Draft. Comments are welcome.

MEASURING CELLS – LABORATORY WORK AND BIOINFORMATICS

Emanuele Ratti[1]

Department of Philosophy, University of Bristol

Abstract. The goal of this article is to construct an account of measurement in molecular biology, with an emphasis on bioinformatics practices. The reasons for constructing this account are two. First, I fill a lacuna in philosophy of biology and philosophy of science, where measuring practices in bioinformatics and molecular biology have been neglected. Second, I argue against a popular idea in molecular biology, according to which experimentalists are in a better position to characterize biological phenomena than bioinformaticians, because of their material access to experimental systems. By arguing that bioinformaticians can measure things that experimentalists cannot, I show that this claim is unwarranted.

1. INTRODUCTION

The goal of this article is to provide a comprehensive characterization of measurement practices in molecular biology and, in particular, its bioinformatics[2] side. Here, molecular biology overlaps with Morange's macromolecular biology (2008), which includes disciplines stemming from the molecular vision, such as systems biology, the various -omics, etc. This paper will show what does it mean for bioinformaticians to measure, what kinds of measurement they construct, and how their measurements relate to the measuring activities of so-called experimentalists or wet-lab biologists[3] (Strasser 2017).

The reasons for characterizing these measurement practices are mainly two. First, in philosophy of biology there has not been enough attention to the epistemic practices of bioinformatics. Attention to bioinformatics has been directed towards the processes of curating and processing data (Leonelli 2016; Strasser 2017), and to how the 'data-intensive' turn can be connected to the mechanistic *ethos* of molecular biology (Lopez-Rubio and Ratti 2021; Bechtel 2020). But the discourse on data has not been properly connected to other bioinformatics

[1] mnl.ratti@gmail.com

[2] We follow Ratti and D'Agostino's usage of the term(2025), which covers also 'computational biology'

[3] The term 'experimentalist' refers to biologists working in laboratories – so-called 'wet-lab' biologists. I will use 'wet-lab biologist' and 'experimentalist' interchangeably.

practices which emerge in the measuring procedures to which bioinformaticians contribute. But even more important, the literature in philosophy of science is, to our knowledge, surprisingly silent on the measuring practices in molecular biology, a discipline which, arguably, is particularly obsessed with measurements. This article can fill these lacunae in both fields.

The second reason lies in debates internal to molecular biology. A number of articles published a few years ago (Bartlett et al 2017; Bartlett et al 2016; Lewis et al 2016; Lewis and Bartlett 2013) have characterized the relation between bioinformaticians and 'wet-lab' biologists. Through surveys (more than 300 participants) and ethnographic investigations (almost 100 interviews), they have provided extensive evidence of the hardships that bioinformaticians or computational biologists are subjected to in biology. In particular, they have shown that a prevalent attitude is that "bioinformatics should remain in a symbiotic, subsidiary relationship to biology" (Lewis and Bartlett 2013, p 249), and that bioinformaticians deliver "neither good biology, nor good computer science, but, rather, (…) a service provider to biology" (Bartlett et al 2016a, p 188). When bioinformatics' work has a more obvious biological connotation (e.g. data curation and data analysis), bioinformatics is "treated as infrastructural support" (Lewis et al 2016, p 479). The consequences of these views are nefarious: bioinformaticians are often perceived "*institutionally* peripheral" (Bartlett et al 2018, p 5), and "the legitimacy of their entire research programmes (…) are being called into questions" (Bartlett et al 2016b, p 3). These studies have been corroborated by recent insights (Markowetz 2017; Grabowski and Rappsilber 2019; Way et al 2021). What is important for this article is that this view of bioinformatics is a consequence of specific epistemic ideas. In particular, it is often drawn a contrast between laboratory work and in-silico work. What is argued against bioinformaticians is that they "have next to no understanding of the biological significance of their findings, never mind the laboratory processes that produce the data" (Bartlett et al 2016a, p 201). This is sometimes connected to the 'materiality' culture of wet-lab biology, and the nature of concrete 'object-processing' characterizing disciplines such as molecular biology (Knorr Cetina 1999). It is said that "bioinformaticians do not perform experiments, at least not in the way that biologists do" (Lewis and Bartlett 2013, p 249), and this seems to imply that wet-lab biologists benefits from a sort of epistemic priority over bioinformaticians. The studies by Bartlett and colleagues describe the idea of 'not doing proper experiments' as the view that bioinformaticians do not have the right sort of access to biological phenomena that biologists have. By adapting the notion of 'inscription' coined by Latour and

Woolgar (1979), they formulate a distinction that, they think, plays a pivotal role in the *ethos* of contemporary molecular biology. Wet-lab biologists think that they have an epistemic priority over bioinformaticians because they are the ones first translating "the matter of life into data (producing 'primary inscriptions'), after which bioinformaticians, working in the dry-lab, carry out further transformations (producing 'secondary inscriptions')" (Bartlett et al 2017, p 3). One reason to write this article is to argue against the view that experiments and primary inscriptions provide a privileged access (called 'epistemic priority') to biological phenomena. I will argue against this idea by constructing an account of measurement that will show that this idea of epistemic priority is simply misleading.

The structure of the article is as follows. First, I formulate more precisely the idea of epistemic priority of experimentalists (Section 2). I call this the 'epistemic priority account' (EPA), and I distinguish two important components (i.e., EPA-a and EPA-b). In Section 3, I construct an account of measurement in molecular biology that applies to how both wet-lab biologists and bioinformaticians measure, and their joint and collaborative practices. With this account of measurement, and by comparing experimentalists' and bioinformaticians' contributions to measurement processes, I address the views on epistemic priority, by showing that EPA-b rarely applies (Section 4), and that EPA-a is false (Section 5). Analyses of EPA-b and EPA-a will be carried out through a detailed engagement with classic measurement activities in biology, such as sequencing, flow cytometry, and gene set enrichment. An upshot of these analyses (Section 6) is, first, that the hierarchical divide between bioinformaticians and wet-lab biologists is, epistemically, unmotivated, and that one might even conclude that bioinformatics practices provide a richer access to biological phenomena. Second, my analysis also suggests show bioinformatics measurement practices provide a vantage point for theorizing.

2 DEFINING EPISTEMIC PRIORITY

In their studies (see Introduction), Bartlett and colleagues conceptualize the epistemic divide between experimentalists and bioinformaticians in two ways. First, a claim is made that experimentalists perform material experiments while bioinformaticians do not, and this confers them a priority on the basis of a material access to biological phenomena that bioinformaticians have not. Second, the distinction between primary and secondary inscriptions is an alternative formulation of the divide. Tending to primary inscriptions confers epistemic priority to wet-lab biologists over bioinformaticians. This is another recurring theme: the importance of who materially generates data is central. This ‘priority’ has consequences for who claims ‘discoveries’. It is worth noticing that Bartlett and colleagues are agnostic with respect to the validity of the consequences drawn from the distinction. In fact, for them the distinction is primarily a lens to conceptualize the epistemic underpinnings of the predicaments of bioinformaticians.

These observations about ‘experiments’ and ‘inscriptions’ can be summarized by saying that, according to the current *ethos*, experimentalists have ‘epistemic priority’ with respect to bioinformaticians. Let us now characterize this idea more in detail. Here we take the priority of experimentalists lying in the fact that, in virtue of doing experiments and constructing primary inscriptions, they have access to more information about biological phenomena. The access to more information is a function, under this conception, of the proximity of biologists to the actual biological material. Because of this proximity, experimentalists modulate the information ‘extracted’ from biological materials: they are gatekeepers, and they constrain and shape knowledge claims. I do not use any specific account of information, but the idea is that, because of their material access to phenomena, experimentalists have a vantage point for not only identifying relevant properties of biological phenomena, but also for determining what kind of properties can be identified in the first place. I call this “the epistemic priority account’ (EPA), defined as follows:

> EPA = experimentalists have a vantage point over knowledge claims that can be made about biological phenomena because, in virtue of their material access and proximity to biological phenomena through experiments, they either (a) establish which kinds of properties of the phenomenon can be identified, or (b) concretely identify such properties.

Please note the ‘either/or’. What we mean is that at least one of the two conditions must apply for EPA to be justified. EPA characterizes the contested battlegrounds emerging from Bartlett and colleagues’ studies, and it reflects my own experience in talking to bioinformaticians. In what follows, I will put to test both *a* (EPA-a) and *b* (EPA-b).

3 MEASUREMENTS IN BIOLOGY AND BIOINFORMATICS

In order to see if EPA is justified, I propose to look at measurement practices in contemporary molecular biology. There are two advantages to use this angle to address EPA over Bartlett and colleagues’ approaches based on experiments and inscriptions.

One advantage of focusing on measurements is that it is much more neutral than discussions about experiments. What counts as an experiment is a highly contentious issue. There are indeed a number of different types of experiments, from confirmatory to exploratory (Radder 2003), and it is not clear how to count them. In fact, some think (Parker 2009) that we should talk about experimental activities rather than experiments. But even if we select a general class of experiment or experimental activity, there are still problems. One way to address EPA would be, for instance, to argue that bioinformaticians indeed perform some form of experiment and then, by comparing wet-lab experiments and bioinformatics experiments, see if proximity is indeed a function of 'more information'. Ratti and D'Agostino explore this path (2025). But because there are strong and polarized opinions about what counts as an experiment, one can always arbitrarily restrict the level of proximity required for something to count as an experimental activity. In the case of measurements, there are indeed competing accounts, but they share a number of important components, as I will show.

Next, unlike the distinction between primary and secondary inscriptions, 'measurement' does not assume, *a priori*, that proximity to the material substance gives advantage to wet-lab biologists. It leaves open this question for scrutiny, which is what I want to do here. But it nonetheless assumes that an interaction is necessary, since measuring requires one between an apparatus and an object of study.

By using this more neutral lens, we can actually compare what biologists and bioinformaticians do, and whether the measurements they construct justifies EPA. In this section, I provide a characterization of measurement practices in molecular biology in general, which will apply to both experimental and bioinformatic activities. With this account in hand, in Sections 4 and 5 I will see what is the contribution of wet-lab biologists and bioinformaticians to specific cases of measurement practices, and whether these contributions motivate EPA-a and EPA-b.

3.1 A Basic Account of Measurement in Molecular Biology

In recent years, there has been an increasing attention to methodological and epistemological issues related to measurements in science (Tal 2020). There are a number of competing accounts defining the central characteristics of the measurement process and what does it mean to measure something. In what follows, I rely on Parker's account of measurement (2017), which is an integration of two recent accounts, one information-theoretic from van Fraassen (2008), and one model-based developed by Tal (2017). The starting point is her distinction (similar to others in the literature) between instrument indication, instrument reading, and measurement outcome.

‘Instrument indication’ refers to the physical state of an apparatus used to measure (which may be indicated in a quantitative way like a numerical readout, or non-quantitative way like a colour), while ‘instrument reading’ is the indication ‘read’ according to the conventions of the apparatus. For the sake of simplicity, here I use ‘instrument reading’ to cover both readings and indications. While the distinction is certainly important, it is not particularly relevant here. I also use the term ‘apparatus’ to refer to one or more instruments, such that an apparatus can be one instrument, or a particular setting where one or more instruments are used and coordinated. The use of this term might differ from others in the literature (see, e.g., Harre 2010).

A measurement outcome is the actual ‘state’ assigned to the measured object, which is inferred from one or more instrument readings. Bokulich defines a measurement outcome as “a knowledge claim that attributes a particular value of a variable of a property to the object or event being measured” (2020, p 429). ‘Measurement outcome’ is conceptualized by Parker in informational terms, meaning that the measurement outcome is *information about the measured object inferred from instrument readings*. I take Parker’s and Bokulich’s formulations to be equivalent: the information inferred from instrument indication can be expressed as a knowledge claim about a property of the object measured. The process of inferring an outcome from a reading is called ‘calibration’ (Bokulich 2020, p 429). While Parker does not use any specific account of information (I will not do either), ‘informative’ is connected especially to models and representations. This emerges clearly if we look at what Parker takes from the two other accounts mentioned above.

In van Fraassen’s account, measurement is an activity based on a physical interaction between an instrument and an object, such that agents setting up the interaction will gather information about the state of object itself (before the interaction). In his account, a measurement outcome is “a representation of what is measured[4]” (2008, p 179). This means that measuring provides a representation of an entity, in such a way that some physical parameters that characterize the measured object are displayed. It is an information-gathering activity, where the information is expressed as a selective representation of the object. Central

[4] To be fair, Van Fraassen’s account of measurement outcome is much more complex than this, given it encapsulates six specific characteristics. However, for this article there is no need to go much in depth about it

to Van Fraassen's account is the idea of a logical space. Measuring activities are about a measured object, where this is an item that is already classified in the domain of a given theory[5]. In his view, when an object is measured, it is located within "a space common to a whole family of models provided by the theory" (p 164). This 'space' is the logical space. Because this logical space is constructed within a given theory, then the representation constructed as an act of locating will necessarily incorporate a number of auxiliary assumptions, calculations, and theoretical and modeling inputs: *via* the representation, the measured object is located "in a certain logical space, with a location that it does not have *a priori*" (2008, p 177). In a simplified case, measuring bodies of gas is the practice of locating the object measured at the intersection of three dimensions, consisting of volume, temperature, and pressure. Tal proposes measurement as a form of model-based inference, where the model is an abstract and idealized representation of the measurement process (i.e., of how an apparatus provides information to establish the value of a parameter), and measuring consists in "inferences from the final state(s) of a physical process to value(s) of a parameter in the model" (2012, p 17). Because the model of the measurement process is idealized, steps to correct deviations from ideal conditions (typical of 'idealizations') must be specified. This happens especially in so-called 'white-box calibration', where the total uncertainty must be clearly specified to properly calibrate the instrument, while in 'black-box calibration' this estimation is already accounted for in the instrument reading. Despite formal differences (see Parker 2017, p 278), these two accounts have a lot in common, Parker argues. In both cases measurement is relative to models or theories, and measurement outcomes are selective representations of some sort.

Parker concludes that measuring is an empirical information-gathering quest, where through the physical interaction of an apparatus with an object, one infers characteristics of the measured object (i.e. a measurement outcome) by locating the object in a logical space, structured according to a given theoretical background. By locating the measured object in a logical space, measuring activities lead to the construction of a selective representation of the object, where the representation itself coheres with relevant background theory and other auxiliary assumptions about the interfering factors, characteristics of instruments, etc.

My account of measurement in biology has a lot in common with Parker's. For instance, I adopt the general idea that measuring is an empirical information-gathering activity based on

[5] "A Claim of the form 'This is an X-measurement of quantity M pertaining to S' makes sense only in a context where the object measured is already classified as a system characterized by quantity M. To so describe an object is already to classify by theory" (2008, p 144)

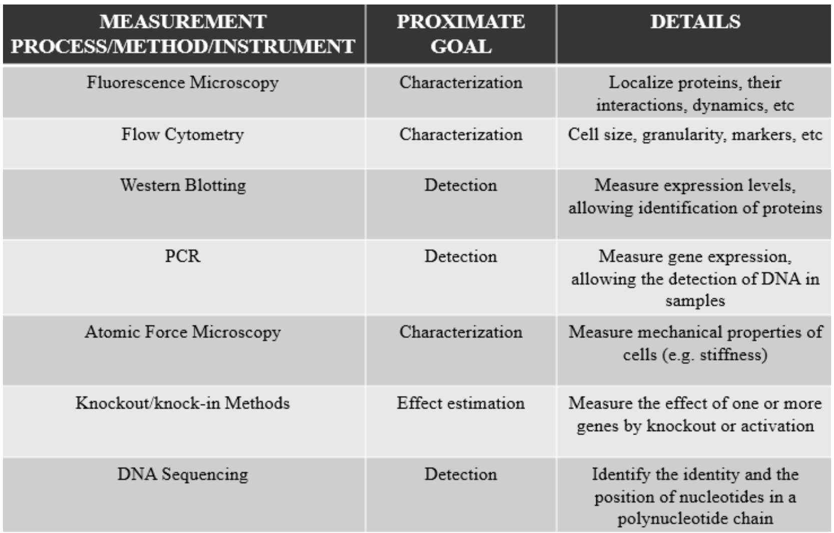
the interaction between an apparatus and an object of interest, and that locates an object in a logical space. However, there are some caveats that will lead my account of measurement in biology in a slightly different direction.

First, the nature of measurement outcomes is broader in biology than it is in the cases cited by Parker. In fact, measurement outcomes can be quantitative or qualitative. While cases discussed by Tal, van Fraassen, and Parker are quantitative[6], in molecular biology you have a bit of everything, from quantitative (such as qPCR to measure the amount of DNA in a sample), semi-quantitative (e.g. agarose gel electrophoresis to determine the presence of a DNA sample in a standard PCR), qualitative (e.g. cell morphology characteristics measured through electron microscopy).

Second, Parker's taxonomy of measurement outcomes does not fit well the peculiarities of molecular biology. Parker distinguishes different types of measurement on the basis of the inferences required to go from instrument readings to measurement outcomes. In a 'direct' measurement, there is no need to "transform the raw instrument reading into a value for a different parameter" (2017, p 28), and the instrument reading is the outcome, while in 'derived' measurement at least one additional layer of inference is required to calculate (using "reliable principles or definitions" p 281) the outcome. Finally, 'complex' measurement is when measurements outcomes are derived by integrating different direct/derived measurements. In my understanding, the distinction is about how sophisticated the process of calibration is. While this taxonomy has its merits, and in fact coheres well with my emphasis on who (between bioinformaticians and experimentalists) is in charge of transforming readings into an outcome, in molecular biology it is difficult to have such clear cut distinctions on the basis described by Parker. For instance, it is difficult to identify cases of direct measurements given that, even basic measurements, will require transformations of materials and representations. For this reason, I distinguish measurement outcomes in molecular biology on the basis of the proximate goal that they serve. There are various proximate goals, such as detection; effect estimation; and characterization. Correspondingly, detection measurements 'record' the presence or absence of a given biological entity or activity; effect estimation measurements gather information about how the presence or absence of an entity correlates with the state of a system; 'characterization' measurements' goal is to integrate different kinds of data modalities on an entity or process. Examples of classic measurement types can be found in Table 1. In this

[6] To be fair, Tal explicitly acknowledges the possibility of 'qualitative measurements' (2017, p 34), but he does not discuss them

article, I discuss measurement especially for detecting and characterizing, but my arguments apply more broadly.



MEASUREMENT PROCESS/METHOD/INSTRUMENT	PROXIMATE GOAL	DETAILS
Fluorescence Microscopy	Characterization	Localize proteins, their interactions, dynamics, etc
Flow Cytometry	Characterization	Cell size, granularity, markers, etc
Western Blotting	Detection	Measure expression levels, allowing identification of proteins
PCR	Detection	Measure gene expression, allowing the detection of DNA in samples
Atomic Force Microscopy	Characterization	Measure mechanical properties of cells (e.g. stiffness)
Knockout/knock-in Methods	Effect estimation	Measure the effect of one or more genes by knockout or activation
DNA Sequencing	Detection	Identify the identity and the position of nucleotides in a polynucleotide chain

Table 1. Different kinds of measurements in biology

Third, Parker stresses the role of ‘inference’ for constructing measurement outcomes. But this can sometimes be misleading in molecular biology. It is certainly true that readings are ‘interpreted’ in light of, e.g., background domain knowledge and integrated with other instrument indications to ‘infer’ the outcome (i.e., calibration). However, restricting calibration to just inferring is too narrow, because it underestimates the materiality of the processes of constructing measurement outcomes from readings. Or better: calibration is not just an inferential process; it is also a construction process. Van Fraassen points out, *en passant*, something along these lines, when he says that measurement activities can be *destructive* (e.g. a photon absorbed, a metal sample vaporized), with the consequence that the measurement outcome does not necessarily reveal the final state of the object, but rather its state *before* the interaction. In molecular biology, the majority (if not all) measuring activities are of this kind. Maybe the term ‘destructive’ is too strong, but certainly radical modifications and manipulations of the object measured are the rule, rather than the exception. This is to say that one might talk about calibration in a broader sense as ‘transformations in the pipeline’, rather than ‘inference’, because ‘interpreting’ and ‘resolving uncertainties’ require material transformations. As reported by Stevens (2013), in biology the word ‘pipeline’ is used to refer to “the series of processes applied to an object in order to render it into some appropriate final format” (p 109). ‘Rendering’ means exactly ‘selectively representing’, and ‘the final format’ is the location of the object in a specific logical space. But this pipeline is something much broader than just an inferential process interpreting an instrument reading: the ‘processes’ can be physical transformations happening in a laboratory (Rheinberger 2023), as well statistical and computational manipulations (which can count as ‘inferring’) happening in a virtual experimental system (Ratti and D’Agostino 2025). Looking only at ‘inferences’ in molecular biology misses this rich tapestry of transitions and transformations.

Fourth, a point about physical interactions with the apparatus is in order. No one can deny the importance of bioinformatics tools in contemporary biology, even if you think that EPA is justified. Therefore, an excessive focus on the initial physical interaction can obfuscate a rich tapestry of practices. While a physical interaction between an apparatus and a biological object is indeed necessary for measurement, this interaction is in many cases materially abstracted and made it virtual, in various digital media. For this reason, the interactions between the object and the apparatus are not only physical, but also ‘virtual’.

Finally, Van Fraassen and Parker use the term ‘logical space’. In the case of molecular biology, the more general term ‘conceptual space’ is preferable, given the heterogeneous nature of the space where items will be located, with theories coming from different disciplines (chemistry, physics, biology, etc), as well as know-how and tacit practical knowledge with respect to instruments and the specificities of the measuring context.

Based on these considerations, I define measurement in biology as:

Measurement in biology (MB) = an empirical activity aimed at gathering information about biological phenomena (e.g., their properties) with the following characteristics:

- The activity happens in a *pipeline*, which specifies the (physical and virtual) *interactions* between one or more instruments and an object (typically an experimental system[7])
- Throughout the pipeline, the object is materially and computationally *processed* to generate instrument readings, which are *transformed* until a final representation is reached
- this final representation is the ‘measurement outcome’; the representation is located in a *conceptual space* defined by a network of auxiliary assumptions, which include domain knowledge coming from various disciplines, and knowledge of the instruments involved;
- the measurement outcome serve proximate goals (e.g., detection; effect estimation; characterization)

[7] See (Rheinberger 1997)

3.2 Connecting MB to EPA

Let us connect MB to EPA. It should be noted that EPA is epistemic because it is about knowledge claims that can be generated as a result of material proximity. MB is also defined in terms of knowledge claims (represented as measurement outcomes), so EPA and MB speak the same language. Justifying EPA by investigating what biologists and bioinformaticians do in terms of MB, would mean to show one of two things (or, ideally, both). First, it means to show that material access and proximity is what enables biologists to turn instrument readings into measurement outcomes, and that bioinformaticians' role is only to refine readings, rather than transforming them into outcomes. This would prove EPA-b: given that information concerning properties of biological phenomena are expressed, *per* Bokulich's definition, as knowledge claims represented as measurement outcomes, and experimentalists are responsible for the outcomes, then EPA-b is justified. Second, it means to show that material access and proximity is what enables biologists to conceptualize measurement outcomes, and hence the corresponding type of knowledge claims that can be made *in principle*. This would prove EPA-a. In both cases, experimentalists have a vantage point over information/knowledge claims because of their material proximity and material access to biological phenomena.

4 EPA-b AND BIOINFORMATICS

In this section, we look into the merit of EPA-b. The strategy is to the take common examples of measurement in biology, where both bioinformaticians and experimentalists contribute to, and find who transforms readings into outcomes, rather than only providing or refining readings. If experimentalists are in charge of this crucial step, and this is because of their material access and proximity to experimental systems, then they have the vantage point described by EPA-b. I introduce two common examples in 4.1 (one type of detection measurement, and one type of characterization measurement), and discuss them in 4.2.

4.1 Common Examples of Measurements

4.1.1 Detection Measurements: Sequencing Technologies

A detection measurement has the goal of detecting the presence of one or more biological objects and/or processes. The information gathered is the presence or absence of something. The process of measuring *qua* detecting activities is common in molecular biology – a classic example is PCR (REF), where the goal is to establish whether a given fragment of DNA is present in a DNA sample. Here I discuss sequencing technology, which is a more complex case.

Sequencing refers to the measurement processes that lead to the detection of a primary structure of polypeptides (e.g. proteins) or polynucleotides (e.g., DNA or RNA molecules). The term 'sequencing' covers a wide array of methods and approaches. My focus is on DNA sequencing. It is usually said that there are three generations of DNA sequencing approaches (van Dijk et al 2018). The first-generation sequencing approach is known as Sanger-sequencing. Second-generation sequencing addresses a number of limitations of Sanger sequencing, in particular by allowing massively parallel sequencing. Finally, third-generation sequencing improves the length of reads by facilitating long-reads sequencing. Often, second- and third-generation sequencing are lumped in the label 'next-generation sequencing' (NGS). In this short description, I will focus especially on second-generation.

There are three main steps in second-generation sequencing (Hu et al 2021). My descriptions of these steps is idealized, as there are many ways to go through the three phases[8]. These steps make up what is known as the 'sequencing pipeline', which refers to those procedures used to transform samples of DNA into measurements concerning the primary structure of, e.g., polynucleotides such as DNA.

The first step of this pipeline is called 'library preparation'. This is when samples of DNA are prepared in such a way that they are amenable to be 'processed' by the sequencing platform which will then produce instrument readings. In addition to quality and purity checks, DNA molecules are fragmented by means of laboratory procedures (e.g., enzymatic, chemical, physical methods, etc) into short pieces. This is because, in second-generation sequencing, sequencing platform cannot 'read' long sequences of DNA, but they can only detect shorter ones, where the length is determined by the requirements of a given sequencing platform. These short pieces, taken together, constitute the totality of a segment of DNA one wants to sequence (e.g., a chromosome; an entire genome; etc). After fragmentation, DNA shorter pieces are 'repaired' by preparing the fragments for so-called 'adaptor ligation', which is when the fragments are attached oligonucleotides that can be 'recognized' by the sequencing machines' surface: these 'short fragments' with adaptors will bind to the surface of the sequencing machine to facilitate massive parallel sequencing. This is a noteworthy progress with respect

[8] See (Metzker 2010) for a comprehensive overview of these differences in second-generation sequencing

to first-generation sequencing, where one had to fragment DNA sequences anyway, but then could only sequence one fragment at the time.

The second step in the sequencing pipeline is sequencing *proper*. Different kinds of sequencing can be distinguished depending on the chemistry used. The most common is 'sequencing by synthesis', in which the DNA library prepared in the first step is amplified (usually via clonal amplification), in order to produce a stronger signal that can be more easily detected. After amplification, nucleotides tagged with specific fluorescent dyes are added to the library and incorporated in the DNA molecules. To simplify a complex process, the nucleotides tagged with fluorescence (called 'probes') hybridize with their complementary sequences. After this phase, the sequencing machine excites fluorescent dyes with a laser, and a CCD camera detects the excitation (i.e. it literally takes photos!). In the rawest instrument reading, the four possible nucleotides are represented by lines of different colours on a chart (Metzker 2010; Stevens 2013).

Finally, the third step is 'analysis'. This step includes a number of automated, partially-automated, and manual procedures. The most basic is called 'base calling', which is when, on the basis of the intensity of the signal detected by the CCD camera, a software provides a score as to how confident the machine is that a given base is indeed a specific nucleotide. Next, bases called (i.e., 'reads') from the fragmented DNA are assembled into a linear sequence. After assembly, sequences must be checked for quality control to, e.g., fill gaps. Base calling, assembly, and quality control are typically called 'primary analysis' (Hu et al 2021). After the raw sequence is stored in a file (in e.g., FASTQ format), it is subjected to secondary analysis, which is read alignment and variant calling. This kind of analysis is done by comparing the raw sequence to a Reference Sequence, in order to understand commonalities and differences. Tertiary analysis corresponds to variant annotation (e.g. differences with respect to the Reference Sequence considered), and functional annotation of variants (e.g., whether they are SNP, INDEL, CNV, etc), which can be used to "determine their biological and pathological functions" (Hu et al 2021, p 805). Secondary and tertiary analyses require the experienced use of software and statistical tools.

Sequencing is a form of MB. First, it is an empirical information-gathering activity: what is gathered is information (broadly conceived) about the sequence of nucleotides constituting genomes. Second, this information-gathering activity happens in a pipeline (as defined in 3.1): it involves the transformation of objects (e.g. a sample of DNA fragmented, its pieces amplified, etc), as well as representations (e.g. coloured points in a chart turned into letters). The measurement outcome – the sequence – is a representation of the polynucleotides chain arranged in a linear sequence. This representation is located in a wide conceptual space, structured around a dense theoretical background, auxiliary assumptions, and technical assumptions about the instruments used. Theoretical background comes from molecular biology (e.g., the nature of nucleotides, the complex mechanistic machine underlying amplification), as well as chemistry (e.g., the chemical basis governing complementary bases). The functioning of the sequencing platform relies significantly on physics (e.g., the way the laser excites fluorescent dyes). Analyses in the third step are based on best software practices, as well as applied statistics techniques. Finally, sequencing can be understood as 'detection' in a number of ways. For instance, it is the process through which the order of nucleotides and their precise arrangement in a polynucleotide chain is detected, or in general it is a process aimed at detecting the primary structure of a polynucleotide chain by identifying its constituent nucleotides. One could even say that what is detected is the presence and identity of nucleotides at every position in a polynucleotide chain. Tertiary analysis admittedly goes beyond mere detection, because it characterizes also the functional dimension of portions of the chain. But sometimes tertiary analysis is considered not part of sequencing proper. Moreover, it is appropriate to say that, after the secondary analysis in the third step, biologists have nonetheless a measurement outcome, namely the identity and position of nucleotides on a chain.

4.1.2 Characterization Measurements (Flow Cytometry)

In characterization measurements, the goal is not simply to detect an object or a process; rather, it is to reveal different aspects of an object or a process. While 'detecting', by establishing the presence of an object or process, might do this on the basis of a specific dimension (e.g., the presence and identity of nucleotides at every position in a polynucleotide chain), in 'characterization' the object or process is specified on the basis of more than just one dimension. Common examples of 'characterization measurements' are the measurement outcomes resulting from microscopy, where morphology, position, and quantities of objects are characterized through visual means. In this section, we focus on a different (but analogous) type of measurement outcome, based on 'cytometry'.

Cytometry is the measurement of the features of cells - the word comes from Greek, where 'kytos' means 'container' (i.e., cell) and 'metron' means 'measure'. 'Flow' refers to the fact that cells in flow cytometry are in 'suspension' in a fluid, rather than attached to a surface like in microscopy. Flow cytometry is performed with the help of a flow cytometer (Robinson et al 2023), which is made of a number of components, including a fluidic system controlling the flow of cells; an optical system equipped, typically, with a laser that is supposed to capture the emitted fluorescence or simply the signals; and a data acquisition system which today are sophisticated software armed with cutting-edge data science tools, but in the early days of flow cytometry were analogue instruments that displayed signals (that is, instrument indications).

Most flow cytometers are fluorescence-based, but there are also alternatives using, e.g., metal isotopes. Flow cytometers detect fluorescence signals emitted by dyes, which are fluorescent molecules (e.g. fluorochrome probes) used to measure a number of parameters of cells by 'flagging' them. In a typical flow cytometry measurement process (Aghaeepour et al 2013), cells are stained with these fluorochrome-dyed molecules that bind to cell surface and intracellular components. When passed through the flow cytometer, cells are scanned through a laser beam existing the fluorochromes, and the emitted light (which is proportional to the density of the molecules dyed that are bind to cells) is measured. Given this general characterization, flow cytometry has the goal of 'cell sorting' (i.e., isolation and recovery of a given cell population) and 'analysis'. The latter, understood as "the recording of many readouts for each individual cell" (Robinson et al 2023, p 6) is of particular interest here.

Analyses that are done in flow cytometry are numerous (Robinson et al 2023). One can do 'phenotyping', which is the identification and classification of cells on the basis of multiple parameters. In immunology, one can analyze the fluorescence associated to given markers to track their expression and dynamics during disease progression. Another noteworthy analysis is 'viability assays'. This procedure is based on the idea of using fluorescent dyes that selectively label cells with certain known characteristics. These 'dyes' have specific hydrophobic properties, which might penetrate intact cellular membranes or only enter compromised membranes, to even more specific that enter live cells but become fluorescent only when interacting with specific macromolecules (Robinson et al 2023). In this way, one characterize cell populations by distinguishing different types of states of cells, and their features. One can also combine flow cytometry for cell cycle analysis to, e.g., monitor p53 cell cycle arrest, or measure multidrug resistance. Finally, one can also measure cellular function, e.g. the oxidative potential of granulocytes (i.e. a type of white blood cell) using dyes sensitive to oxidation states.

Flow cytometry is indeed a measurement practice and, as sequencing, falls neatly into MB. First, flow cytometry is an empirical information-gathering activity, where the information gathered pertains to different dimensions of cells. For instance, in immunophenotyping quantitative information concerning activation, proliferation, or functional changes of cells populations can be gathered. Second, this information-gathering activity happens in a pipeline (as defined in 3.1), where cells go through a number of transformation steps such that they can be visualized in the proper way. The measurement outcome is a representation of specific characteristics of cells populations. As in any measurement, this representation is located in a conceptual space, encompassing theoretical backgrounds, and auxiliary assumptions concerning the instruments used. This wide conceptual space include background knowledge of hydrodynamic, optics, and chemistry for the functioning of the instruments, as well as knowledge about the biological underpinnings of cell component. Consider an example of how diverse is this conceptual space is: in commenting on viability assays, Robinson et al (2023) point out that the most commonly used dyes (i.e., propidium iodide and 7-aminoactinomycin D), while binding to DNA, can only enter compromised membranes. This means that the instrument reading 'DNA is dyed with this fluorescent molecule', combined with knowledge about cell membranes, and the chemistry underpinning the dyes themselves, will be turned into the measurement outcome 'these cells are dead'. Finally, flow cytometry is not simply about detecting; rather it characterizes cells across many dimensions. When measuring cells in the immune systems through flow cytometry, activation, proliferation, and functional changes give us information of temporal and spatial nature. Through this complex tapestry of information, we get a picture of what those cells are, rather than just saying that they are present.

4.2 Discussion

Let us start by discussing detection measurements, in particular sequencing technologies. As we have noticed, there are at least three generations of sequencing. For first-generation sequencing technologies, no sophisticated bioinformatics tool was needed, and, to my knowledge, no sophisticated tool was available when Sanger designed the method. In this case, the property that is measured (i.e. the presence and identity of nucleotides at every position in a polynucleotide chain) is indeed detected because of wet-lab methodologies manipulating and transforming biological material. In other words, experimentalists, in virtue of experimental activities and material access to the biological system, transform a set of instrument readings, into a specific measurement outcome. Indeed, fragments were assembled by interpreting polyacrylamide gels which separate DNA fragments by size. Let us now turn to second-generation sequencing. Investments in alternative sequencing methods were motivated by the fact that, with Sanger sequencing, it was simply not possible to sequence more than one fragment at the time. Given that one requirement implicit in the idea of sequencing is exactly to fragment DNA to handle long sequences better, this meant that it became increasingly difficult to measure the presence and identity of nucleotides at every position in longer polynucleotide chains. The Human Genome Project was a 2-billions dollars, 13-years, consortia-based, massive effort to use this type of sequencing, at the least in the public consortium (Stoeger and Ratti 2025). The innovation brought by second-generation sequencing is to sequence millions of fragments in parallel. In order to transforms those millions of instrument readings coming from millions of fragment into a measurement outcome, bioinformatics tools are necessary (Pereira et al 2020). However, their necessity lie at two rather different levels. First, bioinformaticians provide support to refine readings. In this respect, bioinformatics tools are necessary in primary analysis, which is the detection and analysis of signals generated by the sequencing platforms (e.g., fluorescence reads) which leads to so-called base calling. While base-calling is increasingly automated, the computational procedures followed by algorithmic tools are, indeed, the result of bioinformaticians' insights into how reads with given characteristics, can be indeed processed and analyzed to be turned into 'sequencing reads'. For instance, Illumina platforms signal detection relies on fluorescence, and algorithms convert a fluorescence signal into a sequence by giving a score to the intensity of the four 'fluorescence dyes' that are attached to nucleotides. This also involves specific choices as to how to quantify the uncertainty, which is then made explicit by providing a score for each nucleotide in the sequence – and this process of calibration is done via bioinformatics means. Other procedures in primary analysis involves quality control, which is again based on algorithmic tools whose choice "is highly dependent on the dataset, downstream analysis, and parameters used" (Pereira et al, p 9). In all these cases, bioinformatics provides support for experimentalists, but it is support to refine readings into more precise readings. At a second level, bioinformatics is necessary in a more radical sense. In secondary analysis, the readings are further transformed into outcomes, as a result of tasks carried out by bionformaticians. This is especially evident in sequence alignment and variant calling. Alignment against a reference genome requires using algorithms such as, e.g., Burrow-Wheeler transform algorithm, which are tools for data transformation that restructure data to be more compressible – and the choice of what and how to compress, will rely on computational choices that require a bioinformatics mindset, rather than an 'experimentalist' one. Variant calling will then validate a number of positions in the chain. This means that with alignment and variant calling, the readings concerning the positions in the chain are confirmed, by quantifying and addressing uncertainties, and finally turning them into measurement outcomes. The point is that, to detect/measure the presence and identity of nucleotides at every position in a long polynucleotide chain such as an entire genome, you do not just need bioinformaticians to automate tedious part of your transformations in the sequencing pipeline; it is the bioinformatician who transforms a set of instrument indications (e.g., fluorescence signals first, bases called next) into a precise measurement outcome. While for short polynucleotide chains experimentalists can construct the measurement outcome, in the case of longer chains one can (wickedly) see them acting as technicians preparing the samples, and it is the bioinformatician who constructs the large-scale measurement of long chains.

Similar considerations apply also for characterization measurements, exemplified by flow cytometry. There was indeed a time when flow cytometry made use only of analogue instruments exploiting principles behind Coulter particle counters. Moreover, for simple tasks one might not need to use bioinformatics tools to turn the readings into an outcome. For instance, one can simply do what is called ‘manual gating’. This refers to the process of “visually inspecting multidimensional plots of the data, and drawing boundaries (gates) around populations of interest” (Liu et al 2024, p 11). Bioinformatics tools are here necessary, but in the sense of just refining readings, e.g. by generating plots, and experimentalists then will turn these instruments readings into a measurement outcome by locating the object of interest (i.e. the population of cells) into a well-defined conceptual space. Liu and colleagues (2024) characterize informally the ‘conceptual space’ used to interpret plots (and hence to turn readings into outcomes) as “experience-based, time-consuming, and relies on prior knowledge and arbitrary cutoffs to assign cell populations” (p 2). However, scalability is a problem, as manual gating is unreliable for large and multidimensional datasets. In these large-scale cases, bioinformaticians turn instrument readings into measurement outcomes. It is not a coincidence that, from early 2000s, flow cytometry has been able to measure an increasing number of fluorescent markers at a time, as a result of the increased capacity to generate multidimensional data, and develop bioinformatic tools for their analysis (O’Neill et al 2013). When the number of markers increases, so do the scatter plots that need to be investigated for ‘gating’. Because of such high dimensionality, manual gating is just not possible. Gating in the era of highly-dimensional data requires the design and correct implementation of a variety of algorithms, such as dimensionality reduction tools, combinatorial gating algorithms, clustering algorithms, etc (O'Neill et al 2013). All these algorithms require extensive bioinformatic work to do the gating, meaning that it is a task of the bioinformatician to turn the multidimensional instrument readings (e.g., fluorescent traces coming from the flow cytometer) into a measurement outcome, which will locate the object measured (e.g. a population of cells) within a conceptual space, whose dimensions would not only concern background knowledge and assumptions concerning the biology and the chemistry of the cell and the physics of the flow cytometer, but also the assumptions and standards of the bioinformatics tools involved. In other words, in these cases where highly multidimensional data is involved, bioinformaticians transform readings into outcomes.

These two examples of MB show that, for small-scale cases, EPA-b is valid. Experimentalists turn readings into measurements, and while bioinformatics tools might play important role in the transformation pipeline, the crucial transformation into outcome can be done (in principles and in practice) by experimentalists, in virtue of their access to the object measured, their experimental activities, and the familiarity with the experimental system developed through constant material interactions. However, in large-scale cases, it is the bioinformatician who provides the crucial transformation into outcomes: there is no whole-genome sequencing *measurements* without a bioinformatician crucially transforming a set of disconnected and uncertain readings, into a precise outcome; there are no cytometric measurements, without bioinformaticians turning the utter complexity of multidimensional instrument readings into a precise outcome. It is not just that bioinformaticians contribute to the measurement process in instrumental (though often essential) ways, and then experimentalists integrate the readings into an outcome: it is the task of bioinformaticians to provide the crucial transformation. Therefore, we conclude that EPA-b works only in simple cases, while for large-scale cases it does not. But given that a contemporary molecular biology is indeed large-scale, then one might be skeptical of the relevance of EPA-b, and conclude that measuring in the era of data-intensive biology does not necessarily depend on material proximity. In other words, material manipulations and how close we are to material experimental systems do not provide a vantage point over the identification and characterization of properties of biological phenomena.

5 EPA-a AND BIOINFORMATICS

Experimentalists can still rebut that EPA-a is not undermined by the examples discussed above. This is because, while it is true that the bioinformatician transforms a set of readings into an outcome, the type of outcome (e.g., position of nucleotides; characteristics of the cell) is still defined in terms that are, in fact, amenable to the experimentalist *ethos*: the nucleotides and the cells are entities that are isolated and defined by means of the experimental activities of experimentalists. In other words, nucleotides and cells are physical entities whose properties are conceptualized in certain ways because of how they are materially accessed and manipulated. In this section, we show that the most exciting works of bioinformatics concern the construction of measurement outcomes that, in fact, violates EPA-a. We will exemplify this dimension of bioinformatics measurement through the case of gene set enrichment analysis (GSEA).

5.1 GSEA Measurements

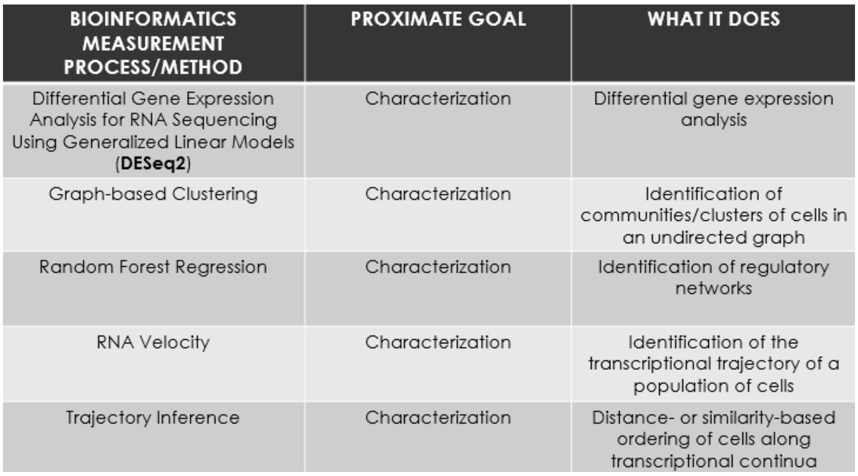
GSEA is a computational method to construct measurement outcomes (Mootha et al 2003; Subramanian et al 2005) related to the altered state of gene pathways in a given phenotype. The outcome can be interpreted as knowledge claims relating expression of groups of genes to higher-level mechanisms, in such a way that these outcomes can be used, in concert with other outcomes, to build mechanistic models. GSEA has been developed in the context of the explosion of data generated in genomics. Back in 2003, mRNA expression profiles (mostly generated by microarrays) were constituted by lists of thousands of genes coming from samples belonging to two classes, where genes were ordered on the basis of their differential expression in the two classes. A common approach was to look at the top and bottom of the list, select a handful of genes showing the largest difference, and then discern biological clues on the basis of their characteristics. But this approach had obvious drawbacks. For instance, cellular processes affect sets of genes, rather than individual genes, and a slight increase or decrease in groups of genes might do more than more significant fold changes in individual genes. If we just look at individual genes, we might miss these subtler dynamics (Subramanian et al 2005). GSEA was developed exactly to identify set of genes that are connected to the emergence of a given phenotypes. From this perspective, GSEA is a detection measurement process.

This is how GSEA works. Take, for instance, the study where it was first introduced (Mootha et al 2003). Mootha and colleagues compare the expression profiles of patients with Type 2 diabetes mellitus (DM2) and normal glucose tolerance (NGT). First, a list L of genes ranked on the basis of their differential expression in the two groups is created, where the ranking is established by fold changes (namely, the ratio between the expression of the same gene in DM2 and NGT samples). Next, bioinformaticians curate an *a priori* list of different sets of genes. Let's call these set of genes *Ss*. Now, there are many ways in which one can choose *Ss*. In Subramanian et al (2005), it is said that these sets will depend on the specific question at hand. In the case of Mootha et al, they created 149 *Ss*, some collating genes from pathways (113) that might be related to DM2, and gene clusters coregulated in a mouse expression atlas (36). These sets were curated by consulting databases and by doing an extensive literature review. Statistically speaking, the null hypothesis is that the rank of the genes is random with respect to the diagnostic characterization of samples. The alternative hypothesis is that the rank reflects one or more *Ss*. GSEA has two instrument readings and one measurement outcome. The first reading is what is called *enrichment score* (ES), which is the strength of the association between the rank of genes in *L* with one or more *Ss*. Another reading is a preliminary Maximum ES (pMES), which is across all *Ss*. What was found was that pMES was detected for genes involved in oxidative phosphorylation (OXPHOS). In ranking *Ss* in pMES, there was also another interest set called cluster20 (c20), which overlapped partially with OXPHOS. This overlap was characterized as a new subset of OXPHOS involved in a specific pathway, and they call this new set OXPHOS-CR. The measurement outcome of GSEA was, in this case, the detection of OXPHOS-CR as a set of genes tightly co-regulated in DM2 and significantly differential overexpressed with respect to NGT. This characteristic of this set of genes is expressed in terms of MES (namely, information about OXPHOS-CR is represented in the form of MES).

GSEA is a measurement process. It is an empirical activity aimed at gathering information about the properties of biological phenomena, in particular the propensity of two groups of samples with different conditions to have genes of specific pathways differentially expressed. The pipeline of GSEA starts in the wet-lab realms, with the collection of gene expression profiles readings. In contemporary cases, collection of these readings comes from RNA-sequencing, which goes through the stages typical of this method. It moves then to a 'virtual' platform, where those readings are transformed by means of the computational and statistical operations described above. The measurement outcome is MES, namely the detection of a differentially expressed set of genes that are coregulated within one or more pathways affecting the phenotype of interest. The conceptual space is structured around domain knowledge concerning the mechanism of gene expression and gene functions. Domain knowledge is even precisely formalized in the way it is represented in databases providing annotations about genes and pathways (Leonelli 2016). Moreover, statistical and computational aspects are particularly important.

5.2 Discussion

How does GSEA put pressure on EPA-a? GSEA measures how differences in two conditions (say, DM2 and NGT) can be associated to genes belonging to specific pathways. MES is a property of the differential expression of a set of genes – you cannot construct MES starting from the expression profiles of one group only. MES is something that is measured only when comparing differences between two conditions (DM2 and NGT). But most important, MES is not a property like the position and identity of nucleotides in a chain, properties of cell membranes, the length of telomeres, or the stiffness of a living tissue. MES is a statistical property of a group of genes and, as such, it is conceived in bioinformatics terms, by using concepts and methods from statistics such as fold change and permutation-based tests, and computational tools such as databases: it is a statistical transformation that lead to the conceptualization of a new kind of measurement outcome. MES is conceivable only through computational and statistical means: the conceptualization of MES *qua* property characterizing a dimension of biological phenomena, is entirely established by bioinformaticians and, as such, undermines EPA-a. In other words, having material access to gene expression dynamics does not provide any vantage point for conceptualizing MES in the first place – but having access to statistical properties and databases does. MES is not an isolated case: as shown in Table 2, there are also other well-known cases violating EPA-a.



BIOINFORMATICS MEASUREMENT PROCESS/METHOD	PROXIMATE GOAL	WHAT IT DOES
Differential Gene Expression Analysis for RNA Sequencing Using Generalized Linear Models (**DESeq2**)	Characterization	Differential gene expression analysis
Graph-based Clustering	Characterization	Identification of communities/clusters of cells in an undirected graph
Random Forest Regression	Characterization	Identification of regulatory networks
RNA Velocity	Characterization	Identification of the transcriptional trajectory of a population of cells
Trajectory Inference	Characterization	Distance- or similarity-based ordering of cells along transcriptional continua

Table 2. Examples of bioinformatics tools violating EPA-a

6 FINAL REMARKS

In this paper, I have provided an account of measurement in biology, where this account allowed me to differentiate the ways in which so-called ‘experimentalists’ (i.e., wet-lab biologists) and bioinformaticians measure the properties of biological phenomena. I have shown that bioinformaticians are responsible for the construction of large-scale measurement outcomes, as well as entire new kinds of measurements that, despite having biological relevance, can only be conceptualized from the standpoint of bioinformatics work and mindset. My analysis has two main consequences, one for philosophy of biology and philosophy of science, and the other for molecular biology as a discipline.

Regarding philosophy of science and biology, this article is the first attempt (to my knowledge) to build a comprehensive account of measurement in molecular biology, and it raises the question of what is the place of such an account within the dominant epistemology of molecular biology, characterized in mechanistic terms. Furthermore, by characterizing measurement practices also in bioinformatics, this article raises questions about the relations between measurements and data, which is an underexplored topic[9].

By putting pressure on EPA, the relevance of this article for molecular biology is twofold. By arguing against EPA-b, it supports a more balanced relation between bioinformaticians and wet-lab biologists. One can even take a more extreme stance, and say that, because of the large-scale nature of contemporary biology dealing with multidimensional data, we are going towards a future where bioinformaticians will have a vantage point for formulating knowledge claims over experimentalists in general. This is very speculative, and yet one can be easily lured into such thoughts. Second, by arguing against EPA-a, bioinformaticians are put in a privileged position for theorizing about aspects of biological phenomena, given how their bioinformatics-based measurement outcomes can indeed reveal properties that might be difficult to even conceptualize by experimentalists.

REFERENCES

Aghaeepour, N., Finak, G., Hoos, H., Mosmann, T. R., Brinkman, R., Gottardo, R., Scheuermann, R. H., Dougall, D., Khodabakhshi, A. H., Mah, P., Obermoser, G., Spidlen, J., Taylor, I., Wuensch, S. A., Bramson, J., Eaves, C., Weng, A. P., Fortuno, E. S., Ho, K., … Vilar, J. M. G. (2013). Critical assessment of automated flow

[9] To my knowledge, this topic is explicit (briefly) addressed only by Van Fraassen (2008, p 166)

cytometry data analysis techniques. *Nature Methods*, *10*(3), 228–238. https://doi.org/10.1038/NMETH.2365

Bartlett, A., Lewis, J., Reyes-Galindo, L., & Stephens, N. (2018). The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -omic biology and high-energy physics. *Big Data and Society*, *5*(1). https://doi.org/10.1177/2053951718768831

Bartlett, A., Penders, B., & Lewis, J. (2017). Bioinformatics: Indispensable, yet hidden in plain sight? In BMC Bioinformatics (Vol. 18, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12859-017-1730-9

Bartlett, Andrew, Jamie Lewis, and Matthew L. Williams. (2016). "Generations of Interdisciplinarity in Bioinformatics." *New Genetics and Society* 35 (2). Taylor & Francis: 186–209. doi:10.1080/14636778.2016.1184965.

Bokulich, A. (2020). Calibration, Coherence, and Consilience in Radiometric Measures of Geologic Time. *Philosophy of Science*, *87*. http://www.stratigraphy

Grabowski, P., & Rappsilber, J. (2019). A Primer on Data Analytics in Functional Genomics: How to Move from Data to Insight? In Trends in Biochemical Sciences (Vol. 44, Issue 1, pp. 21–32). Elsevier Ltd. https://doi.org/10.1016/j.tibs.2018.10.010

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Knorr-Cetina, K. (1999). *Epistemic Cultures*. Harvard University Press.

Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Construction of Scientific Facts* (2nd editio). Princeton University Press.

Leonelli, S. (2016). *Data-centric Biology*. University of Chicago Press.

Lewis, Jamie, and Andrew Bartlett. (2013). "Inscribing a Discipline: Tensions in the Field of Bioinformatics." *New Genetics and Society* 32 (3): 243–63. doi:10.1080/14636778.2013.773172.

Lewis, Jamie, Andrew Bartlett, and Paul Atkinson. (2016). "Hidden in the Middle: Culture, Value and Reward in Bioinformatics." *Minerva* 54 (4). Springer Netherlands: 471–90. doi:10.1007/s11024-016-9304-y.

Liu, P., Pan, Y., Chang, H.-C., Wang, W., Fang, Y., Xue, X., Zou, J., Toothaker, J. M., Olaloye, O., Santiago, E. G., McCourt, B., Mitsialis, V., Presicce, P., Kallapur, S. G., Snapper, S. B., Liu, J.-J., Tseng, G. C., Konnikova, L., & Liu, S. (2024). Comprehensive evaluation and practical guideline of gating methods for high-dimensional cytometry data: manual gating, unsupervised clustering, and auto-gating. *Briefings in Bioinformatics*, *26*(1). https://doi.org/10.1093/bib/bbae633

López-Rubio, E., & Ratti, E. (2021). Data science and molecular biology: prediction and mechanistic explanation. *Synthese*, *198*(4), 3131–3156. https://doi.org/10.1007/s11229-019-02271-0

Markowetz, F. (2017). All biology is computational biology. PLoS Biology, 15(3). https://doi.org/10.1371/journal.pbio.2002050

Metzker, M. L. (2010). Sequencing technologies the next generation. In *Nature Reviews Genetics* (Vol. 11, Issue 1, pp. 31–46). https://doi.org/10.1038/nrg2626

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., … Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3). http://www.nature.com/naturegenetics

Morange, M. (2008). The Death of Molecular Biology? *History and Philosophy of the Life Sciences*, *30*(1), 31–42. https://www.jstor.org/stable/23334314

O'Neill, K., Aghaeepour, N., Pidlen, J. S., & Brinkman, R. (2013). Flow Cytometry in Bioinformatics. *PLOS Computational Biology*, *9*(12). https://doi.org/10.1371/journal.pcbi

Parker, W. S. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, *169*(3), 483–496. https://doi.org/10.1007/s11229-008-9434-3

Parker, W. S. (2017). Computer Simulation, Measurement, and Data Assimilation. In *Brit. J. Phil. Sci* (Vol. 68). https://about.jstor.org/terms

Radder, H. (2003). Technology and Theory in Experimental Science. In H. Radder (Ed.), *The Philosophy of Scientific Experimentation*. University of Pittsburgh Press.

Ratti, E., & D'Agostino, G. (2025). Beyond "Trapped Pets" and "Red Buttons": Bioinformatics as an Experimental Discipline. *Perspectives on Science*, *33*(2), 158–201. https://doi.org/10.1162/posc_a_00638

Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthetizing Proteins in the Test Tube*. Stanford University Press.

Robinson, J. P., Ostafe, R., Iyengar, S. N., Rajwa, B., & Fischer, R. (2023). Flow Cytometry: The Next Revolution. In *Cells* (Vol. 12, Issue 14). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/cells12141875

Stevens, H. (2013). *Life out of sequence - A data-driven history of bioinformatics*. Chicago University Press.

Strasser, B. (2017). *Collecting Experiments - Making Big Data Biology*. The University of Chicago Press.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* doi10.1073pnas.0506580102

Tal, E. (2017). Calibration: Modelling the measurement process. *Studies in History and Philosophy of Science Part A*, *65–66*, 33–45. https://doi.org/10.1016/j.shpsa.2017.09.001

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. In *Trends in Genetics* (Vol. 34, Issue 9, pp. 666–681). Elsevier Ltd. https://doi.org/10.1016/j.tig.2018.05.008

van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press.

Way, G. P., Greene, C. S., Carninci, P., Carvalho, B. S., de Hoon, M., Finley, S., Gosline, S. J. C., le Cao, K. A., Lee, J. S. H., Marchionni, L., Robine, N., Sindi, S. S., Theis, F. J., Yang, J. Y. H., Carpenter, A. E., & Fertig, E. J. (2021). A field guide to cultivating computational biology. In PLoS Biology (Vol. 19, Issue 10). Public Library of Science. https://doi.org/10.1371/journal.pbio.3001419