



February 11, 2001, Sunday

Genome Analysis Shows Humans Survive on Low Number of Genes

By NICHOLAS WADE

Opening a new era in human biology and medicine, two rival teams of scientists this week present their first interpretations of the human genome, the set of DNA-encoded instructions that specify a person.

The two teams report in articles to be published on Thursday and Friday that there are far fewer human genes than thought -- probably a mere 30,000 or so -- only a third more than those found in the roundworm.

One team, Celera Genomics, has compiled a parts list of the proteins needed to make a person. The other team, a publicly funded consortium, has traced the history of how the "junk" regions of the genome accumulated and has found that small elements of the junk may play a useful role. They also discovered that human genes have been derived directly from bacteria.

The two teams announced last June that they had assembled the human genome, but it has taken them until now to analyze their findings.

The interpretation of the genome -- identifying the genes, their functions and controls, and how they relate to human physiology and disease -- is expected in time to revolutionize medicine by clarifying the mechanism of many diseases and generating new tests and treatments.

Physically, the genome is minuscule -- two copies of it are packed into the nucleus of every ordinary human cell, each one of which is about a fifth the size of the smallest speck of dust the eye can see. But the genome is vast in terms of its informational content. Composed of chemical symbols designated by a four-letter alphabet of A's, T's, C's, and G's, the human genome is some 3.2 billion letters in length. If printed in standard type, it would cover 75,490 pages of this newspaper.

The enormous task of decoding the genomic message began in 1990 and is now substantially complete, although both teams' versions of the genome are riddled with gaps.

With so much effort and scientific glory at stake, members of each team remain highly critical of the other's approach, believing that their own strategy for decoding the genome is likely to produce the better and more accurate version. Since last June, however, both have been muting criticism and observing a limited truce. The pact called for a joint announcement, made at the White House on June 26 last year, that each side had finished assembling its version of the genome, and for joint publication of their findings, which is occurring later this week.

The joint publication, however, is about as separate as a union could be, with each side's articles appearing in rival scientific journals issued on different sides of the Atlantic. The findings were to be announced tomorrow, but the embargo was lifted by the two journals after The Observer of London broke it.

One team is a consortium of academic centers, mostly in the United States and Britain but with members in France, Germany, China and Japan. The consortium is financed largely by the National Institutes of Health and the Wellcome Trust of London. Its version of the human genome is described in a 62-page article in Nature, based in London. The principal author is Dr. Eric Lander of the Whitehead Institute in Cambridge, Mass.

The other team is led by Dr. J. Craig Venter, president of Celera Genomics in Rockville, Md. Its report appears in a 48-page article in Science, based in Washington.

Despite the two team's many differences, they largely agree on their findings about the human genome. There is the first overall look at a genetic document of extraordinary strangeness and complexity. No one expected it to be comprehensible at first glance and the two teams have so far mapped only the principal features of its terrain.

Their principal discovery is how few human genes there seem to be. Textbooks have long pegged the number of human genes at around 100,000, but with the sequence of human DNA units in hand the two teams have found far fewer than expected. Dr. Venter says he has identified 26,588 protein-coding genes for sure and another 12,000 possible genes. The consortium says there are 30,000 to 40,000 human genes. Both sides prefer the lower end of their range, since their methods of gene discovery tend to predict more genes than they believe exist.

The low number of human genes -- say 30,000 -- can be seen as good for medicine because it means there are fewer genes to understand.

The impact on human pride is another matter. Of the only two other animal genomes sequenced so far, the roundworm has 19,000 genes and the fruit fly, also a standard laboratory organism, 13,000. Both teams devote part of their huge articles to discussing how it is that humans are more complicated than simple invertebrate animals even though they possess not that many more genes.

Despite these face-saving efforts, human self-esteem may be in for further blows as genome analysis progresses. Dr. Venter said he could find only 300 human genes that had no recognizable counterpart in the mouse. The mouse, though a fellow mammal, last shared a common ancestor with people 100 million years ago, time in which many more genetic differences might have been expected to develop.

Given the minor difference between man and mouse, Dr. Venter said he expected the chimpanzee, which parted company from the human line only five million years ago, to have an almost identical set of genes as people but to possess variant forms of these genes.

The consortium, taking its own jab at anthropocentric pomp, identified 113 human genes, and possibly scores more, that have been acquired directly from

bacteria.

In the journal articles, the two sides also sketch out major features of the genome's architecture, of which genes are only a small part. More than half the genome consists of repetitive DNA that has no genetic meaning. Much of the repetitive DNA is formed by a couple of rogue genes that millions of years ago learned to copy and insert themselves into new sites in the genome. Because mutations clock up in these repeated segments at a fairly regular rate, their origins can be dated.

The consortium has found that the main families of repetitive DNA fell extinct long ago and no longer add clutter to the genome. But one family is still active, and since its members are often found near active genes they may benefit the genome in some way.

Both teams' versions of the genome now seem to be in a good enough state to be of great use to biologists. The consortium's genome is available for free and Celera's through subscription. But Celera provides extra services, such as the ability to compare the human genome sequence with that of the mouse. Mouse DNA has retained a very similar sequence to human DNA both in its genes and in the DNA regions that control the activity of genes, but has diverged through mutation in all the nonessential parts of the genome. Laying mouse DNA on top of human DNA shows at a glance which regions evolution has thought worth conserving.

The consortium, however, is also working on the mouse genome and intends to put that and other important tools for interpreting the human genome in the public domain.

Experts are likely to debate which team's method for sequencing the human genome is better. Dr. Venter's article includes a comparison chart that shows that the consortium's version of the genome has many more gaps than Celera's and that the gaps are larger. But in an interview Dr. Venter complimented the consortium's efforts. "We are really impressed at how good the public paper is, given their input data," he said.

But Dr. Lander said Celera's strategy was a grand experiment that failed because it produced more than 100,000 assembled pieces that could not be anchored to the genome sequence. Dr. Mark Adams of Celera said that the statement was inaccurate and that the company had assembled more than 95 percent of the genome into 2,845 large pieces that were well anchored to the genome.

Despite their different strategies, both sides borrowed heavily from the other. Dr. Venter used not only the snippets of DNA decoded by the consortium but also important information about their position generated by Dr. Robert H. Waterston of Washington University in St. Louis. The consortium belatedly copied two of Dr. Venter's innovations, a clever method of linking DNA sequence data by "paired-end reads," and reliance on heavy-duty computing to assemble data. The consortium had not prepared an assembly program, even though much of the analysis in the report depends on it, until a graduate student at the University of California at Santa Clara, James Kent, stepped in and wrote one for them at the last minute.

The rivalry between the two sides takes many petty forms -- speaking time for each side at a news conference to be held tomorrow was negotiated to the minute, and academic scientists including Dr. Lander tried strenuously to prevent Science from publishing Celera's article except under terms unacceptable to Dr. Venter. But the competition has proved enormously beneficial overall. The consortium was on a leisurely track to finish the genome by 2005 until Dr. Venter jumped into the race in May 1998, saying he would complete the genome by 2000.

"I think the publicly funded group has brought off something extraordinary," said Dr. Donald Kennedy, editor of Science and former president of Stanford University. "Imagine trying to do this job in a number of places with academic scientists -- it's like herding cats. They deserve all kinds of credit, but so does Venter and Celera. There is no doubt the world is getting this well before it otherwise would have if Venter had not entered the race."

The closeness of the finish has now become apparent. Dr. Venter said in his article that he completed his first assembly of the human genome on June 25, just the day before. Mr. Kent completed his first assembly of the consortium's data on June 22, just three days before Celera's.

Both sides have in substantial measure achieved their goals. Celera went from a concept to building a new plant from scratch to completed genome sequence in just 25 months, despite the predictions of the consortium's experts that its DNA sequencing strategy was bound to fail. "This is something I felt I had been driving for for a decade," Dr. Venter said last week, in commenting on his decision to place his name first on the Celera report's list of authors. "No small amount of this was the politics and psychology of being to stay with this and stick with it. If there was any way to stop this, it was tried, down to the end of trying to block our paper being published in Science. If we weren't resistant and somewhat defiant this never would have gotten done."

Dr. Venter's principal partners include the scientific manager of Celera's team, Dr. Adams, his computer program designers, Dr. Eugene W. Myers and Dr. Granger G. Sutton, and Dr. Hamilton O. Smith, who prepared the genome for analysis.

The consortium's goal was to place the human genome in the public domain for unfettered use by the world's biologists, and it has now done so four years ahead of its original schedule. The architects both of this policy and the DNA sequencing strategy were Dr. John Sulston of the Sanger Centre near Cambridge, England, and Dr. Waterston. Their centers completed roughly a quarter each of the genome sequence, and Dr. Lander's center at the Whitehead Institute did another quarter. Dr. Lander was also chairman of the group that analyzed the completed genome sequence. The consortium was led by Dr. Francis S. Collins of the National Institutes of Health.

Both teams believe that the sequencing and interpretation of the human genome is a historic event and expressed pride in their accomplishments. But both groups expressed humility at the minute steps they have so far taken in exploring the human genome's vast repository of knowledge.

"In principle," the consortium's biologists concluded in their report, "the string of genetic bits holds long-sought secrets of human development, physiology and medicine. In practice, our ability to transform such knowledge into understanding remains woefully inadequate."

Dr. Venter said simply that the effort to sequence and interpret the human genome had been "mentally exhausting, in part because we are not mentally equipped to absorb all this."

"We feel like midgets describing the universe and we can't comprehend it all," he added.

Organizations mentioned in this article:

Related Terms:

You may print this article now, or save it on your computer for future reference. [Instructions for saving](#) this article on your computer are also available.

Copyright 2001 The New York Times Company