

New tool helps researchers identify DNA patterns of cancer, genetic disorders

May 19 2009



This symbolic scatter plot reveals the structure of a portion of the gene responsible for Huntington's disease. (Note the section of repeated 3-mers). Credit: David Cox, North Carolina State University

A new tool will help researchers identify the minute changes in DNA patterns that lead to cancer, Huntington's disease and a host of other genetic disorders. The tool was developed at North Carolina State University and translates DNA sequences into graphic images, which allows researchers to distinguish genetic patterns more quickly and efficiently than was historically possible using computers.

David Cox, a Ph.D. student in computer science at NC State, devised the "symbolic scatter plot" tool to provide a visual representation of a DNA sequence. Cox explains, "The human visual system is more adept at identifying patterns, and differentiating between patterns, than existing computer programs such as those that try to identify repetitions of DNA sequences." In other words, the naked eye sees patterns better than computers can.

Identifying patterns in a sequence of DNA is important because it can

help researchers identify the minute genetic variations between subjects that suffer from a disease, such as cancer, and subjects that do not.

"Improved identification of relevant DNA sequences will hopefully expedite the development of successful treatment for a range of diseases," Cox says, "by allowing researchers to focus on the components of DNA that are related to the disease and improving our understanding of the genetic mechanisms of these diseases. For example, what turns specific genes on and off?"



This symbolic scatter plot reveals patterns in the human Y chromosome. Credit: David Cox, North Carolina State University

So, how does the symbolic scatter plot create a visual representation of DNA? DNA is composed of a series of nucleotides. There are only four types of nucleotides, represented by the letters A, T, G and C. Each three-letter string of these nucleotides, such as AAA or ATG, is called a 3-mer. Cox explains, "There are only 64 possible 3-mers, thus each 3-mer maps to a number from zero to 63. The symbolic scatter plots take a very long string of letters representing a DNA sequence and split it into a bunch of 3-mers. It then plots a point for each 3-mer, zero through 63, with that number serving as the y-coordinate." The x-axis is the order that the 3-mer appears in the genetic sequence.

"If this seems really simple," Cox says, "that's because it really is simple. Even so, the resulting scatter plots reveal interesting patterns in the original DNA. I can also string these scatter plots together to produce animations for the purpose of comparing [DNA sequences](#)."

Cox chose to focus on 3-mers because they correlate to codons, which are the genetic codes the body uses to specify the insertion of a specific amino acid during the creation of proteins. In other words, they oversee the creation of proteins - which are themselves the basic building blocks of the human body. "There are 64 3-mers, but only 20 [amino acids](#)," Cox says, "so each amino acid corresponds to multiple 3-mers." Cox designed the symbolic scatter plot so that those 3-mers that correspond to the same amino acid are adjacent to one another.

"This way," Cox says, "it is easier to determine when a difference in 3-mers is significant - from one amino acid to another - rather than a difference in 3-mers that still results in the production of the same amino acid. A change in a single amino acid can be the difference between a relatively harmless disease and a fatal one," Cox says.

More information: Cox will present the research this July at BIOCOMP '09 - The 2009 International Conference on Bioinformatics and Computational Biology in Las Vegas.

Source: North Carolina State University ([news](#) : [web](#))

Citation: New tool helps researchers identify DNA patterns of cancer, genetic disorders (2009, May 19) retrieved 26 April 2024 from <https://phys.org/news/2009-05-tool-dna-patterns-cancer-genetic.html>

<p>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.</p>
--