CONTENTS

Is the genome the "book of life"?

Denis Noble
Secretary-General IUPS

Introduction

The human genome is a vast database of information containing of the order of 30,000 genes. Each of these is used to encode the amino acid sequence of a particular protein, or set of proteins. The complete sequence and structure of the proteins is sometimes referred to as the proteome. Understanding how the information in the genome is used to create the proteome is a major challenge.

First, because we need to identify all the genes, which we are getting closer to doing, at least in the sense of identifying which parts of the DNA code correspond to genes. It is important though to note that this is still a long way from knowing what each gene does, what is its functionality.

Second, because we need to understand the control processes that determine which proteins are expressed and to what extent. We now know that there are many more proteins than genes, so something like a feedback control has to determine which protein is made when. This is beginning to look like a complex interaction between genes and organisms and

their environment, rather than a simplistic view of genes as 'dictating' the organism and its functions.

And *third*, because predicting three-dimensional structure and chemical function from the amino acid sequences of the corresponding proteins is very difficult.

Genome to proteome to physiome

The recent discovery that there are many fewer genes in the human genome than we thought (original estimates were of the order of 150,000) adds to the difficulty since it means that multiple functionality (i.e. each gene being involved in many different biological functions) must be very common indeed, perhaps even universal.

But even these formidable challenges pale in significance when we consider the complexity of the next stage: understanding the interactions of tens of thousands of different proteins as they generate functionality at all levels through cells to organs and systems. This is the task of quantitative analysis of physiological function, which in its entirety is sometimes now called the physiome. Computational modelling and bioinformatics will play an increasingly important role in all these stages of unravelling the way in which the information contained in the genome is 'read' to create living systems. We will be able to say that we have really read "The Book of Life" when we have succeeded in going all the way from the genome, through the proteome to the physiome. Achieving this requires the marriage of mathematics with medical science.

On this view, the genome is not understandable as "the book of life" until it is "read" through its "translation" into physiological function. My contention is that this functionality does not reside at the level of genes. It can't because, strictly speaking, the genes are "blind" to what they do. Moreover, a lot of what their products, the proteins, do is not dependent on instructions from the genes. It is dependent on the poorly understood chemistry of self-assembling complex systems. It is as though the genes specify the components of a computer, but not how they should be put together. They just do this by doing what is chemically natural to them. Some people are predicting that this is also the way in which we will build computers in the future, particularly what are called molecular computers.

So, if the genome is "the book of life" it is a book with enormous gaps, which nature takes for granted since it never had to work out how to code for such natural phenomena. This "missing information" is implicit in the properties of the environment in which genes operate. It is not explicitly coded for. Moreover, this environment crucially determines which genes are expressed and to what degree. The passage of information is not simply one-way, from genes to function.

The "French Omelette"

I like the analogy with the story of the French omelette. The story is that a little family bistro outside Paris acquired a reputation for the great

lightness, flavour and delicacy of its omelettes. So, the connoisseurs decided that, in writing up a compendium of French cuisine it was essential to include the recipe of these famous omelettes. The mother of the house kindly obliged with a detailed recipe of ingredients and the order in which they were incorporated as the omelette was prepared. There was only one snag. When the Parisian chefs tried it, they got a different result. Delicious flavours but totally lacking the light quality of the family's achievement. Frustrated, they experimented with various interpretations of the recipe and forms of omelette pan until they eventually decided that there must be a trick. The mother had surely not revealed all her secrets! So, off they go to the Bistro to find that the mother has died. Her daughter now cooks the omelettes. After finding that her omelettes were just as good as her mother's, they show her the mother's recipe and ask whether it is correct. She reads it carefully and says, yes, it is a marvel of accuracy, down to the last milligram of ingredient. This is exactly what she follows. "There's nothing missing?" they ask. "Of course not", she replies, "mother has written it all down". She knows because this is also what mother wrote down for her before she died. She is following exactly the same recipe as they were trying to follow. So, they naturally ask her whether they can watch as the omelette is prepared. Of course, she says, there is nothing to hide. So they watch carefully to try to detect the slightest difference between the written recipe and what she does. And they are amazed at what they see at the very beginning of her preparation for, as she breaks the eggs, she separates the whites and the yolks, only folding the beaten whites in at the end before cooking. They upbraid her for her mother's inaccuracy, for she never reveals this crucial fact in her written recipe. The daughter takes offence at their 'stupidity' and 'arrogance' as she looks at the assembled dignitaries and asks them, in all innocence, "how else do you think anyone prepares an omelette!?"

What does the genome lack?

My contention is that nature has not coded for what is chemically natural to what the proteins do. It does not need to. I also contend that this information has as much of a claim to be called the "book of life" as does the genome. Finally, I claim that it is going to be much more difficult to work out this side of the story than to sequence the genome.

So, in brief, my view is that the genome:

1. Is most like tedious machine code for the construction of the key players in the game of life: the proteins.
2. Is incomplete in a major respect: how these proteins behave chemically in the cells of the body, how they fold, combine and interact.
3. Is completely lacking in functionality. It does not even tell us whether a particular gene plays a role in 1, 2, 3, a dozen, or a hundred functions.
4. Relies time and again on Mother Nature's ability to know "how to make omelettes."

Analogy with languages

One defence of the "Book of Life" view against my attack is to say that all books are like this to some degree. All languages function in a context of implicit knowledge that the language itself does not need to spell out.

Of course this is true. One way to see this is to note that languages differ in what they take for granted. Some languages do not use plurals, for example (though equally obviously the users of these languages – which include Chinese, Japanese, Korean, Polynesian languages, Maori – have the concept of plural and know when there is more than one of the thing being referred to). And there are innumerable examples of words, even when very similar in different languages, that function differently simply because of the cultural context in which they are used. To call a French woman, in French, "séduisante" is to compliment her highly. To call an English woman "seductive" could be highly dangerous! And to avoid the cultural difference by translating "séduisante" as "beautiful", "appealing" etc is to remove precisely the sexual frisson that the French language intends. And it begs the question of what the cultural context of the word "sex" is in the two languages.

So, I accept the thrust of this argument. This kind of "implicit" knowledge is an irreducible feature of all languages (none are, or can be, neutral with respect to culture since they are themselves the creatures of that culture). But the relationship between the language and its culture is not like that of the relationship between the genome and nature. One way of seeing the difference is to ask what the language is *trying* to do. Human languages aim to describe the world as it is (or rather, as it is seen by the speakers). It is an aim of each language to try to avoid ambiguity. There is even a discipline, called philosophy, that – working as it must within the constraints of a particular language – tries to step back, as it were, and see beyond the limits of the culture, to question those limits, the bounds of sense (to echo Kant) and of meaning.

By contrast, the genomic language does not have anything like this as an aim. Of course, strictly speaking, it must have nothing as an aim – evolution is blind – but if we take this line then, *a fortiori*, the genome is not a book. In order to pursue the book analogy we have at least to ask what it could be said to be a book *about*. So is it about "life"? Well, I would say no, not really. Crucially, it does not describe functionality. The code for gene XYZ does not spell out that XYZ makes a protein that enables synapses in brains to function, testicles to produce sperm, pancreatic cells to secrete insulin etc etc. This is a bit like imagining a book that does not spell out that X is the King, Y the archbishop, Z the villain, etc. But, worse than that, from reading the "book" we don't even know what the relationship between a King and an Archbishop might be. These crucial interactions are outside the scope of what the genes specify.

Thought Experiment

So let's do a thought experiment.

Silmans are an intelligent species that evolved in a world in which silicon functions instead of carbon. They

advance to the point of rapid space travel and find the planet earth. But, they have a serious problem. They can't live on earth. The earth's environment is terribly hostile to silmans as a form of life. Moreoever, they can't bring humans, or any other earth creatures into their spaceships, for what they require to live is equally hostile to earth forms of life. But they know from their own evolution that there must be something equivalent to silman code, there must be earth-life code. They have also worked out that, in their case, this is imprinted in chemical sequences that are inert. Their code molecules do not breathe, or need siligen (the equivalent for them to oxygen) etc etc. They therefore reason that they might be able to send robots down to the planet's surface that will extract earth-code. OK, let's call it DNA. To their joy, they find that this DNA is also chemically inert. It can be taken into their spaceships and it can be subjected to analysis. So, they start reading the complete DNA of a human.

Let's give them too the intuition that this 'inert' molecular sequence codes for another sequence, that of the proteins, and that these are highly reactive. So, having worked out what the code means (which DNA sequence corresponds to which amino acid) they set about determining all the proteins – all 100,000 or more of them – that go to make a human.

But, then they are stuck. Because theirs is a silicon world, they do not have water, they do not have lipids. But, they guess from analogy with their own world that there must be such things. They try to see in the DNA code any clues to what these substances might be. To their intense frustration they find there is nothing. All the DNA code does is to specify one type of molecule, proteins. No other information is there. What a bore!

So, they start to conclude that earth-life must be a very strange thing. Just tens of thousands of proteins thrown together. Perhaps humans are a kind of soup! Perhaps earth-life is extremely primitive. So, they prepare to time-travel on to the next inhabited planet.

But then one of them says, "Hey, wait a minute. You could say that we are "just" a bunch of silicon sequences. But we know we are not: that we think, reproduce etc etc. There is more to life than molecular sequences. What we should do is to send down another robot with a carefully isolated capsule into which it will put all this strange stuff called water, air, lipids etc that we know are down there. Let's bring it all up in the capsule and see what happens when we let the DNA do its thing in this environment."

So, they then watch with incredulity through the capsule as cells are formed, then divide into the early embryo, that then goes through all its wonderful transformations until 9 months later a human emerges (OK – I am simplifying – they are just cloning – there's no mother, no womb etc, and I have to assume that they picked up at least one cell to provide the environment in which the nuclear DNA can work – but putting all these into the picture will only reinforce the point I am making). Just like the Parisian chefs, they feel cheated. The "recipe" doesn't specify all this. It just happens!

Some omelette!