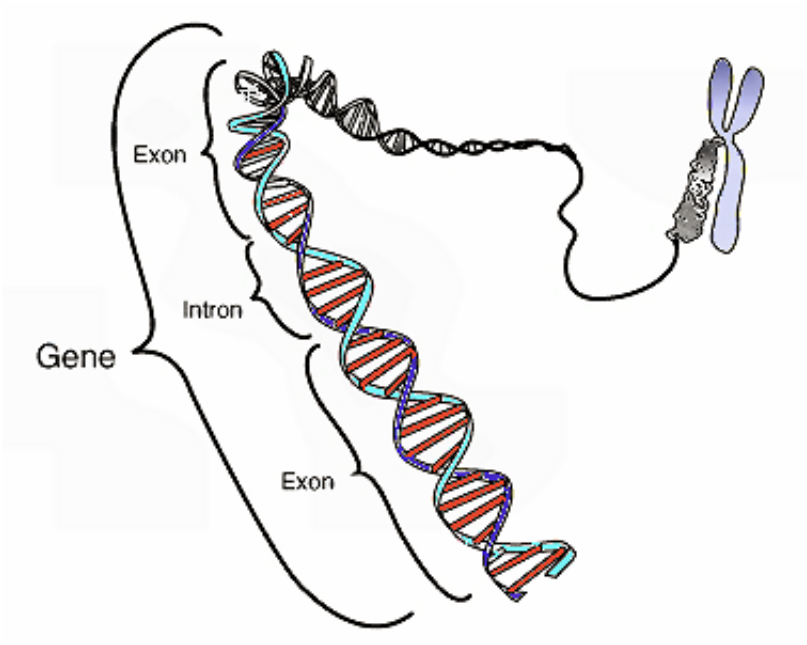
Scientists develop comparative search engine that helps to predict human gene function

February 13 2015, by Kimberlee D'ardenne



This image shows the coding region in a segment of eukaryotic DNA. Credit: National Human Genome Research Institute

The Human Genome Project wrapped up over a decade ago, yet around a third of the genome remains mysterious, its function unknown. Now, School of Medicine researchers have developed a comparative search engine that uses evolutionary correlations between human and other species' genes to help identify human gene function.

"After the human genome was sequenced, scientists thought it would be

a very short time before we knew what all the genes are doing," said Tobias Meyer, PhD, professor and chair of chemical and systems biology. "It turned out not to be so easy, and we are currently in a holding pattern before we can really make use of all the genomic information."

Mapping how the human genome functions is like a completing a giant jigsaw puzzle. Such a map has been called the "interactome," and having some idea about what a gene does helps identify where that gene fits in the puzzle.

"Identifying gene function is important for medicine because how genes interact with each other affects disease," said graduate student Gautam Dey.

The search engine relies on "big data," drawing from an international database that contains genomic sequences of hundreds of species, and is accessible via a Web page that is free and available to the public. The Web page went live Feb. 12, the same day that the paper describing the researchers' method for gene-function mapping was published online in *Cell Reports*. Dey is the lead author of the paper, and Meyer is the senior author.

The search engine is at web.stanford.edu/group/meyerla … MAPServer/index.html.

Where to begin?

About 6,000 of the human genome's roughly 20,000 genes have unknown or poorly characterized function. "The reason we don't know much about these genes is because they do not have an obvious starting point for investigation," Dey said.

To computationally identify the function of a gene, scientists have a few options. The easiest is finding another human gene with a similar sequence for comparison. Another option is searching for human genes with shared ancestry for comparison. But sometimes there is no human gene available for comparison, and scientists have to compare human genes to those from other species.

Importantly, their method does not require that scientists have prior information about gene function, said Meyer, who is also the Mrs. George A. Winzer Professor in Cell Biology. The comparative search engine narrows the myriad possible starting points for identifying human gene function through a process called phylogenetic profiling. Phylogenetic profiling links human genes to genes from other species based on shared ancestry.

To generate the phylogenetic profiles, the search engine queries RefSeq, an online database of human and other species' genomic sequences maintained by National Center for Biotechnology Information. At the end of 2014, there were genomes from more than 200 eukaryotic species—animals whose cells have a nucleus—contained in the RefSeq database. The Stanford researchers compared the human genome to the genomes from 176 species, including birds, fungi and single-celled organisms.

Generating phylogenetic profiles is a complex process, but Meyer and his colleagues made visualizing them easy with the search engine. A Web page asks for the name of a human gene and outputs a phylogenetic profile and a list of genes with shared ancestry. The search engine translates the phylogenetic profile into a color-coded and labeled map in which each group of species has its own color

A piece of the interactome jigsaw puzzle

Meyer and colleagues analyzed the entire human genome and focused on generating functional predictions for a subset of the 6,000 human genes with unknown function.

The phylogenetic profile is not just aesthetically pleasing; it helps researchers probe functional similarities among genes across 176 different contexts, in species ranging from birds and fish to plants and single-celled algae.

Comparing phylogenetic profiles from two different human genes whittles down the gene's likely functions, which means researchers can nail down the actual function with fewer laboratory experiments. If phylogenetic profiles from two human genes are similar, it indicates shared evolutionary history, and the genes might have similar functions.

The researchers validated the search engine on 14 human genes of previously unknown function. They used the phylogenetic profiles, which identify shared ancestry between human and non-human genes, as a starting point to identify the function of the 14 genes, which were found to contain instructions for building proteins important to intracellular transport and signaling.

By constraining the possibilities of gene function, the phylogenetic profiling method starts to demystify those parts of the human genome that are poorly understood. Dey estimates that their technique is capable of making useful predictions about function for about 600 genes with unknown function, or approximately 10 percent of the 6,000 human genes with unknown function.

The researchers' work is an example of using freely available data sets to make it easy for scientists to ask and answer difficult questions. The use of such data sets is growing in the biological sciences.

Provided by Stanford University Medical Center

Citation: Scientists develop comparative search engine that helps to predict human gene function (2015, February 13) retrieved 25 April 2024 from https://phys.org/news/2015-02-scientists-human-gene-function.html

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.