Theory Biosci. (2009) 128:171–177
DOI 10.1007/s12064-009-0069-9

SHORT COMMUNICATION

Response to commentaries on our paper gene and genon concept: coding versus regulation

Klaus Scherrer · Jürgen Jost

Received: 6 June 2009 / Accepted: 15 June 2009 / Published online: 26 September 2009

We have been glad to see that our paper (Scherrer and Jost 2007) solicited such insightful or supportive commentaries as those of Noble, of Gros, of Prohaska and Stadler, of Forsdyke, and Billeter, as well as the alternative proposal of Stadler et al., and we hope that this will trigger further conceptual discussions about the definition of the gene and inspire further research about programs of gene expression, in the light of recent advances in molecular biology and bioinformatics (Billeter 2009; Forsdyke 2009; Gros 2009; Noble 2009; Prohaska and Stadler 2008; Stadler et al. 2009). The commentaries raise some important issues. We agree with some of them, but disagree with others, whereas still others reflect terminological decisions that could be taken so or otherwise. In the sequel, we shall try to address these issues in a systematic manner and motivate the terminological decisions that we have taken. This will also give us the opportunity to emphasize some points that were not explicitly laid out in our original paper.

Since the initial work of Gregor Mendel in the middle of the nineteenth century and the introduction of the term *Gene* by Wilhelm Johannsen in 1909, the basic idea of the concept was to express a discrete unit of a heritable phenotypic property, or perhaps better, a heritable unit of some discrete phenotypic property. As a discrete unit, it can be modified by discrete events, the mutations. The units are passed on to offspring in sexual reproduction via recombination. They are not inherited entirely independently, but the phenomenon of genetic linkage suggested already to Thomas Morgan a linear arrangement. As became clear with the advances of biochemistry, what is inherited is not the trait itself, but rather pieces of DNA that encode information about the biochemical substrate and for the production of that phenotypic trait under specific intra- and extracellular circumstances. Since about 1960, those pieces of DNA were often taken for the genes themselves, and these genetic data then are annotated in the framework of—more or less—suitable ontologies. The phenotypic traits themselves are caused by proteins that themselves are complexes of polypeptides, or complexes of proteins, or by other functional molecules in the cell among which small functional RNAs are presently receiving particular attention. In order to account for emerging biological complexity, the definition of the gene has undergone several substantial changes over the last 100 years. Nevertheless, (Gerstein et al. 2007) request that an updated definition of the gene should attempt to be backward compatible with the earlier ones. For us, this means that we should address the three issues of inheritance, coding, and function that in one way or another underly all thinking about the gene concept, and analyze how or to what degree earlier models are compatible with present molecular biological knowledge and experimental and bioinformatical capabilities.

In our paper, we have therefore analyzed the difficulties of previous gene definitions and proposed one that takes the above requirements into account to a degree that is feasible, compatible with present molecular biological and bioinformatical knowledge, and logically consistent. With this aim we have separated product information (the gene) from information regulating its expression; to designate this program the term of *genon* was introduced. We should also emphasize that our concepts of gene and genon are

K. Scherrer (✉)
Institut Jaques Monod, CNRS, 2 place Jussieu,
75251 Paris Cedex 05, France
e-mail: Scherrer@ijm.jussieu.fr

J. Jost
MPI Mathematik in den Naturwissenschaften,
Inselstraße 22-26, 04103 Leipzig, Germany
e-mail: jjost@mis.mpg.de

essentially based on eukaryotic cells and inspired by the recent progress in understanding eukaryotic gene regulation. Therefore, our concepts do need certain modifications or adaptations when applied to prokaryotes, as we hope to develop in a future paper.

Clearly, the *gene* is one of the most basic concepts of biology, perhaps the fundamental concept that constitutes biology as a science different from chemistry. The rough idea underlying classical gene concepts, as briefly summarized above, was that some structure that can be inherited or is transmitted to offspring by structure copying encodes some biological function. Consequently, a gene should be neither viewed as a structure nor as a function exclusively but, because what is at stake is a relation between a structure and a function. This relation includes the information for a product, at the basis of a function, as well as for the regulative program governing its expression. Therefore, we do not see why we should be forced into the alternative developed by Prohaska and Stadler between a structural and a functional gene definition. In any case, a purely structural definition would risk to only capture the molecular aspects, and not arrive at what makes a cell a biological system. An exclusively functional definition, in contrast, would forget that the functions are coded for by heritable structures.

Modern molecular biology has shown, however, that this relationship between structure and function is not as straightforward as originally thought. While it is a testimony to the genius of Gregor Mendel that he arrived at the essence of this relationship, this has the consequence that a modern gene definition can no longer be as easy or simple as originally thought. The structure that is inherited has to be transformed by specific programs into the structure that encodes a function. This transformation is complex, but we try to capture its essence by the genon concept. It seems to us that this concept captures the essential aspect of gene regulation much better than the general distinction between inputs and parameters suggested by Stadler et al. because it separates information for products and programs necessary for their expression.

As Forsdyke points out, mutations can yield some discriminatory criterion, in the sense that whether some piece of structure contributes to a gene can be checked by whether the function of that gene will be affected by a mutation of the structure. He quotes many good examples, but in our opinion, his examples precisely suggest that one should distinguish between those mutations that change the functional product by varying the coding information and those that affect the programs of its expression. In our terminology, a mutation can change a gene or affect its genon. In principle, a single mutation could also do both, as the genon motives can be superimposed onto the coding sequence. From our point of view, however, this distinction between the two possible effects of a mutation is fundamental. Therefore, we find that our distinction between product and program is at the same time conceptually simpler and offers deeper insight into what is really at stake than his suggestion to define a gene at 3 levels [(1) The DNA sequence that is transcribed. (2) The latter plus the immediate 5′ and 3′ sequences that, when mutated, specifically affect the function. (3) The latter two, plus any remote sequences that, when mutated, specifically affect the function]. His 3 levels may be well designed for his mutation test, but do not distinguish between the two different possible (but not necessarily mutually exclusive) effects of mutations that we want to distinguish.

The discussion by Forsdyke of "placeholder-bases" in the genetic material is appreciated as it nicely illustrates and complements an important point underlying our genon concept. A particularly pertinent example of Forsdyke is the occurrence of glycine–alanine repeats in an EPV protein without influence on the amino acid sequence that, however, interferes with antigen processing. This might be a perfect example of mutation in the program of the genon; in this case program information seems to have preference over product information. Somehow the effect on antigen processing may play not on gene function but on when, where and how the same protein has to be made. He also points out the existence of mutations beyond the apparently transcribed areas which do not bear on the protein product. At this point, we remind the reader that the transcribed area no longer seems to correspond and, in fact, never did correspond to the areas designated by ordinary Northern blot and microarray techniques. The ENCODE project has shown that most of the genome is transcribed, as we already pointed out in our paper on the base of earlier data but, furthermore, the recently developed technique of RNA-Seq show that all the upstream and downstream areas Forsdyke is talking about are transcribed after all. Indeed, it is likely that most of the attachment sites of transcription factors are transcribed and, therefore, mutation may bear on RNA expression and correct processing and transport of the derived RNA at precise times and places in space. The placeholder concept is also in line with the role of intronic sequences and mutations therein and the fact that, in some cases, parts of exonic sequences occur in introns, stopped by termination codons, and that some intronic sequences show higher conservation than the neighbouring exons. All these phenomena can easily be attributed to the genon and its precursors at pre-mRNA and DNA level.

In fact, in our opinion, a basic problem of earlier gene definitions was that they neglected the expression process that links the coding in the DNA with the functional molecules (in a sense to be made precise below) in the cell. In other words, previous gene definitions were about DNA and proteins, but did not involve RNA taking part in the

expression process. For us, as already at the onset of molecular biology in the sixties, a gene is a sequence consisting of nucleotides that is either directly functional or codes for a functional product. In the case that we have discussed in most detail in our paper, when the functional product is a polypeptide, the coding sequence then is realized in the mRNA prior to its translation into the polypeptide. This coding sequence is assembled from transcribed fragments of DNA in a process that arises from the interaction of factors from outside that sequence with specific signals or motifs in the sequence to be processed itself, the collection of which we call the pre-genon carried by the pre-mRNA and, encoded in the corresponding genomic domain, the information of the proto-genon. Our central concept, the genon, is the final product of this process in the mature mRNA; in fact, it is the program that controls the expression of a specific gene as a product. We thus strictly separate information for product and regulation of its expression process.[1]—In that direction, we find the definition proposed in (Gerstein et al. 2007) of a gene as "a union of genomic sequences encoding a coherent set of potentially overlapping functional products" too vague. If one were to adopt that definition, a gene would neither correspond to a unit at the coding nor at the functional level, but only relate two possibly rather diffuse sets of biomolecules.

Concerning the expression process at molecular level, the question of Billeter about the real length of the primary transcripts is most pertinent in the context of actual molecular biology and genetics. As he mentions, for a long time the TATA box was supposed to mark the start of transcripts. Questionable for a long time, to the same extent as promotors and transcription factor attachment sites, the latest results of genomics and in particular the already mentioned ENCODE project show that genomic domains of DNA are fully transcribed. In other terms, all these signals including the TATA boxes, the latter being placed immediately upstream of the first exon of many genes, are transcribed and may, hence, not serve as transcription starts; defined as signals involved in gene expression, they may rather serve as sites involved in RNA processing. The basic mechanism of full domain transcription does not exclude that, in particular cases, mRNA-size molecules are transcribed, as is the case of pseudo-genes inserted in the DNA by reverse transcription or some integrated viral genomes. Furthermore, the length of the primary transcripts is not trivial in the frame of this discussion because it defines the extent of the pre-genon. Since according to latest results, at one time or another, 95% of the genome is transcribed, it follows that the regulative information by far exceeds information for products. This points to an eventual solution for a long-term paradox and may explain the apparently absurd size of the eukaryotic genomes.

In order to arrive at a gene definition that on one hand reflects the contribution of the expression program to the construction of a functional product and on the other hand is still compatible (at least to some degree) with the original intentions and the prior formulations of the gene concept, we needed to put certain limits on what we count as a gene. A basic limit is that we take into account only events up to the synthesis of the polypeptide, as the basic unit of function, and exclude post-translational processes as, e.g. chemical modification and formation of higher order complexes, as pointed out in the commentary by Denis Noble.

Furthermore, it is important that our concept only reflects *genetic* (nomen est omen) inheritance, or more precisely the transmission of genomes or parts of genomes, including somatic modifications like methylation patterns, and not other aspects of epigenetic inheritance in general. This comes about because the elements that ultimately lead to the coding sequence in the mRNA prior to translation into a polypeptide are derived from specific nucleotides in the DNA, as are the regulatory contributions of the genon and its metabolic precursors. As such, that is, as constituents of a coding sequence or as contributions to its expression materialized in the biochemical identity of specific nucleotides, they can be transmitted to offspring by genetic inheritance. The situation is different for the ensemble of factors controlling a specific genon in an mRNA, the basic transgenon, however. The latter includes various biochemical agents that affect the expression of a particular gene, but that need not all be genetically inherited. Examples are cofactors as vitamins or metals, acting at the level of enzymes or regulative proteins, or small molecules acting as allosteric effectors that modify the status of factors involved in regulation.

In any case, however, since a gene is assembled in a sequence of steps during the expression process from various pieces that originate from stretches of DNA, for us, a gene, while heritable, is not a *unit of inheritance* in the sense of being a unit of genetic transmission. More precisely, the precursor pieces at DNA level are inheritable, because they are replicated in genetic transmission to offspring, but as the phenomena of DNA rearrangement under the influence of antigens or other somatic effectors, or alternative splicing show, the way how these pieces are combined into a gene depending on the expression process is not necessarily genetically inherited. Perhaps it is deplorable that a gene is no longer a physical unit of

[1] We note that this is a conceptual distinction, not a material one, as a sequence of coding triplets may well be at the same time the attachment site for some regulatory protein or RNA. We do not see a logical problem in making such a conceptual distinction. In particular, we prefer this to the distinction between input and parameters as proposed by Stadler et al. which we find too general and vague.

inheritance, that is, a contiguous piece of DNA. We believe, however, that our analysis has shown that this is inevitable as there is no simple one-to-one correspondence between pieces of DNA and functional molecules in the cell. In fact, our effort centers about the problem to conceptually account for the relation between the two via the expression process, with contributions both from the genon, coming from the same region in *cis*, and the holotransgenon, including all factors and influences external to the *cis* region in question. In other words, one can have one or the other, a unit of inheritance (in the sense of genetic replication or transmission) or a unit of function (affecting the phenotype), but not both. In the light of present research in molecular biology, we have opted for the latter (with an important caveat, as shall be clarified below), and not the former. Of course, there exist many examples where a unit of inheritance is at the same time a unit of function, like certain small RNAs, but the point is that for understanding cellular biology, we need to separate the two aspects. Anyway, it seems that, about this point, there is little disagreement between Prohaska and Stadler and us. At least, we read their proposal of genes as heritable elementary functional units in the same sense, namely, that they think of a unit of function without implying that this has to be a unit of inheritance at the same time. The proposal of Stadler et al. to make the various fragments at DNA level part of the definition of a gene then links the gene definition more tightly to inheritance, but for the reasons just described, a gene in their sense cannot be a unit of inheritance either.

Another aspect is that, in evolution, selection operates on phenotypic functions and not directly on replicated molecules. Therefore, since, as we have discussed above, function and inheritance in the sense of DNA replication are not congruent, one might not even expect to be able to identify a useful unit of inheritance in terms of replicated genetic material. Recombination mixes the genetic material of the parents, and mutations can affect individual nucleotides, but may also consist of large scale reorganization of the DNA. Individual nucleotides are too small to have functional significance by themselves, and also, changes of individual nucleotides, as well as other mutations, may be functionally neutral. In order to reconcile our definition of a gene as the sequence coding for a function with the issue of replication of DNA which is important for phylogenetic analysis, we could in principle project the coding sequence at mRNA level back to the DNA. In that case, a gene would become a collection of DNA fragments that can be bound together through an expression process in a coding mRNA, or some other functional RNA. Because of alternative splicing and other phenomena, different such genes would in general not be materially disjoint, but rather overlap. We have therefore refrained from taking that step, but our analysis has developed the tools for such backtracking, that is, for relating a coding or functional RNA and its corresponding pieces of DNA.

The alternative proposal of Stadler et al. to include in the definition of a gene not only the function, but also what they call the genomic footprint, that is, the fragments at DNA level out of which the functional sequence is assembled during the expression process is certainly worth of a careful consideration. We have opted instead to make this assembly part of our analysis instead of our gene definition. In fact, we distinguish between the forward analysis of what becomes of a fragment at DNA level, that is, in which functional products it can be represented how often, and the backward analysis of tracing the origins at DNA level of the coding sequence at mRNA level underlying a gene; backward analysis allows us also to trace back programming information contained in the genon. Either choice seems reasonable to us. Stadler et al. achieve a really comprehensive gene concept, which, however, may be somewhat complex in practice, whereas ours may allow for a more flexible analysis and easier application of information theory as we have described in our paper.

In any case, our conceptual scheme also leads to the question whether there do exist mechanisms that ensure the coordinated transmission of all those pieces and regulatory motives at DNA level that together constitute a gene and its genon. At pre-genon level the question arises as to the evolutionary significance of bundles of genes co-transcribed into one pre-mRNA and separated by differential processing. We are presently investigating this question. On one hand, we study combinatorial mechanisms for the coordinated expression of specific sets of genes. On the other hand, we develop new conceptual tools for analyzing the regulatory and functional significance of spatial arrangements.

Still another, at least equally important, limit that we had to impose concerns what we count as a basic function in our definition. Stadler et al. propose a general definition for the function of an object (a biomolecule in the case of interest here) as the set of input–output relations, or more precisely, transformations, in which it participates as a parameter. One then has to be careful to avoid circularity resulting from defining a parameter in terms of its functions in input–output processes. Stadler et al. distinguish inputs as being traces of encoded output letters from parameters. Since in our conceptual framework, a (proto-, pre-)genon can be superimposed to a coding sequence, the underlying string of nucleotides would then have to count both as an input and as a parameter in their sense. They then speak of an autocatalytic reaction when an object appears both as an

input and as a parameter in the same input–output relation. We think that our more specific terminology captures the essential point here much better.

We are clearly aware that from a physiological perspective, our notion of function is inadequate, as Noble points out. Physiological functions typically arise from the cooperation of several proteins and other biologically relevant molecules. Polypeptides are only the building blocks of proteins. So, why do we still consider the mRNA coding sequence for a particular polypeptide as the paradigmatic unit of our analysis? Well, having dispensed of the issue of inheritance, the gene then expresses the relation between coding and basic function in the cell. As already discussed, in line with the historical development of the concept and because this can be clearly distinguished from other forms of coding, we restrict ourselves to coding by nucleotides.[2] At RNA level, a specific sequence of nucleotides can directly be functional—an issue to which we need to return—or consist of triplets coding for a sequence of amino acids, that is, a polypeptide, or both, because nothing prevents an mRNA coding for a polypeptide to also have some other function, for instance regulating the expression of other RNAs. In either case, since this derived from some coding nucleotides in the DNA via the process of transcription, RNA processing, splicing and other modifications, we consider this as coding for a genetic function, and this then is where our notion of a unit of function is originating. Thus, more precisely, we should speak of a coding unit of function, or of a *unit of coding for a function.*

Without this modification, we loose the coherence of the gene concept. In particular, our concept of a gene deliberately excludes all post-translational processes and modifications, like protein folding, with or without the assistance of chaperones. Likewise, for instance, lipids and their biosynthesis are not contained in our concept. Lipids are biosynthesized through the activity of certain proteins, and this process is not directly coded for by nucleotides, but only indirectly, as those proteins are the result of such coding. Also, DNA by itself can have some functional role, for instance for the spatial arrangement of gene regulation as exposed in the unified matrix hypothesis of Scherrer or the solenoid model of Kepes; but again, we decided to not include that in our gene concept—even though there exist important connections. The reason for this exclusion is that, apparently, there is no transcription step involved, or only indirectly via, e.g., a matrix protein recognizing DNA motifs. Whereas the functional RNA is derived and assembled from pieces of coding DNA, so that there is a non-trivial relation between coding and function mediated through an expression process,[3] there is no such mediation for the functional DNA. Therefore, here no concept establishing a direct relation between coding and function via the expression pathway is needed, and we then do not speak of a gene.[4] This implies that there is genetic information transmitted by DNA without implying a gene directly, for instance the mere DNA *length* in between sites where regulatory proteins or RNAs may attach. In our opinion, rather than trying to stretch the concept of a gene to include all functionally relevant molecular structures in the cell, it is better to limit the gene concept by the requirement of coding through nucleotides, and to formulate new concepts for other types of systematic relationships between molecular structures and cellular functions. In this direction, we are presently working on the conceptualization of the functional roles of spatial arrangements in the cell.

For the protein coding genes, we took the polypeptide as the unit of function (and our concept then requires that we take any sequence coding at mRNA level for a polypeptide as a gene, even though there do exist polypeptides that have no cellular function and simply arise as the by-product of some regulation mechanism, as Peter Stadler pointed out to us). For the directly functional RNA genes, it is not as easy to come up with a coherent definition of function. Since experimental research on regulation by small RNAs, for instance, is presently in rapid expansion, we can offer at best some tentative proposal. That would consist in considering as a functional RNA to which we assign a gene any RNA that regularly occurs with a precise sequence identity in a given cell. Thus, for instance when an RNA segment has a mere spacer function so that only its length, but not its composition is relevant for a specific cellular task, we do not assign it a gene. This emphasizes the coding aspect at the expense of the structural one. Furthermore, when the function of an RNA segment is the recognition of some regulatory RNA or protein, this does not represent a gene for us, but rather contributes to a genon. In any case, these are terminological decisions which are not all strictly logically necessary and which

[2] The code for the interaction between nucleotide combinations and regulatory proteins or RNAs is relevant for the genon concept, but this is not at issue here.

[3] We fully agree with Prohaska and Stadler about the importance and relevance of functional RNAs, even though we have not yet worked out the details for RNA coding genes to the same extent as for protein coding ones.

[4] In this sense, the concept of a "genomic phenotype" suggested by Bernardi and Bernardi, as quoted by Forsdyke, may be taken into consideration but even though possibly correct in principle, does not seem so helpful in our opinion. Of course, the spatial arrangement of the genome in a functional cell can be considered as a phenotypic character, but it is not conditioned by the DNA per se, being determined, as everything else, from the interaction of structural physical laws, gene products, and external factors; but here, the role of individual genes is particularly difficult to analyze.

therefore could be made differently (and Stadler et al. assign a function only to those physical objects that influence the transformation of other objects), but which are motivated by the desire to arrive at a concept that takes into account distinct biochemical structures and the relations between them. This relation extends our gene concept beyond an exclusively functional definition.

But how can we then account for actual biological functions in the cell that are based on the cooperation of several polypeptides or genes in our sense? Well, the answer should be obvious. What might be an elementary function at the physiological level can nevertheless be structurally composite. The task in analyzing such a function then is twofold: Identify the separate constituents and describe their interactions. We admit that sometimes this approach can encounter serious difficulties, when a process produces its own constituents. Nevertheless, in the situation relevant for the present discussion, the constituents emerge from the expression of a coding sequence and as such can be identified independently of the processes they are involved in. Our concern then is the identification of the constituents, as opposed to their interaction networks.

This brings us also to another issue raised by Prohaska and Stadler as well as by Gros. Genes are not expressed in isolation from each other in a static environment, but rather in turn interact with and modify their environment and may regulate each other's expression. Thus, obviously one should think of an interactive regulatory network of gene expression. True enough. One may consider this, however, as a second step, the first one (the one that we were concerned with in our paper) being the identification, description and analysis of the elements that constitute such networks. Again, such a procedure may not always be possible, and many networks constitute their own elements through their operations. Here, however, again we can independently identify the elements and follow the pathway from the pieces distributed across the DNA to the final uninterrupted coding sequence at mRNA or functional RNA level. The various steps in the pathway may interact and interfere with each other, and cannot always be clearly distinguished from each other. Thus, the cascade of regulation as described in our paper represents an abstraction (as does the concept of a network, for that matter), but we do not see a conceptual problem here. Here, we view it as an operative elementary system included in such networks. For instance, genon and pregenon and their respective contributions may well overlap. The RNA molecule at the various stages of the expression process can interact with itself, for instance directly through RNA folding, or indirectly through the cellular environment, by causing or inhibiting the production of elements needed for later stages, or simply by using up certain cellular resources. As the existence of retrogenes shows, sometimes the expression process can also be reverted into the opposite direction.

Obviously, our analysis of the expression of individual genes needs to be complemented by an analysis of the regulatory interaction of different genes. In fact, it seems that our conceptual approach is also useful for modelling the coordinated co-regulation of ensembles of genes. In this direction, presently, we are working out the co-regulation of groups of genes via the combinatorics of the genon and the *trans*-genon. In more abstract terms, it remains to be seen whether our analysis of the temporal process of the expression of single genes can be complemented by an analysis of the simultaneous interaction of several genes and genons that are possibly at different stages of expression. In our opinion, however, the regulation and expression of individual genes, and the interactions between different genes and their regulation are complementary aspects, and in contrast to Prohaska and Stadler, we do not think that it is useful to play these aspects out against each other. Here, perhaps the comparison with a rather different science, linguistics, might be insightful. Already de Saussure clarified the relationship between diachronic and synchronic approaches to linguistic phenomena. One can analyze the diachronic development of the pronunciation and the meaning of words, or one can investigate the synchronic relationship between the words in a language. Also, for the latter, one can consider the syntagmatic relationship and the functional interaction between the different words in a sentence, or one can consider the paradigmatic relationship between those words that can assume the same syntactic or semantic role in a given sentence. Similar principles constitute the basis of the structuralist approach to phonology. The former is about mutual influences, as in regulatory networks, the latter is about mutual exclusion as in the expression of individual genes. In linguistics, one therefore needs, and linguists have developed concepts for both aspects individually, instead of requesting that a single concept should capture both of them simultaneously.

Also, a gene, as represented by the mRNA coding sequence prior to translation into a polypeptide or by a functional RNA sequence, and its genon need not be materially distinct. One and the same nucleotide in an RNA can contribute both to the coding and the regulation. Nevertheless, since these are distinct roles and since these roles are (usually) exercised at different times, they can be conceptually separated. The same also applies to the fact known at least since the operon model of Jacob and Monod that DNA regulatory elements (which would be part of a proto-genon in our terminology) may well be contained in transcribed regions.

Since we do not identify a gene with any kind of locus in the DNA determined by spatial proximity or any other

criterion apart from the sequence coding for the final functional product, we do not see any difficulty with *trans*-splicing or similar phenomena. The coding sequence in the mRNA is assembled from different pieces with different provenience in the DNA, as also emphasized by Stadler et al. These pieces may or may not be contained within a single ORF. The characteristic aspect of our formal analysis is that it operates in both directions, forward, by asking to what functional products a specific piece of DNA is contributing, and backward, by asking where the pieces that together constitute a sequence coding for a functional product are originating from in the DNA. Therefore, in most cases, a gene in our sense cannot be identified with a localized stretch of DNA. We realize that this poses a problem for gene annotations as these are typically based on DNA sequence analysis. This problem, however, is not the result of arbitrary terminological decisions on our side, but is already made inevitable by the mechanism of alternative splicing. In fact, already the old phenomenon of giant transcripts discovered by Scherrer poses problems for gene annotations, and this issue is now receiving attention in the context of the ENCODE project. Therefore, Stadler et al. also address this issue in detail.

The problem that remains is the analysis of the relation between the linear arrangement of coding and noncoding pieces of DNA and the regulatory networks that produce the functional molecules in the cell, an issue raised by Prohaska and Stadler and by Gros. Our analysis hopes to lay some conceptual foundations that will help in this direction, but we are certainly aware that this is a real problem that cannot be solved by terminological proposals alone.

This point was not extensively commented on, although by necessity underlying this discussion, but it may foster further investigation and comprehension of still enigmatic facts concerning genome and gene expression. In conclusion the question may be asked to what extent the commentaries received and the discussion arising shall influence further elaboration of the gene and genon concept. A major point not contested seems to be the possibility to apply information-theoretic analysis to gene expression on the basis of a separation of information for product and regulation, although our mathematical elaboration was not discussed extensively in the comments received. Within the frame of actual molecular genetics it may be important that the gene and genon concept gives hints to eventual comprehension of the size of genome and transcripts in higher eukaryotes.

Acknowledgments We also thank Peter Stadler, Martin Billeter and François Gros for insightful discussions about our arguments; since we did not agree on each point, however, they are not responsible for the views presented here.

References

Billeter MA (2009) Comments on "Gene and Genon Concept" by K. Scherrer and J. Jost. Theory Biosci. doi:10.1007/s12064-009-0068-x

Forsdyke DR (2009) Scherrer and Jost's symposium. The gene concept in 2008. Theory Biosci. doi 10.1007/s12064-009-0071-2

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

Gros F (2009) Comments on the paper by K. Scherrer and J. Jost. "Gene and Genon" concept: coding versus regulation. Theory Biosci. doi:10.1007/s12064-009-0070-3

Noble D (2009) Commentary on Scherrer and Jost (2007) Gene and genon concept: coding versus regulation. Theory Biosci. doi: 10.1007/s12064-009-0073-0

Prohaska SJ, Stadler PF (2008) Genes. Theory Biosci 127:215–221. doi:10.1007/s12064-008-0025-0

Scherrer K, Jost J (2007) Gene and genon concept: coding versus regulation. A conceptual and information-theoretic analysis of genetic storage and expression in the light of modern molecular biology. Theory Biosci 126:65–113

Stadler PF, Prohaska SJ, Forst CV, Krakauer DC (2009) Defining genes: a computational framework. Theory Biosci. doi: 10.1007/s12064-009-0067-y