A defense of syntax-based gene concepts in postgenomics:
'Genes as modular subroutines in the master genomic program'

Tudor M. Baetu

Abstract

The purpose of this article is to update and defend syntax-based (conserved DNA-sequence motifs) gene concepts. I show how syntax-based concepts can and have been extended to accommodate complex cases of processing and gene expression regulation. In response to difficult cases and causal parity objections, I argue that a syntax-based approach fleshes out a deflationary concept defining genes as genomic sequences and organizational features of the genome contributing to a phenotype. These organizational features are an important part of accepted molecular explanations, provide the theoretical basis for a large number of experimental techniques and practical applications, and play a crucial role in in annotating the genome, deriving predictions and constructing bioinformatics models.

Introduction

One of the consequences of the elucidation of the structure of DNA was the identification of genes with DNA base sequences (Watson and Crick 1953). Shortly after, the Central Dogma (Crick 1958) allotted DNA the role of an archive containing information for building proteins (the 'blueprint' gene) via yet to be elucidated mechanisms. With the elucidation of the mechanisms of genome expression during the 1960s and 70s, the notion that genes are DNA sequences coding information for proteins was reframed in terms of syntax-like conserved motifs providing instructions for transcription and translation, leading to the formulation of the 'transcription unit' and the 'open reading frame' molecular gene concepts.

For example, the basic prokaryotic transcription unit is characterized by two regulatory regions responsible for recruiting the RNA polymerase (protein that synthesizes an RNA tran-

script from the DNA template), a transcribed coding sequence and a transcription termination sequence. Within the coding sequence, one or more open reading frames (sequences translated into protein gene products) are marked by a site responsible for recruiting ribosomes (the Shine-Delgarno box), a 'begin translation' site (AUG codon) and an 'end translation' site (termination codon).

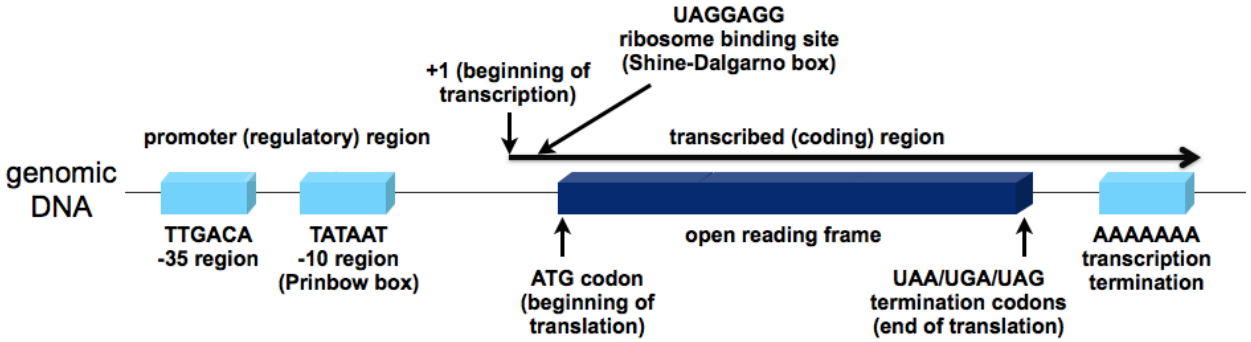

Figure 1. The prokaryotic transcription unit

Such findings suggested that what makes the difference between a region of the genome that eventually contributes to a phenotype and one that does not has something to do with the presence of conserved sequences responsible for interacting with the transcriptional and translational machinery of the cell in order to generate gene products from a DNA template. The assumption, amply verified by *in vitro* binding studies, is that if a given conserved sequence motif is present in a stretch of DNA, it will most likely serve as a binding site for some component of a mechanism of genome expression and regulation, and therefore contribute to some aspect of genome expression. The syntax approach capitalizes on this assumption by treating conserved DNA sequence motifs as instructions, which, if processed in the right order, will ultimately help predict which parts of the genome will be expressed and as what gene products.

Ever since their discovery, molecular biologists have relied on conserved sequences as a means to define and discover genes, formulate hypotheses and make predictions about the regulation of gene expression and the function of gene products, guide genetic engineering, as well as design *in vitro* transcription and translation protocols. Sequence alignment algorithms such as BLAST (Altschul et al. 1990) are used on a daily basis in labs around the world to identify putative genes based on the presence of syntax-like motifs typically required for transcription and translation. The establishment of bioinformatics as a new discipline aiming, among other things,

to further develop genome annotation protocols (Mandoiu and Zelikovsky 2008) suggests that such practices will continue in the near future.

Strangely enough, philosophers of biology do not partake of this enthusiasm. Griffiths and Stotz (2007) claim that the 'nominal gene' defined by current annotation practices will soon be superseded by 'postgenomic' concepts of the gene. In their view, the discovery of gene rearrangement, nested genes, alternative promoters, alternative splicing, trans-splicing, RNA editing, frameshifting, alternate stop codons, polyproteins, and various other complications due to regulatory, post-transcriptional/translational processing mechanisms posit an insurmountable difficulty to a syntax-based approach. Mechanisms of gene expression regulation seem to determine which, when and where coding sequences are transcribed and translated, while mechanisms of post-transcriptional/translational processing seem to control how the information contained in original DNA sequence is used. Given the shortcomings of current syntax based gene concepts, there is a growingly popular tendency to treat the genome as a set of sequences that can be 'read' (transcribed and translated) and 'processed' (RNA splicing and editing, post-translational modifications of peptides, etc.) in a variety of ways depending on developmental (Griffiths and Neumann-Held 1999), cellular (Stotz forthcoming), and environmental contexts (Stotz 2006). According to this view, genes are best described as "things you can do with your genome" (Griffiths and Stotz 2006, 500).

This raises the question "Is this the end of syntax-based gene concepts?" According to these authors, the answer seems to be "Yes":

> The 'same' DNA sequence potentially leads to countless different gene products, different sequences might code for identical products, and the need for a rare product asks for the assembly of a novel mRNA sequence. Hence *the information for a product is not sufficiently encoded in the targeted DNA sequence* but has to be supplemented through sequence information provided by elements outside the coding sequence, such as transcription, splicing, or editing factors. (Stotz 2006, 905, my emphasis)
>
> The sequence of the DNA can [...] be compared to a sequence of letters without spaces or punctuation marks. [...] A different developmental system imposes a different scheme over the letters, that is, over the DNA sequence. *It is therefore misleading to think of*

> *functional descriptions of DNA, such as 'promoter region', as explicable solely in terms of structural descriptions of DNA, such as 'sequence.'* [...] the gene is identified not with these DNA sequences alone but rather with the process in whose context these sequences take on a definite meaning. (Griffiths and Neumann-Held 1999, 661, my emphasis)

In this paper I challenge this conclusion. I argue that current syntax-based concepts can and have been extended to accommodate complex cases of processing and gene expression regulation. The syntax-based approach proposed in this paper has to its advantage three virtues - modularity, retro-compatibility and the ability to explain why the genomic information is expressed and processed in an orderly, predictive fashion - which rival accounts cannot bolster. In response to difficult cases and causal parity objections, I argue that a syntax-based approach fleshes out a deflationary concept defining genes as genomic sequences and organizational features of the genome contributing to (rather than determining) a phenotype.

The paper is organized as follows: In section 2, I show how an updated version of current syntax-based gene concepts can accommodate complications due to mechanisms of genome processing and genome expression regulation. In section 3, I elaborate a three-level account of genome organization, and, in section 4, discuss its advantages. In section 5, I discuss the limits of syntax-based approaches and possible avenues of future improvement. Finally, in section 6, I summarize my conclusions and arguments.

2. Expanding the syntax-based approach

The critiques of current syntax-based gene concepts are fueled by two main concerns:

- gene expression regulation (the role mechanisms of gene expression regulation play in specifying which DNA sequences are expressed)
- the breakdown of the gene-gene product sequence collinearity (due to post-transcriptional and post-translational processing).

Nevertheless, the fact that, in many cases, the mechanisms of gene expression and gene expression regulation work in concert with conserved DNA sequences suggests that one way to cope with the complexities brought about by these two issues is to update and extend available

syntax based gene concepts. A step in this direction is illustrated by Gerstein's et al. attempt to define genes as "subroutines in the genomic operating system." The authors propose that the structure of the genome should be described

> in very much the same way that grammars are used to describe computer programs - with a precise syntax of upstream regulation, exons, and introns. (2007, 671)

Thus, instead of taking into account only the basic syntax of transcription and translation underlying the transcription unit and the open reading frame gene concepts, Gerstein suggests an extended syntax including sequences recognized by transcription factors [enhancers, promoters, activator/repressor binding sites associated with regulatory networks (Lee et al. 2002; Levine and Davidson 2005)] and sequences signaling splicing [reviewed in (Black 2003)].[1]

It is worth noting however that nothing prohibits us extending this approach to any sequence-specific aspect of genome expression, including chromatin regulation [matrix binding and nucleosome assembly [reviewed in (Turner 2001)], RNA protein binding associated with translational regulation [reviewed in (Mazumder et al. 2003)], and post-translational modifications (glycosylation, phosphorylation, cleavage of polyproteins).[2] While there are still many gaps to be filled, there is ample evidence that conserved DNA sequences play a necessary role in specifying sites within the genome (or its transcribed/translated counterparts) where various components of mechanisms of genome expression, regulation and processing bind and initiate activities leading to expression of specific portions of the genome and specific modifications of transcribed/translated products. I argue therefore for a more general syntax-based approach that goes beyond the immediate scope of Gerstein et al., who are concerned solely with transcriptional regulation and RNA splicing.

3. The three-level organization of the genome

Following the analogy suggested by Gerstein et al., I argue that just like computer programs are organized in subroutines, that is, readymade sets of instructions that can be accessed

---

[1] Gerstein et al. elaborate two additional approaches in their paper. One involves the mapping of gene products onto genomic DNA (and therefore would be best described as post-proteomics rather than post-genomics), while the other relies on a functional interpretation of such a mapping.

[2] E.g., programs like the ExPASy Proteomics tools developed by the Swiss Institute of Bioinformatics allow for probabilistic predictions of post-translational modifications.

on demand in a variety of contexts, the genome is organized into modular genes. At each site where transcription is initiated, a modular 'subroutine is run'; each of these 'subroutines' counts as a gene. The genome behaves like a 'master program' relying on a set of specific syntax-like sequences playing a role in specifying where in the genome and which transcription factors bind in order to 'recruit' the transcriptional machinery. Each gene behaves like a 'subroutine' because sequences within the transcribed DNA mark sites (by providing binding sites for proteins or in virtue of specific three-dimensional folding) for further processing of the initial RNA transcript sequence, leading to the synthesis of various gene products.

The gist of the analogy is that the genome is organized as three nested levels of syntax-like DNA sequences. The genomic level is the realm of transcription regulation. Regulatory sequences distributed at various sited throughout the genome play a role in specifying where in the genome and which transcription factors bind in order to allow for (or prohibit) transcription. The gene level corresponds to the transcribed DNA (providing the sequence of the primary RNA transcript). For the most part (the one known exception will be discussed in more detail shortly), genes act like independently processed modules because, once transcribed, their sequence is processed in accordance to the conserved sequences contained within their boundaries alone. The sub-gene level is the realm of translation, translational control and post-transcriptional/ translational processing. Sequences within the transcribed/translated DNA mark sites for eventual RNA translation regulation and further processing of the RNA transcript/peptide leading to the synthesis of one or more final gene products.

Thus delimited, a gene is shorter than a standard transcription unit but more extended than an open reading frame. The transcription regulatory regions (promoters, enhancers) are part of the 'master genomic program', while the gene itself is rigorously delimited as a DNA sequence contained between transcription start and stop sites or by homology with the primary transcript. At the same time, a gene is not exactly an open reading frame either, since it includes the 5' and 3' untranslated regions (UTR), as well as introns and alternate reading frames.

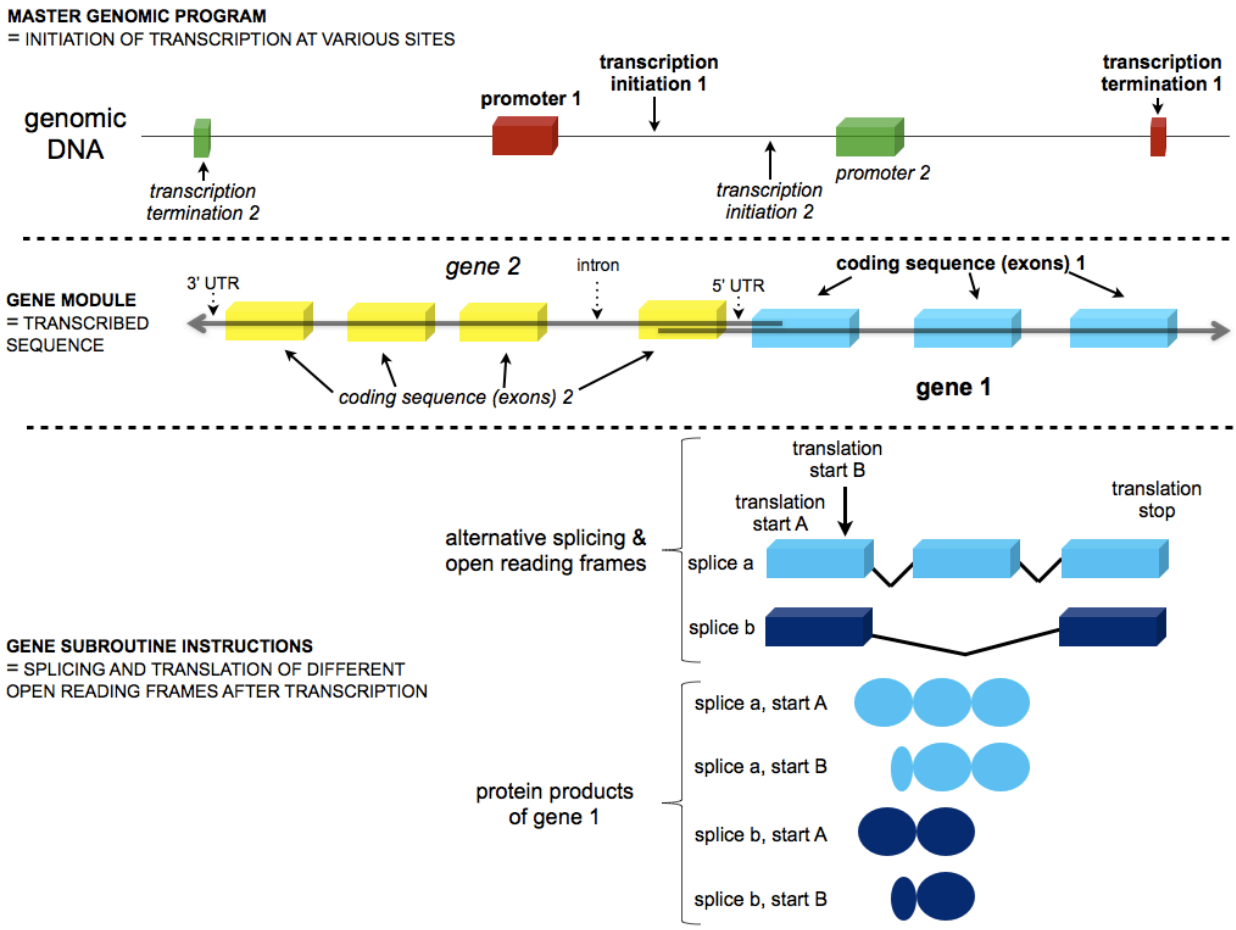

Figure 2. The modular organization of the genome

4. The three-level organization approach in practice

It is interesting to note the similarities between the approach proposed in this paper and the genes as 'things you can do with your genome' approach advocated by Griffiths and Stotz. The signaling pathways mediated activation of different transcription factors allow cells to use the same genome in different ways. This accounts for cellular differentiation and the induction of gene expression in various environmental contexts. Also, alterations and rearrangements of the genome (e.g., gene rearrangement in immune cells, chromosomal deletions/insertions) result in a different 'master genomic program' that ultimately contributes to the synthesis of different gene products. Each gene can also be used in a variety of ways, as dictated by its internal syntax-motifs. This accounts for cases of post-transcriptional and post-translational processing. Fur-

thermore, since the instructions contained in the gene subroutines may prompt events such as splicing, the gene-gene product collinearity assumption is not required.

I also want to stress the differences. The approach advocated by Griffiths and Stotz downplays the specificity and organizational roles played by DNA conserved sequences. Their approach has the advantage of taking into account cases of genome expression regulation and processing that are not sequence-specific, such as some aspects of chromatin regulation and some forms of RNA editing. Unfortunately, the reader is systematically left with the impression that DNA sequence is merely the unformed matter on which the mechanisms of genome expression impose whatever form they want. Griffiths and Stotz give no clues as to why there is order and regularity in the way the genome is processed. In contrast, a three-level syntax-based approach has no trouble explaining the origin of this order and regularity: the genome is processed and expressed - either constitutively (e.g., house-keeping genes) or in response to regulatory stimuli (e.g., activation of transcription factors via signaling pathways) - as dictated by instructions, or syntax-like motifs contained in its sequence.

Furthermore, Griffiths and Stotz's views on the role of DNA breaks the continuity with current molecular gene concepts. While these authors see this break as innovation, most scientists will most likely see it as an unrealistic proposal. After all, syntax-based concepts are *de facto* in use as we speak. Every time the term 'gene' is mentioned in a scientific paper, the occurrence of the term refers either to a transcription unit, to a specific open reading frame, or, as it is almost always the case, to a continuous DNA sequence containing some of the proximal upstream regulatory sequences and at least one open reading frame. A cursory look at the Methods and Techniques section of the paper in question or a survey of the sequences daily submitted to the NCBI GenBank should suffice to prove my point. In its present use, the term 'gene' is not used in reference to spliced mRNA or gene product sequences, it does not refer to epigenetic contributors to inheritance and it certainly does not mean 'things' someone or something could do with its genome. Breaking with the current scientific practice might simply result in the term 'gene' having two distinct meanings, one for practicing scientists and another for philosophers of biology. In contrast, an extended syntax-based approach emphasizes the modular organization of the genome. By preserving discrete gene-modules, the compatibility with previous syntax based

gene concepts is ensured by providing a substantial overlap in terms of the DNA sequences to which these concepts refer.

Beyond retro-compatibility, modularity is a very important methodological requirement in itself (Callebaut and Rasskin-Gutman 2005). If a system is modular, it can be carved into functional modules that can be characterized and investigated on a largely independent basis (Lauffenburger 2000). Instead of collapsing all the distinct ways in which the genome is expressed and regulated under the rather uninformative term 'things one could do with the genome,' a nested syntax-based account successfully conveys the notion that genome expression can be divided in two distinct processes: gene expression regulation, achieved at the level of transcriptional control, and gene sequence processing, achieved at the level of transcription, post-transcriptional modifications, translation and post-translational modifications.

In practice, these three virtues of syntax-based gene concepts - modularity, retro-compatibility and the ability to explain why the genomic information is expressed and processed in an orderly, predictive fashion - are reflected as follows:

1. Reference continuity is ensured by providing a significant overlap with the transcription unit and open reading frames molecular gene concepts, as well as BLAST-generated and GenBank sequences. This means that molecular biologists can continue to identify genes and putative genes via currently available molecular techniques and genome annotation protocols. Current syntax gene concepts are based on a solid and very successful experimental methodology and are not likely to be abandoned any time soon.
2. Since a gene is not a piece of genomic DNA, but a transcribed DNA sequence, there is no restriction on overlapping or nested genes. Also, distinct genes are allowed to share promoter (operons, overlapping promoters) and distal regulatory elements (common enhancers). In contrast to traditional transcription units, alternative promoters are also accommodated.
3. Unlike traditional open reading frames, it is not required that protein end-products are generated. RNA products playing structural (ribosomal, transfer and small nuclear RNAs, ribozymes) and regulatory (microRNAs, RNA interference) roles are also coded by genes. Also, by keeping the 3' and 5' untranslated regions as part of the gene, certain mechanisms

of translation regulation are accommodated as part of the gene 'subroutine'; this allows DNA templates for RNA species that are not immediately processed for translation (such as maternal factors) to count as genes.

1. Since the instructions contained in the gene subroutines may prompt events such as splicing, perfect homology/collinearity between DNA and gene products is no longer required. Waters (1994, 78) proposed that molecular genes are sequences of DNA coding for product sequences generated at some point during gene expression, such as peptides or ribozymes. Since DNA is said to serve as a template for RNA and peptide synthesis in virtue of the mechanisms of transcription and translation, gene products sequences are collinear and homologous to template DNA sequences. This conception was criticized on the grounds that that gene rearrangements, splicing, alternative splicing, trans-splicing, and RNA editing can generate significant divergences between the original DNA sequence and the final gene product sequence (Falk 1986; 2003; Portin 2002; Stotz et al. 2006).
2. More than one gene product can be associated with any given gene. The fact that the same gene can be involved in more than one function/phenotype was acknowledged a long time ago by classical geneticists (Morgan 1935). Splicing, alternative reading frames, polyproteins are newly discovered molecular mechanisms that further contribute to the generation of functional diversity.
3. Trans-splicing (splicing of distinct RNA species) can be accommodated as a special case. I acknowledge that even if most 'gene subroutines' are modular (i.e., once initiated, they are not influenced by other 'gene subroutines'), trans-splicing is an example of non-modular processing. However, there are a number of attenuating circumstances. First, the genes implicated in trans-splicing are unambiguously differentiated since each has its own well-defined transcription initiation and termination sites; the proposed gene definition still applies [e.g., Finta and Zaphiropoulos (2002) explicitly and unambiguously distinguish between the various cytochrome genes involved in trans-splicing]. Second, trans-splicing seems to be driven by conserved sequences required by splicing in general, and therefore is an agreement with a syntax-based approach. And third, naturally occurring trans-splicing involves either transcripts originating from overlapping transcription units (Caudevilla et

al. 1998) or highly homologous transcripts [genes sharing "a high degree of similarity", such as duplicated genes (Finta and Zaphiropoulos 2002, 5882)]; we are thus dealing with a local, homology driven cross-modular processing involving identical or very similar modules, not with a generalized breaking of modularity.

4. Finally, this conception is compatible with an important aspect of Griffiths and Stotz's (2006; Stotz et al. 2006) postgenomic gene: each gene, defined in virtue of transcription initiation/termination, can be used in a variety of ways.

5. Difficult cases and causal/explanatory parity objections

The scope and intended domain of application of syntax-based concepts

An immediate concern is that some forms of regulation and processing don't rely on conserved sequences. For example, it has been argued that some aspects of chromatin regulation involve sequence-independent mechanisms (Fox-Keller and Harel 2007; Stotz forthcoming). Epigenetic contributions to inheritance and development, posit another legitimate source of concern (Fox-Keller 2001). Finally, it has been argued that even if conserved sequences are useful in predicting certain biological outcomes, they don't suffice to explain these outcomes (Fox Keller 2000; Griffiths and Neumann-Held 1999).

My answer to the above objections hinges on a deflationary view of what genes are and how they are defined. In contemporary scientific practice, gene concepts are not substitutes for explanation and therefore don't have to account for every single causal determinant of inheritance. Quite illustrative in this sense, the Human Genome Nomenclature Committee defines a gene as a

> DNA segment that contributes to phenotype/function. In the absence of a demonstrated function a gene may be characterized by sequence, transcription or homology. (Wain et al. 2002, 464)

The first half of the definition accommodates a situation where a segment of DNA is shown to be associated with a phenotype even if this segment is too short (e.g., a point mutation) or too long (e.g., a large DNA segment found between two chromosomal break-points) to count as a molecular gene. The second half of the definition is explicitly meant to accommodate a situation where

putative genes are identified via genome annotation/syntax-based techniques (e.g., promoter-like sequences preceding open reading frames, homology with gene product sequences and known genes).

The extended syntax-based approach defended in this paper spells out in more detail this second half of the HGNC definition in an attempt to accommodate instances of expression regulation and post-transcriptional/translational processing: genes are *genomic sequences and organizational features of the genome contributing to a pattern of genome expression and, if known, its associated phenotype*. Nowhere in this (or the HGNC) definition is it stated that genes are unique causal determinants, that they contribute to all known instances of inheritance, and that they explain or suffice to explain why and how inherited phenotypes occur.

The limits (and areas of future improvement) of syntax-based concepts

I also want to acknowledge the fact that the correlation between a given conserved sequence and its associated regulatory/expression/processing outcome is not perfect (e.g., it is probabilistic rather than deterministic). There is a general agreement that DNA sequence motifs play a crucial role in specifying binding sites for various non-DNA components (usually proteins) of mechanisms of genome expression and regulation. Typically, DNA-protein interactions requires that, (a) certain chemical moieties, specified by the sequence of the DNA motif in question, are present; and (b) that these key moieties are exposed at the right distance and position in respect to each other, as dictated by the chemical structure of the DNA double-helix. In as much as the DNA molecule is not subjected to any stress (torque, bends, super-/under-coiling), it has been shown that if a given conserved DNA sequence motif is present, it serves as a binding site for some specific component(s) of a mechanism of genome expression and regulation, and therefore contribute to some aspect of genome expression in a predictable way.

One complication which present syntax-based concepts fail to take into account is that whenever a protein binds DNA, it creates bends and torques in the DNA double-helix, thus altering to various degrees the spacing and position of nucleotides and chemical moieties required for the binding of other proteins; such changes in spacing and positioning can be important enough to result in an enhancing or repression of the binding of these other proteins. Even if the DNA

sequence will not completely change its 'meaning' due to deformations of the DNA double helix, its specificity/affinity for a given protein target may increase or decrease (sometimes to biologically insignificant values), thus reducing the accuracy of the predictions about genome expression.

It is important however to realize that this shortcoming of syntax-based approaches does not preclude future remedies. For instance, nucleosomes (a key component of chromatin structure) are particularly troublesome because of the coiling and deformation they induce in the DNA double-helix. There is evidence suggesting that the binding and assembly of the nucleosomes is to a large extent sequence-dependent, evidence which, in conjunction to a more and more detailed knowledge about the way in which DNA coiling around nucleosome affects the structure of the double-helix, is likely to allow for a fine-tuning of syntax-based approaches in the near future.

6. Conclusion

I have argued that an expanded syntax-based approach can handle most cases of regulation and processing, while providing a number of key advantages, such as retro-compatibility with molecular gene concepts and current genome annotation protocols, a step-by-step modular methodology for investigating inheritance phenomena, as well as the ability to explain certain regularities characterizing genome expression. In response to objections, I proposed a deflationary view of what genes are and how they are defined, and discussed possible ways in which this approach may be improved. Finally, I acknowledged the limits of syntax-based concepts, and discuss possible avenues of improvement.

Acknowledgments

I would like to thank Lindley Darden, Erika Milam, Eric Saidel, Pamela Henson, Joan Straumanis, Lane DesAutels, Carl Craver, and the DC History and Philosophy of Biology group for helpful discussion and comments on earlier drafts. This work is supported by 'Fonds de la recherche sur la société et la culture', Québec, Canada [grant number 127231].

Bibliography

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman (1990), "Basic Local Alignment Search Tool", *Journal of Molecular Biology* 215 (3):403–410.

Black, D. L. (2003), "Mechanisms of Alternative Pre-Messenger RNA Splicing", *Annual Reviews of Biochemistry* 72 (1):291-336.

Callebaut, W., and D. Rasskin-Gutman (2005), *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. Cambridge, MA: MIT Press.

Caudevilla, C., D. Serra, A. Miliar, C. Codony, G. Asins, M. Bach, and F. G. Hegardt (1998), "Natural Trans-Splicing in Carnitine Octanoyltransferase pre-mRNAs in Rat Liver", *Proceedings of the National Academy of Science* 95 (21):12185-12190.

Crick, F. H. (1958), "On Protein Synthesis", *Symposia of the Society for Experimental Biololgy* 12:138-163.

Falk, R. (1986), "What is a Gene?", *Studies in the History and Philosophy of Science* 17:133-170.

——— (2003), "Linkage: From Particulate to Interactive Genetics", *Journal of the History of Biology* 36 (1):87-117.

Finta, C., and P. G. Zaphiropoulos (2002), "Intergenic mRNA Molecules Resulting from Trans-Splicing", *Journal of Biological Chemistry* 277 (8):5882-5890.

Fox Keller, E. (2000), *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Fox-Keller, E., and D. Harel (2007), "Beyond the Gene", *PLoS ONE* 2 (11):e1231. doi:1210.1371/journal.pone.0001231.

Gerstein, M. B., C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman, and M. Snyder (2007), "What is a Gene, Post-ENCODE? History and Updated Definition", *Genome Research* 17 (6):669-681.

Griffiths, P., and E. Neumann-Held (1999), "The Many Faces of the Gene", *Bioscience* 49:656-663.

Griffiths, P., and K. Stotz (2007), "Gene", in M. Ruse and D. Hull (eds.), *The Cambridge Companion to the Philosophy of Biology*, Cambridge: Cambridge University Press.

Griffiths, P. E., and K. Stotz (2006), "Genes in the Postgenomic Era", *Theoretical Medicine and Bioethics* 27 (6):499-521.

Lauffenburger, D. A. (2000), "Cell Signaling Pathways as Control Modules: Complexity for Simplicity?", *Proceedings of the National Academy of Science* 97:5031-5033.

Lee, T., N. Rinaldi, F. Robert, D. Odom, Z. Bar-Joseph, G. Gerber, N. Hannett, C. Harbison, C. Thompson, I. Simon, J. Zeitlinger, E. Jennings, H. Murray, D. Gordon, B. Ren, J. Wyrick, J. Tagne, T. Volkert, E. Fraenkel, D. Gifford, and R. Young (2002), "Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*", *Science* 298 (5594):763-764.

Levine, M., and E. Davidson (2005), "Gene Regulatory Networks for Development", *Proceedings of the National Academy of Science* 102 (14):4936–4942.

Mandoiu, I., and A. Zelikovsky (2008), *Bioinformatics Algorithms: Techniques and Applications*. Hoboken, NJ: Wiley & Sons.

Mazumder, B., V. Seshadri, and P. L. Fox (2003), "Translational Control by the 3'-UTR: The Ends Specify the Means", *Trends in Biochemical Sciences* 28 (2):91–98.

Morgan, T. H. (1935), "The Relation of Genetics to Physiology and Medicine", *Les prix Nobel en 1933. Imprimerie Royale*:1-16.

Portin, P. (2002), "Historical Development of the Concept of the Gene", *Journal of Medicine and Philosophy* 27 (3):257-286.

Stotz, K. (2006), "With 'Genes' Like That, Who Needs an Environment? Postgenomics's Argument for the 'Ontogeny of Information'", *Philosophy of Science* 73:905–917.

——— (forthcoming), "2001 and All That: A Tale of a Third Science", *Biology and Philosophy*.

Stotz, K., A. Bostanci, and P. Griffiths (2006), "Tracking the Shift to 'Postgenomics'", *Community Genetics* 9:190–196.

Turner, B. (2001), *Chromatin and Gene Regulation: Mechanisms in Epigenetics*. Oxford: Blackwell.

Wain, H. M., E. A. Bruford, R. C. Lovering, M. J. Lush, M. W. Wright, and S. Povey (2002), "Guidelines for Human Gene Nomenclature", *Genomics* 79 (4):464-470.

Waters, C. K. (1994), "Genes Made Molecular", *Philosophy of Science* 61:163–185.

Watson, J. D., and F. H. Crick (1953), "Genetical Implications of the Structure of Deoxyribonucleic Acid", *Nature* 171 (4361):964-967.

Word count: 4995