

Decoding RNA-protein interactions

June 8 2018, by Raleigh McElvery



Researchers have taken an in-depth look at RNA-protein interactions, determining the sequence, structure, and context preferences of RNA-binding proteins in a high throughput manner. Credit: Daniel Dominguez/Burge Lab

Thanks to continued advances in genetic sequencing, scientists have identified virtually every A, T, C, and G nucleotide in our genetic code. But to fully understand how the human genome encodes us, we need to

go one step further, mapping the function of each base. That is the goal of the [Encyclopedia of DNA Elements \(ENCODE\)](#) project, funded by the National Human Genome Research Institute and launched on the heels of the Human Genome Project in 2003. Although much has already been accomplished—mapping protein-DNA interactions and the inheritance of different epigenetic states—understanding the function of a DNA sequence also requires deciphering the purpose of the RNAs encoded by it, as well as which proteins bind to those RNAs.

Such RNA-binding proteins (RBPs) regulate gene expression by controlling various post-transcriptional processes—directing where the RNAs go in the cell, how stable they are, and which proteins will be synthesized. Yet these vital RNA-[protein](#) relationships remain difficult to catalog, since most of the necessary experiments are arduous to complete and difficult to interpret accurately.

In a new study, a team of MIT biologists and their collaborators describes the binding specificity of 78 human RBPs, using a one-step, unbiased method that efficiently and precisely determines the spectrum of RNA sequences and structures these proteins prefer. Their findings suggest that RBPs don't just recognize specific RNA segments, but are often influenced by contextual features as well—like the folded structures of the RNA in question, or the nucleotides flanking the RNA-binding sequence.

"RNA is never naked in the cell because there are always proteins binding, guiding, and modifying it," says Christopher Burge, director of the Computational and Systems Biology Ph.D. Program, professor of biology and biological engineering, extramural member of the Koch Institute for Integrative Cancer Research, associate member of the Broad Institute of MIT and Harvard, and senior author of the study. "If you really want to understand post-transcriptional gene regulation, then you need to characterize those interactions. Here, we take advantage of deep

sequencing to give a more nuanced picture of exactly what RNAs the proteins bind and where."

MIT postdoc Daniel Dominguez, former graduate student Peter Freese, and current graduate student Maria Alexis are the lead authors of the study, which is part of the ENCODE project and appears in *Molecular Cell* on June 7.

A method for the madness

From the moment an RNA is born, it is coated by RBPs that control nearly every aspect of its lifecycle. RBPs generally contain a binding domain, a three-dimensional folded structure that can attach to a specific nucleotide sequence on the RNA called a motif. Because there are over 1,500 different RBPs found in the [human genome](#), the biologists needed a way to systematically determine which of those proteins bound to which RNA motifs.

After considering a number of different approaches to analyze RNA-protein interactions both directly in the cell (*in vivo*) and isolated in a test tube (*in vitro*), the biologists settled on an *in vitro* method known as RNA Bind-n-Seq (RBNS), developed four years ago by former Burge lab postdoc and co-author Nicole Lambert.

Although Lambert had previously tested only a small subset of proteins, RBNS surpassed other approaches because it was a quantitative method that revealed both low and high affinity RNA-protein interactions, required only a single procedural step, and screened nearly every possible RNA motif. This new study improved the assay's throughput, systematically exploring the binding specificities of more than 70 human RBPs at a high resolution.

"Even with that initial small sample, it was clear RBNS was the way to

go, and over the last three-and-a-half years we've been gradually building on this approach," Dominguez says. "Since a single RBP can select from billions of unique RNA molecules, our approach gives you a lot more power to detect the all those possible targets, taking into account RNA secondary structure and contextual features. It's an extremely deep and detailed assay."

First, the researchers purified the human RBPs, mixing them with randomly-generated synthetic RNAs roughly 20 nucleotides long, which represented virtually all the RNAs an RBP could bind to. Next, they extracted the RBPs along with their bound RNAs and sequenced them. With the help of their collaborators from the University of California at San Diego and University of Connecticut Health, the team conducted additional assays to glean what these RNA-protein interactions might look like in an actual cell, and infer the cellular function of the RBPs.

The researchers expected most RBPs to bind to a unique RNA motif, but to their surprise they found the opposite: Many of the proteins, regardless of structural class, seemed to prefer similar short, unfolded nucleotide sequence motifs.

"Human cells express hundreds of thousands of distinct transcripts, so you might think that each RBP would bind a slightly different RNA sequence in order to distinguish between targets," Alexis says. "In fact, one might assume that having distinct RBP motifs would ensure maximum flexibility. But, as it turns out, nature has built in substantial redundancy; multiple proteins seem to bind the same short, linear sequences."

Redundant motifs with distinct targets and functions

This overlap in RBP binding preference suggested to the scientists that there must be some other indicator besides the sequence of the motif

that signaled RBPs which RNA to target. Those signals, it turned out, stemmed from the spacing of the motifs as well as which nucleotide bases flank its binding sites. For the less common RBPs that targeted non-linear RNA sequences, the precise way the RNA folded also seemed to influence binding specificity.

The obvious question, then, is: Why might RBPs have evolved to rely on contextual features instead of just giving them distinct motifs?

Accessibility seems like one of the more plausible arguments. The researchers reasoned that linear RNA segments are physically easier to reach because they are not obstructed by other RNA strands, and they found that more accessible motifs are more likely to be bound. Another possibility is that having many proteins target the same motif creates some inter-protein competition. If one protein increases RNA stability and another decreases it, whichever binds the strongest will prevent the other from binding at all, enabling more pronounced changes in gene activity between cells or cell states. In other scenarios, proteins with similar functions that target the same motif could provide redundancy to ensure that regulation occurs in the cell.

"It's definitely a difficult question, and one that we may never truly be able to answer," Dominguez says. "As RBPs duplicated over evolutionary time, perhaps altering recognition of the contextual features around the RNA motif was easier than changing the entire RNA motif. And that would give new opportunities for RBPs to select different cellular targets."

This study marks one of the first in vitro contributions to the ENCODE Project. While in vivo assays reveal information specific to the particular cell line or tissue in which they were conducted, RBNS will help define the basic rules of RNA-protein interactions—so fundamental they are likely to apply across many cell types and tissues.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Decoding RNA-protein interactions (2018, June 8) retrieved 24 April 2024 from <https://phys.org/news/2018-06-decoding-rna-protein-interactions.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.