Author-created version (final draft)
January 2011

Are all genes regulatory genes?

Rosario Michael Piro

Note: this is an author-created version of the article (final draft). The final publication is available at www.springerlink.com (Biology & Philosophy 26(4):595-602, 2011).

Abstract Although much has been learned about hereditary mechanisms since Gregor Mendel's famous experiments, gene concepts have always remained vague, notwithstanding their central role in biology. During over hundred years of genetic research, gene concepts have often and dynamically changed to accommodate novel experimental findings, without ever providing a generally accepted definition of the 'gene.' Yet, the distinction between 'regulatory genes' and 'structural genes' has remained a common theme in modern gene concepts since the definition of the operon-model. This distinction is now challenged by recent findings which suggest that, at least in eukaryotes, structural genes may in many situations have a regulatory function that is independent of the function of the gene product (protein or non-coding RNA molecule). This brief paper discusses these new findings and some possible implications for the notion of the 'regulatory gene.'

Keywords Gene concepts · Regulatory genes · MicroRNAs · Pseudogenes · Transcript function

1 Introduction

The gene, being a "concept in tension" (Falk, 2000), has inspired many debates amongst both biologists and philosophers of biology. Several authors have regarded the gene as "no more than a handy term that acquires a precise meaning only in some specific scientific context in which it is used" (Griffiths and Stotz, 2007, p. 85), and all efforts to find a generally accepted definition have so far been unsuccessful. Rather than

Rosario Michael Piro
Molecular Biotechnology Center and Department of Genetics, Biology and Biochemistry, University of Torino, Italy
E-mail: rosario.piro@unito.it

Present address:
Department of Theoretical Bioinformatics, German Cancer Research Center (Deutsches Krebsforschungszentrum), Heidelberg, Germany
E-mail: r.piro@dkfz.de

reaching a consensus, the scientific findings and the philosophical debates have lead to a variety of different and sometimes even contradictory gene concepts (Burian, 2005; Griffiths and Stotz, 2007; Portin, 2009; Rheinberger and Müller-Wille, 2009).

The recent historical overviews by Griffiths and Stotz (2007) and by Rheinberger and Müller-Wille (2009) describe the evolution of gene concepts in response to experimental practices and their discoveries. The advent of molecular biology, for example, has completely revolutionized the 'classical gene' as a largely abstract unit of transmission, recombination, mutation, and function. The transmission from one generation to the next turned out to regard the genome as a whole and the unit of mutation to be as small as a single nucleotide.[1] Molecular mechanisms of post-transcriptional modifications such as splicing and, more recently, heritable epigenetic mechanisms like heritable DNA methylation patterns further complicated the picture. Some authors have even proposed to give up the notion of a gene altogether.[2]

Since the following discussion regards particularly RNA transcripts, I use the term 'gene' to stand for a DNA segment from which one or more (not necessarily protein-coding) functional transcripts originate. This notion is strongly related to what Lenny Moss (2003) calls a 'Gene-D'—defined by DNA sequence, but indeterminate with respect to phenotypes—and Griffiths and Stotz (2007) a 'post-genomic molecular gene'—the gene as a structural and functional unit of contemporary molecular biology.[3]

It is important to note that gene concepts do not only regard the question of what a gene *is*—e.g. what structural or sequence properties it has, or what it is made of—but include also notions about what a gene *does*. These two issues are, of course, intertwined and cannot be clearly separated from each other because, as Burian (2005) notes, functional criteria of delimitation are built into gene concepts even at the molecular level. He emphasizes the constructive interaction between the two notions: while what counts as a gene depends on what one chooses as a phenotype, the choice of a phenotype itself is somewhat constrained by what one knows about genes.

It is regarding this broad interpretation that the distinction between regulatory and structural genes can be seen as an important aspect of gene concepts, although these notions are usually not explicitly employed when defining the 'gene.'

2 Regulatory genes and structural genes

Although "recent results seriously call into question the further applicability of straightforward 'gene-for' talk" (Rheinberger and Müller-Wille, 2009), one of the notions most often used for descriptive and explanatory purpose is the distinction between 'regulatory genes' and 'structural genes'.

[1] This is clearly a too simplistic view. As for example Gatherer points out, no concept in molecular biology has proven to be able to capture everything that is implied by the term 'gene' in Mendelian genetics (Gatherer, 2010).

[2] Keller and Harel (2007), for example, argue for the greater flexibility of their concept of a *genetic functor* or *genitor* $G = (O, D, B)$ that, for a given organism O, describes the relationship between a *dene* D, i.e. a predicate about the DNA (including epigenetic properties like DNA methylation states), and a *bene* B, i.e. a statement about an associated functionality or behavior (including complex modal and temporal characteristics). This notion, however, may turn out to be too flexible in order to be of practical use, because—as the authors themselves underscore—"anything goes" (Keller and Harel, 2007, p. 6).

[3] As Griffiths and Stotz (2007) point out, the conceptual space of the 'Gene-D' does also include what they call a 'nominal' gene (annotated sequences as used, for example, for databases and bioinformatics tools).

A first clear distinction between the structural gene ("gène de structure") and the regulator(y) gene ("gène régulateur") was made by Jacob and Monod to explain the relationship between the (regulatory) *lacI* gene and the (structural) *lacY* and *lacZ* genes involved in β-galactosidase activity in *Escherichia coli*: while structural genes define the polypeptidic structures of β-galactosidase (*lacZ*) and galactosidase-permease (*lacY*), the regulatory gene governs their expression through an intermediary repressor molecule (Jacob and Monod, 1959; Pardee et al, 1959).[4] A remarkable property that distinguishes regulatory genes from structural genes is their pleiotropic effect that is simultaneously exerted on the rate of protein synthesis of multiple other genes (Jacob and Monod, 1959, 1961).

This distinction became an important component of Jacob and Monod's operon-model (Jacob et al, 1960; Jacob and Monod, 1961) and has since persisted. The *Oxford Dictionary of Biology*, for example, defines regulatory genes as "genes that control development by regulating the expression of structural genes responsible for the formation of body components. They encode transcription factors, which interact with regulatory sites of other genes causing activation or repression of developmental pathways" (Hine and Martin, 2004, p. 551). Evelyn Fox Keller explicitly refers to this functional separation of genes when describing an operon as "a linked cluster of regulatory elements and structural genes whose expression is coordinated by the product of a regulator gene situated elsewhere in the genome." (Keller, 2000, p. 57). A more generalized separation underlies the reasoning of Peter and Davidson in the context of gene regulatory networks (GRNs). They argue that "by definition, structural genes do not possess [...] regulative capacity. [...] Their expression is controlled by the GRN, but they do not contribute to the GRN" (Peter and Davidson, 2009, p. 3948).

Models of regulatory mechanisms have been substantially refined and extended since the pioneering work of Jacob and Monod. Gene regulation has been found to regard multiple levels ranging from transcriptional control (e.g. through transcription factors and co-factors) over post-transcriptional mechanisms (e.g. post-transcriptional silencing through microRNAs; see below) and translational regulation (e.g. translation initiation factors) to post-translational modifications (e.g. protein cleavage). For the purpose of this paper, it suffices to consider those genes as regulatory that affect the expression (transcript) levels of other genes by means of transcriptional or post-transcriptional control—that regards most, if not all, genes—although the specific translational and post-translational regulation of the activity of a subset of genes is important for many cellular functions.

The widely accepted view that distinguishes between regulatory genes and structural genes is challenged by recent findings that indicate that in eukaryotes even structural genes may have a regulatory function that is independent of the roles of their gene products. I describe these findings and briefly discuss in what biological context they may apply.

3 Potential regulatory function of all exported transcripts

MicroRNAs are short RNA molecules of a length of about 20 to 22 nucleotides that are matured by cleavage from longer precursor-microRNAs after their export from a cell's

[4] Jacob and Monod (1959) supposed that the *lac* repressor, i.e. the gene product of *lacI*, was not a protein but it was identified as such a few years later by Gilbert and Müller-Hill (1966).

nucleus to the cytoplasm. Such precursor-microRNAs can originate, for example, from specifically transcribed DNA segments with own promoters or from introns of spliced protein-coding genes.

Exploiting the base-pairing capabilities of these single-stranded microRNAs, the RNA-induced silencing complex (RISC) can recognize specific binding sites on messenger RNA (mRNA) transcripts and downregulate their function by either promoting their degradation or blocking their translation (Bartel, 2009; Kawamata and Tomari, 2010). Hence, the combined system of microRNAs and RISC complexes forms an important mechanism of post-transcriptional silencing.

Pier Paolo Pandolfi and his co-workers have hypothesized that long RNAs (mRNAs, transcripts of pseudogenes[5] and maybe other non-coding RNAs) possess an additional biological role—other than an eventual protein-coding function—that relies upon a competition for microRNA binding (Poliseno et al, 2010). They transfected cells with a retroviral vector expressing the 3' untranslated region (3'UTR)[6] of either the PTEN tumor suppressor gene or its pseudogene PTENP1 (that have several microRNA binding sites in common) and found that in both cases the overexpression of the 3'UTR transcript led to an upregulation of the wildtype PTEN and PTENP1 mRNA levels. This upregulation was not observed when the microRNA maturation of the cells was impaired. These astonishing results suggest that the 3'UTR, that is believed to contain many if not most of the functional microRNA binding sites of an mRNA (Gu et al, 2009; Forman and Coller, 2010), is capable of diverting microRNAs that otherwise would downregulate PTEN and its pseudogene and other unrelated target genes.

Consequently, PTENP1, although not being able to code for a functional protein, was shown to have an active biological role because increasing expression levels of the pseudogene indirectly cause an increase in PTEN expression due to a less effective microRNA-mediated downregulation ("derepression"). Similar observations were made for the oncogene KRAS and its pseudogene KRAS1P.

Therefore, an upregulation of a microRNA target transcript, even if it does not encode for a transcription factor, can in theory increase the cellular levels of other targets of the same microRNA. Even RNA molecules that have been believed to be non-functional relics of evolutionary processes, like those originating from many pseudogenes, can thus be modulators of gene expression.

How prevalent this post-transcriptional mechanism of non-coding regulatory functions of both mRNAs and processed pseudogenes is, has still to be determined, but the fact that it depends only on the presence of microRNA binding sites on the transcripts (apart from the presence and molecular concentrations of the respective microRNAs) indicates that it may be widely diffused.

This reasoning is supported by another recent study that showed that an increase in total target abundance dilutes the regulatory activity of many microRNAs—it reduces their average effect on each individual gene—which in theory allows for crosstalk between targets (Arvey et al, 2010).

Moreover, artificial microRNA decoys termed 'microRNA sponges' (Ebert et al, 2007), that rely on the same principles, have been successfully used to divert microR-

[5] Pseudogenes resemble copies of classical protein-coding genes, but are considered mostly biologically inactive due to mutations or premature stop codons that impair their translation into functional proteins. Yet, many pseudogenes appear to be under selective pressure and represent a significant portion of the 'transcriptome' (Harrison et al, 2005).

[6] The region of an mRNA transcript that follows the stop codon and is thus not translated into amino acids during protein synthesis.

NAs from their endogenous mRNA targets causing their derepression. Several natural microRNA sponges have been recently found or are suspected to be present in animals, plants, and viruses (reviewed in Ebert and Sharp, 2010).

4 In what biological contexts can these results apply?

It is important to realize that this intriguing, novel regulatory 'decoy mechanism' (Poliseno et al, 2010) largely depends on the relative cellular concentrations of three major components: the microRNAs, the RISC complexes and their target transcripts. Only this inter-dependence between the abundances of the three components confers the targets their indirect regulatory capacity.

The indirect upregulation of other targets of the same microRNAs can potentially apply to many gene transcripts that are exported from the cell's nucleus—*independent of whether the transcripts possess any other regulatory function or not*—even to genes usually thought of as structural genes. However, the presence of microRNAs is mostly tissue-specific, hence whether or when a transcript acquires such a microRNA-mediated regulatory function depends on the internal environment of the cell.

The extent of this mechanism depends also on the genome-wide activity of microRNAs. To date most interactions between microRNAs and binding sites on target transcripts have only been predicted and experimental confirmation for most of them is still lacking (Rajewsky, 2006), but genome-wide measurements of gene expression after microRNA transfection or endogenous microRNA knockdown suggest that a microRNA can, by direct or indirect effects, tune protein synthesis of thousands of genes (Selbach et al, 2008).

It is also clear that these considerations can hold only with respect to organisms that possess a microRNA-mediated (or similar) silencing pathway. Interestingly, although microRNAs are thought to be active mostly in the cytoplasm, a related ribonucleoprotein complex directed by short interfering RNAs (siRNAs)—closely related although not identical to microRNAs (Tang, 2008)—can cause post-transcriptional gene silencing in the nucleus (Guang et al, 2010; Hoffer et al, 2011).

Taken together, these recent findings question whether a distinction between regulatory and structural genes is always appropriate. This distinction may still be valid in procaryotes but is likely not to have a generic applicability. A distinction at the protein level can, however, still be justified, since the regulatory mechanisms discussed here are independent of the encoded gene products and proteins can have exclusively regulatory or structural functions.

5 Final remarks

The above illustrated microRNA decoy mechanism does not only provide a novel layer of post-transcriptional control of gene expression, significantly increasing the complexity of an already complex picture of gene regulation; it also has implications of both conceptual and practical nature.

As already pointed out in the introduction, gene concepts integrate functional criteria of delimitation because what counts as a gene depends on what phenotype is chosen for its definition; the choice of phenotype in turn reflects our knowledge about genes

(Burian, 2005). The knowledge that transcripts may play a biological role if they contain target sites that can compete for the binding of microRNAs, even if they have no other obvious biological function, must be taken into account when deciding whether a particular DNA segment is to be considered a gene. This issue is related to the status of transcribed pseudogenes (of both structural and regulatory protein-coding genes) that are considered to be mostly biologically inactive. Many of them, like PTENP1, may turn out to have kept a regulatory function by providing natural microRNA sponges and might be classified as regulatory genes rather than pseudogenes.

Keeping in mind that the distinction between structural genes and regulatory genes may not always be appropriate, that is, that structural genes may have a regulatory function, is important. Many structural genes may turn out to be part of gene regulatory networks, perhaps having more significant roles in development than appreciated today (see also Ebert and Sharp, 2010).

An indirect regulatory function of transcripts that is independent of the molecular function of the structural gene products may affect also other biological concepts that are related to gene concepts. For example, concepts of genetic traits, i.e. accounts of what it is for a phenotypic trait to be genetic. One possible criterion to judge whether a trait is genetic is the 'proper individuation criterion' (PI) that requires that a trait is specifically caused by the relevant gene or genes (Gifford, 2000). Intuitively, proteins, or the encoding of proteins, should be considered as genetic traits in this sense. However, like Gifford (2000) realized, if genes cause else besides the "direct gene product"—like it is the case if transcripts affect expression levels of other genes via the microRNA decoy mechanism—then applying PI may in many cases not yield the conclusion that a gene causes a protein specifically. Hence, the encoding of a protein could, at least in some cases, not be considered as a genetic trait according to PI anymore.

However, these considerations are not meant to undermine the importance of the notion of the structural gene *per se*. It does still make sense to talk about structural genes, e.g. to indicate that they do not encode for transcription factors, etc. It is only the strict separation from the notion of the regulatory gene that is troublesome because this separation suggests, not always correctly as we have seen, that structural genes are not involved in the regulation of the expression levels of other genes.

Whether the microRNA decoy mechanism illustrated here can be considered as genuinely regulatory will likely be a matter of debate. According to the above reported definition given by the *Oxford Dictionary of Biology*, for example, regulatory genes are those that encode for transcription factors, which would rule out indirect regulatory capabilities such as those described here.[7] On the other hand, taking the definition used by Peter and Davidson, structural genes do not possess regulative capacity, which would rule out not only transcription factors, as we have seen.

Are hence all genes regulatory genes? It is very unlikely that all genes will turn out to be regulatory genes, but a gene can in principle have a role in the regulation of gene expression even if it encodes for a structural protein. A rigorous distinction between structural genes and regulatory genes tends to overlook both the complexities inherent in gene regulatory mechanisms and the context-sensitivity of biological definitions.

[7] But then, most regulatory capabilities are indirect as they must generally rely on the existence of appropriate cellular processes, being these the synthesis of transcription factor proteins, the splicing of nascent transcripts, or the silencing mediated through microRNAs.

Acknowledgements I would like to thank the anonymous reviewers for their considerations and suggestions that have helped in significantly improving the manuscript.

References

Arvey A, Larsson E, Sander C, Leslie CS, Marks DS (2010) Target mRNA abundance dilutes microRNA and siRNA activity. Molecular Systems Biology 6:363

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233

Burian RM (2005) Too many kinds of genes? some problems posed by discontinuities in gene concepts and the continuity of the genetic material. In: The Epistemology of Development, Genetics, and Evolution, Cambridge University Press, New York, pp 166–178

Ebert MS, Sharp PA (2010) Emerging roles for natural microRNA sponges. Current Biology 20:R858–R861

Ebert MS, Neilson JR, Sharp PA (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nature Methods 4:721–726

Falk R (2000) The gene: A concept in tension. In: Beurton P, Falk R, Rheinberger HJ (eds) The Concept of the Gene in Development and Evolution, Cambridge University Press, Cambridge, pp 317–348

Forman JJ, Coller HA (2010) The code within the code: microRNAs target coding regions. Cell Cycle 9:1533–1541

Gatherer D (2010) So what do we really mean when we say that systems biology is *holistic*? BMC Systems Biology 4:22

Gifford F (2000) Gene concepts and genetic concepts. In: Beurton P, Falk R, Rheinberger HJ (eds) The Concept of the Gene in Development and Evolution, Cambridge University Press, Cambridge, pp 40–66

Gilbert W, Müller-Hill B (1966) Isolation of the lac repressor. Proc Natl Acad Sci USA 56:1891–1898

Griffiths PE, Stotz K (2007) Gene. In: Hull DL, Ruse M (eds) The Cambridge Companion to the Philosophy of Biology, Cambridge University Press, New York, pp 85–102

Gu S, Jin L, Zhang F, Sarnow P, Kay MA (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. Nat Struct Mol Biol 16:144–150

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature 465:1097–1101

Harrison PM, Zheng D, Zhang Z, Carriero N, Gerstein M (2005) Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Research 33:2374–2383

Hine RS, Martin E (eds) (2004) Oxford Dictionary of Biology, 5th edn. Oxford University Press, Oxford, UK

Hoffer P, Ivashuta S, Pontes O, Vitins A, Pikaard C, Mroczka A, Wagner N, Voelker T (2011) Posttranscriptional gene silencing in nuclei. Proc Natl Acad Sci USA 108:409–414

Jacob F, Monod J (1959) Gènes de structure et gènes de régulation dans la biosynthèse des protéines. Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences 249:1282–1284

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3(3):318–356

Jacob F, Perrin D, Sanchez C, Monod J (1960) L'opéron: groupe de gènes à expression coordonnée par un opérateur. Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences 250:1727–1729

Kawamata T, Tomari Y (2010) Making RISC. Trends in Biochemical Sciences 35:368–376

Keller EF (2000) The Century of the Gene. Harvard University Press, Cambridge, MA

Keller EF, Harel D (2007) Beyond the gene. PLoS ONE 2(11):e1231

Moss L (2003) What Genes Can't Do. MIT Press, Cambridge, MA

Pardee A, Jacob F, Monod J (1959) The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β-galactosidase by *e. coli*. Journal of Molecular Biology 1:165–178

Peter IS, Davidson EH (2009) Modularity and design principles in the sea urchin embryo gene regulatory network. FEBS Letters 583:3948–3958

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465:1033–1038

Portin P (2009) The elusive concept of the gene. Hereditas 146:112–117

Rajewsky N (2006) MicroRNA target predictions in animals. Nature Genetics 38:S8–S13

Rheinberger HJ, Müller-Wille S (2009) Gene. In: Zalta E (ed) Stanford Encyclopedia of Philosophy, Spring 2009 Edition (revised from the Winter 2004 Edition)

Selbach M, Schwanhusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. Nature 455:58–63

Tang G (2008) siRNA and miRNA: an insight into RISCs. Nature 455:58–63