

Team reduces the size of the human genome to 19,000 genes

July 3 2014



Unlike the two protein coding genes shown (top, middle), the gene ACO074091.13 (below) is predicted to produce no protein and is likely to be removed along with an additional 1,700 genes that are currently in the human genome annotation. Credit: CNIO

How nutrients are metabolised and how neurons communicate in the brain are just some of the messages coded by the 3 billion letters that make up the human genome. The detection and characterisation of the genes present in this mass of information is a complex task that has been a source of ongoing debate since the first systematic attempts by the Human Genome Project more than ten years ago.



A study led by Alfonso Valencia, Vice-Director of Basic Research at the Spanish National Cancer Research Centre (CNIO) and head of the Structural Computational Biology Group, and Michael Tress, researcher at the Group, updates the number of human genes –those that can generate proteins– to 19,000; 1,700 fewer than the genes in the most recent annotation, and well below the initial estimations of 100,000 genes. The work, published in the journal *Human Molecular Genetics*, concludes that almost all of these genes have ancestors prior to the appearance of primates 50 million years ago.

"The shrinking <u>human genome</u>," that's how Valencia describes the continuous corrections to the numbers of the protein-coding genes in the human genome over the years that has culminated in the approximately 19,000 human genes described in the present work. "The coding part of the genome [which produces proteins] is constantly moving," he adds: "No one could have imagined a few years ago that such a small number of genes could make something so complex."

The scientists began by analysing proteomics experiments; proteomics is the most powerful tool to detect protein molecules. In order to determine a map of human proteins the researchers integrated data from seven large-scale mass spectrometry studies, from more than 50 human tissues, "in order to verify which genes really do produce proteins " says Valencia.

Fewer than ten new genes separate mice and men

The results brought to light just over 12,000 proteins and the researchers mapped these proteins to the corresponding regions of the genome. They analysed thousands of genes that were annotated in the human genome, but that did not appear in the proteomics analysis and concluded: "1,700 of the genes that are supposed to produce proteins almost certainly do not for various reasons, either because they do not exhibit any protein



coding features, or because the conservation of their reading frames does not support <u>protein</u> coding ability, "says Tress.

One hypothesis derived from the study is that more than 90% of human genes produce proteins that originated in metazoans or multicellular organisms of the animal kingdom hundreds of millions of years ago; the figure is over 99% for those genes whose origin predates the emergence of primates 50 million years ago.

"Our figures indicate that the differences between humans and primates at the level of genes and proteins are very small," say the researchers. David Juan, author and researcher in the Valencia lab, says that "the number of new genes that separate humans from mice [those genes that have evolved since the split from primates] may even be fewer than ten." This contrasts with the more than 500 human genes with origins since primates that can be found in the current annotation. The researchers conclude: "The physiological and developmental differences between primates are likely to be caused by gene regulation rather than by differences in the basic functions of the proteins in question."

Doing more with less

The sources of human complexity lie more in how genes are used rather than on the number of genes, in the thousands of chemical changes that occur in proteins or in the control of the production of these proteins by non-coding regions of the genome, which comprise 90% of the entire genome and which have been described in the latest findings of the international ENCODE project, a Project in which the Valencia team participates.

The work brings the number of <u>human genes</u> closer to other species such as the nematode worms Caenorhabditis elegans, worms that are just 1mm long, but apparently less complex than humans. But Valencia



prefers not to make comparisons: "The human genome is the best annotated, but we still believe that 1,700 genes may have to be reannotated. Our work suggests that we will have to redo the calculations for all genomes, not only the human genome."

The research results are part of GENCODE, a consortium which is integrated into the ENCODE Project and formed by research groups from around the world, including the Valencia team, whose task is to provide an annotation of all the gene-based elements in the human genome.

"Our data are being discussed by GENCODE for incorporation into the new annotations. When this happens it will redefine the entire mapping of the human genome, and how it is used in macro projects such as those for cancer genome analysis ", says Valencia.

Provided by Centro Nacional de Investigaciones Oncologicas (CNIO)

Citation: Team reduces the size of the human genome to 19,000 genes (2014, July 3) retrieved 5 May 2024 from <u>https://medicalxpress.com/news/2014-07-team-size-human-genome-genes.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.