



Bellazzi, F. (2023). The Gene as a Natural Kind. In J. M. Viejo, & M. Sanjuan (Eds.), *Life and Mind: New Directions in the Philosophy of Biology and Cognitive Sciences* (1 ed., pp. 259-278). (Interdisciplinary Evolution Research; Vol. 8). Springer, Cham.
<https://doi.org/10.1007/978-3-031-30304-3>

Peer reviewed version

License (if available):
CC BY

Link to published version (if available):
[10.1007/978-3-031-30304-3](https://doi.org/10.1007/978-3-031-30304-3)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Springer at <https://doi.org/10.1007/978-3-031-30304-3>. Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

The gene as a natural kind

Dr Francesca Bellazzi

Accepted manuscript, published in *Life & Mind - New Directions in the Philosophy of Biology and Cognitive Sciences* (2003, Springer Publishing)

Abstract What is a gene? Does it represent a natural kind or is it just a tool for genomics? A clear answer to these questions has been challenged by post-genomic discoveries. In response, I will argue that the gene can be deemed a natural kind as it satisfies some requirements for genuine kindhood. Specifically, natural kinds are projectible categories in our best scientific theories and they represent nodes in the causal network of the world (as in Khalidi 2013, 2018, 2020, 2021). In §2, I will present a brief history of the gene and the controversy over its status. In §3, I will introduce the account of natural kinds considered in this paper. In §4, I will first present the relevant definition of genes and how they can be classified. Then, I will argue that the gene can be considered a natural kind as it satisfies the criteria for natural kindhood. §5 concludes.

Keywords: gene – natural kinds – projectibility – causal nodes – genetics

1. Introduction

What is a gene? Is it a natural kind or is it just a tool for genomics? A clear answer to these questions has been challenged by the developments of genetics of (at least) the last

Francesca Bellazzi
Department of Philosophy
University of Bristol
Bristol, UK
Email: francesca.bellazzi@bristol.ac.uk

20 years (Griffiths, Stotz 2006, 2013). On the one hand, the identity relation between individual genes and precise stretches of DNA has proved impossible. This might support a deflationary or nominalist view of the gene (Griffiths, Stotz 2006, El-Hani 2007, Fogle 2010). On the other hand, some have been arguing that we should maintain a realist approach to the gene while understanding better the cellular context in which it operates and embracing its complexity (El-Hani 2007 et al., Griffiths, Stotz 2013, Bellazzi 2022).

The controversy is not easily settled and one of the questions risen is whether the gene is a category “neat” enough to count as a natural kind or whether it is instead too ambiguous. Here, I will consider this question and I will argue in favour of a positive answer as the gene satisfies some requirements for natural kindhood, following Khalidi’s account of natural kinds (2013, 2018, 2020, 2021). These are genuine projectibility and being a node within causal networks. In a nutshell, the thesis defended here is that, under the considered approach to natural kinds, the category “molecular gene” used in scientific practice corresponds to a natural kind despite the complexity of the properties characterising it and it captures some objective features of reality. This is a captivating case study because it illustrates how natural kinds can be found even within complex and highly interactional systems.

This inquiry has some conclusions of interest. First, understanding whether something is a natural kind or not is important because the naturalness of a given category can provide us with a further justification for why we can make more robust inferences from it. In doing so, the identification of something as a natural kind can support the justification of a theory that presents such a kind. Second, a natural kind is more than a theoretical entity whose properties are postulated for practical purposes and this can direct research into discovering (rather than mere postulating) features about it. This supports the role that they have *also* in the process of discovering new information about such category. A natural kind is correspondent to something objective in the world, meaning that some properties could be discovered about it, and some could not. Lastly, and more generally, this project represents an instance in which biology and biological practice inform the philosophical debate on what counts as a natural kind. The structure of this paper will be the following. In section §2, I will provide a brief overview of the history of the gene, in

order to present the status of the controversy. Then, I will consider a definition of the gene within the contemporary debate. In section §3, I will present the relevant account of natural kinds considered, as presented by Khalidi. In this framework, natural kinds are projectible categories and nodes in causal networks. Having provided the metaphysical ingredients, in section §4 I will move on to consider whether the gene is a natural kind. I will argue that this is the case as it satisfies the criteria for natural kindhood aforementioned. Section §5 concludes.

2. A brief history of the gene

The last century has been rightly called *The Century of Gene* by Fox-Keller (2000). After the disruptive discovery and spread of the theory of evolution, the 20th century started with the aim of solving the puzzle of stability of traits and their transmission. In 1900, three journal articles were published by d. Vires, Correns, v. Tschermak exposing the laws of genetics, further developing and re-interpreting Mendel's studies on hybridization¹ (El-Hani 2015). It was clear that something was transmitted discretely from parents to offspring in a way that respected precise probability distributions. What was unclear was *the nature* of the entity under discussion. There were many candidates at the time, such as *gemmules* proposed by Darwin, d. Vires' *pangenes* or Weismann's *determinants*. These were possible different entities that could have been transmitted respecting the laws of genetics (Fox-Keller 2000, Beurton 2010, Falk 2010). In response to this conceptual unclarity, Johannesen decided to introduce the new term *gene* to disentangle the discussion and to facilitate scientific practice. In 1909, he presented the gene as what explains the transmission of traits from parents to offspring and should have been “free from any hypothesis”, with no theoretical pre-assumptions. It was considered only a “convenient notational concept” (Falk 2010, 321). First introduced as an instrumental theoretical entity, one whose observation wasn't needed, but whose postulation could improve the explanations and predictions of a theory, the gene changed our way of doing biology.

¹ Three papers were published in the *Proceedings of the German Botanical Society* in 1900, by d. Vires, Correns, v. Tschermak (Fox-Keller 2000). For the role of Mendel's 1866 paper in the birth of genetics see El-Hani 2015.

Nevertheless, not all scientists were aligned with an instrumentalist view of genetics. With its development, genes started to have properties in their own right, even if they were detectable only indirectly. The tensions between concrete experimental demands and the specification of the phenomenon under consideration led towards a more specific material identification of the gene (Fox Keller 2000, El-Hani 2007, Falk 2010). In particular, Muller, a student of Morgan, supported a realist hypothesis of genes as material units, possibly chemical molecules, in disagreement with most scientists and his master (El-Hani 2007). This realist understanding supported the research for a material molecular basis of the gene. In the early 1940s, there was the rising of the one-gene-one-enzyme hypothesis, supported by the discovery in 1944 of DNA as the substance of heredity by Avery and his team (Beurton 2010). Finally, in 1953 Watson and Crick built a model for the structure of the DNA molecule and its replication, thanks to x-ray diffraction images of DNA taken by Gosling, Franklin and Wilkins (Clark, Pazdernik 2012). The molecular basis of the gene was found. From this moment, the history of the gene should be considered two-fold. On the one hand, the gene remained the so-called Mendelian gene, the unit of trait transmission and the object of classical genetics; on the other hand, the gene became identifiable with a precise chemical molecule (Kitcher 1982). The relations between the gene as the unit of transmission and the molecular gene are complex, and I will not enter into the details here. Suffice to say that Mendelian genes are not easily reducible to the molecular genes and they can be considered as something distinct with a different function and properties (Kitcher 1984, Okasha 2019). Here, I will not consider the Mendelian gene and I will focus only on the molecular gene, the object of molecular genetics (Waters 2007).

In the 1960s, it seemed clear that genes were nothing more than segments of DNA located on a chromosome that gave rise to a particular amino acids sequence. The search for a material basis for the gene led scientists to postulate a correspondence between genes, segments of DNA and amino acids. This correspondence was formulated as the *Crick-sequence hypothesis*: each codon, sequence of three bases, specifies only one amino acid and a gene is a sequence of codons that specify for a polypeptide. The “material molecular gene” was born. Soon, a molecular understanding of the gene was accompanied by the

idea that they were just open reading frameworks: DNA frameworks open to be read (ORFs). This facilitated research at the time, as the gene was identified as a well-defined and structured stretch of DNA, with clear borders and a singular function. The success of molecular genetics was read in terms of eliminativist reductionism. In 1969, Schaffner proposed to apply the models for epistemic reduction used in physical sciences to the case of the gene. If an identity relation between genes and DNA molecules had been found then an eliminativist reduction would have been accomplished. The conclusions of this reductionist interpretation were more than epistemic, as they aimed at the *elimination* of the gene, and there was empirical and experimental support. A gene was claimed to be identical to a precise and defined stretch of DNA, $gene1=DNA1$, as a simple 1:1 co-linear eliminativist reductive relation. “In the light of the Watson-Crick model, Benzer considered the possibility of translating his biological genetics into chemical terms” (Schaffner 1969, 339) and Schaffner regarded the entire reduction of molecular biology to chemistry and physics as something not only possible but very close in time.

However, things turned out to be more complicated than what Schaffner and many biologists thought in the 1970s. The production of new technologies to sequence genomes and the advances in molecular biology of the last forty years have disclosed peculiar genetic phenomena (Fox-Keller 2000; Hall 2001; El-Hani 2007; Griffiths, Stotz 2013; Meyer *et al.* 2013). The sequence of entire genomes and the study of eukaryotic genomes revealed that the Crick-sequence hypothesis was simplistic and further developments of genetics have made it impossible for genes to be a *merely* contiguous DNA segment co-linear with the product derived (Fogle 2010; Perini 2011; Rheinberger *et al.* 2015). These results compromised the material identity of the gene as a discrete stretch of DNA and showed the inefficiency of its identification in mere material terms (Falk 2010)². They also made clear how the gene operates within a system of complex and different mechanisms and processes. Particularly relevant to the beginning of the new phase of molecular biology is 2001 and the publication of the draft human genome sequence. This can be

² This does not imply that a given gene can never be considered within actual scientific practice as a precise and contiguous stretch of nucleotide sequences, such as in prokaryotes. Nevertheless, a strict identity relation between genes and genomic stretches excludes many genetic phenomena (see Fogle 2010; Griffiths, Stotz 2013).

considered a “threshold year” as, from this point onwards, genetics entered the “post-genomic era” (as in Griffiths, Stotz 2013). In 2007, El-Hani spoke of the crisis of the gene that finds itself between “the cross and the sword” because of the identification of a series of complex phenomena, such as split genes, alternative splicing, overlapping and nested genes.

2.1 The gene in the “postgenomic” world

Various attempts have been made to (re)-define the gene concept in the “postgenomic” world³, trying to accommodate both practical and theoretical requirements (as Beurton 2010). Generalising, we can point out two main ways to re-think the gene concept within the molecular context (Bellazzi 2022). The first is a deflationary instrumentalist approach that allows to retain gene speech and use without any further ontological commitment. The second is a realist approach that tries to define the gene embracing its complexity and context-dependency. While both these approaches are informative about gene individuation, i.e. how to identify individual genes, here we are concerned with the gene definition or gene characterisation, i.e. what does it take for a given entity to be a gene.⁴

The deflationary approach to the gene is often referred to as “the nominal gene” and identifies it in an operational way on the basis of actual scientific practice and conditioned to research needs. According to this account, a gene can be *any* stretch of precise nucleotides sequences that encode a specific product (Burian 2004; Griffiths, Stotz 2006, 2013, 66). This approach leads to what is called the “consensus gene”, as whatever pattern of “biochemical architecture and process” that presents the features of the exemplary gene, according to empirical evidence and scientific pattern (Fogle 2010, Bellazzi 2022). This approach to the gene remains nominalist or deflationary because it does not commit to the existence of the gene as something *sui generis* nor ontologically special: genes are any stretches of DNA that we find *useful* to identify as such given a particular model of

³ Often a pluralist view of genes is sustained, for which there are a variety of valid gene concepts in various disciplines (as Hall 2001, Fogle 2010).

⁴ Havstad (2021) defines three classificatory practices concerning kinds: classificatory characterization or definition, individuation and organisation (2021). The first - which is the one we are concerned with - focuses on the definition of the kind. The second focuses on identifying which tokens belong to a given kind. The last focuses on organizing taxonomies.

gene. Moreover the identification of the gene remains conditional on research interests, maintaining minimal commitment to the entity.

In contrast to this approach, we also find a realist one, presented in the context of the ENCODE analysis⁵. Within this project, Gerstein and colleagues take the *gene to be an existent entity (and so independent from our research interests)* and define it as “a union of genomic sequences encoding a coherent set of potentially overlapping functional product” (Gerstein et. al 2007). This has been further re-elaborated by Griffiths and Stotz, who understand the “postgenomic genes” as existent “images of the target produced molecules” (Griffiths, Stotz 2013, 75). While this latter is not a definition⁶, it represents an helpful metaphor to understand the gene: it should present a non-necessarily contiguous sequence that is similar enough to the one of the transcribed molecule. According to this realist view, the gene is not simply identical to a linear and contiguous sequence, but a union of different ones. This union normally includes the finally transcribed sequence and the promoter region (or TATA box) (Fogle 2010, 6). In addition to it, the gene often comprises those regions essential for its activation and the regulation of its transcription and these might be contiguous or not (Griffiths, Stotz 2013)⁷. Following this realist approach in the formulation presented by Bellazzi (2022), a gene is an entity composed of those parts of the relevant nucleic acids that are transcribable (or involved in transcription) and encode a given mRNA. This second aspect, being transcribable, represents a functional component in the definition as the gene is a union of sequences with the function of encoding a target molecule. While the gene has a material component, according to this approach it is not only characterizable materially or as a material entity. The functional characterisation of the gene also allows us to embrace the context-dependency

⁵ The ENCODE project is a project with the aim to identify all the “functional elements” in the human genome sequence. It represents an important project for the gene concept as it elucidated some complex phenomena that compromised the simple identification of the genes with contiguous stretches of DNA. Reference: <https://www.encodeproject.org/about/2012-integrative-analysis>.

⁶ I thank reviewer 1 for the helpful comments on this section.

⁷ We can identify 4 main models that aim at identifying the precise material basis of genes. Model A presents the material basis of the gene as the transcribed region of the DNA plus all neighboring sequences which play a role in the process; Model B considers only the transcribed region, with introns and exons; Model C further restricts the basis and includes only the set of exons derived from a pre-mRNA; at last, Model D limits the gene to only the coding exons of a primary transcript (Fogle 1990, El-Hani 2007, 3). Here, I leave open the question about the exact material component of the gene as it should be determined by scientific practice.

pointed out by the postgenomic analysis: “a function is always a role in something and a contribution to something” (Germain et al. 2014). Transcription is not a self-subsistent phenomenon, but it is rather a reactive one: it happens only within the right circumstances and thanks to a set of interactions that operate at different levels. Accordingly, the gene can be fully understood only within such a context of action and interactions and it is an entity defined materially and functionally (Falk 2010, Bellazzi 2022).⁸

In this paper, I will start from the second approach, in which the gene category refers to an existent union of sequences that are transcribable (or involved in transcription) and that transcribe precise genomic products (Bellazzi 2022)⁹. This definition is coherent with contemporary genomics that is able to identify some (often not-contiguous) unions of sequences that take part in the transcription of given molecules (either amino acids chains or RNA molecules). But is the gene a *natural* kind? Or is it just a convenient category that groups some existent phenomena for practical need? In order to answer these questions, we should explore what it means to say that a natural kind is *natural*. In the next section, I am going to present an overview of the topic and the relevant account of natural kinds.

3. Natural kinds and biological kinds

Division and classification of things into “sorts” or kind categories is common practice in both science and daily life. I recognise that *bread* is my favourite breakfast and biologists study *proteins*, *amino acids* and find *microscopes* in their laboratories. We notice that some properties or features are co-occurrent in some individuals and we cluster such groups of properties into kinds. These groupings are then associated with certain labels or predicates that allow the classification of individuals or relevant phenomena (Khalidi 2010). They allow us to make useful generalisations, being in our explanations and

⁸ As will be further clarified in section 4.1, the functional component of the gene definition allows for both multiple realisation and multiple composition (in contrast with a materialist only view). A gene would be any entity that is composed of the relevant material aspect, nucleic acids – either DNA or RNA – and that has the relevant function. However, for individual genes the correspondence does not need to be 1 stretch: 1 function, as the function could be multiply realised by any stretch that realises it. For further references, see Bellazzi 2022.

⁹ These sequences can either be of DNA or RNA according to the genes considered.

inductive inferences, and they proliferate from every corner of our life. Moreover, kinds play an especially important role in science. At least a part of scientific practice is based on clustering individuals into different categories and on making explanations and predictions about them (Bird, Tobin 2022). For instance, a biologist is generally not interested in the individual instance of a protein she is studying, or in the individual amino-acids string in her lab, but rather she aims at knowing something about the general *category* “protein”. This would allow her to make generalisations valid across different instances. Moreover, the identification of these categories should provide some explanatory power and is often taken to be informative about the world.

Nevertheless, the ubiquity and variety of kinds brought the status of these categories into question. Philosophers have started to ask first what these categories are and second which of them can be deemed to be *natural* and which instead instrumental or artificial (Khalidi 1998, 2013). In order to explore the topic properly, it is important to distinguish two different enterprises, one concerning the kinds themselves and one concerning the naturalness of the categories that refer to them. More precisely, we can summarise two main questions that reflect the discussion on natural kinds (Magnus 2015, 2018):

- An *ontological question* that asks which kind of entity, if it is an entity at all, a natural kind is. And an answer might come from a theory of *sui generis* universals, cluster of properties or similarity between instances¹⁰.
- A *naturalness question*¹¹ that asks how we can recognise an arbitrary category from one that captures some genuine divisions in nature. What must a category do in order to “carve nature at its joints”?

In this paper, we are inquiring into whether genes are something more than a useful tool to do genomics and we will focus on the second question in order to do so. This allows us to explore the status of the category without considering universals, essences or

¹⁰ For an overview on the debate on universals consider Bird, Tobin 2022.

¹¹ This question can also be referred to as the “taxonomy question”. However, I prefer to avoid this terminology as it seems to constrain natural kinds to biological taxonomies or to identify kinds with taxonomical classifications.

questions within fundamental ontology¹² (Magnus 2018). We can identify two broad strategies to answer the natural question. A first is offered by conventionalism: there are no natural kinds, but only conventions that suit different purposes. A second is offered by a form of realism: at least some of our categories correspond to natural kinds and the objective features of reality. It is common to take an hybrid position. Some kinds that we find in scientific practice or daily life are conventions while others might be correspondent to real features of the world. For the sake of the present analysis, I will assume a form of minimal realism for which at least some categories correspond to genuine features of reality and might be candidates for natural kinds (Khalidi 2013; Bird, Tobin 2022). However, not all the categories that we identify as kinds seem to correspond to such divisions. How do we distinguish “natural kinds” from mere “human categories” though (Khalidi 2013)?

In the biological sciences, this question is particularly relevant as the study of life appears to be the reign of taxonomies and classifications. Historically, species have been taken to be a paradigmatic case of kinds and the concerns on whether all of the *Linnean taxa* correspond to actual divisions in reality have been widely debated¹³. In order to identify natural categories, some accounts have been proposed, among which are essentialism in intrinsic and historical forms, HPC cluster theories and others. However, most of them are concerned with the question of whether any of the *taxa* can be considered a natural kind (Slater 2013). This makes them interesting for the species or higher taxonomies debate but makes them less applicable to other candidates of natural kinds in biology, such as kinds in biochemistry or at other levels (Slater 2013, Khalidi 2013, Kistler 2018). In this paper, we are exploring a category that comes from the domain of genetics and we are asking whether unions of genomic sequences identified as a gene can be considered instantiations of a natural kind or not. Thus, in order to explore an answer to this case of the naturalness question, let me present what it takes for a category to be natural.

¹² See Magnus definition of *deep realism* (2018), or Khalidi definition of *Realism* (2013) for which one has to commit to the existence of some fundamental categories.

¹³ A summary of the status of the species controversy can be found in Ereshefsky (2017).

3.1 Natural kinds

The debate on natural kinds is wide in approaches and topics. Nevertheless, there is a consensus on the fact that kinds should be those that allow to make reliable explanations and predictions across instances (Bird, Tobin 2022). This feature is represented and clearly discussed, together with other accounts, in Khalidi's approach (1993, 1998, 2013, 2018, 2020, 2021). His view presents kinds as those categories present in scientific theories that are projectible and capture nodes in causal networks of the world. This approach stands out because it is a realist account, for which kinds track objective features of reality while at the same time avoiding excessive metaphysical commitments. Moreover, it is applied and applicable to a variety of kinds from the physical and special sciences, being able to take into account both structural and historical or etiological properties.¹⁴ Accordingly, I take his view as a starting point to explore whether the gene could be deemed a natural kind. While my argument is conditional upon such account, I do not think that this compromises its validity as this account has commitments on projectibility and causal efficacy, which are often considered valid criteria for natural kinds (as in Bird, Tobin 2022). Let me present Khalidi's account in more detail.

Khalidi is interested in exploring a way to answer the natural question and to identify and distinguish conventional groupings from objectively existent ones. He starts off with a form of *weak or moderate realism (r)* for which kinds are objective features of reality, but not corresponding to distinct metaphysical categories¹⁵. This means that nature has some joints as objective features of the world (whether we know them or not), and our best theorising should aim at carving up these joints. However, not all of our categories are natural kinds and some can be considered mere conventional grouping. His account aims at providing a way to identify *natural* kinds. To do so, he claims that we should start

¹⁴ Khalidi's account has also been applied to a variety of kinds from the life sciences, such as viruses, cancer cells, biological species and ADHD, making this approach even more suitable to consider the gene case (Khalidi 2013, 2021). The validity of Khalidi's view of natural kinds is discussed also in Tahko (2022).

¹⁵ This view is contrasted with a stronger form of *Realism* in which kinds correspond either to *sui generis universals* or to *second-order universals*.

looking at our best science and scientific practice within both natural and social sciences (2013, 2021). Specifically, he sustains a form of moderate naturalism, for which natural kinds can be taken to be some of the categories revealed by our systematic attempts to gain knowledge of nature. Taking a realist stance towards the discipline, science aims at identifying kinds that are really existent in the world, and not mere and useful theorisations and it has proved so far rather successful in doing so. So, if we want to disentangle and identify the *natural* kinds among all the categories, then a good starting point is to look at different sciences. Moreover, the combination of a weak form of realism and a form of naturalism allows us to clarify an important aspect of the theory of natural kinds: science or the philosophy of natural kinds does not invent natural kinds, but rather discovers them. Scientific theories or established knowledge do not determine the existence or the non-existence of kinds, but rather they represent our *access and guide* to the existence of such kinds (Khalidi 2013).

Nevertheless, even within the best scientific theories, the history and philosophy of science have shown that not all categories present in the discipline can be considered as capturing something in the world. For instance, a category like *hysteria*, which has been used as a scientific category in the past, has proved not to be a natural kind of diseases and was abandoned as a kind (Khalidi 2013, 59). Moreover, some categories can have an instrumental role or cannot be considered stable or robust enough to be really informative about the world. This led Khalidi to add two further requirements for naturalness¹⁶. These are i) genuine projectibility and ii) being a node in a causal network.

Projectibility is often assumed as a feature that natural kinds should display. Kinds are particularly efficacious categories when it comes to framing inductive inferences and they feature in many empirically verifiable generalizations. This means that kinds are projectible, in the sense that they can be projected from one instance to another in a successful way. More precisely, the projectibility of a natural kind can be defined as follows: “when it comes to a natural kind predicate *K*, there is no shortage of other predicates, *P*₁, *P*₂, . . . ,

¹⁶ These are based on a re-elaboration of the simple theory of natural kinds proposed by Craver (2009) for which kinds refer to the causal structure of the world.

P_n , and so on, such that we can reliably assert that if x is K , then x is P_1 , x is P_2 , ..., x is P_n and we can do so with a high degree of generality” (Khalidi 2018, 1385). Kinds provide explanatory and predictive power across different contexts and circumstances because they allow us to project a set of properties from one instance to the other and to predict that such properties will be present (Khalidi 2013, 2018, 2020, 2021).

Moreover, this kind of projectibility requires an explanation: why is it possible to draw these inferences? What is the ontological ground for which the kind can be applied in an explanatory way to many instances? Khalidi answers these questions by adding a second requirement that natural kinds should satisfy: they are “nodes” in causal networks. Projectibility results to be a “reflection” of the causal network in which *instances* of the kinds are involved and some kind categories are particularly successful because the properties of the natural kinds are *causally clustered* (Khalidi 2013, 2018, 2020).¹⁷ The joints that natural kinds carve so successfully are those that can be found in the causal structure of the world. Together with providing an ontological reason for the projectibility of kinds, causal relations also provide an answer to the naturalness question. Specifically, they play two main roles in distinguishing natural from conventional kinds. First, natural kinds do not only present a set of projectible properties but a set of properties that are hierarchically ordered as “causes and effects in recurrent causal processes” (Khalidi 2018). They present a set of “core” properties that cause the instantiation of other properties of the instance of a given kind. Natural kind results then to be those categories with a set of properties discoverable by science and whose co-instantiation *causes* the instantiation of other properties (Khalidi 2013). Second, natural kinds are those categories that represent nodes within broader causal processes: they are causally efficacious on other kinds and are inter-sections within the webs of causal relations. The causal relations among property instances and the causal cores of natural kinds represent the ontological principle in virtue of which we can distinguish natural from unnatural kinds. Moreover, it is the underpinning of the projectibility of such categories.

¹⁷ It is important to notice that the that natural kinds display role in the causal network can be seen when considering instances of the kinds due to the nature of the causal relation. I thank Jessica Wilson for suggesting this important clarification and reviewer 2 for insisting on this aspect.

To further elucidate the account, let me consider briefly an instance of a natural kind presented by Khalidi: the case of viruses (2013, 180). Viruses represent an established category of a proper sub-discipline, virology. Virions, individual particles of the virus, are characterised by the identification of some synchronic causal properties. First, they are protein particles (more or less complex) containing a genome capable of making an mRNA readable by the ribosome of a host cell. Second, the virions of a given virus display a specific infectious cycle comprised of well understood causal relations. Following Khalidi's methodology, we need to consider whether viruses present the three main features of naturalness: i) being present in scientific theories; ii) being projectible; iii) being nodes in causal networks. The first criterion is easily satisfied, as our analysis started with a scientific category present in an established enough discipline. Let me consider the other two. The infection cycle and the life cycle of viruses are understood in terms of common and repeatable causal processes that display a given hierarchy of relations. These causal processes allow virologists to make empirical generalisations on the different instances of viruses and their cycle. Such generalisations can then be projected from viruses already observed to the ones that have not been. This projectibility is broad and stable across contexts: different kinds of viruses in different circumstances present a cycle that is reducible to the general causal one identified by virology. This makes virus a projectible kind: it allows explanatory and predictive power across different instances of the kind. Furthermore, projectibility is sustained by the fact that viruses can be taken to be appropriate nodes in causal networks. First, the core properties of the virus, such as being a protein containing a genome capable of making an mRNA readable by a host cell, *cause* the instantiation of other properties, such as a given infection rate and behaviour within the host cell. This orders the properties that can be ascribed to the virus category according to a causal hierarchy. Second, viruses enter into causal interactions in a uniform or similar pattern and they have a causal impact on the network of relations in which they are embedded. Viruses are causally efficacious categories within the network of relations they are involved in. Accordingly, the category virus is a natural kind as it is the object of a successful part of contemporary science, it is a projectible category and such Projectibility is based on causal relations. Specifically, the properties that pertain to the kind

as a whole are causally related to each other's and lead the members to enter into causal interactions in a uniform way.

The theory presents some advantages compared to the ones previously analysed. First, this account is a good example of the consensus reached by philosophers on how to answer the naturalness question, providing us with a way to disentangle natural from unnatural kinds. This is done without embracing an essentialist theory that tries to provide a specific set of properties for which an individual is a member of a kind. Moreover, it is an application of a form of reflective equilibrium between scientific input and philosophy. It is a combination of convictions on categories generally regarded as paradigmatic kinds, often taken to be stable categories in scientific theories, philosophical discussions on natural kinds and a set of considerations that are drawn from scientific practice. Lastly, this account can be applied to a variety of different types of kinds. It is able to accept etiological and historical kinds as natural kinds, considering a particular origin or genealogical history as the core properties that cause the instantiation of other ones. Moreover, the combination of naturalism with projectibility and causality allows the theory to be applied to concrete case studies within fundamental physics and the special sciences, such as lithium, cancer cells, viruses, ADHD (examples from Khalidi 2013).

Concluding, this account answers the naturalness question and will be used in the next section to explore whether genes can be considered a natural kind.

4. The gene as a natural kind

The naturalness of the gene category has been questioned because of its aforementioned history, the complexity of the genetic phenomena and the context-dependency that is implied by genes' functional aspect. Genes might not seem neat enough to count as a natural category or might be too ambiguous. In this section, I am going to argue for the opposite, defending a view for which the gene can be deemed a natural kind. Before pursuing my thesis, some clarifications are in order. This is because the genomic and biochemical domain is rich in systems of practice and taxonomies and clarity is particularly needed to

avoid ambiguities. According to Havstad (2016, 2021) we need to consider three classificatory practices concerning kinds: classificatory characterization or definition, individuation and organisation. The first focuses on the definition of the kind, already presented in §2.1. The second focuses on identifying which tokens belong to a given kind. The last focuses on organizing taxonomies. Accordingly, before assessing the naturalness status of the gene category, I will consider briefly gene classifications and taxonomies. Then, in §4.2, I will explore whether the gene can be considered a natural kind.

4.1 Genes classifications and taxonomies

As presented in section §2.1, the gene is defined the context of the ENCODE analysis as "a union of genomic sequences encoding a coherent set of potentially overlapping functional product" (Gerstein et. al 2007). According to this approach, genes are those (not necessarily contiguous) sequences that have a specific function in encoding a given molecular product. The classification of tokens or instances of genes into types or families is mostly done according to the product: unions of sequences encoding the same product (or products that are similar enough) are clustered into the same type of genes¹⁸. The specific function of a gene within transcription is what allows to classify them¹⁹. Moreover, such classification of genes is done at different levels. First, there is a broad classification of genes into two functional subtypes: i) genes that encode regulatory RNAs that play different functions within cellular processes; ii) genes that encode an RNA for the amino acid sequence of a polypeptide (Perini 2011). Then, we can find more specific classifications of genes tokens into gene types according to the given molecular product of the gene under consideration. Tokens of the same type of genes are clustered together if they encode the same molecular products or molecular products that are considered similar enough (Gerstain et al. 2007; Fogle 2010; Griffiths and Stotz 2013). In line with the importance of the link between genes and products, protein-coding genes (those encoding an RNA for a polypeptide) are normally named after the protein they encode. In

¹⁸ For the importance of functional similarity of products in genes classification, see Fogle 2010.

¹⁹ Despite the importance of the functional aspects of the genes, genes are not mere functional kinds, because the material component they present as union of genomic sequences is also relevant in determining their identity.

this classificatory system, the *locus* is indicated often in italic capital letters while the name of the gene in capital letters and the protein made in normal characters. An instance of gene classification based on DNA sequences and function is the one of the genome of *Saccharomyces cerevisiae*, sequenced in 1996 and constituted of about 6,275 genes, organized on 16 chromosomes. Of these genes, about 5,800 are identified by their function²⁰. For example, the gene DCS2 encodes the protein dcs2, which takes part in the biological processes that regulate the response of the cell to heat (Liu et al. 2017). Generalising, the link between genes and products, together with the consideration of the relevant union of genomic sequences, is what normally allows gene individuation and gene-talk.

As far as other taxonomies or nomenclatures are considered, genes are classified in a variety of ways often referred to as gene ontologies. In order to aid scientific practice, these ways of classifying the genes have been grouped within the project *The Gene Ontology*, a resource that provides a computational representation of our current scientific knowledge about the functions of genes according to the functions of the product they encode. Overall, this project considers three aspects when classifying genes: i) molecular functions performed by gene products, ii) cellular components in which the gene product performs a function, iii) biological processes in which gene products are involved. For example, the gene for the product “cytochrome c” can be classified according to the molecular function of the product (oxidoreductase activity), the biological process (oxidative phosphorylation), and the cellular component (mitochondrial matrix) (Gene Ontology²¹). Lastly, it is possible to find different ways to organise genes in taxonomies contingent on specific contexts and given scientific goals. For example, in evolutionary developmental biology and comparative genomics, genes are often classified in terms of evolutionary history (Muller 2003). However, these taxonomies tend to be highly-context relative and presuppose the aforementioned characterisation of the gene. First, scientific practice clusters tokens of genes across species into the same type according to a given union of sequences and a given function in encoding a product that takes part in molecular and

²⁰ For the database on *S. cerevisiae* genome and the relevant articles see www.yeastgenome.org.

²¹ For the Gene Ontology www.geneontology.org.

cellular processes. Then, thanks to the similarity of the sequence and the RNA encoded, they can be further classified in terms of evolutionary similarities. An example is constituted by genes that can be classified as *homologous* if they are inherited by two different species from a common ancestor and such classification is done in terms of sequences and RNA encoded.

To conclude, genes can be classified and organised in taxonomies according to different practices and in a variety of ways. Nevertheless, this classification presupposes an initial identification of the gene as a given union of genomic sequences that encodes an amino-acid sequence. Once the gene is individuated, then it can be classified for further purposes. This is possible thanks to a general characterisation of the gene concept, as that union of genomic sequences that are transcribable in a target molecule (Bellazzi 2022). Such characterisation of the gene is the one relevant for the purpose of the paper.

4.2 The gene *is* a natural kind

Having summarised how genes can be organised in taxonomies, I will now move on to consider whether the category gene itself can be deemed a natural kind. To do so, I will start considering whether the properties associated with the natural kind category respect the requirements for naturalness presented in §3.1.

In section §2, I have pointed out the concept of gene relevant for our analysis. Genes result to be the union of sequences transcribable in a target molecule and they play their function only in a wider system of interactions and environments (Griffiths Stotz 2013). According to this definition, the gene presents a two-fold identity: in terms of structure or composition and in terms of function. First, the gene has a structural component as a region or union of regions of nucleic acids, while not every region of the given nucleic acid is a gene (Fogle 2010, Griffiths Stotz 2013). Specifically, the gene is composed of those regions that can be considered images of the target molecule and are actively involved in transcription. This leads us to the second aspect of the gene. It is the union of sequences that has a given function in encoding the primary structure of a polypeptide or of a functional RNA molecule. The functional component of the identity of the gene is

evident during transcription, a reactive process that can be up-regulated and down-regulated thanks to specific intra-, inter- and extra- cellular interactions. This makes the gene a context-dependent entity and a multiply-composable one (Germain et al. 2014, Bellazzi 2022). Specifically, a given instance of the gene kind can be composed of different stretches of nucleic acids while the function characterising the gene is maintained (that is the final transcribed molecule).²² This feature is important to account for the complexity of genetic phenomena. Accordingly, to understand what a gene is, one needs to consider: first, the identification of those sequences that are transcribable in a target molecule and, second, the process of transcription where the gene plays its function. The identification of the main features of the gene category is the starting point in the analysis of whether the gene can be considered a natural kind.

Now, we need to explore whether these properties respect the three main requirements for naturalness: i) naturalism, the presence of this category in our best scientific theory; ii) projectibility, whether the category has explanatory power and can be projected from one instance to another; iii) whether the gene is a node within a causal network.

Let me start with naturalism. Despite their complicated history and the current challenges raised by postgenomic analysis, genes still retain an important role across a different variety of life sciences. Specifically, they are often invoked in the study of protein synthesis and the impact that alterations in protein structure can play at the cellular and organismal levels. Moreover, genes still retain a role in developmental and evolutionary biology. The gene is a category that figures across different scientific disciplines in a stable way, thus meeting the first requirement for naturalness. Genes also present projectibility and being nodes in causal networks. First, we can project the properties of the genes from one instance to another in a successful way. This is particularly evident in protein synthesis, one of the phenomena in which genes play an important role. In this case, scientists project

²² In order to appreciate how the definition allows for multiple realisability and composition, it is important to consider that the gene *type* is the one that can be multiply composed, while each gene *token* is going to be composed by specific nucleic acids sequences. Accordingly, for the gene definition we just need to identify which specific token nucleic acids could compose it while maintaining that the relevant function is realised. For further analysis of the complexity of the relations between the properties of the gene see Bellazzi 2022.

the two main properties of the genes from one instance of protein synthesis to another in order to be able to explain and predict general patterns. At the same time, protein synthesis shows how genes are nodes in causal network: they have a causal role in the synthesis of proteins and the two properties are causally related for this to happen.

To further support my argument and appreciate why genes respect the natural kinds requirements, I will consider a specific case study: the aforementioned gene DCS2 in *Saccharomyces cerevisiae*. This gene encodes the protein dcs2, which takes part in the biological processes that regulate the response of the cell to heat (among other processes) (Liu et al. 2017). Furthermore it presents clearly the two properties of the category “gene” (- being composed of nucleic acids and having a function -), making it a good case study to see how these properties support the naturalness of the kind.²³

Following the characterisation of the gene presented here, this gene is transcribable in a given target molecule, the mRNA for the protein dcs2, and i) it has a specific basis on the chromosome XV of the cell; ii) it has a specific function, i.e. the one of encoding the protein dcs2.

Let me now then consider if this gene is projectible, that is if it allows better explanations and predictions based on the causal properties it presents. The gene DCS2 encodes the mRNA for the protein dcs2, a regulatory protein that acts as a pyrophosphatase regulator in many cellular processes and specifically in the heat response of the cell²⁴. The instances of this gene have a common (type) nucleic acid-base, which makes the relevant gene identifiable on a precise locus on the chromosome XV of the cell, and they encode a given protein within the right circumstance. Also, instances of this gene take part in the process of transcription which leads to the presence of the protein within the cell, thereby taking part in a causal process. The same role is played by distinct instances of this gene across different cell individuals and the identification of this causal pattern allows to better

²³ The purpose of this section is not to show that “DCS2” is a gene, but rather that the properties that genes display (and are present in DCS2) respect the requirements for the naturalness of kinds. Moreover, the causal component required by the naturalness claim supports using an instance of the kind to explore whether it is natural. I thank reviewer 2 for insisting on this clarificatory point.

²⁴ Further reference can be found on yeastgenome.com.

predict and explain heat regulation processes. Moreover, this gene can be deemed homologous to the gene DCS2 in other species such as *Homo sapiens* or *Mus musculus* allowing this category to be projectible not only across instances of *Saccharomyce cerevisiae* but also to other species. Accordingly, the gene DCS2 is projectible in virtue of having the properties of the kind “gene”. This example illustrates how it is possible to project the category “gene” from one instance to another in order to make reliable explanations and predictions about genetic phenomena.

Let me consider now whether genes can be considered nodes in causal networks. First, we need to see if the core properties of the gene can be considered causally related. Second, we need to consider whether the gene plays a causal role within a broader context. I will do so with the help of the previous example. The two properties of the gene DCS2 are i) being composed of a given union of sequences and ii) encoding for the protein DCS2. The two seem to be at least partially causally related as the sequence allows for the interactions necessary for the transcription of the DNA sequence into the mRNA that then encodes the protein dcs2. The "cross-talk" between the sequence, RNA polymerases and the actors of transcription is what causes the second definitional property of the gene to be present. The sequence has a causal role in the manifestation of the functional property together with the relevant factors, allowing the encoding for a specific protein. This supports the presence of a causal link between the two main properties of the gene and allows a hierarchical causal characterisation of the other properties that might be an effect of those two. Furthermore, the gene under consideration plays a role within the general causal network. For instance, the encoding of the protein dcs2 contributes causally to the heat regulation within the cell. This can be generalised even further given that genes can take part in a variety of causal phenomena. First, and most evidently, as aforementioned, they represent a node in the causal network that is identifiable in protein synthesis. Second, contemporary molecular biology has found out that genes can encode a variety of RNA molecules that play different roles. This makes the gene a node within a series of cellular processes. As follows, the gene also respect the second requirement: they have causally related properties that, within the right systems of interactions, can cause further phenomena. Lastly, having considered the features of a specific gene to support the

argument does not compromise the generality of the conclusion because the discussed properties of the gene DCS2 are those characteristic of the kind “gene” - being composed of nucleic acids and being transcribable in a target molecule – and not those specific of DCS2. Accordingly, similar considerations can be made for other instances - even more complex ones – in virtue of the two core properties of the kind.²⁵

In conclusion, the gene category captures a natural kind according to the requirements for the naturalness of a category. First, this category has an important role within our scientific theories, respecting a naturalist approach. Second, it is possible to identify two main properties that are important for gene definition. These properties allow its identification and that can be used as a basis for the analysis of naturalness. The gene category is projectible, as its definitional properties have a predictive and explanatory role, and scientists can project results about one instance to other instances on the base of its core properties. Moreover, these properties can be deemed to be causally related with each other’s and the gene can be considered a node in a causal network in different processes.

5. Conclusions

In this paper, I have argued that the gene can still be retained as a natural kind within the postgenomic context, once we accept Khalidi’s view of natural kinds . In the first part, I have briefly illustrated the history of the gene and why it might be considered a controversial category. The discoveries that followed the ENCODE project have shown that it is difficult to identify the gene with a precise stretch of DNA and that it is important to consider a full system of interactions to understand the complexity of genetic phenomena. A possible reaction to such controversy is to take a nominalist view of the gene and consider it as a conventional or instrumental category. However, the success of the genes

²⁵ I thank both reviewer 1 and reviewer 2 for the suggestion to clarify the relation between the example of gene considered in the paper and the general kind “gene”. The definition of gene defended in the paper allows the generalisation from simple to more complex case studies, as the definition presented here is compatible with both multiple realisation and composition. Accordingly, we can generalise it to cases such as those involving alternative splicing because it is possible to identify the two core properties of the kind “gene” also in those cases (see Bellazzi 2022 for more discussion on this).

category across different sciences and its wide applicability can be considered a hint in support of the naturalness of the kind.

In order to argue in favour of the genes as a natural kind, I have illustrated in section §3 what makes a given category *natural*. To do so, I have used Khalidi's account as a starting point for the analysis. Specifically, a kind is said to be natural when it respects three requirements a) it figures in our scientific theories; b) it is projectible; c) it is a node in the causal network. Then, in section § 4, I have illustrated why the gene is a natural kind. This has also brought in necessary considerations on genes taxonomies and classifications into genes type and tokens. The definitional properties of the gene are individuated mostly synchronically and respect the criteria proposed for the naturalness of the kind. First, the gene category figures in scientific theories, respecting the naturalist requirement. Then, the individuated properties of the gene allow the members of the kind to be part of generalisations and projections concerning the category. This is possible because the properties are causally related between each other's and allow the instances of the kind to enter into causal interactions that are uniform and identifiable.

Concluding, while the argument is conditional upon the acceptance of Khalidi's view, a realist approach to genes as natural kinds has some benefits. First, if we accept that one of the aims of science is to discover what kinds of things are in the world, then an argument in favour of the naturalness of the gene category supports the success of genetics. Second, having a realist or instrumentalist view towards a scientific category might have an impact on scientific practice and scientific discoveries. An example of this can be found in the history of the gene itself, summarised at the beginning of this paper. The realist understanding of the gene supported by Muller offered a theoretical framework for which scientists started looking for the gene as a material entity rather than a mere theoretical instrument. This contributed to the discovery and identification of the gene as specific stretches of DNA or as at least located on the DNA. A realist understanding of the gene category seemed to have had an impact on the direction of research. Accordingly, the consideration of the gene as a natural kind can have an impact on how scientists think about this category, even if they might not change scientific practice on a daily basis. Lastly, conceiving the gene a category that "carve nature at its joints" rather than a mere

24

instrument might bring clarity to the debate and offer a broader framework for future research.

References

- Beurton P E (2010) A unified view of the gene, or how to overcome reductionism. In: Beurton P. J., Falk R., Rheinberger H.J (eds) *The Concept of the Gene in Development and Evolution*. Cambridge University Press, Cambridge, p 286-314.
- Bellazzi F (2022) The emergence of the postgenomic gene. *European Journal for Philosophy of Science*. 12, 17.
- Bird A, Tobin E (2022). Natural kinds. In: Edward N. Zalta & Uri Nodelman (eds) *Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), <https://plato.stanford.edu/entries/natural-kinds/>.
- Burian R M (2004) Molecular epigenesis, molecular pleiotropy, and molecular gene definitions . *History and Philosophy of the Life Sciences* 26 (1-Genes, Genomes, and Genetic Elements): 59-80.
- Craver C (2009) Mechanisms and natural kinds. *Philosophical Psychology*, 22:5, 575-594.
- Clark D, Pazdernik N (2012) *Molecular Biology*. Elsevier Science Technology, Academic Cell.
- El-Hani C N (2007) Between the cross and the sword: the crisis of the gene concept. *Genetics and Molecular Biology*. 30 (2). S. Paulo.
- El-Hani C N (2015) Mendel in genetics teaching: some contributions from history of science and articles for teachers. *Science & Education* 24:173-204.

- Ereshefsky M (2017). Species. In: Edward N. Zalta & Uri Nodelman (eds) Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/entries/species/>
- Falk R (2010). The gene - A concept in tension. In: Beurton P. J., Falk R., Rheinberger H.J (eds) The Concept of the Gene in Development and Evolution. Cambridge University Press, Cambridge, p 317-348.
- Fletcher H L, Hickey I G (2012). BIOS Instant Notes in Genetics. Fourth Edition Garland Science.
- Fogle T (1990). Are genes units of inheritance? *Biology and Philosophy* 5: 349-371.
- Fogle T (2010). The dissolution of protein coding genes in molecular biology, In: Beurton P. J., Falk R., Rheinberger H.J (eds) The Concept of the Gene in Development and Evolution, Cambridge University Press, Cambridge, p 3-25.
- Fox Keller E (2000). The Century of the Gene. Harvard University Press, Cambridge MA.
- Gericke N M, Hagberg M, Carvalho dos Santos V, Joaquim L M, El-Hani C N. (2014). Conceptual variation or Incoherence? Textbooks discourse on genes in six countries. *Science and Education* 23: 381-416.
- Germain P L, Ratti E., Boem F (2014). Junk or functional DNA? ENCODE and the function controversy. *Biology and Philosophy* 29: 807-831.

- Gerstain M B, Bruce C, Rozowsky J S, Zheng D, Du J, Korbel J O, Emanuelsson O, Zhang Z D, Weissman S, Snyder M (2007). What is a gene, post-ENCODE? History and updated definition. *Genome research* 17: 669-681.
- Griffiths P and Stotz K (2006). Genes in the postgenomic era. *Theoretical Medicine and Bioethics* 27:499–521. DOI 10.1007/s11017-006-9020-y
- Griffiths P and Stotz K (2013). *Genetics and Philosophy*. Cambridge University Press, Cambridge.
- Hall B K (2001). The gene is not dead, merely orphaned and seeking a home. *Evolution and Development* 3 (4): 225-228.
- Havstad J C (2016). Protein Tokens, Types, and Taxa. In: C. Kendig (eds) *Natural Kinds and Classification in Scientific Practice*, Routledge, New York.
- Havstad J C (2021). Complexity begets crosscutting, dooms hierarchy (another paper on natural kinds). *Synthese* 198: 7665-7696.
- Khalidi M A (1993). Carving nature at the joints. *Philosophy of Science*, 60 (1), 100-113.
- Khalidi M A (1998). Natural kinds and crosscutting categories. *The Journal of Philosophy*, 95(1), 33-50.
- Khalidi M A (2010). Interactive kinds. *The British Journal for the Philosophy of Science*, 61 (2), 335-360.
- Khalidi M A (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press, Cambridge.

- Khalidi M A (2018). Natural kinds as nodes in causal networks. *Synthese* (2018) 195: 1379–1396

<https://doi.org/10.1007/s11229-015-0841-y>

- Khalidi M A (2020). Are Sexes Natural Kinds, In: Dasgupta S. & B. Weslake B. (eds.) *Current Controversies in Philosophy of Science*. Routledge, New York.
- Khalidi M A (2021). Etiological kinds. *Philosophy of Science*, 88, 1-21.
- Kistler M. 2018. Natural kinds, causal profile and multiple constitution. *Metaphysica*. I (19) 1 De Gruyter. DOI: <https://doi.org/10.1515/mp-2018-0006>.
- Liu W, Li L, Ye H, Chen H, Shen W, Zhong Y, Tian T, He H. (2017). “From *Saccharomyces cerevisiae* to human: The important gene co-expression modules”. *Biomedical Reports* 7: 153-158.
- Magnus P D (2015). “John Stuart Mill on Taxonomy and Natural Kinds”. *The Journal of the International Society for the History of Philosophy of Science* (5): 269-280.
- Magnus P D (2018). Taxonomy, ontology, and natural kinds. *Synthese* 195: 1427-1439
- Meyer L M N, Bomfim G C and El-Hani C N (2013) How to understand the gene in the 21st century. *Science and Education* 22 (2): 345-374.
- Mitchell S D (2008). Exporting causal knowledge in evolutionary and developmental biology. *Philosophy of Science* 75 (5): 697-706.

- Okasha S. 2019. *Philosophy of Biology: a very short introduction*. Oxford University Press, Oxford.
- Perini L (2011). Sequence matters: genomic research and the gene concept. *Philosophy of Science* 78 (5): 752-762
- Rheinberger H S, Muller-Wille S and Meunier R (2015). Genes. *Stanford Encyclopedia of Philosophy* (<https://plato.stanford.edu/entries/gene/>)
- Scherrer K, Jos J. (2007). The gene and the genon concept: a functional and information-theoretic analysis. *Molecular System Biology* 3: 1-11.
- Slater M H (2013) Cell types as natural kinds. *Biological Theory* 7 (2): 170-179
- Tahko T (2022). “Natural kinds, mind-independence and unification principles”. *Synthese* 200:144, 1-23.