

Software reveals the inner workings of the human genome

January 12 2010, by Kurt Pfitzer

(PhysOrg.com) -- A biologist and computer scientist seek sites of RNA editing, a phenomenon that plays a key role in human genetic complexity.

The completion of the <u>Human Genome Project</u>, says Stefan Maas, has allowed scientists to dream on a grand scale. Armed with unprecedented knowledge of the body's genes and their location and function, scientists can contemplate advances in medicine and biotechnology that were hardly conceivable a decade or two ago.

This new genetic blueprint, says Maas, an assistant professor of biological sciences, could shed light on the origins of cancer, ALS (amyotrophic <u>lateral sclerosis</u>) and a host of other diseases and lead to new treatments for these diseases.

The genome project, which represented 13 years of work, also raised questions related to how complexity and diversity arise in humans and in other higher life forms. The approximately 30,000 genes discovered in the human genome are far fewer than the 50,000 to 140,000 scientists had expected to find. Furthermore, some simpler organisms have more genes—or proportionally more—than do humans. The rice genome contains 50,000 genes and the fly 14,000, to cite two examples.

This lack of correlation between genome size and complexity suggests that other phenomena contribute to the complexity and diversity found in higher life forms. Maas and Daniel Lopresti, a professor of computer



science and engineering, have been collaborating for four years in a study of one of these phenomena, <u>RNA</u> editing. Research in Maas' lab is supported by the National Institutes of Health.

RNA editing, says Maas, includes a variety of mechanisms by which gene sequences are altered after DNA is transcribed into RNA and before RNA is translated to the proteins that determine an organism's structural, enzymatic and regulatory functions. The most important of these mechanisms involves the modification of single nucleotides, the molecules that connect to form the structural units of RNA and DNA.

From modified nucleotides to changes in protein function

The human genome contains 3.4 billion nucleotides. Modifications in these molecules can cause changes to the <u>amino acids</u> in the proteins that are synthesized, which can lead in turn to an alteration of protein function. Thus, says Maas, who studies the genomes of humans, rats, mice and zebrafish, RNA editing yields a potentially "exponential" increase in the number of gene products that can be generated from a single gene—and a staggering volume of information to analyze.

"To date," says Maas, "about 300,000 sequences of human RNA have been characterized and are available for study. Each of these sequences encodes one protein.

"Only by examining all of the RNA sequences, can you determine how much RNA editing is going on in the human genome. How much diversity does it generate? How many different genes are subject to RNA editing? Not all genes undergo RNA editing, and there is no simple clue to determine which do and which do not."



"Searching for RNA editing sites is like looking for a needle in a gigantic haystack," says Lopresti. "You cannot go through this haystack manually, and you cannot guess where the editing sites are going to be."

To speed the process of identifying the sites in the genome where editing might occur, Lopresti has developed a software program called RNA Editing Dataflow System, or REDS. REDS identifies the discrepancies that occur when DNA is transcribed into RNA, and then separates out those that occur for reasons other than RNA editing. Maas and his students examine suspected RNA editing sites in the laboratory, isolating DNA and RNA from brain and other tissues and amplifying the sequences of both to determine whether editing has occurred.

"We then take the data we obtain from the lab and feed it to our software to improve on our predictions," says Maas. "The more data we obtain, the more our predictions can be based on machine learning."

A-to-I editing, and RNA folding

In the first stage of their study, Maas and Lopresti align each RNA sequence with its original genomic (DNA) counterpart and compare the two to determine if alterations have occurred. They are particularly interested in a type of RNA editing known as A-to-I editing, in which the nucleotide adenosine changes to the nucleotide inosine. They have further narrowed their focus to A-to-I editing cases in which the protein product contains an amino acid change. It is these amino acid changes that have been implicated in Lou Gehrig's disease, epilepsy, depression and other illnesses.

"Genes in which editing occurs usually have an 'A' that switches to an 'I' after RNA editing," says Maas. "If you isolate this gene and determine its sequence, you see the discrepancy between the RNA sequence and the genomic sequence from which it arose.



"Not all A-to-I changes are relevant to protein changes in a gene. We're most interested in cases where the product of the protein has an amino acid substitution as a result of RNA editing. In mammals, A-to-I modifications appear to be particularly widespread and are known to regulate crucial functional properties of neurotransmitter receptors in the brain."

In the second stage of their investigation, Maas and Lopresti subtract out discrepancies not related to RNA editing. These can occur because of errors in the original genomic sequences or in the RNA sequences. And they can be caused by single-nucleotide polymorphisms (SNPs), or <u>DNA</u> sequence variations that occur in a single nucleotide.

In stage three, the researchers examine RNA folding, the structures that result from this folding, and the correlation between these structures and the incidence of RNA editing. RNA is a dynamic molecule and its structure is in constant flux, like strands of spaghetti that fold and loop over each other. It is at these double-stranded regions where RNA editing is most likely to occur.

"RNA folds into a 3-D structure," says Maas, "to minimize energy consumption. This folding, which can occur in many different ways, causes nucleotides to form bonds to stabilize the overall molecule. Each gene where RNA editing occurs has a different structure. A somewhat stable secondary structure surrounds nucleotides that are undergoing RNA editing."

Deducing structure from sequence

Lopresti has written an algorithm that attempts to deduce RNA's structure from its sequence and then to determine, based on that structure, where RNA editing sites are likely to be found.



"We've developed very fast, quick and dirty computational techniques that simulate folding in order to determine the criteria for folding and to confirm the folding structures that are right for editing," he says. "RNA editing occurs inside double-stranded regions that can look like hairpin loops, interior loops, bulges and multi-loop configurations. We're tuning the parameters of our algorithm to find folding structures that match RNA editing sites. The algorithm is not perfect, but it does rank all potential editing sites based on predicted folding because of structure."

"Our computational tool screens the entire genome looking for editing sites," says Maas. "We look for RNA regions where base-pairing can occur. Our goal is to do this quickly while analyzing RNA folding properties. We're trying to develop a way to do this much faster and still get a meaningful outcome."

In the end, says Maas, it comes down to a numbers game.

"In human beings, we know of 30 genes where RNA editing occurs in which codon [a codon is a set of three <u>nucleotides</u> that code for a specific amino acid] changes cause amino acid changes in the protein. These 30 <u>genes</u> are well characterized. Most were found by chance. We're now trying to systematically find other editing sites in the genome and identify the consequences of these events.

"Each gene we find in which RNA editing occurs opens a new chapter about the significance of editing, the pathways that are involved and potential diseases that result from RNA editing deficiency or overactivity."

Provided by Lehigh University

Citation: Software reveals the inner workings of the human genome (2010, January 12) retrieved



28 April 2024 from https://phys.org/news/2010-01-software-reveals-human-genome.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.