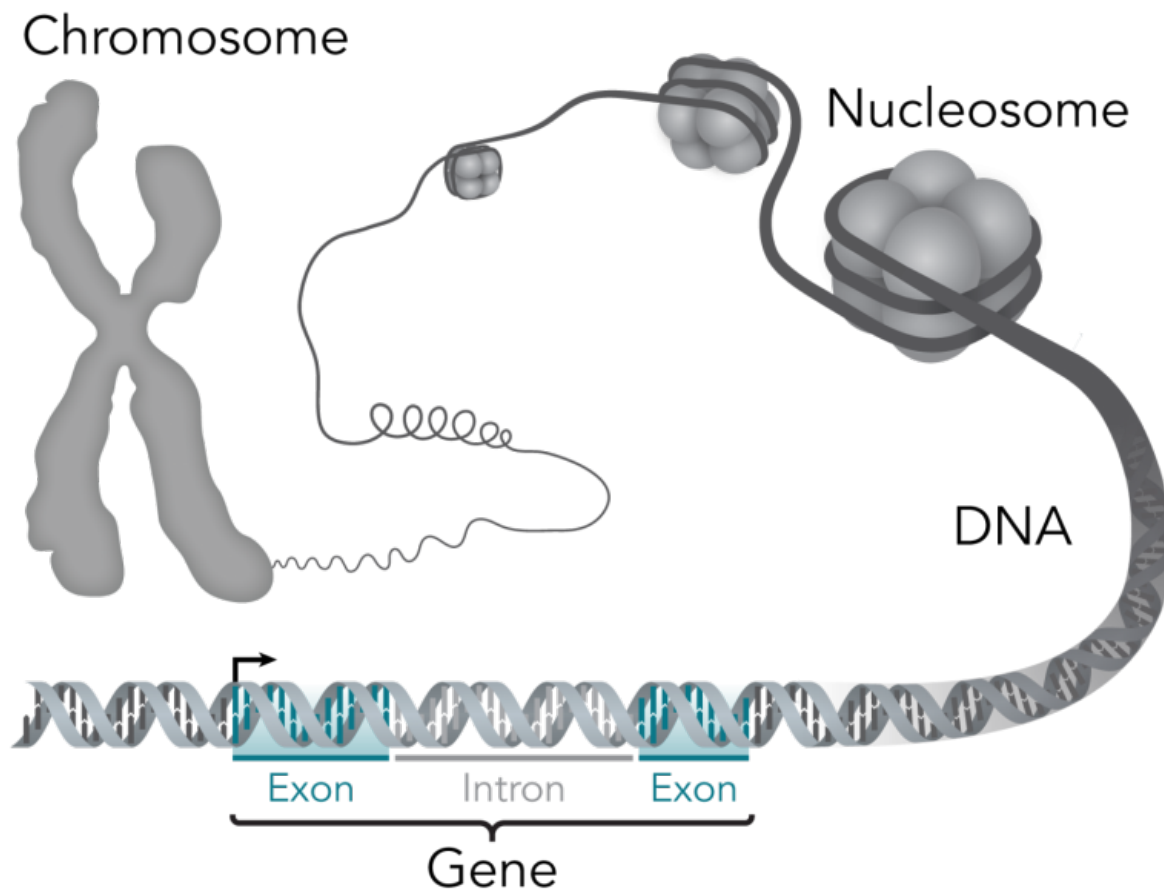


Individuals' medical histories predicted by their noncoding genomes, study finds

February 4 2016



This stylistic diagram shows a gene in relation to the double helix structure of DNA and to a chromosome (right). The chromosome is X-shaped because it is dividing. Introns are regions often found in eukaryote genes that are removed in the splicing process (after the DNA is transcribed into RNA): Only the exons encode the protein. The diagram labels a region of only 55 or so bases as a gene. In reality, most genes are hundreds of times longer. Credit: Thomas

Identifying mutations in the control switches of genes can be a surprisingly accurate way to predict a person's medical history, researchers at the Stanford University School of Medicine have found.

When the scientists used the technique to analyze the whole genome sequences of five individuals, they found that a person with narcolepsy had mutations in the regulatory regions of genes controlling alertness; a person with a family history of [sudden cardiac death](#) had mutations in regions controlling genes associated with [cardiac output](#); and a person with [high blood pressure](#) had mutations in regions controlling circulating [sodium levels](#) in the blood.

"The beauty of having whole genomes available for study is that you can then ask completely agnostic questions," said Gill Bejerano, PhD, an associate professor of developmental biology, of pediatrics and of computer science at Stanford. "We set out to find hidden layers of susceptibility in the regulatory regions of these genomes. We were very pleased that our analysis gave such clear and significant associations between the mutations and medical histories."

Bejerano, a genomicist who is a member of the Stanford Artificial Intelligence Lab, Child Health Research Institute, Neurosciences Institute, Cancer Institute and Bio-X, is the senior author of a paper describing the research, which will be published Feb. 4 in *PLOS Computational Biology*. The first author is Harendra Guturu, PhD, a former Stanford [graduate student](#) who is now a research associate in pediatrics at the university.

Importance of regulatory regions

The researchers focused their analyses on a relatively small proportion of each person's genome—the sequences of regulatory regions that have been faithfully conserved among many species over millions of years of evolution. Proteins called transcription factors bind to regulatory regions to control when, where and how genes are expressed. Some regulatory regions have evolved to generate species-specific differences—for example, mutating in a way that changes the expression of a gene involved in foot anatomy in humans—while other regions have stayed mostly the same for millennia.

"In these cases, evolution has given a clear signal that these regions are important to key biological pathways, and it's important for them to stick around," said Bejerano.

All of us have some natural variation in our genome, accumulated through botched DNA replication, chemical mutation and simple errors that arise when each cell tries to successfully copy 3 billion nucleotides prior to each cell division. When these errors occur in our sperm or egg cells, they are passed to our children and perhaps grandchildren. These variations, called polymorphisms, are usually, but not always, harmless.

GREAT work

Guturu looked for what are called [single nucleotide polymorphisms](#), or SNPs, in the DNA of five people who have made their genomes and information about their own or their family's medical history publicly available for use by researchers worldwide. SNPs are places along a chromosome where the DNA sequence varies from a composite human DNA reference sequence by one letter, or nucleotide.

Rather than search through the whole genome, Guturu focused on SNPs in evolutionarily conserved [regulatory regions](#). Even within these

regions, each person had many SNPs. So Guturu used a software program, Predicting Regulatory Information of Single Motifs, developed in the Bejerano lab, to predict which nucleotide changes were likely to disrupt the conserved binding of a transcription factor.

Guturu then turned to software called Genomic Regions Enrichment of Annotations Tool to determine whether the disrupted binding sites were likely to perturb the expression of groups of genes that together control a particular biological function. GREAT, which was also developed in the Bejerano lab, curates knowledge about the diverse functions of thousands of different groups of genes. For any set of genomic regions a user inputs, GREAT determines the most common set or sets of nearby genes.

Using this approach to study the genomes of the five individuals, Guturu, Bejerano and their colleagues found that one of the individuals who had a [family history](#) of sudden cardiac death had a surprising accumulation of variants associated with "abnormal cardiac output"; another with hypertension had variants likely to affect genes involved in circulating sodium levels; and another with narcolepsy had variants affecting parasympathetic nervous system development. In all five cases, GREAT reported results that jibed with what was known about that individual's self-reported medical history, and that were rarely seen in the more than 1,000 other genomes used as controls.

'Exciting avenue for study'

The researchers would like to create a web portal that would allow others to easily conduct similar studies. However, they concede that, for some diseases, the results may not be so clear-cut.

"We are the sum of billions of transcription-factor-binding events in thousands of cell types throughout our bodies," said Bejerano. "Not

every disease will be amenable to this type of analysis. But this study shows that nature, even the noncoding genome, can be very benevolent when you ask the right questions. And it may help us begin to combine our knowledge about variations, or mutations, that occur throughout the genome. It's a very exciting avenue for study."

The research is an example of Stanford Medicine's focus on precision health, the goal of which is to anticipate and prevent disease in the healthy and precisely diagnose and treat disease in the ill.

Other Stanford co-authors of the paper are graduate student Sandeep Chinchali and former graduate student Shoa Clarke, MD, PhD.

Bejerano and Guturu have filed a patent application on the algorithm used in this study.

Provided by Stanford University Medical Center

Citation: Individuals' medical histories predicted by their noncoding genomes, study finds (2016, February 4) retrieved 28 April 2024 from <https://medicalxpress.com/news/2016-02-individuals-medical-histories-noncoding-genomes.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>
--