The New York Times
nytimes.com

December 5, 2002

Similarities Found in Mouse Genes and Human's

By NICHOLAS WADE

An analysis of the mouse genome by an international consortium of scientists, a landmark event in biology, shows it is so similar to that of people that it should speed efforts to understand the human genome and the genetic roots of disease.

This is the first time that the reasonably complete genomes of two mammals, mouse and man, have become available for comparison. While the genome of a mammal even closer to the human, like the chimpanzee, may someday be decoded, the mouse is both genetically close and also an ideal laboratory animal.

Man and mouse are cousins, each descended from a small mammal that split into two species toward the end of the dinosaur era. Despite 75 million years of separate evolution, only about 300 genes Ñ 1 percent of the 30,000 possessed by the mouse Ñ have no obvious counterpart in the human genome, according to the new analysis published in today's issue of Nature.

This similarity makes the mouse genome an excellent surrogate for studying the human genome, especially for tests that would be ethically impossible in people. To understand the role of any newly found human gene, researchers can identify the counterpart gene in mice, genetically engineer a strain of mouse that lacks the gene, and figure out from the mouse's defects what the missing gene is meant to do.

The analysis has also yielded new insight into the workings of evolution and brought to light the existence of a large class of novel genes that produce a substance related to DNA.

Dr. Robert Winston, a human fertility expert at the Imperial College London, called the analysis a "landmark announcement" that "will undoubtedly further our understanding of the molecular basis for human diseases and the treatment of the widest range of human disorders."

The analysis accompanies the official release of the consortium's version of the mouse genome, which has been available online since May and is considered an essential component of the human genome project.

The international consortium's mouse genome was prepared by the Whitehead Institute in Cambridge, Mass., the Genome Sequencing Center at the Washington University School of Medicine in St. Louis, and the Sanger Institute near Cambridge, England. The leading authors of the consortium's report are Dr. Robert Waterston of Washington University, Dr. Eric Lander and Dr. Kerstin Lindblad-Toh of the Whitehead Institute, Dr. Jane Rogers of the Sanger Institute and Dr. Ewan Birney of the European Bioinformatics Institute, which is housed next to the Sanger. The mouse genome project was financed largely by the National Institutes of Health and the Wellcome Trust of London.

The new analysis, based largely on comparing the mouse's genome Ñ DNA unit by DNA unit Ñ with that of the human genome, has turned up new insights into how evolution shapes a species.

The mouse, for instance, to serve its acuter sense of smell, has developed many more genes related to odor detection than people have. The consortium's genome-scanning computers have identified 25 of these gene groups that are expanded in the mouse, compared with people. Besides smell, the expanded groups underlie such roles as pheromones, hormones that are involved in sexual attraction; degradation of toxins in the diet; and immune defense.

The analysis also suggests that more of the genome is useful than scientists had realized. Genomes are shaped by mutation Ñ random changes in the chemical units of DNA caused by radiation or copying errors Ñ or by the force Darwin called natural selection, preserving the fittest organisms and discarding the rest.

Using a test that compares mutations on the human and mouse genomes, and that takes into account those which cause changes in a protein, the authors of the analysis found that some 5 percent of each genome is under selective pressure.

The figure is provocative because less than 3 percent of the DNA is thought to be occupied by genes that code for proteins, the working parts of the living cell. So another 2 percent of the genome must be doing something of such importance for each organism's survival that it has been conserved for many millennia.

This other 2 percent may include control regions Ñ stretches of DNA upstream of a gene that serve as its on-off switch Ñ as well as a novel class of genes called RNA genes.

RNA, a more ancient chemical version of DNA, performs many basic tasks in a cell, one of which is to form a copy or transcript of a gene and direct the synthesis of the gene's protein. Recently, some of these RNA transcripts have been found to have executive roles all their own, without making any protein. An RNA gene is responsible for the vital task of shutting all the genes on one of the two X chromosomes in each female cell, ensuring that women get the same dose of X-based genes as men, who have just one X chromosome.

The new analysis suggests that a large family of such RNA genes may exist, a point confirmed by a Japanese team under Dr. Yoshihide Hayashizaki of the Riken Genomic Sciences in Yokohama. Dr. Hayashizaki, whose report also appears in today's Nature, has collected almost all the RNA transcripts made by mouse cells taken from every tissue of the mouse's body.

Though most of the transcripts code for mouse proteins, he found as many as 4,280 he could not match to any known mouse protein, suggesting that a large part of the genome consists of nonprotein coding genes. The roles of almost all these RNA genes have yet to be understood.

The mouse genome has been decoded separately by both the international consortium and the Celera Genomics Corporation of Rockville, Md., the same rivals who vied in decoding the human genome. Celera, then directed by Dr. J. Craig Venter, decoded the mouse genome two years ago but made it available by subscription only.

The consortium, which makes all its data public, has also prepared a reasonably complete draft of the mouse genome, which is the basis for the analysis released today. But to do so, it switched to Dr. Venter's method for decoding genomes, known as a whole genome shotgun, which consortium members criticized when he used it on the human genome.

"I feel wonderfully vindicated that they have seen the power of a whole genome shotgun," Dr. Venter said today. Dr. Venter published a study of one mouse chromosome in May, but did not complete a study of the full genome before he left Celera.

The mouse genomes decoded by the consortium and Celera are complete enough to be highly useful to researchers yet are still far from finished. They differ with each other in about 10 percent of the genome, according to a comparison by Dr. Michael Q. Zhang of the Cold Spring Harbor Laboratory on Long Island. In a study published today in the online journal Genome Biology, Dr. Zhang matched Celera's second version of the mouse genome with the consortium's third version, both of which were released online in May.

He found that Celera's had "higher accuracy in base pairs and overall coverage of the genome," but that the consortium's was better in certain regions where it had more recent data. Base pairs are the units of which the DNA molecule is composed. Because of these different strengths, Dr. Zhang advises researchers to use both genomes.

The mouse genome is 2,510 million units of DNA in length, according to Celera, and 2,475 million units according to the consortium. The human genome, by comparison, is 2,900 million units. But since both species seem to possess around 30,000 protein-coding genes, the extra DNA in the human genome is probably of no direct importance.

The availability of two reasonably complete mammalian genomes is of great help to gene finders. The decoded genome is a string of A's, G's, C's and T's, the four units of DNA, with no annotation as to which regions are genes and which are junk. Because humans and mice have inherited much the same overall set of genes from their common ancestor, finding a new gene on one species now often leads to the discovery of its counterpart gene on the other.