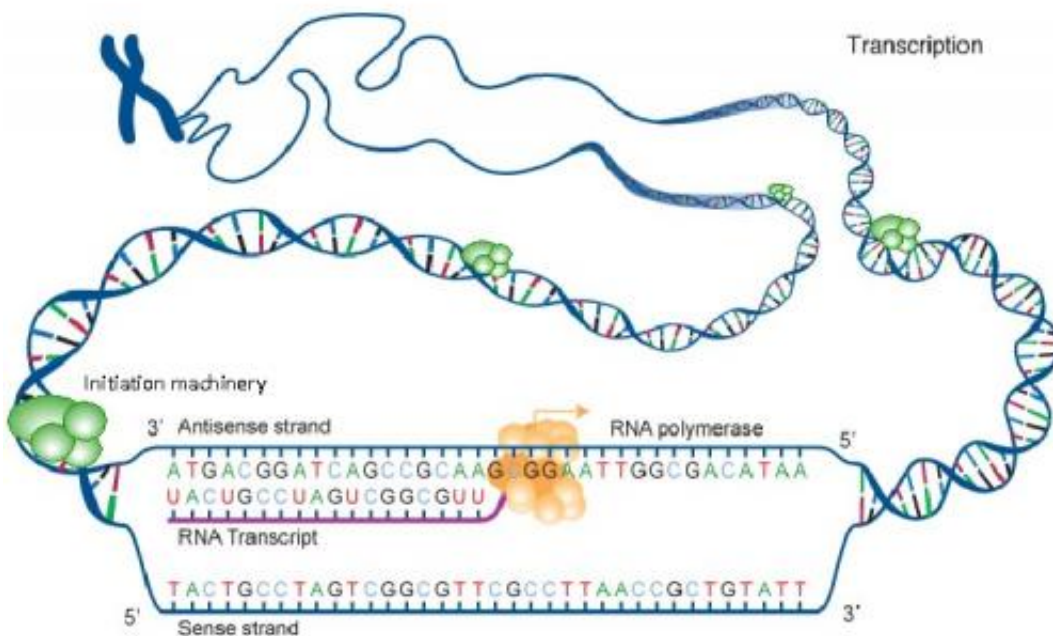


# Origins of genomic 'dark matter' discovered

September 18 2013



A duo of scientists at Penn State University has achieved a major milestone in understanding genomic "dark matter" -- called non-coding RNA. This "dark matter" is difficult to detect and no one knows exactly what it is doing or why it is there in our genome, but scientists suspect it may be the source of inherited diseases. This research achievement may help to pinpoint exactly where complex-disease traits reside in the human genome. This illustration shows, in the upper left corner, a chromosome -- a densely compressed package containing one long, continuous strand of DNA. The DNA is pervasively transcribed into RNA, but only a very small fraction of the RNA has the instructions (or codes) for making proteins. The green circles in this illustration represent places along the strand of DNA where transcription originates. New research led by B. Franklin Pugh of Penn State University shows that essentially all RNA, whether or not it codes for proteins, originates at the same types of locations along the strand of DNA. The

findings eventually may help to pinpoint exactly where complex-disease traits reside, since the genetic origins of many diseases reside outside of the coding region of the genome. Credit: National Institutes of Health and B. Franklin Pugh, Penn State University

A duo of scientists at Penn State University has achieved a major milestone in understanding how genomic "dark matter" originates. This "dark matter"—called non-coding RNA—does not contain the blueprint for making proteins and yet it comprises more than 95 percent of the human genome. The researchers have discovered that essentially all coding and non-coding RNA originates at the same types of locations along the human genome. The team's findings eventually may help to pinpoint exactly where complex-disease traits reside, since the genetic origins of many diseases reside outside of the coding region of the genome. The research, which will be published as an Advance Online Publication in the journal *Nature* on 18 September 2013, was performed by B. Franklin Pugh, holder of the Willaman chair in Molecular Biology at Penn State, and postdoctoral scholar Bryan Venters, who now holds a faculty position at Vanderbilt University.

In their research, Pugh and Venters set out to identify the precise location of the beginnings of transcription—the first step in the expression of [genes](#) into proteins. "During transcription, DNA is copied into RNA—the single-stranded [genetic material](#) that is thought to have preceded the appearance of DNA on Earth—by an enzyme called RNA polymerase and, after several more steps, genes are encoded and proteins eventually are produced," Pugh explained. He added that, in their quest to learn just where transcription begins, other scientists had looked directly at RNA. However, Pugh and Venters instead determined where along [human chromosomes](#) the proteins that initiate transcription of the non-coding RNA were located.

"We took this approach because so many RNAs are rapidly destroyed soon after they are made, and this makes them hard to detect," Pugh said. "So rather than look for the RNA product of transcription we looked for the 'initiation machine' that makes the RNA. This machine assembles RNA polymerase, which goes on to make RNA, which goes on to make a [protein](#)." Pugh added that he and Venters were stunned to find 160,000 of these "initiation machines," because humans only have about 30,000 genes. "This finding is even more remarkable, given that fewer than 10,000 of these machines actually were found right at the site of genes. Since most genes are turned off in cells, it is understandable why they are typically devoid of the initiation machinery."

The remaining 150,000 initiation machines—those Pugh and Venters did not find right at genes—remained somewhat mysterious. "These initiation machines that were not associated with genes were clearly active since they were making RNA and aligned with fragments of RNA discovered by other scientists," Pugh said. "In the early days, these fragments of RNA were generally dismissed as irrelevant since they did not code for proteins." Pugh added that it was easy to dismiss these fragments because they lacked a feature called polyadenylation—a long string of genetic material, adenosine bases—that protect the RNA from being destroyed. Pugh and Venters further validated their surprising findings by determining that these non-coding initiation machines recognized the same DNA sequences as the ones at coding genes, indicating that they have a specific origin and that their production is regulated, just like it is at coding genes.

"These non-coding RNAs have been called the '[dark matter](#)' of the genome because, just like the dark matter of the universe, they are massive in terms of coverage—making up over 95 percent of the [human genome](#). However, they are difficult to detect and no one knows exactly what they all are doing or why they are there," Pugh said. "Now at least we know that they are real, and not just 'noise' or 'junk.' Of course, the

next step is to answer the question, 'what, in fact, do they do?'"

Pugh added that the implications of this research could represent one step towards solving the problem of "missing heritability"—a concept that describes how most traits, including many diseases, cannot be accounted for by individual genes and seem to have their origins in regions of the genome that do not code for proteins. "It is difficult to pin down the source of a disease when the mutation maps to a region of the genome with no known function," Pugh said. "However, if such regions produce RNA then we are one step closer to understanding that disease."

**More information:** Genomic organization of human transcription initiation complexes, [DOI: 10.1038/nature12535](https://doi.org/10.1038/nature12535)

Provided by Pennsylvania State University

Citation: Origins of genomic 'dark matter' discovered (2013, September 18) retrieved 20 April 2024 from <https://phys.org/news/2013-09-genomic-dark.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>
--