

'Dark matter' of the genome revealed through analysis of 29 mammals

October 12 2011, by Haley Bridger

An international team of researchers has discovered the vast majority of the so-called "dark matter" in the human genome, by means of a sweeping comparison of 29 mammalian genomes. The team, led by scientists from the Broad Institute, has pinpointed the parts of the human genome that control when and where genes are turned on. This map is a critical step in interpreting the thousands of genetic changes that have been linked to human disease. Their findings appear online October 12 in the journal *Nature*.

Early comparison studies of the human and mouse genomes led to the surprising discovery that the regulatory information that controls [genes](#) dwarfs the information in the genes themselves. But, these studies were indirect: they could infer the existence of these regulatory sequences, but could find only a small fraction of them. These mysterious sequences have been referred to as the dark matter of the genome, analogous to the unseen matter and energy that make up most of the universe.

This new study enlisted a menagerie of mammals – including rabbit, bat, elephant, and more – to reveal these mysterious genomic elements.

Over the last five years, the Broad Institute, the Genome Institute at Washington University, and the Baylor College of Medicine [Human Genome](#) Sequencing Center have sequenced the genomes of 29 placental mammals. The research team compared all of these genomes, 20 of which are first reported in this paper, looking for regions that remained largely unchanged across species.

"With just a few species, we didn't have the power to pinpoint individual regions of regulatory control," said Manolis Kellis, last author of the study and associate professor of computer science at MIT. "This new map reveals almost 3 million previously undetectable elements in non-coding regions that have been carefully preserved across all mammals, and whose disruptions appear to be associated with human disease."

These findings could yield a deeper understanding of disease-focused studies, which look for genetic variants closely tied to disease.

"Most of the genetic variants associated with common diseases occur in non-protein coding regions of the genome. In these regions, it is often difficult to find the causal mutation," said first author Kerstin Lindblad-Toh, scientific director of vertebrate genome biology at the Broad and a professor in comparative genomics at Uppsala University, Sweden. "This catalog will make it easier to decipher the function of disease-related variation in the human genome."

This new map helps pinpoint those mutations that are likely responsible for disease, as they have been preserved across millions of years of evolution, but are commonly disrupted in individuals that suffer from a given disease. Knowing the causal mutations and their likely functions can then help uncover the underlying disease mechanisms and reveal potential drug targets.

The scientists were able to suggest possible functions for more than half of the 360 million DNA letters contained in the conserved elements, revealing the hidden meaning behind the As, Cs, Ts, and Gs. These revealed:

- Almost 4,000 previously undetected exons, or segments of DNA that code for protein

- 10,000 highly conserved elements that may be involved in how proteins are made
- More than 1,000 new families of RNA secondary structures with diverse roles in gene regulation
- 2.7 million predicted targets of transcription factors, proteins that control gene expression

"We can use this treasure trove of new elements to revisit disease association studies, focusing on those that disrupt conserved elements and trying to discern their likely functions," said Kellis. "Using a single genome, the language of DNA seems cryptic. When studied through the lens of evolution, words light up and gain meaning."

The researchers were also able to harness this collection of genomes to look back in time, across more than 100 million years of evolution, to uncover the fundamental changes that shaped mammalian adaptation to different environments and lifestyles. The researchers revealed specific proteins under rapid evolution, including some related to the immune system, taste perception, and cell division. They also uncovered hundreds of protein domains within genes that are evolving rapidly, some of which are related to bone remodeling and retinal functions.

"The comparison of mammalian genomes reveals the regulatory controls that are common across all mammals," said Eric Lander, director of the Broad Institute and the third corresponding author of the paper. "These evolutionary innovations were devised more than 100 million years ago and are still at work in the human population today."

In addition to finding the DNA controls that are common across all [mammals](#), the comparison highlighted areas that have been changing rapidly only in the human and primate genomes. Researchers had previously uncovered two hundred of these regions, some of which are linked to brain and limb development. The expanded list – which now

includes more than 1,000 regions – will give scientists new starting points for understanding human evolution.

The comparison of many complete genomes is beginning to offer a clear view of once indiscernible genomic regions, and with additional genomes, that resolution will only increase. "The power of this resource is that it continues to improve with the inclusion of more species," said Lindblad-Toh. "It's a very systematic and unbiased approach that will only become more powerful with the inclusion of additional genomes."

More information: Lindblad-Toh et al. "A high-resolution map of human evolutionary constraint using 29 mammals." Nature October 12, 2011 [doi: 10.1038/nature10530](https://doi.org/10.1038/nature10530)

Provided by Massachusetts Institute of Technology

Citation: 'Dark matter' of the genome revealed through analysis of 29 mammals (2011, October 12) retrieved 26 April 2024 from <https://phys.org/news/2011-10-dark-genome-revealed-analysis-mammals.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.