The relativity of biological function

ARTICLE *in* THEORY IN BIOSCIENCES · OCTOBER 2015

Impact Factor: 1.23 · DOI: 10.1007/s12064-015-0215-5

4 AUTHORS, INCLUDING:

Manfred Laubichler

Arizona State University

128 PUBLICATIONS **761** CITATIONS

SEE PROFILE

Peter F. Stadler

University of Leipzig

665 PUBLICATIONS **26,125** CITATIONS

SEE PROFILE

Katja Nowick

University of Leipzig

21 PUBLICATIONS **945** CITATIONS

SEE PROFILE

Theory Biosci.
DOI 10.1007/s12064-015-0215-5

SHORT COMMUNICATION

The relativity of biological function

Manfred D. Laubichler[1,2,3] · Peter F. Stadler[4,5,6,7,8,9] · Sonja J. Prohaska[10] · Katja Nowick[11]

Received: 2 May 2015 / Accepted: 3 September 2015

Abstract Function is a central concept in biological theories and explanations. Yet discussions about function are often based on a narrow understanding of biological systems and processes, such as idealized molecular systems or simple evolutionary, i.e., selective, dynamics. Conflicting conceptions of function continue to be used in the scientific literature to support certain claims, for instance about the fraction of "functional DNA" in the human genome. Here we argue that all biologically meaningful interpretations of function are necessarily context dependent. This implies that they derive their meaning as well as their range of applicability only within a specific theoretical and measurement context. We use this framework to shed light on the current debate about functional DNA and argue that without considering explicitly the theoretical and measurement contexts all attempts to integrate biological theories are prone to fail.

Keywords Biological function · ENCODE · Biological theory · Coarse graining · Theory integration

K.N. is supported by the Volkswagen Foundation within the initiative "Evolutionary Biology". This work has also been supported by the Templeton Foundation (grant # 24332) and NSF SES Grant No. 1243575.

✉ Peter F. Stadler
studla@bioinf.uni-leipzig.de

Manfred D. Laubichler
manfred.laubichler@asu.edu

Sonja J. Prohaska
sonja@bioinf.uni-leipzig.de

Katja Nowick
katja@bioinf.uni-leipzig.de

[1] School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

[2] Marine Biological Laboratory, Woods Hole, USA

[3] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

[4] Bioinformatics Group at the Department of Computer Science, Interdisciplinary Center of Bioinformatics, LIFE - Leipzig Research Center for Civilization Diseases, and German Center for Integrative Biodiversity Research, University Leipzig, Härtelstraβe 16-18, 04107 Leipzig, Germany

[5] Max-Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany

[6] Fraunhofer Institut für Zelltherapie und Immunologie – IZI, Perlickstraße 1, 04103 Leipzig, Germany

[7] Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Wien, Austria

[8] Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, 1870 Frederiksberg C, Denmark

[9] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

[10] Computational EvoDevo Group, Department of Computer Science and Interdisciplinary Center of Bioinformatics, University Leipzig, Härtelstraβe 16-18, 04107 Leipzig, Germany

[11] TFome Research Group at Department of Computer Science, Paul-Flechsig-Institute for Brain Research, and Interdisciplinary Center of Bioinformatics, University Leipzig, Härtelstraβe 16-18, 04107 Leipzig, Germany

Published online: 08 October 2015

Introduction

In light of available whole-genome data, we see a recent reemergence of a debate about "functional DNA". Specifically, the question "how much of the (human) genome has an identifiable function" is discussed controversially. Estimates range roughly from 5 to 90 % (The ENCODE Project Consortium 2004, 2012; Graur et al. 2013; Kellis et al. 2014; Ponting and Hardison 2011). Such divergence cannot be reconciled by more accurate data. Rather it reflects dramatic disagreements about the proper definition of function. In this debate, scientists frequently refer to philosophical positions in support of their claims (Wouters 2003). For instance, a number of papers attacking the conclusion of the ENCODE project that most of the genomic DNA is functional refer to the Selected Effect (SE) concept of function in their refutation of this claim (Graur et al. 2013; Doolittle 2013; Eddy 2013). The SE concept of function defines the function of an item or trait as the effect for which it was selected by natural selection or by which it is maintained (Wright 1973; Millikan 1989; Neander 1991; Griffiths 1993). This definition of a "proper function" asks for the "purpose" of an item: "What is it supposed to do" or "What is it good for". The contributors to the ENCODE project, on the other hand, use their measurements of transcribed and modified genomic regions as proxy for a functional role. They derive their definition of function from the Causal Role (CR) conception (Cummins 1975; Craver 2007). CR function emphasizes *what* an entity does, rather than why it does what it does. The controversy is thus based on different theoretical and measurement assumptions, rather than verifiable observations. Contrary to our colleagues, we see this not as a problem, but as evidence for a more fundamental property of biological concepts: their inherently context-dependent nature.

Ambiguities related to biological function

From the perspective taken here, each of the two major concepts of function discussed in the philosophical literature, the SE and CR account of function, has advantages and shortcomings. To illustrate this claim, we will discuss three concrete examples and demonstrate how each account only captures part of the underlying biological reality.

1. True novelty is well documented for protein coding genes (Bornberg-Bauer et al. 2010; Tautz and Domazet-Lošo 2011; Neme and Tautz 2013), microRNAs (Hertel et al. 2006; Meunier et al. 2013), and regulatory elements (Otto et al. 2009; Bradley et al. 2010). In all these cases (seemingly), random DNA sequences start to produce new products that fortuitously are placed into a pre-existing regulatory network context. The innovations have well-defined CR functions by virtue of their influence on the organism's phenotype. The SE concept, however, runs into a serious problem: Novelty must first have come into existence before selection can act on it. Thus, SE, in its strict form, cannot capture innovations that have not been exposed to natural selection yet, because it insists on selection in the past as a prerequisite for present function. It is often argued that the "time-gap" before selection can begin to act is short enough to be negligible. We propose here that instead of attempting to rescue strict SE function by semantic acrobatics, the underlying biological realities are captured better by embracing different, context-dependent notions of function.

2. Cocaine presumably has evolved as an insecticide in coca plants (*Erythroxylum spp.*) (Nathanson et al. 1993). Human neurons have evolved receptors to recognize endogenous neurotransmitters, yet, they can bind cocaine as well and subsequently elicit a signaling cascade leading to well-documented physiological and psychological effects. Clearly, neither the coca plant's secondary metabolism that synthesizes cocaine nor the structure of mammalian neurotransmitters has evolved for these drug effects. Taking a strict SE point of view, the drug effect of cocaine is not a biological function of either cocaine or the receptors. Under a CR conception, a functional attribution can be made in each context. The cocaine example exposes the fact that biological function, in practice, is dependent on the context under consideration. For a junkie, the function of cocaine is to cause a high; for the coca plant, the function of cocaine is to kill insects that attack it. We argue that neither point of view is a priori more valid then the other.

The very same argument, whether or not the item under consideration is under selection for its (CR) function in a different context, can be made for the (CR) function of a genomic locus or a gene product in geriatric diseases (which for all we know are not subject to selection in the human population).

3. Some macroRNA genes, e.g., mouse Air and human Airn are consistently transcribed. It appears, however, that the RNA product (which is also not translated) is irrelevant; instead the act of transcription itself is crucial in this case for the maintenance of a genomic imprinting (Latos et al. 2012). Here, the issue is the theory invoked to discuss the function of Airn. The Airn locus has no discernible selection on its nucleotide sequence, hence escapes a research program grounded in SE function in practice. A CR-based interpretation would tend to assign function to the gene product, which is also not correct. This example exposes that the concept of function requires a theoretical framework, that in the case of Airn needs to allow

functions, and eventually a selective effect, to be assigned to the process of transcription, rather than a gene product or a DNA location as such.

The role of theory and measurement

As we have seen in our examples all conceptions of biological function include references to (1) a specific type of causal interaction; (2) a specific theoretical context T, such as natural selection, biochemical reaction or regulation; and (3) a corresponding way M of measuring the theoretically specified effect in the system in question. We can thus formalize biological function as $y = f(x; T, M)$. Within a specific research field, the theoretical assumptions T and measurement procedures M generally remain implicit and unspecified as its practitioners understand and share these assumptions.

The measurement context M is much more than just a description of measurement procedure. Rather, it entails particular data types and their interrelations, estimates of accuracy, and a theoretical conception on how the measured data related to the reality of the biological systems. For transcriptome sequencing with NGS methods, for example, sequencing reads are interpreted as fragments of RNA sequence that can be evaluated quantitatively. The theoretical context T, on the other hand, encapsulates a priori assumptions on the mutual relationship of biological concepts. Again using NGS as an example, it entails in particular a model of all the processes involved in transcription and RNA processing. Scientific insights can be deduced only when the measurement context M and theoretical context T are consistent.

The examples also show, however, that these assumptions are often not shared across research fields and that these implicit contextual assumptions are no longer perceived as such, rather they are elevated to the status of unquestionable truth.

The continued debate about the "proper" definition of biological function is thus at the same time pointless (as there is no single right conception of function) and useful (as it has revealed in some detail the specific theoretical assumptions and measurements underlying each conception). For example, it is clear that SE function has a different theoretical framework than CR function, which focuses on mechanistic interactions. At the same time, each conception deploys a different type of measurement that allows to distinguish functional from nonfunctional objects and processes. Attributing biological functions is therefore an example of theory-driven coarse graining within biological theory formation. For instance, the question which phenotypic effects have to be measured, and which would then define a unique, well-grounded notion of biological function utterly depends on the theoretical perspective. Most controversies revolve around the primacy of either evolutionary, or more precisely selective interpretations or causal-mechanistic (mostly molecular) descriptions. The debates triggered by the interpretations of the ENCODE data are a case in point.

The ENCODE controversy

The ENCODE Consortium concluded in their main paper that "80.4 % of the genome is functional". Stated in this form, this claim is as devoid of meaning as competing ways of measuring function necessarily yield different percentages. The statements that (1) "80.4 % of the genome take part in detectable molecular interactions and processes such as transcription and histone modification" (The ENCODE Project Consortium 2012), (2) "8.2 % of the genomic DNA is constrained" (w.r.t. to nucleotide substitutions, insertions, and deletions) (Rands et al. 2014), (3) "13.6 % of the genome is under stabilizing selection for RNA secondary structures elements" (Smith et al. 2013), and (4) "accounting for turnover a steady state value lies between 10 and 15 %" (Ponting and Hardison 2011), all report empirical observations transformed by specific—and different—theoretical models that specify the criteria under which a particular nucleotide belongs to the "positive set" of functional DNA in each study. Barring experimental errors, all of them can be simultaneously true. Not surprisingly, the lowest numbers are obtained using rather stringent signatures of stabilizing selection at nucleotide level, while the highest numbers combine biochemical effects such as transcription or specific histone modifications and integrate over a larger number of individual experiments.

The "ENCODE controversy", see, e.g., (Graur et al. 2013; Kellis et al. 2014; Graur et al. 2015) for pointed expositions of the different points of view, thus has its root in the simple, tacit, and erroneous assumption that there is a unique notion of biological function that is independent of a theoretical context and a frame of measurement, and that everybody necessarily has to agree on.

More generally, the CR and SE concepts are incompatible only if one insists that SE function necessarily has a purposive aspect and requires a "proper history". In a weaker form, the SE concept assigns functions not only to items that have been selected in the past but to all items that are selectable in principle. In this version, SE function becomes a CR endowed with a very specific form of measurement, namely (natural) selection. Whether this SE–CR function is a productive concept depends on the underlying theoretical framework. It is perfectly adequate in the context of constraints of sequence evolution. At the

same time, it reduces to the tautology "what is functional is selectable, and what is selectable is functional" in the context of de novo innovations.

Context dependency and theory integration

Our brief analysis of biological function in the everyday practice of biology revealed that all applications are context dependent and that the relevant context includes both theoretical assumptions and a specific type of measurement procedure. As a consequence, each biological object can have different functions or no function at all depending on the theoretical and measurement context.

While the theoretical context of biological concepts, such as function, is frequently discussed, less attention is paid to the role of measurement, i.e., the assignment of numbers to attributes of the natural world. Of course biologists are well aware that biological objects are not found as such in nature, that there is no natural kind of a biological species or of a gene, but these discussions generally do not include the specific formal aspects of measurement. Rather they refer to measurement problems implicitly, for example, when discussing the limits of the biological species concept and its measurement criteria when going beyond sexually reproducing organisms (Wilson 1999; Wheeler and Meier 2000; Hey 2006; Levy 2010). The same also holds true for any of the gene concepts widely used in current biology, see, e.g., (Gerstein et al. 2007; Gingeras 2007; Prohaska and Stadler 2008).

This suggests that what we need is an appropriate application of measurement theory in biology (as reviewed by Houle et al. 2011) and a more adequate conception of biological function and other biological concepts that explicitly reflect this theoretical and measurement dependency. One consequence of the context-dependent nature of biological concepts is a pluralistic approach. Pluralism in this context must not mean "anything goes", a position that has been advocated by some critics of science (based on a misunderstanding of an earlier philosophical discourse) (Dupré 1993; Rosenberg 1994; Galison and Stump 1996; Feyerabend and Oberheim 2011; Galison and Stump 1996). Rather, we need a constrained pluralism that makes explicit a specific and relevant theoretical context within which a concept can be clearly and unambiguously defined and the relevant quantities can be precisely measured.

This level of specificity in theory building is indispensable in particular for all attempts to integrate the results of different investigations. Several areas of the life sciences, especially those that generate large amounts of data through specific measurements, e.g., the Cancer Genome Atlas and Brain Atlas projects, are now trying to connect their data with other relevant domains. These attempts can only be successful if both the theoretical and measurement contexts are not only clearly specified but also become an explicit part of the integration efforts. The ongoing misunderstandings connected with the interpretation of the ENCODE data can serve as a warning of how a failure to address these issues can result in highly unproductive debates.

Integration of data, or measurements, is, of course, a highly theory-driven problem and is thus connected to issues of theory development in the life sciences (Krakauer et al. 2011). Despite some progress, the life sciences are still a rather under-theorized science. Taking some inspiration from physics, Krakauer et al. identified the issues of coarse graining as one of the basic problems for theory development in the life sciences. But not all cases of theory integration can be solved simply by coarse graining of data and measurements because not all problems are simply an issue of scale and resolution. In each case, theory integration requires us to link different "units of biological theory," defined as collections of data, theory/models and measurement, with each other. This cannot be done without making explicit the various constraints related to theoretical and measurement contexts that are often hidden within data models. Focusing on these formal aspects of data, theory, and measurement is thus an important facet of theoretical biology.

Acknowledgments We thank Jan Aerts and Steve Hoffmann for critical reading of the manuscript.

References

Bornberg-Bauer E, Huylmans AK, Sikosek T (2010) How do new proteins arise? Cur Opp Struct Biol 20:1–7

Bradley RK, Li XY, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB (2010) Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. PLoS Biol 8(e1000):343

Craver C (2007) Explaining the brain: mechanisms and the mosaic unity of neuroscience. Oxford University Press, New York

Cummins R (1975) Functional analysis. J Philos 72:741–765

Doolittle WF (2013) Is junk DNA bunk? a critique of ENCODE. Proc Natl Acad Sci USA 110:5294–5300

Dupré J (1993) The disorder of things: metaphysical foundations of the disunity of science. Harvard University Press, Cambridge, MA

Eddy SR (2013) The ENCODE project: Missteps overshadowing a success. Current Biol 23:R259–R261

Feyerabend P, Oberheim E (2011) Tyranny of science. Polity Press, Cambridge, UK, Malden, MA

Galison P, Stump DJ (1996) The Disunity of science: boundaries, contexts, and power. Stanford University Press, Stanford, CA

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, post-ENCODE? history and updated definition. Genome Res 17:669–681

Gingeras TR (2007) Origin of phenotypes: Genes and transcripts. Genome Res 17:682–690

Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E (2013) On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol 5:578–590

Graur D, Zheng Y, Azevedo RB (2015) An evolutionary classification of genomic function. Genome Biol Evol 7:642–645

Griffiths PE (1993) Functional analysis and proper functions. Br J Philos Sci 44:409–422

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. The Students of Bioinformatics Computer Labs 2004 and 2005 (2006) The expansion of the metazoan microRNA repertoire. BMC Genomics 7:15

Hey J (2006) On the failure of modern species concepts. Trends Ecol Evol 21:447–450

Houle D, Pelabon C, Wagner G, Hansen T (2011) Measurement and meaning in biology. Q Rev Biol 86:3–34

Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford G, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigó R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RC (2014) Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA 111:6131–6138

Krakauer DC, Collins JP, Erwin D, Flack JC, Fontana W, Laubichler MD, Prohaska SJ, West GB, Stadler PF (2011) The challenges and scope of theoretical biology. J Theor Biol 276:269–276

Latos PA, Pauler FM, Koerner MV, Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338:1469–1472

Levy A (2010) Pattern, process and the evolution of meaning: species and units of selection. Biosci 129:159–166

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H (2013) Birth and expression evolution of mammalian microRNA genes. Genome Res 23:34–45

Millikan RG (1989) In defense of proper functions. Philos Sci 56:288–302

Nathanson JA, Hunnicuttt EJ, Kantham L, Scavone C (1993) Cocaine as a naturally occurring insecticide. Proc Natl Acad Sci USA 90:9645–9648

Neander K (1991) Functions as selected effects: the conceptual analyst's defense. Philos Sci 58:168–184

Neme R, Tautz D (2013) Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics 14:117

Otto W, Stadler PF, López-Gialdéz F, Townsend JP, Lynch VJ, Wagner GP (2009) Measuring transcription factor binding site turnover: A maximum likelihood approach using phylogenies. Genome Biol Evol 1:85–98

Ponting CP, Hardison RC (2011) What fraction of the human genome is functional? Genome Res 21:1769–1776

Prohaska SJ, Stadler PF (2008) Genes. Biosci 127:215–221

Rands CM, Meader S, Ponting CP, Lunter G (2014) 8.2 % of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet 10(e1004):525

Rosenberg A (1994) Instrumental biology, or the disunity of science. University of Chicago Press, Chicago

Smith MA, Gesell T, Stadler PF, Mattick JS (2013) Widespread purifying selection on RNA structure in mammals. Nucleic Acids Res 41:8220–8236

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702

The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. Science 306:636–640

The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74

Wheeler QD, Meier R (2000) Species concepts and phylogenetic theory: a debate. Columbia University Press, New York

Wilson RA (1999) Species: new interdisciplinary essays. MIT Press, Cambridge

Wouters AG (2003) Four notions of biological function. Stud Hist Philos Sci Part C Stud Hist Philos BiolBiomed Sci 34:633–668

Wright L (1973) Functions. Philosophical Rev 82:139–168