Biology and Philosophy (2007) 22:293–306
DOI 10.1007/s10539-006-9039-5
© Springer 2006

Reports of the death of the gene are greatly exaggerated

ROB KNIGHT
Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, 80309, USA (e-mail: rob@spot.colorado.edu; phone: +1-303-492-1984; fax: +1-303-492-7744)

Received 16 May 2006; accepted in revised form 2 June 2006

Introduction

The concept of the gene is under attack. Several recent books, including "The Century of the Gene" by Evelyn Fox Keller (Keller 2000) and "Who Wrote the Book of Life? A History of the Genetic Code" by Lily Kay (Kay 2000), argue that the focus on genes in biology stems from historical and cultural factors rather than from empirical factors, and will soon be at an end. More recent work has suggested that the concept of the gene may continue to be useful in empirical research (Waters 2004; Weber 2004; Burian 2005), but that this concept must evolve to become much more open and inclusive than molecular biologists have traditionally assumed (Portin 1993, 2000).

"What Genes Can't Do", by former cell biologist and current philosopher of biology Lenny Moss (Moss 2003b), aims another salvo at gene-centered research. Moss argues that biology, especially cancer research, has been greatly hindered by the conflation of two incompatible gene concepts. If Moss's central thesis is correct, the term 'gene' will soon have about the same relevance to ongoing scientific enquiry as 'phlogiston'. The book has garnered a variety of responses, ranging from "a precise, detailed and conceptually useful work" (Borry 2004) to "a rambling, pretentious synthesis based on a poor understanding of the biologic research literature" (Lusis 2003).

Unlike Keller and Kay, who rely primarily on rhetorical analysis, Moss bases his critique primarily on biological findings. In this review, I focus on three of Moss's main assertions. First, I address the assertion that genetic information is a problematic concept. Second, I ask whether the division of gene concepts into Gene-P and Gene-D accurately reflects the multifarious uses of the term 'gene' in modern biology, and whether the diversity of gene concepts is likely to have hindered current research. Finally, I give a brief overview of some recent findings in cancer research that support the tradition that genes are critical factors in tumorigenesis. I conclude that the gene is still a useful concept, and that reports of its death are greatly exaggerated.

Is genetic information a problematic concept?

One of Moss's central arguments is that the concept that genes carry information is problematic. Chapter 2 begins with an epigraph citing Schroedinger's classic introduction of the concept of the chromosome as a "code-script" (Schroedinger 1944). Moss discusses Schroedinger's derivation from equilibrium thermodynamics of the idea that the central unit of order in the cell must be digital rather than analog, and must therefore be located in the "aperiodic crystal" structure of the chromosome. This metaphor, which led to the proliferation of terms such as transcription, translation, codons, and reading frames, pervades modern molecular biology.

Moss argues that this informational language is misguided. He states that "...the effort here will be to demonstrate and argue the following: *Neither DNA nor any other aperiodic crystal constitutes a unique repository of heritable stability in the cell; in addition, the chemistry of the solid state does not constitute either a unique or even an ontological or causally privileged basis for explaining the existence and continuity of order in the living world*." (p. 76, emphasis in original).

Taken at face value, the first part of this assertion is now empirically well-supported. Many factors besides DNA can be stably inherited, including the classic examples of membrane configuration, methylation patterns and histone modifications that affect transcriptional regulation (Jablonka 1995). More recently, changes in translational silencing mediated by miRNAs have been shown to be heritable for at least a few generations (Fire et al. 1998). Information theory contains a fundamental symmetry. Anything that affects the outcome in a receiver can be considered as either a source or as noise in the channel, and therefore both genes and environmental factors can be considered to convey information about the phenotype (MacLaurin 1998). This symmetry means that anything that affects the phenotype and is stably heritable has as much right to be called genetic information as does a sequence of DNA (Oyama 1985; Griffiths and Gray 1994a, 1994b; Sterelny et al. 1996; Griffiths and Gray 1997; Griffiths and Knight 1998). However, although biologists usually think of genes as consisting solely of nucleic acids, contemporary interest in and funding for studies of other mechanisms of heritable variation are high.

However, Moss also makes the much weaker argument that DNA cannot be the sole source of genetic information because the information contained in the nucleotide sequence is one dimensional. Specifically, he argues against the idea that "...spatial arrangement is somehow prefigured and predetermined by the one-dimensional array of nucleic acids in the genes" (p. 94). This idea that a linear array cannot encode 3D information seems to be founded on some misapprehension of information theory. For example, both audio and data CDs store their information in a single linear spiral. The news that this 1D information cannot predetermine complex 3D spatial arrangements will come as a surprise to anyone who has enjoyed the intricate pattern of waveforms produced by Beethoven's 9th when played through a stereo system, played a

video game that incorporates 3D graphics, or used a CAM (computer-aided manufacturing system) to fabricate machine parts. The finding that the number of points in a line is equal to the number of points in the equivalent plane, volume, etc. dates back to Cantor's theory of transfinite numbers first published in 1895 and 1897 (Cantor 1955). In general, the same information can be represented in any number of dimensions. The fact that a DNA sequence is 1D should not be used as an argument against the idea that it can specify phenotypic information in 3D.

Two strong lines of evidence that suggest that DNA encodes information about fitness come from experiments that evolve either computer programs or nucleic acid sequences from random starting points. In both cases, the "genome" is fixed at a specified number of bits of information, either in a computer's memory or in a length of DNA.

Genetic algorithms that evolve computer programs were pioneered by John Holland and his graduate students in the 1960s (Holland 1975). In this procedure, a language that interprets bit patterns as programs is defined in advance. A collection of random bit patterns is generated. Each bit pattern is fed a set of input data and evaluated according to its ability to produce the correct output. The bit patterns that produce the most correct output are mutated by random changes to generate the next generation, and the cycle begins again.

Genetic algorithms provide a direct mechanism for finding the rare, random bit patterns that can be interpreted to produce interesting behavior, including a sorting algorithm almost as efficient as the best hand-coded sorts (Hillis 1991), specifications for electrical circuits that function as predicted (Rudnick et al. 1994), and even specifications for machines built out of motors, bearings and rods that display a range of different but effective strategies for moving across a tabletop when built in the physical world (Lipson and Pollack 2000). Thus, genetic algorithms can effectively find the extremely rare bit patterns out of a large but finite search space that, in the context of a system for interpreting them, produce interesting dynamical behaviors.

It could be argued that genetic algorithms do not truly mirror biological evolution because the interpretation of each bit pattern is predefined, and because the encoding scheme is artificial. This criticism is much more difficult to apply to SELEX, a procedure for artificially selecting functional nucleic acid sequences (Ellington and Szostak 1990; Robertson and Joyce 1990; Tuerk and Gold 1990). In this procedure, a pool of random-sequence DNA molecules of a defined length is synthesized. These DNA sequences are either transcribed to RNA, or left as DNA. The sequences are chemically screened for their ability to bind to a molecular target, or for their ability to perform a chemical reaction. The pool of sequences that survives the selection step is copied back into DNA and amplified by the polymerase chain reaction to create a new pool enriched in functional sequences. The cycle is then repeated. After 5–15 cycles of this procedure, the final pool is typically highly active. This procedure has been used to select RNA sequences with many different catalytic and binding activities, including binding to a range of amino acids (Yarus et al. 2005),

modification of RNA with amino acids (Illangasekare et al. 1995) and nucleotide cofactors (Huang and Yarus 1997), carbon–carbon bond formation (Tarasow et al. 1997) and metal nanoparticle formation (Gugliotti et al. 2004).

SELEX provides a direct example of the accumulation of information that specifies phenotype in nucleic acid sequences. The only information that each DNA molecule in the pool has is its nucleotide sequence, which can be represented as two bits of information per base. Although two RNA molecules transcribed from the same DNA with the same nucleotide sequence may have differences in activity because they have folded into different shapes, these differences can be reversed by heating and cooling, and in any case are not heritable. The fraction of all molecules in the starting pool that contain the sequence elements required for a particular biochemical activity can be calculated precisely using a combination of mathematical and laboratory techniques (Sabeti et al. 1997; Knight and Yarus 2003; Carothers et al. 2004; Knight et al. 2005; Legiewicz et al. 2005). Shannon's theorem (Shannon 1949) provides a direct conversion between probabilities and bits of information. We can thus calculate the number of bits of information in the nucleotide sequence required for a particular kind of RNA activity (Legiewicz et al. 2005).

The basic principle underlying SELEX is that the ability to perform a particular reaction requires stringent constraints on the sequence, which can be represented as information. These same constraints mean that specific sequence features can be highly predictive of function. Thus, the sequence can reasonably be described as containing information about the phenotype in the traditional sense, and the amount of information required for a specific phenotype can be measured quantitatively.

The SELEX example provides a clear case where the fitness of the genotype is determined by the laws of physics and chemistry, in contrast to the genetic algorithm (which might be considered artificial because the meaning of each bit pattern is predefined). However, the SELEX example is directly analogous to both the genetic algorithm example, in which all the information is clearly stored in the bit pattern, and to the case of biological evolution, in which the genome accumulates changes that increase fitness in the environment. It would be extremely difficult to define information in a way that would rule out the concept of genetic information in the cell without also leading to the absurd conclusion that computer programs stored on a CD do not carry the information required for their function. Arguments about context-sensitivity apply equally to both cases. The runtime environment required for any computer program to function, while not as complex as the cell, is still exceedingly intricate. The behavior of any program depends sensitively on the actions of a large number of other components of the system, as anyone who has tried to run old software on a new operating system can attest. Thus, either the widely accepted view that the genome stores information about the phenotype is correct, or we have to give up the idea that actual computer programs are made of information.

Is Gene-P/Gene-D the right division of gene concepts?

The two central claims of Moss's book, and of many of his other recent writings (Moss 2001; Moss 2002; Moss 2003a), is that there are precisely two gene concepts and that confusion between them has misled biologists. However, neither of these claims withstands close scrutiny.

As has now been well-documented (Waters 1994; Griffiths and Neumann-Held 1999; Beurton et al. 2000; Stotz and Griffiths 2004; Stotz et al. 2004), biologists use many different working definitions of the word 'gene'. In fact, there may not be a strict definition so much as a series of generalizations based on well-studied examples that the biological community agrees should be classified as 'genes' (Waters 1994). For example, ribosomal RNA genes encode no protein product and are present in many identical copies within each genome (Arnheim et al. 1980), genes can overlap one another (Barrell et al. 1976) or be scattered in pieces throughout the genome (Bonen 1993), and a single gene can produce many protein products through alternative splicing (up to 38,016 in the case of Dscam in *Drosophila melanogaster* (Celotto and Graveley 2001)). Each of these cases dramatically violates the classical "one gene, one enzyme" hypothesis (Beadle and Tatum 1941), but the response has always been to expand the definition of gene rather than to classify the newly-discovered objects as non-genes.

These new discoveries about the complexity of transcription and translation introduce clear incompatibilities between gene concepts. For example, G.C. Williams's Evolutionary Gene Concept defines a gene as "that which segregates and recombines" (Williams 1966). This concept closely resembles the classical notion of a locus. In contrast, the Classical Molecular Gene (CMG) concept, as Eva Neumann-Held has termed it, defines a gene as the segments of DNA that contribute to a functional product (Neumann-Held 1997). In the case of trans-splicing, where the pieces of DNA that lead to a single protein are found at many different loci (Bonen 1993), these two concepts are in direct conflict. However, these differences in gene concept do not prevent biologists from being able to communicate with one another about trans-splicing and other complex molecular phenomena.

Moss argues that there are precisely two concepts at work: the preformationist Gene-P, and the developmentalist Gene-D. The discussion is often difficult to follow, in part because in places these concepts of gene appear to refer to the locus, at other times to individual alleles, and at other times to types of alleles.

Briefly, a Gene-P is a type of DNA sequence that allows one to predict a phenotypic state (assuming a typical genetic and environmental background). Moss uses the example of a "gene for blue eyes". The clearest statement defining Gene-P is found on p. 45: "Genes for phenotypes, i.e. Genes-P, can be found, generally – and as Johanssen surmised – where some deviation from a normal sequence results with some predictability in a phenotypic difference".

In contrast, the Gene-D is an individual DNA sequence that acts as a developmental resource for the production of some molecular product. The clearest statement defining Gene-P is found on p. 46: "Quite unlike Gene-P, *Gene-D is defined by its molecular sequence*. A Gene-D is a developmental resource (hence the D) which in itself is *indeterminate* with respect to phenotype" (emphasis in original).

Moss argues that the Gene-P is defined with respect to phenotype but indeterminate with respect to DNA sequence, and that the Gene-D is defined with respect to DNA sequence but indeterminate with respect to phenotype. He states that "...the gene-centered perspective was built of a conflation of two individually warranted but mutually incompatible conceptions of the gene (Gene-P and Gene-D) and that these were held together by the rhetorical glue of the gene-as-text metaphor" (p. 184). In particular, he argues that the same sequence cannot count both as Gene-P and Gene-D.

Biologists may find these definitions to be rather restrictive versions of familiar concepts. The Gene-P is essentially the allele, with the restriction that all alleles with the same phenotypic effect count as the same Gene-P. The Gene-D is essentially the Classical Molecular Gene (Neumann-Held 1997) or the Domain Set for Active Transcription (DSAT) (Fogle 1990), with the restriction that two sequences with even a single-nucleotide difference cannot count as the same Gene-D. The incompatibilities between the two concepts arise primarily from the restrictions that Moss introduces, rather than from the ways in which biologists use the concepts in practice.

A "gene for" a trait is shorthand for "a locus in which sequence variation causes a difference in phenotype, all other things being equal" (Williams 1966). Much of the time, the reason that the sequence variation causes the difference is because it affects the function of a Classical Molecular Gene (CMG). To take Moss's oft-used example of cystic fibrosis, the reason that variation at the Cystic Fibrosis (CF) locus affects the functioning of the chloride ion channel is because it affects the CFTR (Cystic Fibrosis Transmembrane Receptor) CMG, causing a full or partial loss of function (Riordan et al. 1989). The usage in the original paper is fairly clear: the researchers report that they have found the locus in which variation is associated with the cystic fibrosis phenotype. They identify an allele at that locus in which the molecular function of the protein is disrupted. These changes explain the differences in the chloride ion channel activity that, across most genetic and environmental backgrounds, make the difference between normal and cystic fibrosis phenotypes in the homozygote. This is uncontroversial.

What is controversial is Moss's interpretation of the situation. By defining a Gene-D as a single sequence and a Gene-P as a set of sequences associated with equivalent phenotypes, Moss draws a token/type distinction that means that a Gene-D cannot be a Gene-P. He then concludes that the conflation of the two meanings has led to an inappropriate gene-centered focus in biology. However, the situation is easily resolved by paying attention to both the locus and the sequence. A CMG is typically found at a specific locus in all individuals within

a population. If a functional copy of the CMG is not present (on either copy if the organism is diploid), the organism will develop differently or fail to develop at all. Thus, any sequence that does not perform the function of the original CMG is an allele for the alternative state. However, it is only an allele for the alternative state because of its position on the chromosome: a perfectly normal CMG for, say, beta-globin would be an allele associated with cystic fibrosis if it were to replace the CFTR sequence at the CF locus. However, this is nothing more than a restatement of the idea that the phenotypic effect of a sequence depends on its context.

Thus, if Moss's point is that the function of a sequence is context-sensitive and difficult to predict from the sequence alone, he is saying nothing controversial. If we accept that a Gene-P is simply the set of sequences that leads to the same phenotypic result (in the context of a specific locus), and make the small adjustment that a Gene-D is the set of sequences that produces the same molecular product (in the context of a specific locus), then many Genes-P will also be Genes-D and modern molecular biology is saved.

A central point of Moss's argument is that Genes-P that cause differences in phenotype should not be considered developmental resources if their effect results from the inability to produce a particular molecular product. For example, the 'gene for blue eyes' results from the inability to synthesize a brown pigment that would otherwise mask the blue color, and the 'gene for cystic fibrosis' results from the inability to synthesize CFTR.

This case of loss-of-function alleles requires special attention. Moss's difficulties on this point largely seem to stem from a conflation of causal-role functions (Cummins 1975) and selected-effect functions (Neander 1991; Griffiths 1993; Godfrey-Smith 1994). For example, on p. 150 he notes that talk of the function of a gene in cancer cannot be due to etiology, because the ability for a gene to perform these tasks is not adaptive. However, such discussions invoke causal-role, not selected-effect, functions. A similar confusion occurs on p. 141, in which Moss denies that oncogenes carried by viruses can have the function of causing cancer because cancer is maladaptive for the host. Presumably these CMGs are adaptive for the virus and have some selected-effect function related to the ability of the virus to survive and reproduce: their role in cancer is a causal-role function. In general, Genes-P seem to be associated with causal-role functions, while Genes-D seem to be associated with selected-effect functions (because otherwise the sequences of related Genes-D would not be conserved over evolutionary time). Thus, if we accept that most discussion of ''genes for'' diseases refers to causal-role functions that can also be losses of selected-effect functions, the difficulty is resolved. The vast number of different alleles associated with cystic fibrosis can be described as different ways of losing an essential selected-effect function, and have causal-role functions associated with disease.

The argument that a loss-of-function allele can be a Gene-P but not a Gene-D hinges on the idea that the absence of something cannot be a developmental resource. For example, both the presence of oxygen and the absence of cyanide

are key requirements for human survival. Is it reasonable to list 'absence of cyanide' as part of the developmental system? In any case, the inability to form a particular molecular product does not prevent a DNA sequence from being a developmental resource. For example, a mutant of Dscam that could produce 'only' 38,015 of the canonical 38,016 isoforms would still be a considered a CMG. This is not a novel concept: mutations that affect some but not all transcripts produced from a given CMG have been known for over 20 years (Karlik and Fyrberg 1985).

A related question is whether Genes-P/Genes-D does full justice to the range of gene concepts. Many gene concepts, such as Eva Neumann-Held's process molecular gene (PMG) that consists of all the resources required to make a specific molecular product (Neumann-Held 1997), do not fit under either definition. The Gene-P/Gene-D distinction does not address contemporary ambiguities in the literature (Fogle 2001), such as whether regulatory sequences should be considered part of the gene or adjacent to the gene, whether transgenes moved to a different locus or species count as the same gene as the original or as different genes, and whether homologous sequences in different species should be considered the same gene or related genes. I would argue that a complete account of what makes two objects the 'same gene' requires, at minimum, the concepts of locus, allele, CMG, PMG, and ortholog. However, these concepts already exist, and technical discussions where confusion might arise typically use diagrams or descriptions that remove the ambiguities inherent in the definition of 'gene'. Like 'species', which also encompasses a multitude of incompatible definitions, it may not be important to know what the precise meaning of 'gene' is.

Is cancer genetic?

Although many biological examples are discussed throughout the book, Moss focuses much of his argument on cancer. Given his extensive discussion of Johanssen's concerns that alleles involved with disease may not be representative of heredity in general (pp. 28–48), it is surprising that Moss places so much emphasis on a disease himself. The argument rests on three assertions. First, normal mammalian cells have not been transformed into tumors through the addition or deletion of single Genes-D. Second, cancer is associated with changes in membrane structure, which are heritable within a cancer cell line. Moss proposes that these changes in membrane structure are the primary cause of cancer, and that the genomic instability and other changes in the DNA are simply effects. If true, this hypothesis would have substantial implications for cancer treatment. Third, not all carcinogens are genotoxic, and not all genotoxic agents are carcinogens, so the causes of cancer should not be assumed to be changes in DNA (although this does not imply, as Moss seems to think, that changes in DNA cannot cause cancer).

All three of these assertions are substantially misleading. Let us begin with the criticism that single changes in Genes-D are insufficient to transform normal cells into cancerous cells. Cancer is generally seen as a breakdown in cell regulation, and full-blown carcinogenesis requires a well-defined series of steps. These steps include self-sufficiency in growth signals, insensitivity to antigrowth signals, unlimited replicative potential, evasion of apoptosis, ability to sustain angiogenesis, and the ability to invade tissues and metastasize (Hanahan and Weinberg 2000). It should perhaps be unsurprising that no single change is capable of providing all these capabilities.

However, as early as 1999 it had been shown that a combination of three oncogenes (CMGs), the telomerase *hTERT*, the SV40 large-T oncoprotein, and an oncogenic allele of H-*ras*, could transform normal human epithelial and fibroblast cells into cancer cells (Hahn et al. 1999). Subsequent work defined the precise signaling pathways that this combination of Genes-D disrupts (Hahn et al. 2002). Similarly, the difficulties in transforming breast cells into tumor cells through mutations in *BRCA1* and *BRCA2* turn out to be due to differences in the responses of these cells to culture plate conditions and to the 3D matrix in which they are embedded *in vivo*. Advances in culturing techniques have allowed oncogenic transformation of normal mammalian cells using a variety of oncogenes (Elenbaas et al. 2001; Kuperwasser et al. 2004). Thus, the contention that changes in Genes-D alone cannot cause cancer is clearly incorrect.

Moss's second assertion is that changes in membranes are the primary cause of cancer, and that genomic instability and other genetic changes are effects. However, human tumor cells can be derived without genomic instability through the simultaneous introduction of three oncogenes (Zimonjic et al. 2001). This greatly weakens the idea that changes in the membranes are primary.

The final assertion is that not all carcinogens are genotoxic, and not all genotoxins are carcinogens. Therefore, changes other than those in DNA sequence must be able to cause cancer, and the focus on genes involved in cancer is misguided. Cancer begins with a breakdown in signaling, which could occur through many pathways (Hanahan and Weinberg 2000). It is thus unsurprising that compounds that affect signaling pathways by acting on mRNAs or proteins directly can perturb the state of the cell in a manner similar to introducing or deleting the relevant Gene-D. One cautionary note is that it can be difficult to tell whether a particular compound is genotoxic or not, because many non-genotoxic compounds are converted into genotoxic compounds in the liver or elsewhere (Shimada and Fujii-Kuriyama 2004). Compounds can also exert their activity through changes in promoter methylation, which are stably inherited, rather than changing the DNA sequence (Pereira et al. 2004). Many other non-genotoxic carcinogens affect well-known signaling pathways. For example, phorbol esters target members of the Protein Kinase C family (Brose and Rosenmund 2002; Roux et al. 2004). There are many other well-studied examples (Rakitsky et al. 2000; Kaneko et al. 2002; Mally and Chipman 2002;

Iida et al. 2003). Thus, although there are many ways to disrupt signaling other than by mutating or introducing Genes-D, the effects are still mediated by signaling proteins that are translated from DNA templates, and understanding the functions of these molecules through genetic alterations is a useful research strategy. It should also be noted that the abundance of studies on non-genotoxic carcinogens suggests that the allegedly misguided focus on genes in cancer has not prevented the elucidation of the mechanism of action of compounds that do not directly mutate DNA.

Has the Gene-P/Gene-D conflation really impeded biology?

Although many factors other than DNA sequences can be heritable (Jablonka 1995), DNA remains uniquely useful for laboratory studies. The pragmatic benefits of being able to sequence, amplify, insert and express DNA sequences, along with the algorithmic benefits that arise from our ability to treat DNA sequences as character strings, greatly outweigh any theoretical concerns about factors that are being missed due to lack of study. If a researcher's goal is to determine the effect of the gain or loss of a specific molecular product such as a transcript or protein, manipulating the DNA sequence is clearly the best option available in most cases. However, newer techniques, such as RNAi (Fire et al. 1998), are making inroads in organisms where they are available, undermining Moss's arguments that biological researchers are blind to options other than manipulating DNA. The advantage of manipulations of DNA are that they are stably heritable and can be applied to any sequence in the genome. Manipulating patterns of expression by, say, changing methylation patterns or by screening a chemical library for an inhibitor that affects a specific enzyme would be far more difficult and, in most cases, would produce no new information. Thus the focus on DNA can be seen as a primarily pragmatic decision, rather than an ideologically motivated one. One analogy might be that computer programmers choose to rearrange the bits of a program in memory rather than the structure of the CPU to create new system behavior, although both approaches are in principle possible. Few would argue that 'program-centered computer science' results from ideology rather than pragmatism.

By defining a Gene-D as a single sequence and a Gene-P as a set of sequences, Moss ensures that the two concepts are subject to token/type incompatibility. However, the definition of a Gene-D as a single sequence flies in the face of contemporary usage. If we relax the definition of a Gene-D to refer to a set of sequences that produce the same molecular product (perhaps at the same level and with the same regulation), in most cases the Gene-D and the Gene-P become one and the same. This relationship stems primarily from genome organization: nearly every Gene-D occupies a single locus at which its absence would lead to a different phenotype. If Genes-D were arbitrarily scattered through the genome in multiple copies, as is the case for ribosomal RNA sequences and histone sequences, it would be much more difficult to use

the sequence at a single site for prediction because its effect might be complemented by other sites. However, this is simply a statement that epistasis exists, a fact well known to geneticists for almost a century (Bateson 1909).

In the final chapter, Moss advocates a move away from "gene-centered biology". He suggests that researchers should consider heritable factors other than DNA as potential causes for conditions, and seek interventions other than changes in DNA sequence as cures. However, it is difficult to see how this would be different from what biologists and medical researchers actually do. Gene therapy is still an extremely marginal technique compared to drugs or radiation, and researchers typically consider a wide range of causes beyond single-locus defects for disorders (although these causes are often prioritized in terms of experimental accessibility for research purposes). Systems biology (Ideker et al. 2001), although an emerging discipline, already seeks to integrate our understanding of DNA, RNA, protein, small-molecules, post-translational modifications, membranes, signaling interactions, environmental inputs, and so forth into a unified, predictive account of specific cells and organisms.

However, it seems unlikely that attempting technically difficult possibilities before the technically facile possibilities have been exhausted will prove to be a productive research strategy. Thus, the gene is likely to continue its central role in biology for the foreseeable future.

Acknowledgements

I would like to thank Natalie Ahn for providing me with an overview of recent findings in cancer research, Karola Stotz for sharing related work before publication, and Amanda Birmingham for critical discussion of the manuscript.

References

Arnheim N., Krystal M., Schmickel R., Wilson G., Ryder O. and Zimmer E. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proc. Natl. Acad. Sci. USA 77: 7323–7327.

Barrell B., Air G. and Hutchison C. 1976. Overlapping genes in bacteriophage phix174. Nature 264: 34–41.

Bateson, W. 1909. Mendel's Principles of Heredity. Cambridge University Press.

Beadle G.W. and Tatum E. 1941. Genetic control of biochemical reactions in Neurospora. Proc. Natl. Acad. Sci. USA 27: 499–506.

Beurton P.J., Falk R., Rheinberger H.-J. (eds) 2000. The Concept of the Gene in Development and Evolution: Historical and Epistemological perspectives. Cambridge University Press.

Bonen L. 1993. Trans-splicing of pre-mRNA in plants, animals, and protists. FASEB J. 7: 40–46.

Borry P. 2004. Book review: Moss, Lenny. What genes can't do. Theor. Med. 25: 75–77.

Brose N. and Rosenmund C. 2002. Move over protein kinase C, you've got company: alternative cellular effectors of diacylglycerol and phorbol esters. J. Cell. Sci. 115: 4399–4411.

Burian R.M. 2005. The Epistemology of Development, Evolution and Genetics. Cambridge University Press.
Cantor G. 1955. Contributions to the Founding of the Theory of Transfinite Numbers. Dover Publications.
Carothers J., Oestreich S., Davis J. and Szostak J. 2004. Informational complexity and functional activity of RNA structures. J. Am. Chem. Soc. 126: 5130–5137.
Celotto A. and Graveley B. 2001. Alternative splicing of the Drosophila DSCAM pre-mRNA is both temporally and spatially regulated. Genetics 159: 599–608.
Cummins R. 1975. Functional analysis. J. Philos. 72: 741–765.
Elenbaas B., Spirio L., Koerner F., Fleming M., Zimonjic D., Donaher J., Popescu N., Hahn W. and Weinberg R. 2001. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. 15: 50–65.
Ellington A. and Szostak J. 1990. In vitro selection of RNA molecules that bind specific ligands. Nature 346: 818–822.
Fire A., Xu S., Montgomery M.K., Kostas S.A., Driver S.E. and Mello C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391(6669): 806–811.
Fogle T. 1990. Are genes units of inheritance?. Biol. Philos. 5: 349–371.
Fogle, T. 2001. The dissolution of protein coding genes in molecular biology. In: Beurton P., Falk R. and Rheinberger H. (eds), The Concept of the Gene in Development and Evolution. Cambridge University Press.
Godfrey-Smith P. 1994. A modern history theory of functions. Nous 28: 344–362.
Griffiths P. and Neumann-Held E. 1999. The many faces of the gene. Biosciences 49: 656–662.
Griffiths P.E. 1993. Functional analysis and proper functions. Br. J. Philos. Sci. 44: 409–422.
Griffiths P.E. and Gray R.D. 1994a. Developmental systems and evolutionary explanation. J. Philos. XCI: 277–304.
Griffiths P.E. and Gray R.D. 1994. Replicators and vehicles? or developmental systems?. Behav. Brain Sci. 17: 623–624.
Griffiths P.E. and Gray R.D. 1997. Replicator II: judgement day. Biol. Philos. 12: 471–492.
Griffiths P.E. and Knight R.D. 1998. What is the developmentalist challenge?. Philos. Sci. 65: 253–258.
Gugliotti L., Feldheim D. and Eaton B. 2004. RNA-mediated metal-metal bond formation in the synthesis of hexagonal palladium nanoparticles. Science 304: 850–852.
Hahn W., Counter C., Lundberg A., Beijersbergen R., Brooks M. and Weinberg R. 1999. Creation of human tumour cells with defined genetic elements. Nature 400: 464–468.
Hahn W., Dessain S., Brooks M., King J., Elenbaas B., Sabatini D., DeCaprio J. and Weinberg R. 2002. Enumeration of the simian virus 40 early region elements necessary for human cell transformation. Mol. Cell. Biol. 22: 2111–2123.
Hanahan D. and Weinberg R. 2000. The hallmarks of cancer. Cell 100: 57–70.
Hillis D. 1991. Co-evolving parasites improve simulated evolution as an optimization procedure. Artif. Life II SFI Stud. Sci. Complex. 10: 313–324.
Holland J. 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press.
Huang F. and Yarus M. 1997. Versatile 5' phosphoryl coupling of small and large molecules to an RNA. Proc. Natl. Acad. Sci. USA 94: 8965–8969.
Ideker T., Galitski T. and Hood L. 2001. A new approach to decoding life: systems biology. Annu. Rev. Genom. Hum. Genet. 2: 343–372.
Iida M., Anna C., Hartis J., Bruno M., Wetmore B., Dubin J., Sieber S., Bennett L., Cunningham M., Paules R., Tomer K., Houle C., Merrick A., Sills R. and Devereux T. 2003. Changes in global gene and protein expression during early mouse liver carcinogenesis induced by non-genotoxic model carcinogens oxazepam and wyeth-14,643. Carcinogenesis 24: 757–770.
Illangasekare M., Sanchez G., Nickles T. and Yarus M. 1995. Aminoacyl-rna synthesis catalyzed by an RNA. Science 267: 643–647.

Jablonka E. 1995. Epigenetic Inheritance and Evolution: The Lamarckian Dimension. Oxford University Press.
Kaneko M., Morimura K., Nishikawa T., Wanibuchi H., Takada N., Osugi H., Kinoshita H. and Fukushima S. 2002. Different genetic alterations in rat forestomach tumors induced by genotoxic and non-genotoxic carcinogens. Carcinogenesis 23: 1729–1735.
Karlik C. and Fyrberg E. 1985. An insertion within a variably spliced drosophila tropomyosin gene blocks accumulation of only one encoded isoform. Cell 41: 57–66.
Kay L.E. 2000. Who Wrote the Book Of Life? A History of the Genetic Code. Stanford University Press.
Keller E.F. 2000. The Century of the Gene. Harvard University Press.
Knight R., De Sterck H., Markel R., Smit S., Oshmyansky A. and Yarus M. 2005. Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. Nucleic Acids Res. 33: 5924–5935.
Knight R. and Yarus M. 2003. Finding specific RNA motifs: function in a zeptomole world?. RNA 9: 218–230.
Kuperwasser C., Chavarria T., Wu M., Magrane G., Gray J., Carey L., Richardson A. and Weinberg R. 2004. Reconstruction of functionally normal and malignant human breast tissues in mice. Proc. Natl. Acad. Sci. USA 101: 4966–4971.
Legiewicz M., Lozupone C., Knight R. and Yarus M. 2005. Size, constant sequences, and optimal selection. RNA 11: 1701–1709.
Lipson H. and Pollack J.B. 2000. Automatic design and manufacture of robotic lifeforms. Nature 406(6799): 974–978.
Lusis A.J. 2003. What genes can't do, by Lenny Moss. Nature Med 9: 501.
MacLaurin J. 1998. Reinventing molecular Weismannism: information in evolution. Biol. Philos. 13: 37–59.
Mally A. and Chipman J. 2002. Non-genotoxic carcinogens: early effects on gap junctions, cell proliferation and apoptosis in the rat. Toxicology 180: 233–248.
Moss L. 2001. Deconstructing the gene and reconstructing molecular developmental systems. In: Oyama S., Griffiths P.E. and Gray R.D. (eds), Cycles of Contingency: Developmental Systems and Evolution, MIT Press, pp. 85–97.
Moss L. 2002. From representational preformationism to the epigenesis of openness to the world? reflections on a new vision of the organism. Ann. NY. Acad. Sci. 981: 219–229.
Moss L. 2003a. One, two (too?), many genes?. Quart. Rev. Biol. 78: 57–67.
Moss L. 2003b. What Genes Can't Do. MIT Press.
Neander K. 1991. Function as selected effects: The conceptual analyst's defense. Philos. Sci. 58: 168–184.
Neumann-Held E.M. 1997. The gene is dead – long live the gene: Conceptualising the gene the constructionist way. In: Koslowski P. (ed.), Developmental Systems, Competition and Cooperation in Sociobiology and Economics. Springer-Verlag.
Oyama S. 1985. The Ontogeny of Information. Cambridge University Press.
Pereira M., Wang W., Kramer P. and Tao L. 2004. DNA hypomethylation induced by non-genotoxic carcinogens in mouse and rat colon. Cancer Lett. 212: 145–151.
Portin P. 1993. The concept of the gene: short history and present status. Quart. Rev. Biol. 68: 173–223.
Portin P. 2000. The origin, development and present status of the concept of the gene: a short historical account of the discoveries. Current Genom. 1: 29–40.
Rakitsky V., Koblyakov V. and Turusov V. 2000. Nongenotoxic (epigenetic) carcinogens: pesticides as an example. A critical review. Teratog. Carcinog. Mutagen. 20: 229–240.
Riordan J., Rommens J., Kerem B., Alon N., Rozmahel R., Grzelczak Z., Zielenski J., Lok S., Plavsic N. and Chou J. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066–1073.
Robertson D. and Joyce G. 1990. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. Nature 344: 467–468.

Roux P., Ballif B., Anjum R., Gygi S. and Blenis J. 2004. Tumor-promoting phorbol esters and activated ras inactivate the tuberous sclerosis tumor suppressor complex via p90 ribosomal s6 kinase. Proc. Natl. Acad. Sci. USA 101: 13489–13494.

Rudnick E.M., Holm J.G., Saab D.G. and Patel J.H. 1994. Application of simple genetic algorithms to sequential circuit test generation. In: EDAC-ETC-EUROASIC, pp. 40–45.

Sabeti P., Unrau P. and Bartel D. 1997. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. Chem. Biol. 4: 767–774.

Schroedinger E. 1944. What is Life? The Physical Aspect of the Living Cell. Cambridge University Press.

Shannon C.E. 1949. The Mathematical Theory of Communication. University of Illinois Press.

Shimada T. and Fujii-Kuriyama Y. 2004. Inhibition of human cytochrome P450 1a1-, 1a2-, and 1b1-mediated activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1a1 and 1b1. Cancer Sci. 95: 1–6.

Sterelny K., Smith K.C. and Dickison M. 1996. The extended replicator. Biol. Philos. 11: 377–403.

Stotz K. and Griffiths P. 2004. Genes: philosophical analyses put to the test. Hist. Philos. Life Sci. 26: 5–28.

Stotz K., Griffiths P.E. and Knight R.D. 2004. How biologists conceptualize genes: an empirical study. Stud. Hist. Philos. Biol. Sci. C Biomed. Sci. 35: 647–673.

Tarasow T., Tarasow S. and Eaton B. 1997. RNA-catalysed carbon-carbon bond formation. Nature 389: 54–57.

Tuerk C. and Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505–510.

Waters C.K. 1994. Genes made molecular. Philos. Sci. 61: 163–185.

Waters C.K. 2004. What was classical genetics?. Stud. Hist. Philos. Sci. C Med. Sci. 35: 783–809.

Weber M. 2004. Philosophy of Experimental Biology. Cambridge University Press.

Williams G.C. 1966. Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought. Princeton University Press.

Yarus M., Caporaso J. and Knight R. 2005. Origins of the genetic code: the escaped triplet theory. Annu. Rev. Biochem. 74: 179–198.

Zimonjic D., Brooks M., Popescu N., Weinberg R. and Hahn W. 2001. Derivation of human tumor cells in vitro without widespread genomic instability. Cancer Res. 61: 8838–8844.