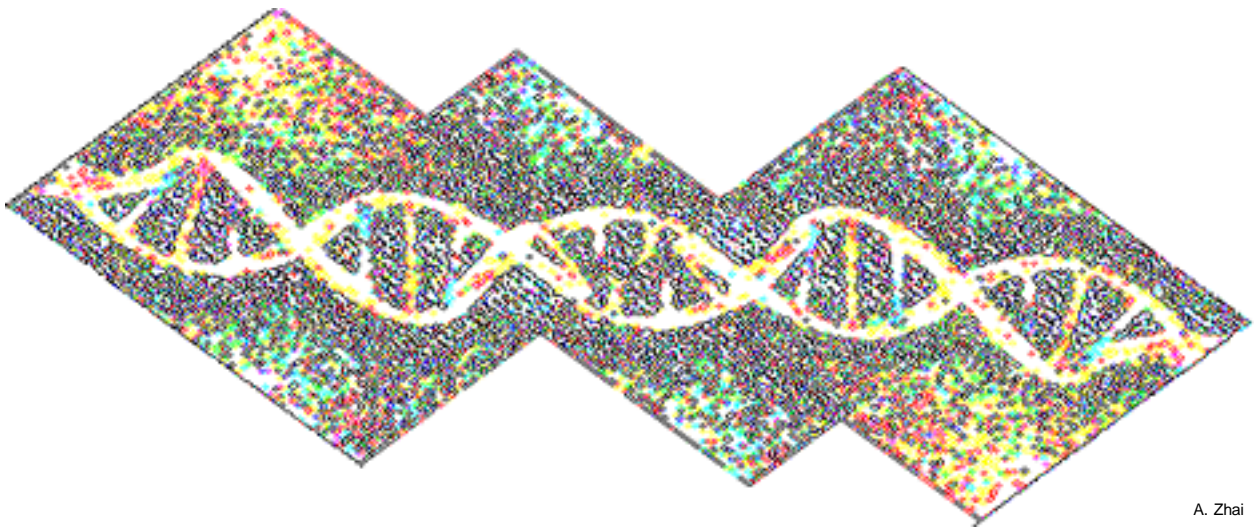


COMMENTARIES

A. Zhai

**THE HUMAN GENOME SEQUENCE:
WHAT CAN WE GET FROM IT?**

After spending millions of dollars on the Human Genome Project, the long awaited sequence of the human genome has finally been retrieved and assembled (1,2). Although some scientists may be skeptical as to the significance and usefulness of the sequence, most others have enthusiastically welcomed the information. Perhaps the response depends on how one views the current state of knowledge: Is this the end of the project or is it merely a beginning?

From the complete sequence, we know the number of genes present in the human genome and we can now acquire a detailed description of our genes and the biological information stored within them. Using a brute force approach, we can analyze the transcriptional regulating elements for all genes and predict the sequence and structure of all gene products using some sophisticated computer programs and research tools. Hence, all proteins encoded directly from our genomic sequence can be determined and compared with those of other species; their similarities and differences will allow us to explore the evolutionary relationship between humans and other species. The diversity in protein structure and function that has arisen through evolution may be a result of new combinations of existing domains as well as the generation of entirely new domains or proteins. Although these analyses can be easily done with the help of bioinformatics, how far can these data go in telling us the true function of a given protein and what results if its level is deregulated or its structure modified?

Not to be overly skeptical or pessimistic, but there are always some things we cannot predict. For instance, before the discovery of DNA recombinase and RNA splicing, one may well have imagined that the enormous diversity of proteins was necessarily explained by a huge number of genes. Without DNA recombinase for the recombination of immunoglobulin genes, our immune system would not be able to protect us from the existing and evolving pathogens in our dynamic environment (3). Without RNA splicing, most proteins could not be produced in a matured form because of the existence of introns (4). The discovery of these two kinds of proteins not only provides the explanation for the diversity of proteins which may vastly outnumber the genes that encode them, it also provides evidence demonstrating the complexity of the system, which is not always easily predicted. Hence, can we validate the predicted results from our genome sequence? The answer may not be so simple. For instance, can we rule out the existence of protein recombinase in our system? Or can we ignore the intronic sequences as if they serve absolutely no biological functions? Furthermore, most proteins undergo post-translational modifications (*e.g.* phosphorylation, methylation, acetylation, cleavage) and their final biological function depends both on their level of expression and their state of modification. Moreover, depending on the functional state of the organism or cell, and the interactions between and balances among multiple proteins, a given molecule may behave very differently; in other words, even after taking into account mechanisms like post-translational modifications, the function of a molecule may not be

completely intrinsic to the molecule's structure but may depend, in part, on its environment. Given these uncertainties on the functionality of unknown molecules within our system, can we know everything about ourselves from our genomic sequence? The answer is probably not what we want to admit.

Many more years of research are required to even start answering the questions alluded to above. What we can assume is that knowing the sequence of the human genome will greatly facilitate this research. Proteomics represents the next generation of multi-million dollar projects that will attempt to identify and characterize all the proteins encoded by our 31,000 or so genes (5,6). As with genomics, the insight eventually gleaned from proteomics will represent another major step towards our understanding of ourselves. But perhaps the major question is still unasked – how many such steps will it take before we

can consider our understanding complete and our interpretations definitive?

Carlos Chan, B.Sc.

Faculty of Medicine, McGill University

REFERENCES

1. Baltimore D. Our genome unveiled. *Nature* 2001; 409: 814-816.
2. Pennisi E. The human genome. *Science* 2001; 291: 1177-1180.
3. Merrell DS, Camilli A. Detection and analysis of gene expression during infection by in vivo expression technology. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 2000; 355: 587-599.
4. Newman A. RNA splicing. *Current Biology* 1998; 8: R903-R905.
5. Chambers G, Lawrie L, Cash P, Murray GI. Proteomics: a new approach to the study of disease. *Journal of Pathology* 2000; 192: 280-288.
6. Banks RE, Dunn MJ, Hochstrasser DF, et al. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 2000; 356: 1749-1756.

THE 'BOOK OF LIFE': A GOLD MINE?

Together, the special issues of *Nature* (1) and *Science* (2) dedicated to the sequencing of the human genome amount to a stack of pages about 2.5 cm thick and well over 1.2 kilograms in weight. Complete with free CD-ROMs, gigantic posters, and a myriad of colourful advertisement are the reports of the (almost completed) human genome sequence. Yet these formidable volumes are merely the introductions to what is now being widely hailed as the Book of Life; actually, make that the 'Books' of Life. There are in fact two versions of this hallowed script, independently transcribed by the publicly funded International Human Genome Sequencing Consortium (published by *Nature*) and by the private venture Celera Genomics (published by *Science*).

The sequencing of the human genome is, even from the most mundane perspectives, an awesome feat. This powerful tool has already transformed the biomedical sciences so much, that some say with this information, we have now entered the 'post-genomic' age. But beyond its purely scientific value, the map of our own genome carries with it great symbolic significance. After all, it is the blueprint of ourselves that is now stored in databases and is accessible with but an internet connection and a few clicks of the mouse. The fifty years between the elucidation of the structure of DNA (3) and the completion of the Human Genome project will prove to be an enduring chapter in our understanding of ourselves.

Yet why would we need two versions of the same thing? Well, they are not exactly the same. For one thing, whereas the publicly funded version of the human genome is freely available in the public database known as GenBank, Celera's version can be accessed too, but with a number of strings attached. How this came to be is a story of meeting the bottom line – companies wanting to cash in on 'genomania'. Those who wish to download DNA sequences from Celera's database for free must be the purely academic ivory tower types who limit themselves to basic (read 'non-profitable') research. On the other hand, if researchers wish to go beyond the advancement of knowledge and apply their findings to develop clinical applications such as anti-cancer therapies, they will have to pay. In short, we didn't need to have two copies of the 'Books of Life'; the extra copy, which Celera claims to be more complete, is simply around because it can make money.

I am not trying to trivialize the achievement of Celera. Completing the massive project of sequencing the entire human genome ahead of schedule was by no means a walk in the park; quite the contrary, what Craig Venter and Celera Genomics did was very remarkable. Despite early criticism that his 'whole genome shotgun' approach to sequencing was impossible and unrealistic, it is now widely accepted as a standard method for future genomic projects. Furthermore, Celera is a company that has pumped millions of dollars into sequencing the human genome, so recouping this money (plus a little in profits) is not unreasonable. However, from an entirely different perspective, it does