Name: Baetu, Tudor
Affiliation: Konrad Lorenz Institute
Title: A Comparative Analysis of Syntax-Based and Mapping Concepts of the Gene

Long Abstract:

The Human Genome Nomenclature Committee (HGNC) 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 refers to a mapping of phenotypes, as it is typically done in forward genetics, as well as to molecular experiments demonstrating a causal link between a genomic sequence and a phenotype (e.g., transgenic/knockout organisms, site-directed mutagenesis and other forms of genetic engineering). The second half of the definition is meant to accommodate a situation where putative genes are identified via genome annotation techniques or homology-driven mapping of gene product sequences onto genomic DNA. It is interesting therefore to observe that the HGNC concept is pluralistic in nature, as it encompasses several approaches to gene identification, most notably the mapping of phenotypes and gene products, and sequence annotation techniques. Similar distinctions between two main classes of gene concepts are found in the philosophical literature - gene-P/gene-D (Moss 2003) and the instrumental/nominal gene (Griffiths and Stotz 2006) -, and several authors defend pluralism about gene concepts (Griffiths and Stotz 2006; Moss 2003; Stotz et al. 2006; Waters 2006; 2008).

The HGNC concept has several advantages. By allowing genes to be defined in terms of their effects on phenotypes alone (as stipulated by the first half of the concept) and independently of any other expectations about structural motifs typically associated with genes (the definition covered in the second half of the concept is strictly required only if the contribution to phenotypes is uncertain), the HGNC concept can accommodate a wide range of findings about genomic contributions to phenotypes, including sequences coding for protein products and regulatory RNA species, sequences specifying post-transcriptional/translational processing (e.g., splicing, glycosylation), miscellaneous DNA motifs (e.g., repetitive sequences increasing the probability of genomic rearrangements associated with certain medical conditions, 'spacer' regions of a definite length but variable sequence required by certain regulatory mechanisms), or sequences that contribute to phenotypes via yet to be elucidated mechanisms. Thus, the HGNC

concept addresses important criticisms concerning the inability of molecular gene concepts to account for newly discovered mechanisms of genome expression (Gerstein et al. 2007; Portin 2009).

The HGNC concept is also sensitive to the issues of genetic determinism and epigenetic contributions to phenotypes (Fox Keller 2001; Griffiths and Stotz 2006; Oyama 2000). While it limits the use of the term 'gene' to genomic sequence contributions to phenotypes, it does not define genes as unique causal determinants, as contributing to all known instances of inherited phenotypes, or as providing sufficient explanations of why and how certain phenotypes occur. Rather, it adopts a deflationary view according to which genes are causally-relevant factors that contribute, along with other factors, to certain, but not all, phenotypes (Baetu 2011). Finally, by accommodating sequences identifiable by a plurality of techniques, the HGNC concept represents an acknowledgment of the fact that there is no unique set of structural motifs that characterize genes (Burian 2004; Falk 2003), and that the genome is a complex network of structural motifs (Griffiths and Stotz 2006; Portin 2009).

On the negative side, the HGNC concept fails to provide a principled way of organizing the overwhelming variety of sequences that may count as genes, leading to a genome annotation problem: it is not clear whether genomic sequences shown to contribute to a phenotype, especially ones that overlap or are immediately adjacent, are [parts of] the same gene, different genes, or if each sequence counts as a distinct gene. As I will show in a moment, the annotation problem is particularly troublesome because different ways of grouping sequences are known and expected to result in different phenotypes.

The goal of this paper is to review newly developed functional-mapping and syntax-based concepts and assess how well they succeed in addressing the annotation problem while preserving the advantages of the HGNC concept. I argue that recent concepts solve the annotation problem by grouping together sequences contributing to a given phenotype via the same genome expression pathway. For example, according to recently proposed syntax-based characterizations, adjacent/overlapping sequences are grouped together if they contribute to a product/phenotype via the same genome expression pathway (i.e., via a common primary transcript). Branchings of expression pathways are allowed both before (e.g., DNA rearrangements), and after transcription (e.g., alternative splicing); what is strictly required for a

genome expression pathway to be recognized as a uniquely identifiable process is the transcription of a single, well-defined genomic sequence.

At the same time, in order to preserve the degree of generality achieved by the HGNC concept, I argue that a pluralism involving a dynamic interplay between mapping and syntax-based concepts is required. An attempt to collapse all concepts of the gene into a unified, over-arching concept is likely to be counterproductive. What seems to be required in contemporary experimental practice is a dynamic interplay between several concepts and their associated investigative practices. Unorthodox sequences identified by mapping approaches - what we may call 'anomalies' (Darden 1991; 2006), the most recent examples being trans-splicing and scrambled genes - play a role in the discovery of new mechanisms of genome regulation and processing. As these new mechanisms are elucidated, they prompt a more or less radical revision of syntax-based concepts, either by adding new conserved sequence motifs or by altering the organizational scheme of the genome. In turn, this transforms syntax-based concepts into more powerful tools, allowing for the generation of novel lab-produced phenotypic outcomes, and yielding predictions about subtle differences in phenotype and gene product expression not accessible to other forms of investigation.

Bibliography

Baetu, T. M. 2011. "A Defense of Syntax-Based Gene Concepts in Postgenomics: 'Genes as Modular Subroutines in the Master Genomic Program'", Philosophy of Science 78 (5):712-23.

Burian, R. M. 2004. "Molecular Epigenesis, Molecular Pleiotropy, and Molecular Gene Definitions", History and Philosophy of the Life Sciences 26 (1):59-80.

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

Fox Keller, E. . 2001. "Beyond the Gene but Beneath the Skin", in S. Oyama, P.Griffiths and R. Gray (eds.), Cycles of Contingency: Developmental Systems and Evolution, Cambridge, MA: MIT Press, 299–312.

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):669-81.

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

Moss, L. 2003. What Genes Can't Do. Cambridge: MIT Press.

Oyama, S. 2000. "Causal Democracy and Causal Contributions in Developmental Systems Theory", Philosophy of Science 67:S332-S47.

Portin, P. 2009. "The Elusive Concept of the Gene", Hereditas 146:112–17.

Stotz, K., A. Bostanci, and P. Griffiths. 2006. "Tracking the Shift to ‘Postgenomics'", Community Genetics 9:190–96.
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):464-70.
Waters, C. K. 2006. "A Pluralist Interpretation of Gene-Centered Biology", in S. Kellert, H. Longino and K. Waters (eds.), Scientific Pluralism, Minneapolis, MN: University of Minnesota Press.
———. 2008. "Beyond Theoretical Reduction and Layer-Cake Antireduction: How DNA Retooled Genetics and Transformed Biological Practice", in M. Ruse (ed.), The Oxford Handbook of Philosophy of Biology, Oxford: Oxford University Press, 238-62.