) Upon transplantation, tumors usually grow into anatomically correct (orthotopic) organs better than in heterotopic ones (Nathanson, Nelson, & Lee 1993). This observation can be interpreted by assuming that an invasive and transplantable tumor, even if quite different from the organ of origin, tends to be more similar to that organ than to others; in consequence, it would respond to a regenerative signal from the former better than to one from the latter, resulting in faster tumor growth.

Conclusions

The hypothesis that we have presented herein could explain the permanence of cancer for hundred million years assuming that it is coupled to the normal regenerative mechanisms of the organisms without which no pluricellular organism could survive. Furthermore, some cases of the record fossil suggest that neoplasms could also be a major component of the evolutionary machinery of pluricellular organisms, taking into account that some extant and extinct animals and plants seem have adopted some kinds of neoplasms as adaptative strategies to survive in hostile conditions. In addition, it could also explain the Peto' Paradox, as long as we assume that the true basic unit of carcinogenesis is the tissue or organ as a whole rather than the individual cell, as it is usually thought when following the SMT paradigm. Apart from its theoretical value, this proposal also might have therapeutic consequences. Namely, all conventional therapies against cancer attempt to kill all cancer cells. However, according to the hypothesis that we have advanced, the problem might not be solved even though all tumor cells were eradicated. In such a case, if the organ failure remained, new tumor cells would emerge and the tumor would reinitiate its progressive growth in response to the permanent regenerative signal of the non-restored organ. The possibility that currents cancer treatments are obsolete and must be changed has been recently suggested (Galmarini 2020).

Therefore, efficient anti-cancer therapy should combine an attack against the tumor cells themselves with the correction of the organ anomaly, which would be in the core of the cancer problem. The possibility that this anomaly, that is, the decline or loss of the organ regenerative ability, may be eventually reversed is suggested by novel experiments in which transplantation of differentiated cells derived from induced-pluripotent stem cells successfully induced functional recoveries in rodent models (Sánchez Alvarado & Yamanaka 2014; Elkashty 2021). Finally, the comparative study of cancer phenomenon and cancer-resistant animals that do not age might unveil common and still unknown routes to immortality.

Acknowledgements

The authors thank Drs. C. Sonnenschein and A.M. Soto for their helpful comments and wise suggestions.

References

Abegglen, LM, Caulin, AF, Chan, A, Lee, K, Robinson, R, Campbell, MS, Kiso, WK, Schmitt, DL, Waddel, PJ, Bhaskara, S, Jensen, ST, Maley, CC, & Schiffman, JD 2015 "Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans" JAMA, vol. 314, pp. 1850–1860.

Aktipis, CA, Boddy, AA, Jansen, G, Hibner, U, Hochberg, ME, Maley, CC, & Wilkinson, GS 2015, "Cancer across the tree of life: Cooperation and cheating in multicellularity", Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 370, no. 1673, art. 20140219. doi: 10.1098/rstb.2014.0219.

Alvin, KL, Barnard, PDW, Harris, TM, Hughes, NF, Wagner, RH, & Wesley, A 1967 "Gymnospermophyta", in: Harland WB *et al.* (eds.) The fossil record, London: Geological Society of London.

Armstrong, WP 1995, "To be or not to be a Gall", Pacific Horticulture, vol 56, pp. 39–45.

Arrieta, MA, Mendonca, OJ, & Bordach, MA 2018 "Differential diagnosis of a neoplastic condition in a prehistoric juvenile individual from La Falda site, northerwest, Argentina" International Journal of Paleontology, vol. 21, pp. 11–120.

Babcock, LE 1993, "Trilobite malformations and the fossil record of behavioral asymmetry" Journal of Paleontology, vol. 67, pp. 217–229.

Bailar, JC III & Gornik, HL 1997 "Cancer undefeated" New England Journal of Medicine, vol. 336, pp. 1569–1574.

Baker, J, & Brothwell, D 1980, Animal disease in antiquity. London: Academic Press.

Bayer, MH, Kaiser, HE, & Micozzi, MS 1994, "Abnormal growth processes in plants and animals: A comparison", In vivo, vol. 8, pp. 3–16.

Bignold, LP 2002, "The mutator phenotype theory can explain the complex morphology and behavior of cancer", Cellular and Molecular Life Sciences, vol. 59, pp. 950–958.

Borges, RM 2021, "Interactions between figs and gall-inducing fig wasps: Adaptations, constraints and unanswered questions" Frontiers in Ecology and Evolution, vol. 9, art. 685542.doi: 10.3389/fevo.2021.685542.

Boveri, T 1929 The origin of malignant tumors, Baltimore: Williams & Williams.

Brent, JN, Franks, DW, Foster, EA, Balcomb, KC, Cant, MA, & Croft, DP 2015, "Ecological knowledge, leadership and the evolution of menopause in killer whales", Current Biology, vol. 25, pp. 746–750.

Bustuoabad, OD, & Ruggiero RA 2017 "Hypothesis of the basic biological sense of cancer revisited: A putative explanation of Peto's Paradox", Cancer Hypothesis, vol. 4, p. 9.

Bustuoabad, OD, Ruggiero, RA, Bustuoabad, V, Montagna, DR, & Speroni, L 2021 El cáncer no es el problema, sólo es la respuesta [Cancer is not the problem, it is just the answer], Buenos Aires: Dunken.

Cabrera, A 1934 "Anomalías patológicas dentarias en algunos ungulados pampeanos" [Dental pathological anomalies in some pampas ungulates]. Notas del Museo de la Plata, Botánica, vol. 2, pp. 183–185.

Cagan, A, *et al.* 2023 "Somatic mutation rates scale with lifespan across mammals", Nature, vol. 604, pp. 517–524.

Campisi, J 2013 "Aging, cellular senescence and cancer", Annual Review of Physiology, vol. 75, pp. 685–705.

Capasso, LL 2005, "Antiquity of cancer", International journal of Cancer, vol. 113, pp. 2–11.

Capasso, L, & Di Tota, G 1996, "The antiquity of osteosarcoma", International Journal of Osteoarcheology, vol. 6, pp. 512–514.

Castle, WN, & McDougal, WS 1984 "Contralateral renal hyperplasia and increased renal function after relief of chronic unilateral uretral obstruction", Journal of Urology, vol. 132, pp. 1016–1020.

Caussinus, E, & Gonzalez, C 2005 "Induction of tumor growth by altered stem–cell asymmetric division in Drosophila Melanogaster", Nature Genetics, vol. 37, pp. 1027–1028.

Chang, M, Wang, X, Liu, H, Miao, D, Zhao, Q, Wu, G, Liu, J, Li, Q, Sun, Z, & Wang, N 2008 "Extraordinarily thick-boned fish linked to the aridification of the Qaidam Basin (northern Tibetan Plateau)", PNAS vol. 105, pp. 13246–13251.

Chen, C, & Appleby, J 1984 "Biology of the cypress twig gall midge, Taxodiomyia cupressiananassa (Diptera: Cecidomyiidae) in central Illinois", Annals of the Entomological Society of America, vol. 77, pp. 203–207.

Chen, MS, Liu, S, Wang, H, Cheng, X, El Bouhssini, M, & Whitworth, RJ 2016, "Genes expressed differentially in Hessian fly larvae feeding in resistant and susceptible plants", International Journal of Molecular Science, vol. 17, p. 1324.

Conkling, S 1990 "News from members" Society of Vertebrate Paleontology, vol. 150, p. 41.

De Baets, K, Fatka, O, Budil, P, & Geyer, G 2021, "Trilobites as host for parasites: From paleopathologies to etiologies", in: De Baets, K & Huntley, JW (eds.) The evolution and fossil record of parasitism: Coevolution and paleoparasitological techniques, Cham: Springer, pp. 173–201.

Czametzki, A, Schwaderer, E, & Pusch, CM 2003 "Fossil record of meningioma", Lancet, vol. 362, no. 408.

Czametzki, A 1980 "Pathological changes in the morphology of the young Paleolithic skeletal remain from Stetten (Southwest Germany)", Journal of Human Evolution, vol. 9, pp. 15-16.

Danthala, M *et al.* 2017 "Generic Imatinib in chronic myeloid leukemia: Survival of the cheapest", Clinical Lymphoma, Myeloma & Leukemia, vol. 17, pp. 457–462.

Dawe, CJ 1969 "Phylogeny and ontogeny", Journal of the National Cancer Institute Monographs, vol. 31, pp. 1–40.

Downs, CJ, Dochtermann, NA, Ball, R, Klasing, KC, & Martin, LB 2020 "The effects of body mass on immune cell concentrations of mammals", The American Naturalist, vol. 195, pp. 107–114.

Dujon, A, Boutry, J, Tissot, S, Meliani, J, Guimard, L, Rieu, O, Ujvari, B, & Thomas, F 2022, "A review of the methods used to induce cancer in invertebrates to study its effects on the evolution of species and ecosystem functioning", Methods in Ecology and Evolution, vol. 13, pp. 1885–1898.

Dumbravá, MD, Rothschild, BM, Weishampel, DB, Csiki-Sava, Z, Andrei, RAA, Acheson, K, & Codrea, VA 2016, "A dinosaurian facial deformity and the first occurrence of ameloblastoma in the fossil record", Scientific Reports, vol. 6, art. 29271.

Dung, CV 2014, "A metabolic perspective of Peto's paradox and cancer", Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 370, art. 20140223.

Dunn, GP, Koebel, C, & Schreiber, RD 2006, "Interferons, immunity and cancer immunoediting", Nature Reviews Immunology, vol. 6, pp. 836–848.

Edgar, A, Mitchell, DG, & Martindale, MQ 2021 "Whole–body regeneration in the Lobate Ctenophore Mnemiopsis leidyi", Genes (Basel), vol. 12, no. 6, p. 867. doi: 10.3390/genes12060867

Effron, M, Griner, L, & Benirschke, K 1977, "Nature and rate of neoplasia in captive wild mammals, birds and reptiles at necropsy", Journal of the National Cancer Institute, vol. 59, pp. 185–198.

Ekhtiari, S, Chiba, K, Popovic, S, Crawther, R, Wohl, G, Wong, AKO, Tanke, DH, Dufault, DM, Geen OD, Parasu, N, Crowther, MA, & Evans, DC 2020 "First case of osteosarcoma in a dinosaur: A multimodal diagnosis", Lancet Oncology, vol. 21, pp. 1021–1022.

Elicki, O, & Geyer, G 2013, "The Cambrian trilobites of Jordan–taxonomy: Systematic and stratigraphic significance", Acta Geologica Polonica, vol. 63, pp. 1–56.

Elkashty, OA *et al.* 2021 "Aging, stem cells and cancer updated", Aging, vol. 13, pp. 20854–20855.

Erenpreisa, J, Vainshelbaum, NM, Lazovska, M, Karklins, R, Salmina, K, Zayakin, P, Rumnieks, F, Inashkina, I, Pjanova, D, Erenpreiss, J 2023 "The price of human evolution: Cancer-testis antigens, the decline in male fertility and the increase in cancer" International Journal of Molecular Sciences, vol. 24, art. 11660. doi: 10.3390/ijms241411660.

Espirito-Santo, M, & Wilson Fernandes, G 2007 "How many species of gall-inducing insects are there on Earth, and where are they?", Annals of the Entomological Society of America, vol. 100, pp. 95–99.

Ewing, J 1940, Neoplastic disease: A treatise on tumors. Philadelphia: Saunders.

Fankhauser G. Maintenance of normal structure in heteroploidy salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. J Exp Zoology 100: 445–455, 1945.

Fatula, S 2020 Detection of metastatic cancers in bone: A comparison of microscopic and radiographic methods. Master's Thesis, University of Tennessee. https://trace.tennessee.edu/utk_gradthes/5634/

Foth, C, Evers, SW, Pabst, B, Mateus, O, Flisch, A, Patthey, M, & Rauhut, OWM 2015, "New insights into the lifestyle of Allosaurus (Dinosauria: Theropoda) based on another specimen with multiple pathologies", PeerJ, vol. 3, art. e940.

Gahn, FJ, & Baumiller, TK 2010 "Evolutionary history of regeneration in Crinoids (Echinodermata)", Integrative and Comparative Biology, vol. 50, pp. 514a–514m.

Galmarini, CM 2020 "Lessons from Hippocrates: Time to change the cancer paradigm", International Journal of Chronic Diseases, Jun 7, art. 4715426.

Gohlke, J, & Deeken, R 2014, "Plant responses to Agrobacterium tumefaciens and crown gall development", Frontiers in Plant Science, vol. 5, p. 155.

González, R, Gallina, PA, & Cerda, IA 2017 "Multiple paleopathologies in the dinosaur Bonitasaura salgadoi (Sauropoda: Titanosauria) from the Upper Cretaceous of Patagonia, Argentina", Cretaceous Research, vol. 79, pp. 159–170.

Grabowski, M, & Koetter, R 2019, Crown Gall, in: University of Minnesota Extension, Minneapolis. https://extension.umn.edu/plant-diseases/crown-gall.

Grupe, G, & Harbeck, M 2014, "Taphonomic and diagenetic processes", in: Henke, W, & Tattersall, I. (eds.) Handbook of paleoanthropology, Berlin and Heidelberg: Springer, pp. 417–439.

Gubin IuM, Petrovichev NN, Solov'ev IuN, Kochergina NV, Luk'ianchenko AB, & Markov SM 2001. "Novoobrazovanie pokrovnyh kostej cherepa rannetriasovoj amfibii" [Neoplasm in cranial bone of early Triassic amphibian], Voprosy Onkologii, vol. 47, no. 4, pp. 449–455.

Hall, F, Morita, M, & Best, JB 1986 "Neoplastic transformation in the planarian: I. Cocarcinogenesis and histopathology", Journal of Experimental Zoology, vol. 240, pp. 211–227.

Hanahan, D, & Weinberg, RA 2000 "The hallmarks of cancer", Cell, vol. 110, pp. 57–70.

Haridy, Y, Witzmann, F, Asbach, P, Schoch, RR, Fröbisch, N, & Rothschild BM 2019, "Triassic cancer-osteosarcoma in a 240-million-year-old stem-turtle", JAMA Oncology, vol. 5, pp. 425–426.

Hublin, JJ, Weston, D, Gunz, P, Richards, M, Roebroecks, W, Glimmerveen, J, & Anthonis, L 2009 "Out of the North Sea: the Zeeland Ridges Neandertal", Journal of Human Evolution, vol. 57, pp. 777–785.

Hunt, K, Kirkpatrick, C, Campbell, R, & Willoughby, J 2017 Cancer Research in Ancient Bodies (CRAB) Database [online]. Pine River. www.cancerantiquity.org/crabdatabase

Hunt, K, Roberts, C, & Kirkpatrick, C 2018 "Taking stock: A systematic review of archeological evidence of cancer in human and early hominin remains", Interntaional Journal of Paleopathology, vol. 21, pp. 12–26.

Hunter, HA, & Langston, W Jr 1964, "Odontoma in northern mammoth", Paleontology, vol. 7, pp. 674–681.

Impson, F, Post, JA, & Hoffmann, J 2013, "Impact of the flower-galling midge, Dasineura rubiformis Kolesik, on the growth of its host plant, Acacia mearnsii De Wild, in South Africa. South African Journal of Botany, vol. 87, pp. 118–121.

Jiang, Y, Ye, J, Veromann-Jurgensön, LL, & Niinemets, U 2021 "Gall- and erineum-forming Eriophyes mites alter photosynthesis and volatile emissions in an infection severity-dependent manner in broad-leaved trees Alnus glutinosa and Tilia cordata", Tree Physiology, vol. 41, pp. 1122–1142.

Joy, D, Feng, X, Mu, J, Furuya, T, Chotivanich, K, Kretti, AU, Ho, M, Wang, A, White, NJ, Suh, E, Beerli, P, & Su, XZ 2003 "Early origin and recent expression of Plasmodium falciparum", Science, vol. 300, pp. 318–321.

Kakiuchi, N, & Ogawa, S 2021 "Clonal expansion in non-cancer tissues" Nature Reviews Cancer, vol. 21, pp. 239–256.

Karin, M, Lawrence, T, & Nizet, V 2006 "Innate immunity gone awry: Linking microbial infections to chronic inflammation and cancer", Cell, vol. 24, pp. 823–835.

Keenan, SW 2021, "Fossilization", in: Elias, S, & Alderton, D (eds.) Encyclopedia of Geology, 2nd Edition, Amsterdam: Elsevier, pp: 1–11.

Kenny, PA, & Bissell, MJ 2003 "Tumor reversion: Correction of malignant behavior microenvironmental cues", International Journal of Cancer, vol. 107, pp. 688–695.

Khatami, M 2018 "Cancer, an induced disease of twentieth century! Induction of tolerance, increased entropy and 'Dark Energy': Loss of biorhythms (Anabolism v. Catabolism). Clinical and Translational Medicine, vol. 7, no. 1, art. 20. doi: 10.1186/s40169-018-0193-6.

Kiriakakis, E, Markaki, M, & Tavemarakis, N 2015 "Caenorhabditis elegans as a model of cancer research", Molecular & Cellular Oncology, vol. 2, art. e975027.

Kitsoulis, CV, Baxevanis, AD, & Abatzopoulos, TJ 2020 "The occurrence of cancer in vertebrates: A mini review", Journal of Biological Research - Thessaloniki, vol. 27, no. 9. doi: 10.1186/s40709-020-00119-0

Kluepfel, M, McLeod, SJ, Blake, JH, & Gorsuch, CS 2017, Galls & outgrowths [factsheet], in: Clemson Cooperative Extension Home & Garden Information Center, Clemson. https://hgic.clemson.edu/factsheet/galls-outgrowths.

Kobayashi, T 1937, "Odontoma in a fossil elephant from the Inland Sea of Japan", J Geol Geogr Japan 14: 59–60, 1937.

Krassilov, VA, & Karasev, E 2008 "First evidence of plant–arthropod interaction at the Permian-Triassic boundary in the Volga Basin, European Russia", Alavesia, vol. 2, pp. 247–252.

Krzeminska, I 2008 "Preliminary characteristics of pathologies found in the skeletons of mammoths at the Krakow Spadzista Street (B) site", Veterinarija ir Zootechnika, vol. 43, pp. 52–57.

Labandeira, CC 2021, "Ecology and evolution of gall-inducing arthropods: The pattern from the terrestrial fossil record", Frontiers in Ecology and Evolution, vol. 9, art. 632449. doi: 10.3389/fevo.2021.632449.

Lacroix, B, & Citovsky, V. 2001 "Crown gall tumors", in: Maloy, S, & Hughes, K (eds.) Brenner's encyclopedia of genetics, Amsterdam: Elsevier, 2nd edition, vol. 2, pp. 236–239.

Luna, L, Aranda, C, Bosio, L, Berón, M 2008 "A case of multiple metastasis in late Holocene hunter-gatherers from argentine pampean region", International Journal of Osteoarcheology, vol. 18, pp. 492–506.

Luna, L, Aranda, C, Santos, AL, Ramundo, P, Rizzuti, C, & Stagno, D 2015 "Probable prostate cancer in a pre-Incaic individual from Pukara de la Cueva, northwestern Argentina", Anthropologischer Anzeiger: Journal of Biological and Clinical Anthropology, vol. 72, pp. 201–222.

Maffini, MV, Soto, AM, Calabro, JM, Ucci, AA, & Sonnenschein, C 2004 "The stroma as a crucial target in rat mammary gland carcinogenesis", Journal of Cell Science, vol. 117, pp. 1495–1502.

Martincorena, I et al 2018 "Somatic mutant clones colonize the human esophagous with age", Science, vol. 362, pp. 911–917.

Mayo Clinic 2023, "Cancer: Diagnosis and treatment", in: mayoclinic.org. Rochester. www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594.

McAuliffe, K, & Whitehead, H 2005 "Eusociality, menopause and information in matrilineal whales", Trends in Ecology & Evolution, vol. 20, p. 650.

McLoughlin, S 2011, "New records of leaf galls and arthropod oviposition scars in Permian-Triassic Gondwanan gymnosperms", Australian Journal of Botany, vol. 59, pp. 156–169.

Miralles, A, & Crusafont Pairo, M 1952 "Algunos casos de paleopatología y de paleoteratología en artiodáctilos del terciario español", Estudios Geologicos - Instituto de Investigaciones Geológicas "Lucas Mallada", vol. 8, pp. 225–231.

Mitchell, AD, & Valk, WL1962 "Compensatory renal hypertrophy", Journal of Urology, vol. 88, pp. 11–18.

Moein, S, Adibi, R, Da Silva Meirelles, L, Nardi, NB, & Gheisari, Y 2020 "Cancer regeneration: Polyploid cells are the key drivers of tumor progression" Biochimica et Biophysica Acta Reviews on Cancer, vol. 1874, art. 188408.

Monge, J, Kricun, M, Rodovcic, J, Radovcic, D, Mann, A, & Frayer, DW 2013, "Fibrous dysplasia in a 120,000+ years old Neandertal from Krapina, Croatia" PLoS ONE, vol. 8. doi:10.1371.

Moodie, RL 1921, "Status of our knowledge of Mesozoic pathology", Bulletin of the Geological Society of America, vol. 32, pp. 321–326.

Moodie, RL 1927, "Tumors in the Lower Carboniferous", Science, vol. 66, p. 540.

Moodie, RL 1929 "Studies in Paleopathology XX. The teeth and jaws of Nothrotherium", Pacific Dental Gazette, vol. 37, pp. 677–680.

Muller, AWJ 2017 "Cancer is an adaptation that selects in animals against energy dissipation", Medical hypothesis, vol. 104, pp. 104–115.

Natarajan, LC, Melott, AL, Rothschild, BM, & Martin, LD 2007 "Bone cancer rates in dinosaurs compared with modern vertebrates", Transactions of the Kansas Academy of Science, vol. 110, pp. 155–158.

Nathanson, SD, Nelson, LT, & Lee M 1993 "A spontaneous subcutaneous tumor in C57BL/6 mice that metastatize to the liver", Clinical & Experimental Metastasis, vol. 11, pp. 45–54.

Nerlich, AG, Rohrbach, H, Bachmeier, B, & Zink, A 2006 "Malignant tumors in two ancient populations: An approach to historical tumor epidemiology", Oncology Reports, vol. 16, pp. 197–202.

Nielsen, ML, & Nielsen, AT 2017, "Two abnormal pygidia of the trilobite Toxochasmops from the Upper Ordovicin of the Oslo region, Norway", Bulletin of the Geological Society of Denmark, vol. 65, pp.171–175.

Nordling, CO 1953, "A new theory on the cancer inducing mechanism", British Journal of Cancer, vol. 7, pp. 68–72.

Norman, DB, & Milner, A 1989, Dinosaurs, Toronto: Stoddart Publishing Company.

Nunney, L 2020 "Resolving Peto's paradox: Modeling the potential effects of size-related metabolic changes, and of the evolution of immune policing and cancer suppression", Evolutionary Applications, vol. 13, pp. 1581–1592.

Odes, EJ, Quinney, PSR, Steyn, M, Throckmorton, Z, Smilg, JS, Zipfel, B, Augustine, TN, De Beer, F, Hoffman, JW,

Franklin, RD, & Berger, LR 2016 "Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa", South African Journal of Science, vol. 112, no. 7/8, art. 5. doi: 10.17159/sajs.2016/20150471

Ostrander, GK, Cheng, KC, Wolff, JC, & Wolfe, MJ 2004 "Sharks, cartilage, cancer and the growing threads of pseudoscience", Cancer Research, vol. 64, pp. 8485–8491.

Owen, AW 1983, "Abnormal cephalic fringes in the Trinucleidae and Harpetidae (Trilobita)", Special Papers in Palaeontology, vol. 30, pp. 241–247.

Pahl WM. 1986 "Tumors of bones and soft tissues in ancient Egypt and Nubia: a synopsis of the detected cases", International Journal of Anthropology, vol 1, pp. 267–276.

Pales, L 1959, "Pathologie de l'ours des caverns (Collection Felix Gariigou au Musée de Foix)", Annales de Paleontologie, vol. 44, pp. 1–44.

Pales, L, & Wernert, P 1953, "Une mandibule pathologique de grand bovidé du loess d'Achenheim (Bas-Rhin)" [A pathological mandible of a large bovid from the Achenheim loess (Bas-Rhin)], Annales de Paleontologie, vol. 39, pp. 21–26.

Patte, E 1937, "Molare pathologique d'une cheval fossile. Interprétation d'un cas actuel" [Pathological molar of a fossil horse: Interpretation of a current case], Mammalia, vol. 1, pp. 97–105.

Perret, C, Gidoin, C, Ujvari, B, Thomas, F, & Roche, B 2020, "Predation shapes the impact of cancer on population dynamics and the evolution of cancer resistance", Evolutionary Applications, vol. 13, pp. 1733–1744.

Peto, R 2015 "Quantitative implications of the approximate irrelevance of mammalian body size and lifespan to lifelong cancer risk", Philosophical Transactions of the Royal Society B, vol. 370, art. 20150198.

Peto, R, Roe, FJC, Lee, PN, Levy, L, & Clack, J 1975 "Cancer and ageing in mice and men", British Journal of Cancer, vol. 32, pp. 411–426.

Petralia, RS, Mattson, MP, & Yao, PJ 2014 "Aging and longevity in the simplest animals and the quest for immortality", Ageing Research Reviews, vol. 16, pp. 66–82.

Pickrell, J 2003 "First Dinosaur Brain Tumor Found, Experts Suggest", *National Geographic*, 24 Nov. 2003 (as cited by Larson P *et al.* 2003 Abstracts of papers. Sixty-third Annual Meeting Society of Vertebrate Paleontology, Science Museum of Minnesota. St. Paul, Minnesota, October 15-18, 2003. *Journal of Vertebrate Paleontology* 23, Issue Sup 003).

Prehn, RT 1997 "Regeneration versus neoplastic growth", Carcinogenesis, vol. 18, pp. 439–1444.

Prehn, RT 2005, "The role of mutation in the new cancer paradigm", Cancer Cell International, vol. 5, pp. 9–14.

Prejean, JD, Peckham, JC, Casey, AE, Griswold, DP, Weisburger, EK, & Weisburger, JH 1973 "Spontaneous tumors in Sprague-Dawley rats and Swiss mice", Cancer Research, vol. 33, pp. 2768–2773.

Przyklady, KH 1965 "Anomalii w uzebieniu slony kopalnych" [Examples of anomalies in the teeth of fossil elephants] Folia Quaternaria, vol. 19, pp. 45–61.

Pugh, TD, Oberley, TD, & Weindruch, R 1999 "Dietary intervention at middle age: Caloric restriction but not dehydroepiandrosterone sulfate increases lifespan and lifetime cancer incidence in mice", Cancer Research, vol. 59, pp. 1642–1648.

Quinney, PSR, Williams, SA, Steyn, M, Meyer, MR, Smilg, J, Churchill, S, Odes, EJ, Augustine, T, Tafforeau, P, & Berger, LR 2016 "Osteogenic tumor in Australopithecus sediba: Earliest hominin evidence for neoplastic disease" South African Journal of Science, vol. 112, no. 7/8, art. 7. doi: 10.17159/sajs.2016/20150470.

Rega, E, Holmes, R, & Tirabasso A 2010, "Habitual locomotor behavior inferred from manual pathology in two Late Cretaceous Chasmosaurine ceratopsid dinosaurs, Chasmosaurus irvinensis (CMN 41357) and Chasmosaurus belli (ROM 843)", in: Ryan, MJ. Chinnery-Allgeier, BJ, Eberth, DA (eds.) New perspectives on horned dinosaurs: The Royal Tyrrell Museum of Paleontology Ceratopsian Symposium, Bloomington: Indiana University Press, pp. 340–354.

Robert, J 2010, "Comparative study of tumorigenesis and tumor immunity in invertebrates and non–mammalian vertebrates", Developmental and Comparative Immunology, vol. 34, pp. 915–925.

Rothschild, BM, & Martin, LD 1993, Paleopathology: Disease in the fossil record, Boca Raton: CRC Press.

Rothschild, BM, Tanke, DH, Helbling 2nd, M, & Martin, LD 2003, "Epidemiology study of tumors in dinosaurs", Naturwissenschaften, vol. 90, pp. 495–500.

Rothschild, BM, Tanke, DH, Hershkovitz, I, & Schultz, M 1998, "Mesozoic neoplasia: Origin of haemangioma in the Jurassic age", Lancet, vol. 351, p. 1862.

Rothschild, BM, Tanke, D, Ruhli, F, Pokhojaev, A, & May, H 2020, "Suggested case of Langherans cell histiocytosis in a Cretaceous dinosaur", Scientific Reports, vol. 10, art. 2203.

Rothschild, BM, Witzke, BJ, & Hershkovitz, I 1999 "Metastatic cancer in the Jurassic", Lancet, vol. 354, p. 398.

Rothschild, BM, & Woods, RJ 1991 "Spondyloarthropathy: Erosive arthritis in representative defleshed bones", American Journal of Physical Anthropology, vol. 85, pp.125–134.

Rozhok, AI, & DeGregori J 2016 "The evolution of lifespan and age-dependent cancer risk", Trends Cancer, vol. 2, pp. 552–560.

Rubin, H 2006 "What keeps cells in tissues behaving normal in the face of myriad mutations?", BioEssays, vol. 28, pp. 515–524.

Ruggiero, RA, & Bustuoabad, OD 2006 "The biological sense of cancer: A hypothesis", Theoretical Biology and Medical Modelling, vol. 3, p. 43.

Ruggiero, RA, Bustuoabad, OD, Chiarella, P, Bruzzo, J, & Meiss, RP 2008 "On the biological significance of tumor growth", in: Kimura, D (ed.) Cell growth processes: New research, New York: Nova Science Publishers, pp. 89–121.

Saló, E 2006 "The power of regeneration and the stem cell kingdom: Freshwater planarias (Platyhelminthes)", Bioessay, vol. 28, pp. 546–559.

Sanchez Alvarado, A, & Yamanaka, S 2014, "Rethinking differentiation: Stem cells, regeneration and plasticity", Cell, vol. 157, pp.110–119.

Schachat, SR, & Labandeira, CC 2015 "Evolution of a complex behavior: The origin and initial diversification of foliar galling by Permian insects", Naturwissenschaften, vol. 102, p. 14.

Scheele, WE 1954, Prehistoric animals. Cleveland: Collins & World.

Schread, JC 1971 "Control of the eastern spruce gall aphid", Circular of the Connecticut Agricultural Experiment Station, New Haven, no. 242, pp. 1–6. Available from: https://portal.ct.gov/-/media/CAES/DOCUMENTS/Publications/Circulars/C242pdf.pdf

Scott, WB 1898, "Notes on the Canidae of the White River Oligocene", Transactions of the American Philosophical Society, vol. 19, pp. 329–415.

Scott, A, Stephenson, J, & Collinson, ME 1994 "The fossil record of leaves with galls", in Williams, MAJ (ed.) Plant galls: Organisms, interactions, populations, Oxford: Clarendon Press, pp. 447–470.

Senter, P, & Juengst, SL 2016 "Record–breaking pain: The largest number and variety of forelimbs bone maladies in a theropod dinosaur", PLoS One, vol. 11, art. e0149140.

Sharpless, NE, DePinho, RA 2004 "Telomers, stem cells, senescence and cancer", Journal of Clinical Investigation, vol. 113, pp. 160–168.

Shimkin, MB 1977 Contrary to nature: Being an illustrated commentary on some persons and events of historical importance in the development of knowledge concerning ... cancer, Washington: U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health.

Shimkin, MB, & Novinsky, MA 1955 "A note on the history of transplantation of tumors", Cancer, vol. 8, pp. 653–655.

Shimosato, Y, Kameya, T, Nagai, K, Hirohashi, S, Koide, T, Hayashi, H, & Nomura, T 1976 "Transplantation of human tumors in nude mice", Journal of the National Cancer Institute, vol. 56, pp. 1251–1260.

Smith, EF 1916, "Further evidence that crown gall of plants is cancer", Science, vol. 43, pp. 871–879.

Smithers, DW 1962 "An attack to cytologism", Lancet vol. 1, pp. 493–499.

Sonnenschein, C, & Soto, AM 1999, The society of cells. New York: Springer Verlag.

Sonnenschein, C, & Soto, AM 2000 "Somatic mutation theory of carcinogenesis: Why it should be dropped and replaced", Molecular Carcinogenesis, vol. 29, pp. 205–211.

Sonnenschein, C, & Soto, AM 2016 "Carcinogenesis explained within the context of a theory of organisms", Progress in Biophysics and Molecular Biology, vol. 122, pp. 70–76.

Sonnenschein, C, & Soto, AM 2020 "Over a century of cancer research. Inconvenient truths and promising leads" PLOS Biology, vol. 18, no. 4, art. e3000670.

Sonnenschein, C, & Soto, AM 2021, "Control of cell proliferation: Is the default state of cells quiescence or proliferation?", Organisms. Journal of Biological Sciences, vol. 5, no. 1, pp. 33–42. doi: 10.13133/2532–5876/17338

Souza Barbosa, FH, Gomes da Costa Pereira, PV, Paglarelli Berggvist, L, & Rothschild, BM 2016, "Multiple neoplasms in a single sauropod dinosaur from the upper Cretaceous of Brazil", Cretaceous Research, vol. 62, pp. 13–17.

Souza Barbosa, FH, Porpino, KO, Rothschild, BM, Da Silva, RC, & Capone, D 2021 "First cancer in an extinct Quaternary non–human mammal", Historical Biology, vol. 33, pp. 2878–2882.

Sparks, AK 1969 "Review of tumors and tumor-like conditions in Protozoa, Coelenterata, Platyhelminthes, Annelida, Sipunculida and Arthropoda excluding Insects", National Cancer Institute Monograph, vol. 31, pp. 671–682.

Stadtman, K 1992, "News from members", Society of Vertebrate Paleontology, vol. 155, p. 45.

Stocum, DL 2017, "Mechanisms of urodele limb regeneration", Regeneration (Oxford), vol. 4, pp. 159–200. doi: 10.1002/reg2.92

Street, HP, & O'Keefe, R 2010 "Evidence of pachyostosis in the cryptocleidoid plesiosaur Tatenectes laramiensis from the Sundance Formation of Wyoming", Journal of Vertebrate Paleontology, vol. 30, pp. 1279–1282.

Strouhal, E 2001 "Malignant tumors in past populations in Middle Europe", in: La Verghetta, M, & Capasso, L (eds.) Proceedings of the XIIIth European Meeting of the Paleopathology Association, Teramo: Edigrafical Publisher, pp: 265–272.

Sung, H *et al.* 2021 "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: A Cancer Journal for Clinicians, vol. 71, pp. 209–249.

Surmik, D, Slowiak-Morkovina, J, Szczygielski, T, Kamaszewski, M, Kalita, S, Teschner, E, Drózdz, D, Duda, P, Rothschild, B, & Konietzko-Meier, D 2022, "An insight into cancer palaeobiology: Does the Mesozoic neoplasm support tissue organization field theory of tumorigenesis?", BMC Ecology and Evolution, vol. 22, no. 1, p. 143.

Tascedda, F, & Ottaviani, E 2014 "Tumors in invertebrates", ISJ-Invertebrate Survival Journal, vol. 11, pp. 197–203.

Tollis, M, Boddy, AM, & Maley, CC 2017 "Peto's paradox: How has evolution solved the problem of cancer prevention?",

BMC Biology, vol. 15, no. 1, art. 60. doi: 10.1186/s12915-017-0401-7

Tomasetti, C, Marchionni, L, Nowak, MA, Parmigiani, G, & Vogelstein, B 2015, "Only three driver gene mutations are required for the development of lung and colorectal cancers", Proceedings of the National Academy of Sciences of the United States of America, vol. 112, pp. 118–123.

Vasilenko, DV 2005, "Damages on Mesozoic plants from the Transbaikalian locality Chernovskie Kopi" Paleontological Journal, vol. 39, pp. 628–633.

Vincze, O, Colchero, F, Lemaitre, JF, Conde, DA, Pavard, S, Bieuville, M, Urrutia, AO, Ujvari, B, Boddy, AM, Maley, CC, Thomas, F, & Giraudeau, M 2022, Cancer risk across mammals. Nature 601: 263–267, 2022.

Vineis, P, Manuguerra, M, Kavvoura, FK, Guarrera, S, Allione, A, Rosa, F, Di Gregorio, A, Polidoro, S, Saletta, F, Ioannidis, JPA, & Matullo, G 2009 "A field synopsis on low penetrance variants in DNA repair genes and cancer susceptibility", Journal of the National Cancer Institute, vol. 101, pp. 24–36.

Vogts, G 2008 "How to minimize formation and growth of tumours. Potential benefits of decapod crustaceans for cancer research", Interntaional Journal of Cancer, vol. 123, pp. 2727–2734.

Waddington, CH 1935 "Cancer and the theory of organisers", Nature, vol. 135, pp. 606–608.

Wang, XM, & Rotschild, BM 1992 "Multiple hereditary osteochondroma in Oligocene Hesperocyon (Canidae: Carnivora)", Journal of Vertebrate Paleontology, vol. 12, pp. 387–394.

Weihua, Z, Lin, Q, Ramoth, AJ, Fan, D, & Fidler, IJ 2011 "Formation of solid tumors by a single multinucleated cancer cell", Cancer, vol. 117, pp. 4092–4099.

Weiss, RA 2004 "Multistage carcinogenesis", British Journal of Cancer, vol. 91, pp. 1981–1982.

Wellings, SR 1969 "Neoplasia and primitive vertebrate phylogeny: Echinoderms, prevertebrates and fishes", National Cancer Institute Monograph, vol. 31, pp. 59–128.

Wheatley, DN, & Clegg JS 1994, "What determines the basal metabolic rates of vertebrates in vivo?", Biosystems, vol. 32, pp. 83–92.

Whitney, MR, Mose, L, & Sidor, CA 2017, "Odontoma in a 255-million-year-old mammalian forebear", JAMA Oncology, vol. 3, pp. 998–1000.

Woodruffs, M 1982 "Interaction of cancer and host", British Journal of Cancer, vol. 46, pp. 313–322.

Zajicek, G 1996 "A new cancer hypothesis" Medical Hypotheses, vol. 47, pp. 111–115.

Zhu, J, Oger, PM, Schrammeijer, B, Hooykaas, PJ, Farrand, SK, & Winans, SC 2000, "The bases of crown gall tumorigenesis", Journal of Bacteriology, vol. 182, pp. 3885–3895.

Zimmer, C 2007, "Evolved for cancer?", Scientific American, vol. 296, pp. 68–74.

SAPIENZA UNIVERSITÀ DI ROMA

In Memoriam

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

DOI: 10.13133/2532-5876/18309

Celebrating Evelyn Fox Keller: "The Toronto Statement"

Rasmus Grønfeldt Winther[a], Eva Jablonka[b], Giuseppe Longo[c], Gregory Radick[d], Sarah S. Richardson[e], Jessica Riskin[f], Gregory Rupik[g], Ana M. Soto[h], Carlos Sonnenschein[h], Sonia Sultan[i], Marga Vicedo[j], Denis Walsh[j], & Anat Zeligowski[k]**

[a]*Humanities Division, University of California, Santa Cruz, USA and Globe Institute, University of Copenhagen, Denmark;* [b]*The Cohn Institute for History and Philosophy of Science and Ideas, Tel Aviv University, Israel;* [c]*Centre Cavaillès, CNRS and École Normale Supérieure, Paris, France;* [d]*School of Philosophy, Religion and the History of Science, University of Leeds, UK;* [e]*History of Science Department, Harvard University, USA;* [f]*Department of History, Stanford University, USA;* [g]*Office of the President, University of St. Michael's College, Canada;* [h]*School of Medicine, Tufts University, USA and Centre Cavaillès, CNRS and École Normale Supérieure, Paris, France;* [i]*Biology Department, Wesleyan University, USA;* [j]*Institute for the History and Philosophy of Science and Technology, University of Toronto;* [k]*Medical doctor and artist, Bari, Italy*

***Corresponding author:** Rasmus G. Winther, Email: *rgw@ucsc.edu*

Citation: Winther, RG, Jablonka, E, Longo, G, Radick, G, Richardson, SS, Riskin, J, Rupik, G, Soto, AM, Sonnenschein, C, Sultan, S, Vicedo, M, Walsh, D, & Zeligowski, A,"Celebrating Evelyn Fox Keller: 'The Toronto Statement'", *Organisms: Journal of Biological Sciences*, vol. 6, no. 2 (2023), pp. 71-76. DOI: 10.13133/2532-5876/18309

With much anticipation, in time and spirit, a group of historians and philosophers of biology, and theoretical and experimental biologists, met at the Institute for the History and Philosophy of Science and Technology at the University of Toronto, Canada, on 22 and 23 September 2023, to pay homage to Evelyn Fox Keller for her contributions to theoretical biology. The conference was supported and sponsored by the Department of Philosophy, Institute for the History and Philosophy of Science and Technology, Faculty of Arts and Science (all at University of Toronto), as well as the Social Sciences and Humanities Research Council of Canada. The workshop was an in-person celebration complementing the publication of a volume celebrating her work (Vicedo & Walsh 2020a, 2020b), which includes the following contributions: (Herrington & Jablonka 2020; Longo & Mossio 2020; Radick 2020; Riskin 2020; Soto & Sonnenschein 2020; Walsh 2020). After introductory remarks by co-organizer Denis Walsh and a short video in which Evelyn Fox Keller thanked us for this homage and wished us well, we learned of her death.

Figure 1. *Evelyn Fox Keller: Making sense of gender and science* (ink and pencil) by Anna (Anat) Zeligowski

Perhaps the overarching message that emerged from the two-day conference "Language, History, Gender, and Science: Celebrating the Work of Evelyn Fox Keller" is that historians, philosophers, sociologists of science and biologists are nowhere near done thinking through the vexing and complex roles of gender, metaphor, reductionism, and mechanism in the life sciences. Conference attendees joined in thanking Keller for doing so much to uncover and critique these challenging—but perhaps also generative—features of the life sciences. And in a closing session, participants reflected on the seemingly unshakeable dominance of genetic determinism in biological theory and practice as co-organizer Marga Vicedo forcefully asked, "But how do we change things?" Thought without action is empty, and in this brief conceptual review of the conference, we summarize some key themes, diagnose outstanding challenges, and report some calls to action brought up by conference attendees.

All literatures are saturated. Even so, we believe that Keller's work is sufficiently important and trenchant to make a public record of this conference worthwhile. In this necessarily brief report, our organizing principle will be to group key themes of different talks in terms of four of Evelyn Fox Keller's books. This is solely a strategy of convenience—there is no intent of denigrating or ignoring her other work. We conclude with some comments on art and science, a topic Keller did not explicitly address.

1. The Mirage of a Space between Nature and Nurture (2010). Durham: Duke University Press

Why do the life sciences—particularly those concerned with development, heredity, and evolution—continue to ground analyses on a strong/stronger nature (genes, DNA) versus nurture (environment, culture) distinction? Yes, we are all aware that these elements interact, and yet…

History matters. Eva Jablonka has long argued that not only is there change in heritable patterns of gene expression, but there are also important effects of epigenetic states of the rate and types of mutations in DNA (Jablonka & Lamb 1995, 2006, 2020; Monroe *et al.* 2022). Giuseppe Longo distinguished dynamic state spaces in biology versus pre-given "transcendental" state spaces in physical theories. In so doing, he detailed different forms of mechanisms in physics and hinted at a close analysis of "physical emergence" versus "production of biological novelty", in order to illuminate biological historicity (Longo & Montévil 2014, Longo 2021; Riskin 2016).

Constant interaction. Gregory Radick presented a triangle, modelled on the fire triangle, with sides labelled "oxygen", "fuel" and "heat". The phenotype triangle that Radick introduced in his talk had sides labelled "genotype", "internal context" and "external context", to give visual expression to the Weldonian-Lewontian-Kellerian perspective in which talk of genotypes causing phenotypes feels as absurd as talk of oxygen causing fires (Radick 2020, 2023).

Plant plasticity. Sonia Sultan's phenotypic plasticity discussion moved norms of reaction from "properties of the genome" (standard view) to dynamic, complex, often adaptive and sometimes multi-generational behaviors of plants which take place without coordination by a brain or consciousness (Sultan 2019, 2021).

Not only a question of semantics: nature/nurture in Daniel Lehrman's work on animal behavior. Reviewing developmental psychologist Daniel Lehrman's contributions to animal behavior, Marga Vicedo talked about his pioneering work that highlighted the active role of organisms in constructing their environment. She also noted how Lehrman made significant contributions by clarifying key concepts in the nature versus nurture debate (Vicedo 2023a; Vicedo 2023b). However, Vicedo argued that semantic analysis ("linguistic hygiene") will never be sufficient to resolve/dissolve the nature versus nurture debate. She suggested

Figure 2. *Five mothers: Inputs to development and heredity* (ink and pencil) by Anna (Anat) Zeligowski

moving beyond reductionist research that is premised on and further entrenches that debate and its (misguided) underlying assumptions. In her view, we need a new type of science that allows for detailed analysis of the complex factors that dynamically interact in biological systems.

2. Refiguring Life: Metaphors of Twentieth-Century Biology (1995). New York: Columbia University Press

> "*Refiguring Life* begins with the history of genetics and embryology, showing how discipline-based metaphors have directed scientists' search for evidence. Keller continues with an exploration of the border traffic between biology and physics, focusing on the question of life and the law of increasing entropy. In a final section she traces the impact of new metaphors, born of the computer revolution, on the course of biological research" (From the original book description).

Towards a theory of organisms. Ana M. Soto and Carlos Sonnenschein argued that metaphors may inspire new concepts but play a different role than theories (Soto & Sonnenschein 2020). Scientific theories are needed to determine observables, frame experiments, and provide understanding (Longo & Soto 2016; Winther 2020b). A "theory of organisms" encompassing the entire life cycle would help clarify the difference between organisms, which are historical purposive agents and non-historical inert objects (Soto *et al.* 2016). Three principles for such a theory were enunciated: constitutive proliferation and motility (Soto, Longo, Montévil, & Sonnenschein 2016), constitutive variation (Montévil *et al.* 2016) and organization by closure of constraints (Mossio *et al.* 2016; Walsh 2015). These principles were also used to frame a theory of cancer, the tissue organization field theory (TOFT) (Sonnenschein & Soto 2020).

Metaphorical reductions. "Lamarckian inheritance of acquired traits", "the Weismannian barrier", and "the Mendelian Gene"—key theoretical biological advances often involve a *hardening* (S. J. Gould) and a kind of *nothing-but* (William James) thinking, where the rich complexity of the key theoretician's framework is productively yet perniciously reified (Winther 2020a). But is Weismannism to Weismann as Mendelism is to Mendel, as Darwinism is to Darwin, as Lamarck is to Lamarckism (Riskin 2023; White, Hodge, & Radick 2021; Winther 2000, 2001; Radick 2023)? Should we be mindful of such metaphorical reifications of the views of single biologists, and of the cross-biologist (dis)analogies? For instance, the Weismannian barrier, which Weismann himself did not necessarily endorse, has been a scientific dogma that originated in a religious dogma, and has stood as a barrier partitioning evolutionary development from the agency of organisms.

3. Secrets of Life/Secrets of Death: Essays on Language, Gender and Science (1992). New York: Routledge

> "Part of the motivation for this book is to distinguish the particular strand of 'Gender and Science' studies concerned with the role of gender ideologies *in* science, and to embed it in a more general historiographic and philosophical pursuit" (Keller 1992, p. 8).

Sex contextualism. In earlier work, Sarah Richardson

> "examine[d] the interaction between cultural gender norms and genetic theories of sex from the beginning of the twentieth century to the present, postgenomic age … using methods from history, philosophy, and gender studies of science" (Richardson 2013, book description).

In her conference talk, Richardson used a philosophical approach to statistics to present multiple *reductiones ad absurdum* of the assumption of intrinsic sex. She indicated how "sex contextualism" was itself a fruitful and important research program (Richardson 2022).

A dialectical feminism? Rasmus Winther argued that a dialectical feminism highlights the promise of approaching contradictions generatively—reason meets intuition, objectivity meets subjectivity, reduction meets pattern, and linear causation meets complex causation (Winther 2021). It is a critical and capacious stance that can produce good normal and revolutionary science, and can also call for an ethical approach to science.

Circuses and octopi. Zeligowski's circus drawings in the onsite Zeligowski IHPST exhibit resonate with Lynn Margulis' octopoid woman incessantly multitasking:

> "A woman must be almost octopoid in her attentions if she is to survive. Holding the infant in one arm, [Mary

> Catherine] Bateson points out, she stirs the pot with the other, while she watches the toddler" (Margulis 1998, p. 24).

4. A Feeling for the Organism: The Life and Work of Barbara McClintock (1983). New York: W.H. Freeman

Keller's classic biography of Barbara McClintock shed light on McClintock's holism, and her focus on chromosomal organization and what McClintock called the "reactive genome".

Towards a philosophy of nature. Philosophy of nature is an integrative, holistic, and inclusive philosophy—it remains open to teleology, emotions, even mysticism, in the interest of giving a more accurate and meaningful portrayal of our world (Winther 2019). It stands in a creative, dialectical contrast to analytic philosophy of science. Following Keller, one could call McClintock a philosopher of nature. In the last chapter of her biography, Keller had spoken of McClintock's "deep reverence for nature [and] a capacity for union with that which is to be known" and cites her in describing McClintock's own "love affair with the world", which included the facts that McClintock gladly called herself a "mystic" and believed that "everything is one" (Keller 1983, pp. 201, 204–205).

A feminine way of doing science? Sonia Sultan, Jessica Riskin, and Sarah Richardson provided different exegetical standpoints from second wave and third wave feminisms in the discussion following Winther's talk. Are empathic or holistic approaches necessarily gendered? Keller's version of feminism was universalist, not identity feminism. As she put it in a 1986 interview with Boston Globe cited in her *The New York Times* obituary, "I am not saying that women will do a different kind of science, I am saying when there are more women in science, everybody will be free to do a different kind of science" (Risen 2023). Moreover, in her 2023 memoirs, *Making Sense of My Life in Science*, she responded to the so-called McClintock Myth (Comfort 2003) by citing a long passage from her 1985 book *Reflections on Gender and Science* (pp. 174–175), which included these sentences:

> "Her … [i.e.,] any scientist who happens to be a woman … alternative is to attempt a radical redefinition of terms. Nature must be renamed as not female, or, at least, as not an alienated object. By the same token, the mind, if the female scientist is to have one, must be renamed as not necessarily male, as gender neutral, and accordingly recast with a more inclusive subjectivity …"

As a final comment, participants considered a topic resonating with, e.g., Keller's work on generative metaphor in science: the productive interface between art and science (Herrington & Jablonka 2020). Jablonka's beautiful talk, and the resonances with the IHPST art installation of Anat Zeligowski highlighted this connection (Ginsburg & Jablonka 2022). What is the role of aesthetic judgment (*à la* Kant, Romanticism, and *Naturphilosophie*) in the context of discovery and in the context of justification of science (Winther & Raffn 2024)? How can art and aesthetic judgment help us think about the theory-data or representation-phenomena relations? How does Waddington's analysis on the resonance between art and science in the 20th century (Waddington 1969) illuminate these questions?

In sum, there is much work to do to move beyond the reductionist and metaphorical "gene thinking" that Evelyn Fox Keller so cogently and eloquently worked to dismantle. Her contributions to gender and science discourse also continue to be influential and of going concern.

References

Comfort, N 2003 *The tangled field: Barbara McClintock's search for the patterns of genetic control*, Harvard, MA: Harvard University Press.

Ginsburg, S, & Jablonka, E 2022 *Picturing the mind: Consciousness through the lens of evolution*, Cambridge, MA: MIT Press.

Herrington, E & Jablonka E 2020 "Inviting a 'gestalt shift' in 21st century evolutionary biology: Roles for metaphor in the conceptual landscape of the extended evolutionary synthesis (EES)", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 360–379.

Jablonka, E, & Lamb, MJ 1995 *Epigenetic inheritance and evolution: The Lamarckian dimension*, Oxford: Oxford University Press.

Jablonka, E, & Lamb, MJ 2006 *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*, Cambridge, MA: MIT Press.

Jablonka, E, & Lamb, MJ 2020 *Inheritance systems and the extended synthesis*, Cambridge, MA: Cambridge University Press.

Longo, G 2021 "Confusing biological twins and atomic clocks. Today's ecological relevance of Bergson-Einstein debate on time", in Camp, A & Gozzano, S (eds.) *Einstein vs. Bergson. An enduring quarrel of time*, Berlin: De Gruyter, pp. 375–405. https://doi.org/10.1515/9783110753707-020

Longo, G, & Montévil, M 2014. *Perspectives on organisms: Biological time, symmetries and singularities*, Dordrecht: Springer.

Longo, G, & Mossio, M 2020 "Geocentrism vs genocentrism: Theories without metaphors, metaphors without theories", *Interdisciplinary Science Reviews*, vol. 45, pp. 380–405. https://doi.org/10.1080/03080188.2020.1798588

Longo, G, & Soto, A 2016 "Why do we need theories?", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 4–10.

Margulis, L 1998 *Symbiotic planet: A new look at evolution*, New York: Basic Books.

Grey Monroe, J, Srikant, T, Carbonell-Bejerano, P, *et al.* 2022 "Mutation bias reflects natural selection in Arabidopsis thaliana", *Nature*, vol. 602, pp. 101–105. https://doi.org/10.1038/s41586-021-04269-6

Montévil M, Mossio M, Pocheville A, & Longo G 2016 "Theoretical principles for biology: Variation", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 36–50.

Mossio M, Montévil M, & Longo G 2016 "Theoretical principles for biology: Organization", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 24–35.

Radick, G 2020 "Making sense of Mendelian genes", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 299–314.

Radick, G 2023 *Disputed inheritance: The battle over Mendel and the future of biology*, Chicago: University of Chicago Press.

Richardson, S 2013 *Sex itself: The search for male and female in the human genome*, Chicago: University of Chicago Press.

Richardson, S 2022 "Sex contextualism", *Philosophy, Theory, and Practice in Biology*, vol. 14, no. 2. https://doi.org/10.3998/ptpbio.2096

Risen, C 2023 "Evelyn Fox Keller, Who Turned a Feminist Lens on Science, Dies at 87", *The New York Times*, September 30, 2023. www.nytimes.com/2023/09/30/science/evelyn-fox-keller-dead.html

Riskin, J 2016 *The restless clock: A history of the centuries-long argument over what makes living things tick*, Chicago: University of Chicago Press.

Riskin, J 2020 "Biology's mistress, a brief history", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 268–298.

Riskin, J 2023 "Questions for Jablonka and Ginsburg drawn from Lamarck's life-made world", *Biosemiotics*, vol. 16, no. 1, pp. 27–34.

Sonnenschein C, & Soto, AM 2020 "Over a century of cancer research: Inconvenient truths and promising leads", *PLoS Biology*, vol. 18, no. 4, art. e3000670. doi: 10.1371/journal.pbio.3000670.

Soto, AM, Longo, G, Miquel, PA, Montévil, M, Mossio, M, Perret, N, Pocheville, A, & Sonnenschein C 2016 "Toward a theory of organisms: Three founding principles in search of a useful integration", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 77–82.

Soto, AM, Longo, G, Montévil, M, & Sonnenschein, C 2016 "The biological default state of cell proliferation with variation and motility, a fundamental principle for a theory of organisms", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 16–23.

Soto, AM, & Sonnenschein, C 2020 "Information, programme, signal: Dead metaphors that negate the agency of organisms", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 331–343.

Vicedo, M 2023a "Epistemological discipline in animal behavior studies: Konrad Lorenz and Daniel Lehrman on intuition and empathy", *History and Philosophy of the Life Sciences*, vol. 45, no. 1, art. 6. https://doi.org/10.1007/s40656-023-00558-7

Vicedo, M 2023b "Beyond the instinct debate: Daniel Lehrman's contributions to animal behavior studies", *Journal of the History of Biology*, vol. 56, pp. 251–284. doi: 10.1007/s10739-023-09706-y.

Vicedo, M, & Walsh, D (eds.) 2020a *Making sense of metaphor: Evelyn Fox Keller and commentators on language and science, Interdisciplinary Science Reviews*, vol. 45, no. 3. www.tandfonline.com/toc/yisr20/45/3

Vicedo, M, & Walsh, D 2020b "Entanglements between language, science, and worldviews", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 234–248.

Waddington, CH 1969 *Behind appearance: A study of the relations between painting and the natural sciences in this century*, Cambridge, MA: MIT Press.

Walsh, D 2015 *Organisms, agency, and evolution*, Cambridge: Cambridge University Press.

Walsh, D 2020 "Action, program, metaphor", *Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 344–359.

White, RM, Hodge, MJS, & Radick, G 2021 *Darwin's argument by analogy: From artificial to natural selection*, Cambridge: Cambridge University Press.

Winther, RG 2000 "Darwin on variation and heredity", *Journal of the History of Biology*, vol. 33, no. 3, pp. 425–455.

Winther, RG 2001 "August Weismann on germ-plasm variation", *Journal of the History of Biology*, vol. 34, no. 3, pp. 517–555.

Winther, RG 2019 "Mapping the deep blue oceans", in: Tambassi, T (ed.) *The Philosophy of GIS*, Cham: Springer, pp. 99–123. https://link.springer.com/chapter/10.1007/978-3-030-16829-2_5

Winther, RG 2020a *When maps become the world*, Chicago: University of Chicago Press.

Winther, RG 2020b "The structure of scientific theories", *Stanford Encyclopedia of Philosophy* [online]. Cited 11 November 2023. https://plato.stanford.edu/entries/structure-scientific-theories/

Winther, RG 2021 "Richard C. Lewontin as master dialectician (rest in power, Dick)", in: *science-forthepeople.org* New York. Cited 11 November 2023. https://magazine.scienceforthepeople.org/lewontin-special-issue/lewontin-as-master-dialectician

Winther, RG 2022 *Our genes: A philosophical perspective on human evolutionary genomics*, Cambridge: Cambridge University Press.

Winther, RG, & Raffn, M 2024 "What If?", in: Ambrosio, C, & Sánchez-Dorado, J (eds.) *Abstraction in science and art: Philosophical perspectives*, New York: Routledge.

Sapienza Università di Roma

Opinions

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

DOI: 10.13133/2532-5876/18307

New Genetically Modified Organisms (GMOs): Towards a "scientific precautionary principle"

Giuseppe Longo[a*]

[a] *Centre Cavaillès, CNRS and Ecole Normale Supérieure, Paris, France*

***Corresponding author:** Giuseppe Longo, Email: *giuseppe.longo@ens.fr*

Abstract

To the ordinary precautionary principle, we should add a more precise "scientific precautionary principle". In short, we cannot act on nature based on 'dogmas' that are either manifestly false or are implicitly adopting an uncritical way of thinking. Science is the invention of a new way of thinking, of new theoretical frameworks, starting from a critical review of the principles mobilized, which are themselves well explained. Without this, technoscience, in all its power, becomes a nightmare, as it is totally unsuited to make us live in an ecosystem with all its complexity. The case of New Genetic Technologies, whose application to agriculture is under discussion in Europe, is paradigmatic and urgent.

**A preliminary version of this paper was published in French in* Philosophy World Democracy*, 1/9/23, www.philosophy-world-democracy.org/articles-1/les-nouveaux-organismes-genetiquement-modifies-ogm-pour-un-principe-de-precaution-scientifique.*

Keywords: scientific precautionary principle, GMO, technoscience

Citation: Longo, G 2023, "New Genetically Modified Organisms (GMOs): Towards a 'scientific precautionary principle'", *Organisms: Journal of Biological Sciences*, vol. 6, no. 2, pp. 77–81. DOI: 10.13133/2532-5876/18307

In recent months, a relentless campaign, with the help of many lobbyists, is moving through the European Parliament and Commission the NGT (New Genetic Technologies) as an eligible variant to grow GMOs (Genetically Modified Organisms) in Europe. The *European Network of Scientists for Social and Environmental Responsibility* (ENSSER: https://ensser.org/, see also the commitment of the AAGT : https://generation-thunberg.org/accueil), with other non-governmental organizations, is conducting a difficult scientific and political battle against these new products.

The motivations for this new commercialization refer to the so-called «naturalness» of these powerful genomics techniques, called CRISPR-Cas9, which are based on an important scientific discovery, about twenty years ago, on how bacteria may affect the DNA of certain viruses. Now, it is one thing to identify the processes that take place in very complex evolutionary contexts, refined by a long biological history, but it is another thing to use them outside of well-confined laboratories. In these laboratories, CRISPR has been shown to be very useful for DNA and RNA analyses, which have allowed us to understand its great power as well as its limitations (see references below and in (Longo 2021)). Admittedly, these tools are anything except "very accurate". Already in the case of GMOs so far banned

in Europe, the pesticides to which they are resistant or the toxins they produce attack many symbionts, well beyond the target parasite, thus disrupting the humus, i.e., the living layer of soil essential to fertility. Indeed, these molecules act on almost everything that is alive, just with lower probabilities than on the target parasite. The same lack of precision and the impossibility of perfect "steering" of the plant in the ecosystem also concerns these NGTs. However, it is claimed that they can allow us to "perfectly control" the development and insertion into the ecosystem of the plants concerned. This conclusion is based on erroneous "scientific" dogmas—points 1 and 2 below—and without accepting a debate on the failures of existing GMOs, which are also grounded on the same dogmas, see (Kranthi & Stone 2020). For example, no mention is made of the side effects of "BT cotton" in India (Gutierrez, Herren, & Kenmore 2020), of the loss of diversity due to GMOs about maize diversity in Mexico and monocultures (Landry 2015; Rodríguez Mega 2018).

Faced with the abuse of these powerful but poorly understood techniques, presented within a nonsensical, dogmatic frame from the scientific point of view, it is necessary to lay down a "scientific precautionary principle", which should accompany and better specify the "precautionary principle" that is often mentioned. In short: one cannot act on nature on the basis of a "theoretical frame" or, in this case, of "dogmas", that are manifestly false and very often recognized as false even by their very promoters, usually in private (see below for an explicit, late acknowledgement). This behavior is a novelty in science, and it falls outside of any scientific ethics.

The application of GMOs and NGT we are talking about is based on two major dogmas of Molecular Biology that justify the application of the NGT in the ecosystems:

1 - the Central Dogma of Molecular Biology (synthetically: the "information" contained in DNA is "complete" as to the development and evolution of organisms (Crick 1970)—or even "development is entirely written in DNA" as in a computer program). Typically, any contribution of epigenetics to this information is excluded.

2 – the dogma that macromolecular interactions are "exact", (stereo-)specific, as they say, "key-lock correspondence" or "hand-glove"... as this would be "necessary to transmit genetic information", citing (Monod 1970). This makes the cell, even the organism, a "Cartesian mechanism" or "a Boolean algebra", according to the latter.

The second dogma is not less important or less obviously false than the first. For decades, physico-chemists have treated these interactions statistically—macromolecules have enormous oscillations, move in a Brownian stream and almost all their chemical affinities depend also on their context. A marginalized minority in biology has been defending this evidence since 1983 (see for an overview (Paldi 2020)), further developed by a more echoed article on this subject in 2002 (Elowitz *et al.* 2002). The two dogmas are at the basis of a mechanistic (Cartesian) vision of the living, as particularly emphasized by Francis Bacon (1561–1626, cited since the 1930s by the promoters of genetic engineering): in this perspective animals and plants must be considered and treated as machines... We may reprogram them at leisure by "editing" the "selfish genes" that completely encode them, claim the promoters of geno-centrism still today (Dawkins 2016). For them, everything is information, encoded in the genes, modifiable at leisure—and it may be "edited", exactly, like an alphabetic text, letter by letter.

It is amazing to hear, from private conversation or secondary publications, the proponents of these dogmas recognize that they are false. E. Fox-Keller closely analyzed this phenomenon. In particular, she quoted Philip Ball, a "former editor" of the journal *Nature* who recognized "the misleading nature" of these dogmas, despite their use in any popularization and in most academic textbooks. These "'misleading' narratives are routinely perpetuated in the teaching of Molecular Biology, indeed in so much of the technical, the lay, and even the philosophical literature", wrote Ball, quoted in (Fox-Keller 2020), who also offered a historical perspective in (Fox-Keller 2003) (for theoretical alternatives to geno-centrism, see (Soto *et al.* 2016)). Indeed, these narratives are at the core of all kinds of promises in genetic technologies and... of the sale of shares on the stock market of the start-ups that work on them.

Based on these dogmas, it can be stated that we have the "power to control evolution", according to the title (and the content) of the 2017 book by J. Doudna (J. Doudna and E. Charpentier, were awarded the 2020 Nobel Prize for the remarkable technique they developed). CRISPR-Cas9, she writes, may reprogram the genome by acting on the DNA "exactly", by "editing"

it, "as with scissors"... while in laboratories these same authors act on large numbers of cells, choosing the cells where the process has worked (cherry-picking) (Bock *et al.* 2022). Gene knockouts, which has been practiced for decades, may not work (Smits *et al.* 2019); CRISPR-Cas9 modifications designed to suppress gene function may fail, and damaged genes may continue to produce proteins (many of which are still functional), just as there are collateral and/or unpredictable effects (Burgio & Teboul 2020), as well as editing-resistance (Mehta *et al.* 2019). The process instability is particularly evident when CRISPR-Cas9 is applied to animal models (Papathanasiou *et al.* 2021). It is therefore quite possible that after a very large number of transgenic manipulations and experiments of many different techniques, the few temporary successes in implantation in the fields of existing GMOs are *less due to the relevance of the genetic manipulations than they are to the great resilience of living organisms*. But this resilience has limits: the transformation of humus into sand in a few years is one of the most serious consequences of existing techniques (Bizzarri 2012)—but not the only one (see the case of Teosinte: uncontrolled diffusion in the fields of this wild maize, inedible, would correlate with an "adaptive crop-to-wild introgression of transgenic maize"—the "noxious weed" (Le Corre *et al.* 2020)).

The book by J. Doudna is a paradigm of the geno-centered approach, based on the two dogmas cited above (the first explicitly, the second implicitly) and on the marketing of NGTs, rich in promises without criticism, without any reflection on the limits and failures of existing GMOs. Application of these old techniques should have solved the problem of hunger in the world (as it was said in 2000), and similarly we should be able to do so today by using NGTs, while adjusting life to the changing ecosystem. This is claimed with no reference to the limits of these new techniques, which are the result of an immensely complex technicality that intervenes on the living on the basis of the same dogmatic imaginary as the old GMOs. Science, on the contrary, is the invention of a new way of thinking from a critical perspective of the principles mobilized, themselves well (and honestly) displayed. Without this, techno-science, in all its power, becomes a "nightmare", like the one we are experiencing as a result of the limitless extractivist engineering techniques that have changed the climate. I am referring here to the role of fossil fuel extraction and its transformation through innovative and very powerful techniques and their a-critical use, for more than a century, without a 'theoretical' unified thinking of the Earth and its atmosphere (Longo 2023).

The life sciences can and must use these NGT in laboratories, including this new and formidable CRISPR-Cas9 technique, and perform genetic manipulations in well-isolated bio-reactors (with enormous vigilance against possible leaks). The production of insulin by genetically modified bacteria is *the* great success of a now mature, 50-years old technique. Insulin, an inert product, is then released from the bio-reactors. Conversely, the insertion of organisms resulting from genetic manipulations into the complexity of ecosystems is a serious error. Both the set of all induced mutations on plants and the side effects on the context, such as the humus, are a priori unpredictable, like the effects of traditional GMOs. More generally, the networks of changing interactions that characterizes the living is anything but a system on which one can think of acting as with a "Swiss army knife". These methods have nothing to do with the patient co-evolution of top-down human techniques (grafts, hybridizations, etc.). Of course, even by these traditional techniques we can do damage: when we create huge monocultures of perfect apples, all identical to Snow White's apple, we have lost the scientific sense of the role of diversity in the resilience and, thus, evolution of the living.

To summarize, these techniques of genetic engineering are without scientific support and are not adapted to help us live in an ecosystem, which we must also or first understand. And we also should acknowledge the scientific limits of these powerful techniques, such as the following: a false or incomplete theoretical framework; often unattainable genetic targets; off-target effects; previous failures in other forms of genetic manipulation, and finally, the inherent unpredictability of many phenotypic and ecosystem consequences—for a review and references, see (Longo 2021).

In this context, accepting GMOs, based on these NGTs, which "do not produce more than 20 mutations" (as proposed in the new European regulation (Nature Plants Editorial Board 2023; ENSSER 2023) is a nonsense: in no case we can predict the exact nature and number of mutations that will be induced by these

techniques, even less their phenotypic and ecosystem consequences.

The argument of the "20 mutations" is based on the observation that a larger number of mutations is very unlikely to be produced by evolutionary chance (Nature Plants Editorial Board 2023). This argument does not imply that the induced mutations, below 20, would be "natural" (the flows in Logic and the abuse of the "differential method" were the first observation of this author, a mathematician, when reading texts of "dogmatic" molecular biology (Longo & Tendero 2007)); it only makes it more difficult to trace artificially induced mutations, against any obligation of transparency. Further, we have just come out of a pandemic where a single mutation, N439K, in SARS-CoV-2, has profoundly modified, and in a largely unpredictable way, the pathological effects of the virus, since it "enhances the binding affinity for the ACE2 receptor and reduces the neutralizing activity of some monoclonal antibodies (mAbs) and polyclonal antibodies present in sera from people who have recovered from infection" (Harvey *et al.* 2021).

Acting on the environment on these bases is equivalent to entrusting to the 11th century great astronomers with missiles capable of reaching Mars. These astronomers were remarkable observers and mathematicians, but they were working within the Ptolemaic, geo-centered, theoretical frame. Thus, they were wasting a lot of time in drawing epicycles, with little predictive effectiveness, while working in... Astrology, that is at making *promises* and *predictions* (Longo & Mossio 2020)— not dissimilar to the 2003 promise to wipe cancer off the face of the Earth by 2015 through gene therapies as claimed by (von Eschenbach 2003), then president of the National Cancer Institute. Not only those missiles would never have reached Mars, but they would have fallen on a nearby city or exploded for excessive acceleration because their preparation would not have taken into account the rotation of the Earth. In addition to asking for caution (the traditional "precautionary principle"), we must insist on calling attention on the false theoretical framework of the old and new genetic technologies mentioned above and the duty of *scientific precaution* not to implement them in the Earth's ecosystem. This is done in some debates, but far too rarely (for some documents on the ongoing battle at European level, in which ENSSER is participating, see: https://ensser.org and www.di.ens.fr/users/longo/files/NGT-public-linksJuly5-2023.zip).

References

Bizzarri M 2012 *The new alchemist: The risks of genetic modification*, Boston: WIT Press.

Bock, C, *et al.* 2022 "High-content CRISPR screening", *Nature Reviews Methods Primers*, vol. 2, pp. 1–23. https://doi.org/10.1038/s43586-021-00093-4

Burgio, G, & Teboul, L 2020 "Anticipating and identifying collateral damage in genome editing", *Trends in Genetics*, vol. 36, pp. 905–914. https://doi.org/10.1016/j.tig.2020.09.011

Crick, F 1970 "Central dogma of molecular biology", *Nature*, vol. 227, pp. 561–563. https://doi.org/10.1038/227561a0

Dawkins, R 2016 *The extended phenotype: The long reach of the gene*, Oxford: Oxford University Press.

Elowitz, MB, *et al.* 2002 "Stochastic gene expression in a single cell", *Science*, vol. 297, pp. 1183–1186. https://doi.org/10.1126/science.1070919

von Eschenbach, AC 2003 "NCI sets goal of eliminating suffering and death due to cancer by 2015", *Journal of the National Medical Association*, vol. 95, no. 7, pp. 637–639. https://pubmed.ncbi.nlm.nih.gov/12911262

ENSSER (European Network of Scientists for Social and Environmental Responsibility) 2023 "Analysis statement by ENSSER on the EU Commission's new GM proposal. Here for Annex 1 on NGT 'equivalence criteria'", in: *ensser.org* Berlin. Cited 11 November 2023. https://ensser.org/press_release/analysis-statement-by-ensser-on-the-eu-commissions-new-gm-proposal-here-for-annex-1-on-ngt-equivalence-criteria/

Fox-Keller, E 2003 *The century of the gene*, Cambridge: Harvard University Press.

Fox-Keller, E 2020 "Cognitive functions of metaphor in the natural sciences", in: Vicedo, M & Walsh, DM (eds.) *Making sense of metaphor: Evelyn Fox Keller and commentators on language and science, Interdisciplinary Science Reviews*, vol. 45, no. 3, pp. 249–267.

Gutierrez, AP, Herren, HR, & Kenmore, PE 2020 "The failure of GMO cotton In India", in: *resilience.org* Corvallis, USA. Cited 11 November 2023. www.resilience.org/stories/2020-09-09/the-failure-of-gmo-cotton-in-india

Harvey, WT, *et al.* 2021 "SARS-CoV-2 variants, spike mutations and immune escape", *Nature Reviews Microbiology*, vol. 19, pp. 409–424. https://doi.org/10.1038/s41579-021-00573-0

Kranthi, KR, & Stone, GD 2020 "Long-term impacts of Bt cotton in India", *Nature Plants*, vol. 6, pp. 188–196. https://doi.org/10.1038/s41477-020-0615-5

Landry, H 2015 "Challenging evolution: How GMOs can influence genetic diversity", in: *sitn.hms.harvard.edu* Cambridge, USA. Cited 11 November 2023. https://sitn.hms.harvard.edu/flash/2015/challenging-evolution-how-gmos-can-influence-genetic-diversity

Le Corre, V, *et al.* 2020 "Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 25618–25627. https://doi.org/10.1073/pnas.2006633117

Longo, G 2021 "Programming evolution: A crack in science. A review of the book by Jennifer A. Doudna and Samuel H. Sternberg 'A crack in creation: Gene editing and the unthinkable power to control evolution'", *Organisms. Journal of Biological Sciences*, vol. 5, no. 1, pp. 5–16.

Longo, G 2023 *Le cauchemar de Prométhée: les sciences et leurs limites*, first edition, Paris: Presses Universitaires de France.

Longo, G, & Mossio, M 2020 "Geocentrism vs genocentrism: Theories without metaphors, metaphors without theories", *Interdisciplinary Science Reviews*, vol. 45, pp. 380–405. https://doi.org/10.1080/03080188.2020.1798588

Longo, G, & Tendero, PE 2007 "The differential method and the causal incompleteness of programming theory in molecular biology", *Foundations of Science*, vol. 12, pp. 337–366. https://doi.org/10.1007/s10699-007-9111-x

Mehta, D, *et al.* 2019 "Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses", *Genome Biology*, vol. 20, art. 80. https://doi.org/10.1186/s13059-019-1678-3

Monod, J 1970 *Le Hasard et la Nécessité: Essai sur la philosophie naturelle de la biologie moderne*. Paris: Éditions du Seuil.

Nature Plants Editorial Board 2023 "EU rethinks genome editing", *Nature Plants*, vol. 9, pp. 1169–1170. https://doi.org/10.1038/s41477-023-01505-x

Paldi, A 2020 "Stochastic or deterministic? That is the question", *Organisms. Journal of Biological Sciences*, vol. 4, pp. 77–79. https://doi.org/10.13133/2532-5876/16963

Papathanasiou, S, *et al.* 2021 "Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing", *Nature Communications*, vol. 12, art. 5855. https://doi.org/10.1038/s41467-021-26097-y

Rodríguez Mega, E 2018 "Small farmers in Mexico keep corn's genetic diversity alive", in: *scientificamerican.com* New York, USA. Cited 11 November 2023. www.scientificamerican.com/article/small-farmers-in-mexico-keep-corns-genetic-diversity-alive

Smits, AH, *et al.* 2019 "Biological plasticity rescues target activity in CRISPR knock outs", *Nature Methods*, vol. 16, pp. 1087–1093. https://doi.org/10.1038/s41592-019-0614-5

Soto, AM, Longo, G, & Noble, D 2016 "From the century of the genome to the century of the organism: New theoretical approaches", *Progress in Biophysics and Molecular Biology*, vol. 122, pp. 1–3. https://doi.org/10.1016/j.pbiomolbio.2016.09.011

Book Review

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

DOI: 10.13133/2532-5876/18310

Process-Philosophical Perspectives on Biology: Intuiting Life (edited by Spyridon A. Koutroufinis and Arthur Araujo)—Philosophical Intuitions for a New Understanding of Life

Mariano Bizzarri[a*]

[a] *Head of the Systems Biology Group Laboratory, Department of Experimental Medicine, Sapienza University of Rome, Italy*

***Corresponding author:** Mariano Bizzarri, Email: *mariano.bizzarri@uniroma1.it*

Citation: Bizzarri, M 2023, "Process-philosophical perspectives on biology: Intuiting life (edited by Spyridon A. Koutroufinis and Arthur Araujo)—Philosophical intuitions for a new understanding of life" [book review], *Organisms: Journal of Biological Sciences*, vol. 6, no. 2, pp. 83–84. DOI: 10.13133/2532-5876/18310

The ontologies of the seminal 20th century philosophers, Alfred North Whitehead and Henri Bergson criticized the intellectual attitudes that dominate modern life sciences. These attitudes particularly influence one of the core problems of contemporary biology and philosophy of biology, i.e. the nature of explanation. In 20th century philosophy of science, Carl Hempel's theory of explanation was the backbone of theorizing about scientific explanation for decades. Hempel differentiates between the deductive-nomological and the inductive-statistical types of explanation but both types have the same logical structure. In philosophy of biology there is a broad consensus that the explanatory relevance of modelling in contemporary biology—especially in mathematically operating systems- and theoretical biology—cannot be captured by Hempel's account. Some influential philosophers of biology agree that in life sciences both kinds of laws—deductive-nomological and statistical generalizations—do not explain, but rather characterize phenomena. As life scientists commonly seek to uncover the 'mechanism' responsible for the phenomenon of interest, in life sciences explanations are based on mechanisms. Leading philosophers of science who advocate a school of thought, which is often described as "New Mechanical Philosophy" or "New Mechanism" argue that in many fields of science what is considered a satisfactory explanation requires providing the description of a mechanism. Indeed, mechanistic explanations form the main theoretical basis of most, if not all, contemporary biological disciplines, and life science practice can be understood in terms of the discovery and description of mechanisms.

The neo-mechanistic school in biology is a specific manifestation of what Whitehead called "scientific materialism"—a metaphysics that emerged from the spirit of the late 19th century. As a result, many life scientists implicitly assume an outdated reductionist metaphysics that does not do justice to the complexity of biological phenomena and leaves many features of living processes unexplained. Scientific materialism—and its latest expression, the "New Mechanism"—can also be seen as a typical product of the technological intellect, which, despite the limitations that Bergson warned about in his works, reduces reality in order to manipulate it.

Some modern criticisms of the biological relevance of mechanistic explanations echo Bergson's warning in *Creative Evolution* that "[t]he [abstract] intellect is characterized by a natural incomprehension of life" (addition by S.K.). Based on this, the new volume *Process-Philosophical Perspectives on Biology: Intuiting Life*, published by Cambridge Scholar Publishing (UK) in 2023, reflects the belief that intuition must assist the life-studying intellect, for only intuition can do justice to those aspects of life, which, for fundamental reasons, transcend the discursive-analytic modes of thought. Intuitive knowledge is not the only conceivable response to neo-mechanistic thinking, but it is certainly one that takes into account essential facts that neo-mechanism simply ignores. The authors of this new volume are convinced that philosophy, and in particular process philosophy, must breathe new life into what has been suppressed by scientific reductionism. Serving this purpose, the present volume is committed to the following maxim: Starting from philosophical intuitions, biophilosophy must unveil any abstractions of biology and overcome them with new metaphysical hypotheses.

This book challenges the reductionist and materialistic metaphysics often adopted by biologists, arguing that it overlooks the intricate complexities and essential characteristics of life. The authors explore the viability of process metaphysics to advance our understanding of fundamental biological concepts such as organism, ontogeny, agency, teleology, environment, and normativity. Based on the metaphysics of Whitehead and other process thinkers, e.g. Bergson, who attribute subjectivity, value, and purposeful striving to all organisms, they ascribe subjective interiority to all living beings, from unicellular organisms to the most complex animals. In doing so, they highlight the uniqueness and intrinsic value of living beings. The book presents a new approach to essential dimensions of the phenomenon of life with the aim of opening new horizons in the thinking of philosophers, philosophers of biology, life scientists, and environmentalists.

The book contains the following chapters:

The easiest way to order the book is directly from the publisher at www.cambridgescholars.com/product/978-1-5275-0450-9

Organisms

ISSN: 2532-5876