

NEWS

Major errors in genome can be harmless

BY EMILY SINGER

27 FEBRUARY 2012

Lost DNA: Healthy people have about 100 mutations in their genomes that alter gene function.

Lost DNA: Healthy people have about 100 mutations in their genomes that alter gene function.

More than 250 genes in the human genome — about one percent of our genes — can be eliminated without serious health effects, according to research published last week in *Science*¹.

The findings are the result of an exhaustive effort to catalog ‘loss-of-function’ mutations — those that disrupt the function of a gene — in 185 people from the **1000 Genomes Project**.

Although genome sequencing studies have shown that apparently healthy people harbor many genetic mistakes, even sometimes possessing two flawed copies of a gene, the new study is unique in its scope.

“This is the first study to look at so many possible types of loss-of-function variants, and also the first to invest in really correcting all of the different kinds of errors that can pop up in high-throughput human genome sequencing,” says lead researcher **Daniel MacArthur**, who completed the work as a postdoctoral fellow at the Wellcome Trust Sanger Institute in Hinxton, U.K., and is

starting a new research group at Massachusetts General Hospital.

The researchers created a catalog of these mutations, which should help scientists interpret the results of whole-genome sequencing studies of disease. When mutations that disrupt the function of genes are identified, they're usually fingered as candidates for the genetic culprit behind the disease under study. But knowing that certain genes can be eliminated without serious consequence will allow researchers to send those to the bottom of the list of those potentially causing the disease.

Catalog of loss:

Ever since scientists began sequencing individuals' genomes, it became apparent that genomes from even healthy adults are full of flaws.

"The challenge has always been how to interpret that variation," says **Nicholas Katsanis**, director of the Center for Human Disease Modeling at Duke University in Durham, North Carolina, who was not involved in the study. "It will help us understand the buffering capacity of the genome, the ability to tolerate deleterious mutations."

MacArthur and his collaborators started by compiling a list of 3,000 possible loss-of-function mutations, which can range from single-letter mutations that make a non-functional protein to larger deletions that take out an entire gene.

The data came from the 1000 Genomes Project, an international collaboration to catalog human genetic variation. Participants are outwardly healthy adults from all over the world.

They then filtered out mutations that were due to sequencing or other errors, such as incorrect labeling of a gene. This process eliminated about half the list.

"Without applying so many filtering steps we would never have been able to generate a high-quality catalog of loss-of-function variants for further study," says MacArthur. "Now that we have this catalog, we can really start to figure out what effects these variants have on human variation and disease risk."

They found that everyone has approximately 100 of these mutations in their genome, about 20 of which affect both copies of a gene. At least one individual in the group had two dysfunctional copies of 253 genes in the genome.

"Even after going through all these filters, there are still a lot of genes that appear to be knocked out in healthy people," says **Mark Gerstein**, professor of biomedical informatics at Yale University and one of the researchers on the study.

The most common of these are genes involved in the olfactory system but also included those that determine blood type and the ability to metabolize drugs.

To confirm that the most common loss-of-function mutations aren't strongly linked to common complex diseases, researchers looked for correlations in data from the **Wellcome Trust Case Control Consortium**, a database of genetic information on 16,000 people who have one of seven different diseases. They found only one exception, a mutation that has previously been linked to Crohn's disease.

The new gene catalog will initially help researchers searching for the source of Mendelian diseases, such as cystic fibrosis and muscular dystrophy, in which disrupting two copies of a gene definitively leads to the disease, to eliminate mutations from the potentially long list of candidate mutations in an individual's genome.

But it should also aid in the search for the genetic mutations that increase risk for common diseases, such as autism, because scientists can use the catalog to prioritize the list of candidate mutations.

Expanding list:

The researchers compared their catalog of genes to known disease-linked genes, and found broad differences between the two classes.

On average, the former "were less evolutionarily conserved and had fewer protein-protein interactions," says **Chris Tyler-Smith**, head of the Human Evolution team at the Wellcome Trust Sanger Institute and a researcher involved in the study. That finding is not surprising; one would expect that genes that can be knocked out in people without serious effects would play a less central role in the cell, and that such genes would not be under strong evolutionary pressure.

While these criteria can't be used to sort out expendable genes from those that may cause disease, the findings should help researchers assess the importance of newly identified loss-of-function mutations.

The team plans to look at the full dataset of the 1000 Genomes Project, due to be completed this year, and aims to find additional loss-of-function mutations. They also plan to look at mutations outside the protein-coding regions of the genome, which comprises 99 percent of our DNA. Such mutations can influence gene expression and other cellular processes.

One of the limitations of the data from the 1000 Genomes Project is that the researchers did not characterize the participants' phenotype or physical traits. Although they broadly defined the participants as healthy adults at the time they donated DNA, some of them may have had minor health problems or gone on to develop medical issues.

“In the longer term, we would like to be able to go back to people carrying loss-of-function variants and look for subtle effects on their phenotype, health-related or otherwise,” says Tyler-Smith.

Most of the loss-of-function mutations identified in the study are rare, found in less than two percent of individuals in the group. This suggests that they were prevented from becoming more common by natural selection, as a result of having mildly or even seriously harmful effects, says MacArthur. “We suspect **rare loss-of-function variants** are more important than common ones in terms of effects on disease risk,” he says.

MacArthur is planning to create a custom genotype array — a tool designed to quickly detect specific mutations — to cost-effectively assess loss-of-function variants in people with specific diseases and controls.

“If there are loss-of-function variants playing a major role in these diseases, this approach should be able to track them down,” he says.

A second limitation of the new catalog is that it doesn’t identify mutations that may only cause problems when **combined with a second genetic mistake**.

“Mutations do not exert their effects in a vacuum but rather in the context of the entire genome,” says Katsanis. Some cases of autism have been linked to mutations in two or even three genes. “If studied in isolation, we might conclude [that these loss-of-function mutations] are tolerated,” he says. But it’s possible that “it’s only when they come together that they give rise to a clinical phenotype.”

References:

1: MacArthur D.G. et al. *Science* 335, 823-828 (2012) PubMed