Are There Genes?[1]

JOHN DUPRÉ

Introduction

Contrary to one possible interpretation of my title, this paper will not advocate any scepticism or ontological deflation. My concern will rather be with how we should best think about a realm of phenomena the existence of which is in no doubt, what has traditionally been referred to as the genetic. I have no intention of questioning a very well established scientific consensus on this domain. It involves the chemical DNA, which resides in almost all our cells, which is capable of producing copies of itself that accurately reproduce a very long sequence of components, and which plays a role in the physiology of the cell which in certain basic respects is quite well understood. This substance has also achieved a remarkable iconic status in contemporary culture. It is seen as fundamental to personal identity both in the practical sense of providing a criterion of identity through DNA testing, and in the much deeper sense of being seen as, somehow, defining who we are. The latter role is illustrated, for example, by the recent debate about the right of children conceived by sperm donation to know who are their fathers. Such people, it is passionately argued, must be able to find out where they came from, who they really are. On a daily basis we are confronted with claims about the discovery of the genetic basis of—or in fact very often the 'gene for'—all manner of psychological and physical characteristics, and all kinds of disorders. This holds out apparent possibilities for curing or preventing diseases or for eugenic control over future generations. But more subtly it contributes to an increasingly general assumption that what we are depends more than anything else on our genetic endowment.

In this paper I want to address the question how we should best understand the phenomena that underlie all these ways of thinking.

[1] The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was part of the programme of the ESRC Research Centre for Genomics in Society (Egenis). I am also grateful to several colleagues in Egenis for advice and comments on earlier drafts.

A general thesis about which I shall say something, is that what scientific experts say doesn't in fact provide much support for these wider general cultural understandings of genetics and of DNA. But my main focus will, as my title suggests, be largely ontological. I want to start not with genes, but with the genome, the totality of our genetic material. What kind of a thing is this, and in what sense may it usefully be considered as composed, at least in part, of genes? As I have already indicated, one thing the genome is is a quantity of a particular chemical, DNA. But presumably there is more to the genome than just what it is made of.

In the end, of course, our account of what such a thing as a genome is can only be derived from the science through which such an object is presented to us. However rather than approach the science directly, in order to bring into contact this scientific picture and more popular understandings of the genome, I shall consider some of the very familiar metaphors through which genomes (and genes, and DNA) are often described by scientific experts and assimilated by a wider public. The genome is often said to embody a code; to be a repository of information; to provide a blueprint for the developed organism; or perhaps a recipe; and so on. All of these are clearly metaphorical, and my proposal is to assess the aptness of these metaphors in relation to our contemporary scientific understanding of what genomes are and what they do. With a better understanding of this question it should be easier to see what turns on the idea that the genome contains or consists of genes, what kinds of things genes are or might be, and whether they provide a useful way of distinguishing components of the genome.

Let me repeat that although I shall be critical of some of the metaphors used to describe genes and genomes, this is not intended as criticism of the science which these metaphors are used to describe—indeed the criticism is largely dependent on taking the science on trust. I certainly do think philosophers can sometimes usefully criticise parts of science, and I have occasionally attempted this myself. But this is not my present aim and in fact the area of science I am talking about is one which, on the whole, I find impressive and admirable. Although the metaphors I'm considering no doubt play a role in the thinking of scientific practitioners, in the present case I suspect the role is often minimal or, at any rate, harmless. Contemporary genomics and molecular genetics is thoroughly, and in this case admirably, mechanistic. It is about molecules fitting together (like locks and keys), molecules being spliced together or cut in pieces, and channels in cell walls through which molecules are pumped. Now of course mechanism is itself a

metaphor—the cell is not a machine—but it is a very different metaphor from the ones I am considering. I don't want to talk about these little, local metaphors—locks and pumps—but about the much larger metaphors in terms of which the whole project of genetics has been presented—codes, blueprints, secrets of life, and so on.

These big metaphors matter not so much to the scientists on the coal face working on the mechanical details of the cells but rather to the people away from the frontiers of molecular biology who assimilate them. I mentioned the recent proposal that people conceived through sperm donation in the UK should in the future have a right to know the identity of their biological father. In defending this measure, scientists, experts of various kinds, and members of the public, speak of the importance to people of knowing their genetic origin. Since most of these people are not familiar with the technical details of the science—would probably not know a cytosolic channel partner from a translocation substrate—they are appealing to some image of the science which, I suggest, is very substantially formed by the metaphors promulgated by experts.

And something similar happens within science, now broadly conceived. One of the most effective disseminators of genetic metaphors has been Richard Dawkins[2], not a geneticist but an evolutionist. On the murkier fringes of evolutionary theory, new wave sociobiologists, or evolutionary psychologists as they are now known, often appeal to Dawkins's image of genetics to justify stories they tell about the evolution of the human mind. And these stories, unlike the work on the genomic coal face are often the subjects of international best sellers, presumably effecting images of the genetic among the general public. My claim is that these metaphors often misrepresent the science not usually, to repeat, to the scientists themselves, but rather to a range of scientific and non-scientific consumers of these metaphors.

Genetics and Genomics

As the director of a research centre with the word 'genomics' in the title, I am often made aware that many people do not know what this is. Most people have heard of the human genome project, but it is not my impression that a high proportion know what it was. And it

[2] See especially *The Selfish Gene* (Oxford: Oxford University Press, 1976).

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

is not trivial to say what the genome is. The first definition I found on Google defines the genome as the complete set of genes, which would cover about 3% of the genome on a typical estimate. Much better would be the complete set of chromosomes, which at least includes those parts that nobody thinks constitute genes. It is, however, more than a quibble to point out that a set is, on the face of it, an abstract object. It is more than a quibble because there is a real issue whether a genome should be considered as something abstract or something concrete. The common idea that the genome just is a sequence of base pairs (or even letters representing base pairs) suggests an abstract object—perhaps a canonical or standard genotype. But it is much better, I suggest, to think of the genome as something concrete—an object that occupies part of the nucleus in the centre of most organic cells. To mention just one reason for this preference, it seems quite likely that chromosomes themselves are not fully independent free-floating objects, but are structurally related to one another in functionally significant ways. Whether the genome is a distributed object or a single connected object is an empirical issue that will not be crucial for this paper; but I shall mention some reasons why it should be thought of as an object.

By contrast to this solidly material picture of the genome, genes turn out to be a much trickier matter. As already mentioned, only a small proportion of the genome is thought by anyone to be composed of genes. And whereas there are some potential pitfalls in attempts to define the genome, when we come to definitions of the gene we encounter fundamental disagreement. Anyone who doubts this should look at the Representing Genes project at the University of Pittsburgh[3], which has involved empirical investigation of the reaction of biologists' to various definitions of genes. The study confirmed the expectation that biologists with different interests tended to understand quite different things by this term. At any rate, for now I shall mean by 'gene' some—to be determined—principled subdivision of parts of the genome. Whether there are any principles adequate to motivate such a subdivision is a question this paper will address.

My own predominant opinion is that the concept of a gene is a misleading one. I do not think it misleads scientists who work on these things so much as the various other specialists and members of the general public who hear about them and derive more or less accurate pictures of the workings of our cells and our bodies. But these misunderstandings are deep and important. So even if it is

[3] See http://www.pitt.edu/~kstotz/genes/genes.html

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

unlikely that talk of genes will be abandoned altogether, we can hope that it may be treated with a healthy pinch of salt.

So what is a gene, or, if there are none, what would one be if there were any? The concept is often associated with the name of Mendel, though as is well known Mendel's work was discovered only posthumously in 1900, some years after his death, and the word 'gene' was not introduced until 1909 by Johannsen[4]. The crucial idea, associated rightly or wrongly with Mendel, was of discrete units of inheritance, discrete all or nothing causes of phenotypic traits. The familiar so-called Mendelian ratios between traits such as smooth or wrinkled and green or yellow seeds in peas were widely taken as evidence that these traits were caused by specific heritable factors. The hypothesis was developed that these factors came in pairs and that alternative variants (alleles) could make up these pairs. Particular variants could be preferentially expressed—or dominant—over others, so that just one copy of the dominant allele for green seeds would suffice to produce green seeds, whereas two yellow seed alleles would be required for yellow seeds. The additional assumption that these alleles were randomly assorted during sexual mating generated the simplest classical Mendelian ratios, those for two factors with complete dominance.

Essentially this picture underlay the classic studies on inheritance in fruit-flies (*Drosophila melanogaster*) undertaken by Thomas Hunt Morgan, his students Alfred Sturtevant and Herman Muller, and others. One of the most salient outcomes of this research was the observation of departures from Mendelian ratios that could be attributed to the association, or perhaps physical proximity, of genetic factors. This phenomenon, genetic linkage, was used by Sturtevant in 1911 to construct the first genetic linkage maps, proposing an ordering of the genes identified by a phenotypic difference[5]. Inevitably interest developed in the search for a physical realisation of the gene, and this early mapping work was important in developing support for the view that chromosomes, identifiable as fibrous structures in the cell, were the physical basis of genes. This theory was eventually accepted universally in the aftermath of the disclosure of the structure of DNA in 1953.

The crucial point in this story is the progression from purely theoretical entity—the hypothetical cause of inherited differences in

[4] Willhelm Johannsen, *Elemente der exakten Erblichkeitslehre*. (Jena: Gustav Fischer, 1909).

[5] A. H. Sturtevant, 'The Linear Arrangement of Six Sex-Linked Factors in *Drosophila*, as Shown by their Mode of Association', *Journal of Experimental Zoology* 1 (1913), 43–59.

phenotype—to a well-described structure, the DNA molecule. What I want to argue is that this transition actually involved the discovery that nothing fitted well with the concept underlying the original theoretical entity, 'gene'. Thus the famous results of Watson, Crick (and others) as well as being a natural culmination of the project of classical genetics, proved also to be the beginning of its end.

To avoid excessive circumlocution, I shall use the term 'gene' for the time being to mean any region of the genome containing coding sequence. Something like this is assumed when, for example, we are told how many genes there are in the human genome, though estimates vary from a now popular figure of about 30,000 to as many as 100,000. One of many difficulties is that genes are assumed to contain non-coding regions (introns), but clearly there must be a question whether a non-coding sequence is an intron or the gap between two genes. At any rate, it will become clear that this very rough definition fits poorly with many other assumptions that are made about the gene.

Another conception of the gene, not entirely a straw man, is as the material cause of a phenotypic feature. I suppose nobody quite believes that there are strings of bases that can be properly understood as the full and sufficient cause of a Roman nose or an artistic temperament. However there is considerable pressure towards beliefs of this sort apparent in, and perhaps in part stemming from, some of the most widely familiar metaphors in terms of which the genome is described. It is still common to hear the genome described, for instance, even by eminent experts, as a blueprint for the organism. And one point about blueprints is that there is a systematic mapping from parts or features of the blueprint to parts or features of the thing for which it is a blueprint. And talk of genes for this or that phenotypic trait might naturally be taken to give us the mapping from genome to phenotype. Perhaps not many people will defend the blueprint metaphor very far these days if pushed, however. A common retreat is to the metaphor of a recipe. Certainly this overcomes the immediate objection: one doesn't expect distinct parts of the cake that correspond to the flour or the sugar, for instance. But this metaphor is still quite inadequate. With due allowance for an element of assumed common knowledge, the recipe is a complete set of instruction for how to make the cake. The massive insufficiency of the genome, let alone merely the genome sequence, to determine the development of the phenotype points to deeper ways in which standard metaphors for describing the relation of the genome to the phenotype are inadequate. To explain

why this is so, it will be useful to return to my highly schematic and simplistic historical narrative.

Investigations following the discovery of the structure of DNA led eventually to the unravelling of the so-called genetic code, the mapping of triplets of the bases composing the DNA molecule on to specific amino acids, and hence the ability to correlate stretches of DNA with complex amino acid sequences, or proteins. An immediate worry—if hardly a surprising one—arising from the identification of the mode of action of DNA is that this action, the production of specific proteins within cells, is at a considerable causal distance from the phenotypic traits with which the story began. This causal distance immediately explains the classic philosophical objection to the attempt to identify classical Mendelian genes with parts of the chromosome: the relations between bits of chromosomes and Mendelian, or just phenotypic, traits are many/many.[6] Even the sketchiest conception of the processes connecting the production of a protein to the shape of a nose, let alone to, say, a sensitivity to violation of social contracts, makes it obvious that such a process will interact with many other proteins on the way. And it is at least likely that the ramifications of the production of a protein will be felt at many different points on the phenotype. These phenomena are referred to as pleiotropy—the multiple effects of a single gene—and polygeny—the multiple genes involved in producing a phenotypic effect.

The consequence of this on which I want to focus here is that polygeny and pleiotropy make phenotypic features generally poorly suited to distinguishing particular genes. An expression such as 'gene for measuring waist-to hip ratio', in the unlikely event that it has any referent at all, must trace back to many segments of the genome[7]. Conversely, many segments of genome will trace forward to a phenotypic feature of interest. What we so far lack, therefore, is a principle for identifying bits of the genome as individual genes. An obvious solution, and one that still appeals to many thinkers, is merely to move the effect much closer to the gene, and identify genes as the templates for particular proteins.[8] The problem with

[6] This point was clearly established thirty years ago by David Hull, *The Philosophy of Biological Science* (Englewood-Cliffs, NJ: Prentice-Hall, 1974).

[7] This improbable but widely cited candidate for an evolved product of genetic processes is due to Devendra Singh, 'Adaptive Significance of Waist-to-Hip Ratio and Female Physical Attractiveness', *Journal of Personality and Social Psychology* 69 (1993), 293–307.

[8] A sophisticated attempt of this sort is Kenneth Waters, 'Genes made molecular', *Philosophy of Science* 61 (1994), 163–85.

this, however, is that it has become increasingly clear that even the relationship between bits of DNA sequence and proteins is many/many. A typical stretch of DNA that appears to correspond more or less to what used to be thought of as a gene contains a series of coding sequences, called exons, separated by non-coding sequences, called introns. When the gene is transcribed into messenger RNA, this can be done in a variety of ways selecting from the available exons to produce various different RNAs. These RNA sequences in turn may subsequently be spliced on to other sequences, perhaps deriving from DNA from other parts of the genome. The RNA is then translated into polypeptide sequences which may themselves be spliced to further polypeptides after translation. The upshot of all this mess is then that the final protein product may contain sequences derived from DNA from diverse parts of the genome; and coding sequences of DNA may contribute to the production of a range of final protein products. So the genome cannot be classified into parts based on the proteins that it generates any more than it can in terms of phenotypic traits, and for just the same reason: in both cases the processes intervening are too complex and diverse.[9]

Both the many/many relations and the diversity of process can be elegantly illustrated by comparing the genetic basis of sensory mechanisms in the ear and in the nose. A gene that is involved in the production of the hairs in the cochlear cells in the ear produces several hundred distinct protein products that provide the variable-lengthed hairs sensitive to different sound frequencies.[10] By contrast, the many different cells in the nose sensitive to distinct molecules each employ a protein product from a distinct gene in tuning their sensitivity to a different molecule.[11] Interestingly, both these systems have undergone substantial recent evolution, suggesting that these are two different mechanisms for providing a certain kind of (relatively) rapid evolutionary response. I must confess,

[9] For details of these difficulties, see Thomas Fogle, 'The Dissolution of Protein Coding Genes in Molecular Biology', in *The Concept of the Gene in Development and Evolution*, Peter Beurton, Raphael Falk, and Hans-Jorg Rheinberger (eds.) (Cambridge: Cambridge University Press, 2000), 3–25 ; Lenny Moss *What Genes Can't Do* (Cambridge, Mass.: MIT Press, 2003).

[10] See Douglas L. Black, 'Splicing in the Inner Ear: a Familiar Tune, but what are the Instruments? *Neuron,* 20 (February, 1998), 165–8.

[11] See Shou Serizawa et al., 'Negative Feedback Regulation Ensures the One Receptor—One Olfactory Neuron Rule in Mouse', Science 302 (19 December, 2003), 2088–94.

however, that whereas there is no doubt that scientists describe these systems in these quite different ways, in the absence of greater clarity about the interpretation of the word 'gene' it is difficult to be sure how fundamental this difference is. What is clear, though, is that these findings illustrate the great diversity and complexity in the relations between genotype and phenotype.

A more familiar problem is that most of the genome is not even a candidate for analysis into genes in the way currently being considered, simply because it does not consist of sequences that provide information for protein production. As is now widely known, many active parts of the genome function by promoting or inhibiting the transcription of other sequences. Other parts are transcribed into RNA that has various cellular functions, but is not translated into polypeptides. And most of the genome appears to have functions of neither kind, and has been widely assumed to have no function at all. It has sometimes been referred to as 'junk DNA', and is thought to constitute the very large majority of most genomes, though in a moment I shall suggest that this junk maybe less junky than often supposed.

Before pursuing the question of genes any further, it will be better to return to the genome—the entire collection of nuclear DNA in an organism. I shall approach the genome this time through another metaphor, equally familiar in presentations of genetics, that of information.[12] The first thing to explain here is the sense in which this term is a metaphor. There is a technical sense of 'information' in which information is the inverse of uncertainty. A source is said to carry information about a target if knowledge of the source reduces uncertainty about the state of the target. In this sense, the genome is, undoubtedly rich in information. The outcome of human development is, in general outline, stunningly predictable. And if we scramble the genome a bit the outcome will very rapidly become less predictable. The only problem is that in this thin sense of information there is no special sense in which the genome is more a bearer of information than is any other essential developmental resource. As is often remarked, DNA can do nothing without a cell replete with the mass of chemical machinery need for its transcription and translation, and this machinery is as rich in information (in the present sense) as is the DNA.

[12] The critique here has been developed extensively in the context of developmental systems theory. See Susan Oyama, *The Ontogeny of Information* (New York: Cambridge University Press, 1985); Paul Griffiths and Russell Gray, 'Developmental Systems and Evolutionary Explanation', Journal of Philosophy 91 (1994), 277–304.

Of course this also implies that there is this thin sense in which bits of the genome carry information about features of the phenotype, and this will be appealed to when I discuss a very minimal sense in which the concept 'gene' still does useful work. A certain repetitive DNA sequence on human chromosome 4 carries the information that a person is almost certain to suffer the devastating neurological deterioration characteristic of Huntington's disease in middle age, for instance. Similarly, of course, the deposits of fatty material in a person's arteries carries information about the probability of sudden death. It is not obvious what is uniquely information-bearing about genes.

Part of the attraction of the informational view of the genome is that it appeals to a much richer idea of information and suggests a view of the sequence as something quasi-linguistic—linguistic not so much by analogy to the informal everyday languages of human chatter, but to the much more formal instances of the machine languages and programming languages of information technology. This parallel carries with it a certain temptation to abstraction. Though a sequence of machine code requires a machine to implement the programme it contains, it is often suggested that there are indefinitely many possible such machines. All the information is in the sequence and the programme it contains. When distinguished scientists display the rightly celebrated achievement of having determined, more or less, the sequence of nucleotides in the human genome; and when they go on to claim that this contains the blueprint for a human, or all the information necessary to build a human, we naturally think of such parallels from computer science. And naturally, too, we think of the rest of the cell as relatively undistinguished hardware implementing the programme in the genome.

One important move away from this picture is to recall the suggestion that the genome, rather than an abstract string of information, is a concrete material object occupying space.[13] An initial observation that might encourage this perspective is that there is about a 2m length of DNA in the human genome, whereas the cell is about 20 micrometres. This difference of about 8 orders of magnitude requires some fairly serious crunching up (or, technically, condensation) of the DNA. This is increasingly proving relevant to one of the most fundamental issues in molecular

[13] See, for instance, Timothy O'Brien et al., 'Genome Function and Nuclear Architecture: From Gene Expression to Nanoscience', Genome Research (2003), (http://www.genome.org/cgi/doi/10.1101/gr.946403) 1029–41.

biology, what determines which parts of the genome are being transcribed in any particular cell. The common computer analogy may make us think of this in terms of a sequential programme, one piece of transcription following another in a (somehow) predetermined sequence. But this, apart from bringing in an assumption of predetermination that sounds improbable in view of the general ability of organisms to respond to conditions of their environment, seems largely to beg the question of what determines changes in the sequence of activities. A sensible development of the computer metaphor will have changes determined by concentrations of products from previous stages, and this is no doubt an important part of the story. But it is also increasingly clear that part of the relevant mechanism is structural. As the chromosome condenses into more concentrated forms, the accessibility for transcription decreases, and particular parts of the structure become more or less available. More subtly, it is clear that chemicals, for example chemicals that degrade RNA or DNA, cannot be allowed to go just anywhere within the cell, or even the nucleus, and must be restricted to particular locations. Complex structures, membranes, barriers, and mechanisms are gradually being revealed within the cell. The cell is, in short, a highly structured space. It sometimes appears as if the cell is imagined as consisting of a nugget of information floating in a homogeneous chemical soup. In reality, this is as promising as throwing the components of a car into a vat of oil and expecting to drive the resultant mess down the motorway. A speculative thought about the structure of the genome that I won't explore here, is that it disposes of the notion that most of the DNA in most genomes is junk. Clearly large stretches of repetitive sequence, even if they have no coding function, will make a difference to the shape of the molecule and hence, very probably, have some effect on the functioning of the mechanism. Moreover stretches of junk will alter the distance between coding sequences and it is quite possible that this will also have functional consequences. My point here is not, of course, to advocate any such empirical thesis. It is rather to emphasise that the notion of DNA as junk is very much dependent on the picture of DNA as information bearer. From the perspective of DNA as part of a spatially integrated mechanism the metaphor of 'junk' actually makes little sense. And this is, apart from anything else, an excellent illustration of the sometimes unexpected implications of metaphors in this area.

Emphasis on the structural complexity of the genome motivates metaphorical appeals to mechanism rather than information

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

technology. And when one starts to think of the cell as a piece of exceedingly complex and intricate machinery, the question what is so special about the genome arises with new force. A mechanism depends on the interactions between parts, and however ingenious a particular component may seem to be, it is difficult to see how it can have any ontologically special status. It is tempting to argue that it is merely a historical accident, consequent on the fact that it was possible to learn something about genomes long before the techniques were developed for detailed investigation of cellular mechanisms, that leads us to attribute a special significance to the genome.

I do, as a matter of fact, think it is therapeutic to take very seriously the downgrading of the genome that has been proposed by some thinkers (including, on occasion, myself). The genome is part of the cellular mechanism and is entirely devoid of function apart from its meshing with the rest of the cellular machinery. In addition, contrary to what was for a long time known as the 'central dogma' of molecular biology, that information flowed only from DNA to RNA to proteins, it is becoming increasingly clear that the interaction between DNA and the rest of the cell is thoroughly interactive. A successor to the human genome project, the epigenome project, is devoted to the mapping of the sites at which the best understood of the mechanisms of action on the genome, methylation, can act.[14] It is, at any rate, no longer possible to think of DNA as an executive molecule, handing down instructions to its cytological minions. However, it is still worth exploring the intuition that sees something unique about the DNA molecule.

It is sometimes argued that there must be something like DNA for organisms to be possible at all (DNA as the conclusion of a transcendental argument). There are, in fact, two versions of such an argument, the phylogenetic and the ontogenetic. The supposedly necessary features in question are richness of information content, stability, and perhaps a capacity for self-replication. Nothing I have said critical of the unique informational status of DNA should be taken to deny the obvious fact that the DNA molecule is capable of storing enormous quantities of information—and here 'information' can be understood in the technical sense of reduction of uncertainty. Each of the three billion bases in the human genome has the potential to make a difference to some developmental process by virtue of its selection from the four possible bases, and hence each base at least in a coding part of the sequence potentially carries information about

[14] For details, see http://www.epigenome.org/index.php

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

the development of the organism. It is also indisputable that a great deal of information must be deployed in ontogeny. It may well be said that the central problem of biology is the reproduction of form—the ability of organisms to produce descendant organisms of the right kind, and one approach to that problem is to look for the information that guides ontogeny. On the other hand, and this is the point that drives much criticism of overly gene-centred views of development, in principle the necessary information can be dispersed across a variety of locations. In species deploying parental care, most notably our own, a good deal of development may be dependent on this. But more generally, the transmission of an entire cell—the minimum that is physically passed from parent to offspring—will involve the passing on of a great deal of structure and material in addition to the DNA sequence.

What about stability? It is often noted that DNA is a very stable molecule, especially for one of such high molecular weight, and certainly it is important that a good deal of developmental information is transmitted reliably. Stability of individual molecules may be important, but an even more interesting feature of DNA is its ability to replicate—to produce molecules that are accurate copies of itself. Richard Dawkins has notoriously elided these two properties, describing DNA molecules as achieving, through stability and copying fidelity, immortality. Without worrying too much about immortality, it seems to me implausible that self-replicating parts are a priori required for a replicating whole. One might perhaps imagine a lineage of pieces of paper, which achieve reproduction through photocopiers that copy instructions for the construction, use and maintenance of photocopiers, and distribute these instructions to symbiotic humans. Moreover, complex structures in the cell replicate in cell-division without apparently being composed of self-replicating parts. Nonetheless, there is a very strong intuition that such a molecule is an extremely good idea, an intuition worth some further exploration.

An analogy that seems to me potentially useful here is one that subtly but importantly modifies the standard information metaphor, and that is the idea of data storage. Storage is immediately suggested as a function by the reference to stability: when one stores things one generally hopes that they can be retrieved in a condition very similar to that in which they were put away. (An interesting, but not presently relevant, exception is provided by such things as wine or cheese, that may even come out better than when they went in.) One also stores things while one is not using them. Since most coding sequences are not being actively

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

transcribed at any time, they may reasonably be seen as in storage, and as just noted, it is to be hoped that they will exhibit a high degree of stability in this condition.

The conception of the genome as the cell's main device for data storage is rather different from most of the contemporary metaphors used for thinking of the function of the genome. First, it should be stressed that what is in question is long-term storage. If we spell out the inevitable parallel with computer technology, we are thinking of the genome as something like a hard disk. The parallel with active memory in a computer (RAM) is rather with RNA or protein molecules. But it will also occur immediately to the listener that programmes as well as data reside on the hard disk. So does the genome contain the developmental programme after all?

Of course the genome doesn't run any such programme, but in this respect there is no disanalogy with the computer. The hard disk stores the programme but doesn't run it. And this points to the important, if trivial, general observation about any programme, that it is always required in addition to a programme that there be a system capable of implementing it. A much more significant disanalogy stems from the phenomena of epigenesis. As noted above, the central dogma, of a one way flow of information from DNA to RNA to protein is increasingly untenable. It is now well-established that other elements in the cell can make permanent or transitory changes to the DNA that affect the likelihood of transcription. Of course it could be that these are instructions originating in programme parts of the DNA and implemented on data storage. But there are more basic reasons for thinking the programme metaphor, if not entirely misguided, is liable to mislead.

For one thing, the metaphor is deeply deterministic. There is little reason to think that development is a deterministic process. Note that this is not to say that development is an unreliable process. Quite the contrary: the process is too reliable in outcome to be plausibly modelled as a predetermined series of steps. This is perfectly familiar and unproblematic in our experience of human action. If I ask someone to go to the shop and buy me a loaf of bread, and they agree, I am fairly confident that the outcome will be as I intend. If I provide a deterministic programme—take 12 paces north-west, raise hand, turn knob, push, etc., there are too many unanticipated interventions that can derail the process for me to have much hope of success. Teleology is much better than deterministic causation at getting things done, and development is much too reliable to be seen as anything but teleological. I don't mean to monger mysteries here, or deny that the 'teleological'

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

process is analysable in terms of a sequence of efficient causes. I am only claiming that the kind of algorithmic sequence of deterministic causes employed by (most) computers is not a plausible model. (No doubt the same can be said for similar models of the brain.)

This point is perhaps more obvious when we reflect that development is a very long process. Indeed human development has no obvious terminus short of death. While it is important that a normal human life-course takes place against a backdrop of species-typical developmental stages—infancy, childhood, puberty, adulthood, middle age, senescence, etc.—the diversity of the lives that take place within this broad framework is such as to make the notion of a programme strikingly inappropriate. It will no doubt be suggested that if one abstracts the biological dimension of the life course one will find a programme at the common ontogenetic core. Though there is surely something in this, I still consider it more misleading than helpful. First, the length of time over which this programme is alleged to run makes the point that it must be thought of teleologically that much more compelling. But second, I doubt whether it makes sense, certainly for a human life, to abstract the biological from the social, cultural, and merely environmental. (And many non-humans have a social dimension to their lives, all have an environmental dimension.)

Let me summarise what I have suggested about the genome. It is a concrete structure rather than an abstract pattern, and as such it may be seen as part of the mechanism that constitutes the cell. No familiar mechanical metaphor very well characterises the genome, though perhaps that of a data storage device is the most useful. (There is also something important to be said along these lines about the relation between the genome and evolution.) The most accurate general description is perhaps a semi-technical one from theoretical biology, a developmental resource, one of a number of such that is required to solve the problem of organic reproduction. Let me turn, in conclusion, to the question whether this leaves us anything useful to say about genes?

Back to genes

Suppose we think of the genome as a material part of the cell, and as something like a data storage device. Might its parts be genes? The analogy suggests not. The parts of a hard disk are not chunks of data but mechanical components. Similarly the genome

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

certainly has structural parts—chromosomes, and within those, such things as centromeres and telomeres. Anyone tempted to believe that, say, my latest Kylie Minogue album was part of my hard disk need only be reminded that the relevant electronic data are likely to be scattered around different parts of the disk, an observation interestingly parallel with the distribution of the code required to make proteins around the genome. If someone claimed that they might just be spatially distributed parts, I would then point out that the same chunks of data on the same bit of disk might be parts of many different files, here reminding us both of the IT equivalent of pleiotropy and the lamentable ease of contemporary plagiarism.

Might the gene concept nevertheless prove to be useful in characterising information, in the sense simply of significant sequence, in the genome? The first thing to say is that this is in the end a question for genomic scientists rather than philosophers. Having said that, my impression is that the useful concepts for this task are likely to be at a lower level and more specific. The project of counting genes seems doomed to incoherence and this speaks unpromisingly for the future of the concept. The second thing to say, or to repeat, is that if such a concept does prove useful, it will prove incommensurable with the entire tradition of thinking of genes in relation to their phenotypic consequences. This does, at the least, present severe dangers of misunderstanding.

What residuum is there of the traditional concept? The first point here is that, in keeping with its Mendelian roots, it is a concept that applies only to the explanation of difference. It was a natural hope, perhaps, that the explanation of differences would have provided a path towards the general explanation of developmental outcomes, but this has not been the case. In the sense in which there is a gene for brown eyes there are no genes for eyes, tout court, because everyone has them. Those parts of the genomes in which variation is not permitted, that is, those in which it causes inviability, contain no genes. Clearly this is something that can be made no sense of within anything remotely related to the blueprint conception of the genome, and again there are obvious dangers of misunderstanding as this usage of the word 'gene' disseminates to the general public.

The irreconcilable tension between the molecular and the Mendelian rumps of the gene concept suggest that we would do well to canvass for the abolition of the word altogether. However there are contexts in which the Mendelian concept still lives a more active life, which we may have to respect. The first of these is in evolutionary theory. Since most evolutionary theory is focused on

natural selection, and selection, by definition, is concerned only with differences, it is not surprising that evolutionists have generally found the Mendelian concept admirably suited to their needs. I think, as a matter of fact, that this has led to very serious problems as evolutionists have at the same time failed to take adequate account of the limitations of developmental interpretations of genetics. But this is something I have discussed at length in other places and will not go into here.

More relevant to present concerns is the case of medical genetics. Without wishing to speculate on the likely future achievements of this growing branch of medicine, it would be careless, at least, to legislate it out of a subject matter. And indeed it surely has a perfectly good subject matter: familiar and often devastating diseases such as cystic fibrosis and Huntington's are maladies directly caused by genes, and genes about which we often have detailed knowledge. We should at least be able to make sense of this kind of knowledge.

This is not, in fact, particularly difficult. Classic monogenetic diseases are diseases caused by errors in the genome. Typically this is the failure to make a functional protein due to an error in the DNA sequence. Unlike text, small errors can be fatal in DNA transcription and translation—one deleted base, for instance, will shift the whole reading frame to nonsense. 'Genes' for such diseases are, therefore, a set of possible errors in a particular area of the genome that produces a particular developmental or metabolic failure and a characteristic syndrome of symptoms. This is all quite unmysterious if tragic. Just two simple points should be stressed. First there is no specific physical structure that is, say, the gene for cystic fibrosis. Almost any disease gene will be a set of possible errors rather than a particular sequence. It is, perhaps, best seen as an abstract object. Second, a genetic disease implies no phenotypic characterisation of the normal, healthy state of that piece of the genome. It is part of the genome that contains information needed for building a protein, and the lack of the protein causes disease. One could, technically, I suppose, say that this was the gene for not having cystic fibrosis, but to do so would be, to say the least, misleading. If the subject matter of medical genetics turns out to be a set of abstract characterisations of some sites in the genome, it hardly looks likely to legitimate a general reinstatement of the phenotypic characterisation of genes. The dangers here, incidentally, are well illustrated by the widespread tendency to use genetic diseases such as phenylketonuria, which have serious effects on mental development, as evidence for the importance of

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

behavioural genetics. I hope it is clear that this provides no such evidence.

Conclusion

The history of the term 'gene' has seen the increasing erosion of the assumptions with which its use has often been associated. This has led to a set of rather poorly defined and perhaps unnecessary uses in molecular genetics that reflect the contemporary understanding of the phenomena out of which the tradition of genetics arose, and a fringe of uses in such areas as evolutionary theory and medical genetics, essentially similar to traditional meanings but liable to carry strongly misleading implications to the unwary.

The concept is a wonderfully rich source of potential insight into the historical development of scientific concepts, and into the processes by which such concepts travel from one technical context to another and disseminate into public discourse. The present paper is intended to provide an introductory survey of some of the issues that arise within these projects.

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press

https://doi.org/10.1017/S1358246105056092 Published online by Cambridge University Press