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.