

Scientists uncover new class of non-protein coding genes in mammals with key functions

February 1 2009



The RNA found in cells can go on to make protein, or in the case of lincRNAs, have a function of its own.

A research team at the Broad Institute of Harvard and MIT and Beth Israel Deaconess Medical Center has uncovered a vast new class of previously unrecognized mammalian genes that do not encode proteins, but instead function as long RNA molecules. Their findings, presented in the February 1st advance online issue of the journal *Nature*, demonstrate that this novel class of "large intervening non-coding RNAs" or "lincRNAs" plays critical roles in both health and disease, including cancer, immune signaling and stem cell biology.

"We've known that the human genome still has many tricks up its sleeve," said Eric Lander, founding director of the Broad Institute and co-senior author of the *Nature* paper. "But, it is astounding to realize that there is a huge class of RNA-based genes that we have almost entirely missed until now."

Standard "textbook" genes encode RNAs that are translated into proteins, and mammalian genomes harbor about 20,000 such protein-coding genes. Some genes, however, encode functional RNAs that are never translated into proteins. These include a handful of classical examples known for decades and some recently discovered classes of tiny RNAs, such as microRNAs.

By contrast, the newly discovered lincRNAs are thousands of bases long. Because only about ten examples of functional lincRNAs were known previously, they seemed more like genomic oddities than critical components. The new *Nature* study shows that there are actually thousands of such genes and that they have been conserved across mammalian evolution.

"The challenge in finding these lincRNAs is that they have been hiding in plain sight," said John Rinn, a Harvard Medical School assistant professor at Beth Israel Deaconess Medical Center and an associate member of the Broad Institute of Harvard and MIT. "The human and mouse genomes are already known to produce many large RNA molecules, but the vast majority show no evolutionary conservation across species, suggesting that they may simply be 'genomic noise' without any biological function."

To uncover this large collection of new genes, the Broad scientific team looked not at the RNA molecules themselves but at telltale signs in the DNA called chromatin modifications or epigenomic marks. They searched for genomic regions that have the same chromatin patterns as

protein-coding genes, but do not encode proteins. By surveying the genomes of four different types of mouse cells (including embryonic stem cells and cells from various tissue types), they found an astounding 1,586 such loci that had not been previously described. The researchers also found that the vast majority of these genomic regions are transcribed into lincRNAs, and that these are conserved across mammals.

"The epigenomic marks revealed where these genes were hiding," said Mitch Guttman, a MIT graduate student working at the Broad Institute. "Analysis of their sequence then revealed that the genes are highly conserved in mammalian genomes, which strongly suggested that these genes play critical biological functions."

By correlating the expression patterns of lincRNAs in various cell types with the expression patterns of known critical protein-coding genes in those same cells, the scientists observed that lincRNAs likely play critical roles in helping to regulate a variety of different cellular processes, including cell proliferation, immune surveillance, maintenance of embryonic stem cell pluripotency, neuronal and muscle development, and gametogenesis. Further experimental evidence from several of the identified lincRNAs verified these observations.

Because of the stringent experimental conditions imposed by the researchers in identifying the 1,600 lincRNAs in the Nature study, it is likely that there are many more lincRNA genes hiding in plain sight in the genome, as well as other RNA-encoding genes that are as important to genome function as their better-recognized protein-coding counterparts.

Paper: Guttman et al. 2009 "Chromatin signature reveals over a thousand highly conserved, large non-coding RNAs in mammals." *Nature* DOI 10.1038/nature07672

Source: Broad Institute of MIT and Harvard

Citation: Scientists uncover new class of non-protein coding genes in mammals with key functions (2009, February 1) retrieved 9 April 2024 from <https://phys.org/news/2009-02-scientists-uncover-class-non-protein-coding.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.