A remarkable achievement for mankind

Scientists release a map of the human genome

Frank Gaglioti
28 February 2001

On February 12, scientists from the publicly-funded Human Genome Project (HGP) and Celera Genomics, a privately-funded biotechnology company, released what they termed “an initial working draft sequence” of the human genome. It is the first detailed map of the most significant human genetic structures, covering 90 percent of the gene-rich sections of human DNA. The HGP published its results in the Internet edition of *Nature* and Celera on the *Science* web page. A print edition of the journals was released on February 15.

An initial analysis by scientists on the Internet provides a stunning verification of evolution and overturns a number of long-held theories. In one sense, the completion of the mapping of the human genome is just the beginning, as it opens up many new avenues with the potential to revolutionise our understanding of how humans evolved, as well as fruitful new lines of medical research.

Dr. Francis Collins, head of the Human Genome Project, commented: “This is the first time for having in front of us the human book of life and realising that it's actually three books. It's a history book that tells us about where we came from, looking back hundreds of millions of years. It's a shop manual that contains within it a list of the parts that we're made of. And it's a textbook of medicine that contains within it clues to the causes of diabetes, heart disease, cancer and a variety of other disorders that we only at the moment barely glimpse.”

James Watson and Francis Crick in 1953 first unravelled the makeup of DNA or deoxyribonucleic acid, the chemical in the nucleus of each cell which contains the genetic information for the maintenance and reproduction of any organism. They found that DNA consists of two connected twisted strands in the shape of a helix. The DNA strands are made up of different combinations and permutations of just four different chemical building blocks known as “bases”. The base sequence provides the genetic code for each living thing.

It is a testimony to human ingenuity that less than 50 years after that original ground-breaking discovery, scientists have been able to devise techniques to laboriously “read” the code and sequence most of the 3 billion bases that comprise the human genome. The HGP began the undertaking in October 1990 and was joined by Celera in 1998 in what has been one of the biggest scientific projects in history, involving the efforts of hundreds of scientists from the US, Japan, France, United Kingdom, Germany and Canada.

The technique adopted by HGP involved cutting the genome into small segments, then inserting these pieces into a bacteria which mass reproduces the DNA sequences into sufficient quantities for analysis. The segments are cut up with overlapping ends so that the various segments can be placed in order once sequencing is complete. In the course of the last decade the sequencing technology has been completely revolutionised, with costs being halved every 18 months.

Celera was able to exploit the work done by the HGP. Established in May 1998, the company announced it would sequence the entire human genome in just three years. As well as having enormous resources from the pharmaceutical industry, Celera was able to outpace HGP because it used a less rigorous method, which was only viable because it was able to verify its work against the HGP data. Celera cut the genome into random bits and then used a powerful computer and the HGP discoveries to order the sequences.

Although the genome map is still incomplete, HGP and Celera have sequenced about 2.7 billion base pairs. At the same time, the two research teams have sequenced the genomes of other organisms, both to perfect their techniques and as points of comparison for analysing the human genome. The genetic sequencing for organisms such as bacteria, fruit flies and roundworms has already been completed and a map of the mouse genome was published almost simultaneously with that of human beings.

Last June, HGP and Celera published the sequence of bases of the human genome but this still left the task of identifying the genes. These are particular aggregates of bases that provide the code for the production of proteins, which are the building blocks of the cell and regulate all cellular processes. By comparing the DNA sequencing of different organisms, scientists found that the basic structure of most genes has been conserved through evolutionary history. In other words, once a gene is identified in simpler organisms, one only has to find a similar sequence in the human genome to identify a gene. The sequence of bases for genes of fundamental processes such as DNA duplication and respiration are almost the same in all cells. The identification of genes still remains a very difficult process and many of the genes have an uncertain status.

Initial scientific findings

Scientists have found between 25,000 and 40,000 genes in the human genome, which is a surprising result—it is only about double the number of genes in a fruit fly or worm. It was previously expected that the human genome would be made up of at least 120,000 genes. With approximately 30,000 genes, the human genome is capable of producing 250,000 proteins required for the successful functioning of the human body. Each gene codes for on average three proteins and for as many as 10 proteins. Human complexity appears to be the outcome of a capacity for the regulation and fine-tuning of the protein-producing abilities of genes which is lacking in simpler organisms.

The comparison of genetic structures in different organisms reveals that human evolution involves the variation of systems first evolved in the simplest organisms. The human genome added new functions to old proteins, which are found in organisms such as worms and flies. One genetic system in flies and worms used to fight off infecting microbes was copied a number of times and rearranged to evolve into the highly complex human immune system. A similar process produced the human blood clotting and hormone systems.

The analysis of the human genome has totally dispelled the notion propounded since the 19th century by the social Darwinists and others that there is a biological basis for race. Celera compared the genetic material from people who described themselves as Caucasian, Hispanic Mexican, Asian Chinese and African American and found that they were 99.9 per

cent identical in their genetic make up. Not one gene could be found that could be linked to race. Genetic variation is as wide or wider within racial groups as between them. In fact, humans have a great deal in common with organisms such as the mouse, with which we share 80 to 95 per cent of our genetic structure.

Scientists have also discovered that the processes of evolution are more complex that previously thought. It has been found that the human genome contains over 200 ancient genes, which can be traced to genetic material inserted from bacteria into a primitive human ancestor such as fish. However, these bacterial sequences are not present in more primitive organisms such as worms and insects, and thus do not occur in human beings as a result of the normal processes of evolution from simpler to more complex organisms. It is not known how this remarkable transfer of genetic information took place. Scientists know that bacteria can swap genetic information from bacteria to bacteria but have never observed genetic transfer from bacteria to higher organisms. Even more surprisingly, most of these bacterial genes have evolved to play vital functions in the human genome such as in metabolism, the maintenance of cell structures known as mitochondria and embryonic development in humans.

Another surprising discovery is that the genes may be regulated and controlled by sections of the genome previously known as “junk” DNA. Only just over one percent of the genome provides the codes for proteins and most of the rest was dismissed as so much “junk”. There are very long sections of the genome with endlessly repeating base sequences that have no apparent function. The reality is much more complex than previously imagined. It is becoming clear that these sequences, especially when located near the end of a chromosome or near a constriction in the chromosome, play a vital role in the evolution of new functions for genes.

A large proportion of the “junk” DNA has remained stable for some time and seems to consist of parasitic genetic elements which have moved from one section of the genome to another through evolutionary history. The movement of some of these elements through the genome corresponds to some key developments in human evolution. The reading of this complex history will provide a rich vein for scientists to mine.

One section of the “junk,” known as ALU sequences, makes up 14 percent of the genome. These are scattered throughout the genome and seemed not to do much except to make more of themselves. Recent studies indicate that ALUs located near genes are preserved while those located away from genes are weeded out, thus indicating they have some positive function. At this stage it appears that they have a role in activating protective genes in response to certain conditions.

Commercial potential

By revealing the detailed structure of the basic mechanisms for the functioning and reproduction of cells, the mapping of the human genome has opened up new avenues for medical research. Scientists believe that the data will provide fresh insights into disease processes, particularly in the case of inherited diseases such as muscular dystrophy, and into the ways in which the genetic makeup of individuals may make them more or less susceptible to different types of disease. Such knowledge may in turn lead to genetic therapies to treat various diseases.

Pharmaceutical companies are also keen to exploit the genetic data to develop new drugs, which is one of the reasons why big business has shown such a keen interest in developments in biotechnology. Along with computer, communication and Internet firms, speculation in biotech companies has led to huge rises in their share values in expectation of substantial future profits. Celera is just one of a number of companies that have sought to exploit the commercial potential. Its shares shot up by 15 percent on the New York Stock Exchange on the basis of the latest scientific announcement.

Celera has published a free version of the genome on the Internet but sells subscriptions of its updated data bases and search engines. Although actual subscription rates have not been disclosed, it has been reported that pharmaceutical and biotechnology companies have paid up to $US15 million for a subscription and university researchers pay US$10,000. Companies such as Pfizer and educational institutions such as Harvard are known to be subscribers.

An article in the *Wall Street Journal* endorsed the boasting of J. Craig Venter, celera's CEO, that his company could produce a better map of the human genome far faster than “legions of tax-payer funded academics who had almost a decade-long head start.” It commented: “It turns out that for all his braggadocio, Dr. Venter was right. Three-year-old Celera, it now is clear, has produced a map that drug and biotech companies, hungry for gene information that will help them find new treatments, are plunking down millions of dollars a year for the rights to sift through.”

But as the newspaper was forced to admit, towards the end of its lengthy comment, Celera was only able to achieve its results by capitalising on the information made freely available by the “legions of tax-payer funded academics” working on the Human Genome Project. Far from assisting scientific investigation, the creation of private monopolies on sections of the human genome will hinder further scientific and medical research, which, by its very nature, involves the collaboration of researchers internationally.

Significantly, a number of scientists have publicly opposed Celera's use of the human genome data for profit. John Sulston, head of a British team involved in sequencing the genome, stated: “It would have been criminal to prevent the access to this information.” Commenting on the Human Genome Project data, he said: “There are no patents filed... no subscription fees, no licenses and no documents to sign. All you need is an Internet connection. The human genome is our shared inheritance. It belongs to all of us."

Others are also seeking to cash in on the genetic discoveries. The US Patent and Trademark Office estimates that it has issued 1,000 patents for full-length genes and that tens of thousands of applications are pending. Robert Waterston, a DNA mapping expert at Washington University in St. Louis commented: “I think there are lots of [law] suits to be filed, and this [announcement of the genome map] will make it more so.”

The publication of results by HGP and Celera represent a truly remarkable human achievement, produced by the collective efforts of hundreds of scientists and technicians. It beautifully reveals that the human genome and humanity itself is the product of billions of years of evolutionary history in which even the lowly bacteria has played its part in the most surprising way.

A huge amount of research still needs to be done. The rest of the human genome has to be mapped and the genes on the genome definitively identified, along with those on the genomes of other animals and plants. There is no doubt that this work will provide more surprising insights into the twists and turns of the complex evolutionary process. Even hampered by the constraints imposed by the profit system, the stunning speed at which this research is proceeding—the amount of genetic data is estimated to be doubling every six months—demonstrates the potential of mankind for progress.