European Review, Vol. 9, No. 3, 241–243 (2001) © Academia Europaea, Printed in the United Kingdom

Editorial: what about the human genome?

ARNOLD BURGEN

Downing College, Cambridge CB2 1DQ, UK. E-mail: asvb@cam.ac.uk

There are few weeks in which the news media do not announce some great breakthrough in medical or biological science, usually inferring some immediate benefit to mankind. Alas, the promise of these benefits is nearly always premature and leads to disappointment. However, the frequency of the reports does indeed reflect an astonishing rate of new discovery in a field of great complexity.

Two of the leading general science journals, *Nature* and *Science*, have devoted the main part of their issues, in February this year, to the announcement of the 'complete' description of the human genome. This represented a huge scientific achievement since the genome is a linear sequence of about 3 billion letters and the result (which is almost complete) has been achieved in a surprisingly short time by an international effort. It has required the development of industrial, automated sequencing methods and complex computation. However, we have to be a little careful about what is meant by the genome. It refers primarily to those genes that are transcribed through RNA into proteins. There is agreement that these number only about 30,000 and constitute only a few percent of the total length of the DNA in the cell nucleus. It was previously believed that the number of genes was considerably larger, but this has now decreased to what now seems a very modest number, considering the huge task of specifying and creating a complex organism. In the past few years, the genomes of an increasing variety of organisms have been sequenced, ranging from viruses with only a handful of genes, through to the fruit fly, *drosophila*, with about one-third of the number in humans, to plants such as maize whose genome is nearly twice the size of ours. The description of the complete genome of the mouse, which was also announced in February, is very similar to that of humans, differing by only about 2%, mainly in the genes concerned with the development of the nervous system. It is indeed remarkable that many genes, for example those for many metabolic enzymes and structural proteins are found to be the same or very closely related throughout living organisms, indeed even between bacteria, plants and animals. There are also clear signs that some 'old' genes have been modified over time to provide new functions. The existence of such a family of genes crossing species and time provides stunning support for Darwin's great insight of evolution by natural selection.

This description of the genome has been described as the 'Book of Life', suggesting that it contains a kind of encyclopaedia and that if one knows how to look it up, the whole of life would be revealed. However, not surprisingly, there are still lots of complications to be

https://doi.org/10.1017/S1062798701000229 Published online by Cambridge University Press

overcome before that can be the case. For instance, even the prescription that one gene equals one protein is not adequate. It is now known that a single gene can lead to the formation of several proteins and that the total number of distinct proteins is likely to be at least ten times as large as the number of genes. This multiplicity is due partly to the constitution of the genes in higher organisms in which the gene is broken up into pieces separated by stretches of DNA that are not transcribed into proteins. These non-translated parts are called *introns* as opposed to the translated parts, which are called *exons*. The presence of the introns allows exons to be spliced together in different ways, giving a variety of proteins, in different cells and stages of development. A great deal of further effort will be needed to identify the proteins that the genes define and to discover their functions (this search has been given the name *proteomics*).

You will certainly be asking at this point 'whose genes are they talking about?' since we are all aware of individuality, which is certainly a matter of genes. The answer is that the sequence of the genome reported was derived from a rather large group of people whose DNA was lumped together so as to smooth out the individual differences, so it is a kind of 'common' genome. We are all aware that the police and the law courts now commonly use DNA testing to identify individuals in rape and other criminal acts from trace materials, semen or blood left at the scene of a crime. Interestingly, this is achieved by examining part of the non-genomic DNA known as 'minisatellites', which are highly repetitive sequences, but nevertheless very individual. Because of their repetitive constitution they are easy to characterize. However, there are also frequent variations in the sequence of nucleotides between individuals that occur, on average, one per 1–2000 bases. So far, over 2 million of these *single nucleotide polymorphisms* (SNPs) have been discovered. A small proportion of these, perhaps only 60 000, are within the genome, the rest in non-coding areas. However, this does mean that there are two or more variations in each gene. The SNPs provide a possible basis for our individuality and family resemblances. They will be valuable in studies of genetic disease and in ethnic origin, and there is some hope that they may help in sorting out susceptibility to disease and the side effects of drugs.

An interesting scientific and philosophical problem is how this linear one-dimensional sequence of letters in the genes can lead to the three-dimensional complexity that one finds in multicellular organisms. A first answer to this can be seen in the proteins. Many of these are simple linear chains of amino acids, but the interactions between the side chains of the amino acids leads the protein to assume a stable, complex, unique three-dimensional shape that is characteristic for each protein: thus, we have here a one-dimensional sequence generating three-dimensional information. At another level, there is a further interesting question. The mammalian egg is a regular sphere, yet it will develop into a three-dimensional animal that has a differentiated front, back, etc. The key orientation is provided by the site of entry of the sperm into the egg, which defines the posterior pole (i.e. it is not determined by the DNA sequence) – but then back and front and left and right have also to be defined! Some of the arrangements of the body and organs are provided by genes that have specific effects in development, called *homeotic* genes. These have been extensively studied, particularly in *drosophila*. However, it is very likely that a large part of developmental information is passed locally between individual neighbouring cells. It is one of the very interesting features of animals that each cell in the body contains the complete set of genes, and it seems that this allows the instructions for development to be provided locally. A fine

https://doi.org/10.1017/S1062798701000229 Published online by Cambridge University Press

example of decentralization of control!What about the large majority of the DNA that does not provide ‘genes’? It is hardly believable that the efficient processes of evolution have maintained all this material if it is useless, and we have seen that part of it is certainly genetically conserved in the minisatellites. There is much to be learned about this non-genomic DNA which also contains ‘historical’ sequences related to evolution and to virus infection.

This editorial can provide only a very sketchy account of the state of our knowledge about how genetic information gives rise to our very complex bodies. Although the description of the human genome is indeed a landmark, it is obvious that it cannot be read like an engineer’s blueprint. It is rather like having an alphabet and just beginning to find how it makes words, as yet having little idea of how to string the words into sentences, let alone to produce the *story of life*.

The prospect of the uses that will be made of our rapidly increasing knowledge of human genetic and developmental processes is indeed daunting. Science is said to be neutral, after all it is just about our understanding of nature, but the uses to which that knowledge can be put are far from neutral. Alas, the history of mankind does not give any assurance that this new knowledge will invariably be used wisely or rationally.

Arnold Burgen is Editor-in-Chief of the *European Review*.

https://doi.org/10.1017/S1062798701000229 Published online by Cambridge University Press