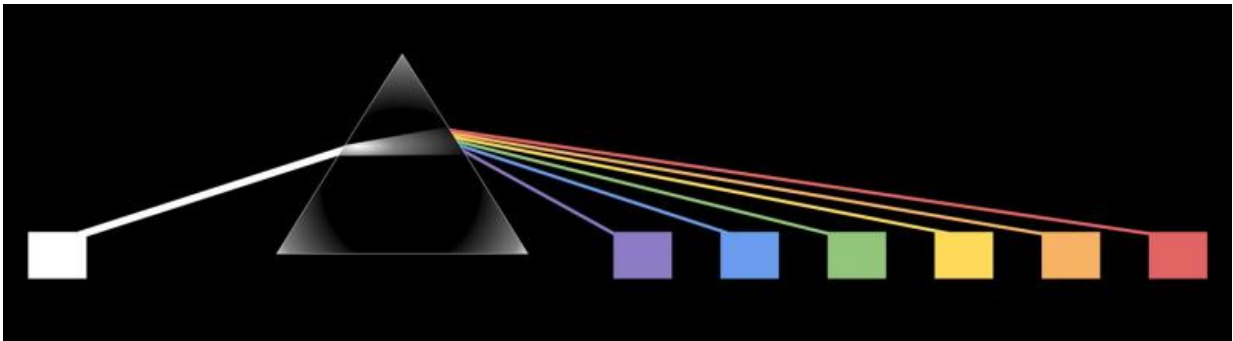


Researchers illuminate 'dark side' of the transcriptome

February 9 2016



A new method to detect, quantify and visualize RNA splicing helps to shed light on the "dark side" of the transcriptome: complex splicing variations that have been known to exist but have not been well studied. Credit: Matthew Gazzara in the Barash lab, Perelman School of Medicine, University of Pennsylvania

A new way of mapping the "transcriptome"—the collection of RNA read-outs that are expressed by a cell's active genes—has been devised by researchers from the Perelman School of Medicine at the University of Pennsylvania. RNA is both the molecular bridge between DNA and the production of proteins that carry out the functions of life and the molecular toolbox that collectively helps those proteins do their work. As such, RNA exists in a variety of forms, each with a particular role and purpose, not all of which are fully understood.

Using the new method to shed additional light on the role of RNAs in

cells, the team identified RNA variants in mammals that had been largely invisible to previous techniques. The researchers also demonstrated that these "dark" variations in RNA are strikingly common in mammalian cells and likely have roles in gene regulation across tissues, development, and in human diseases. The team plans to perform the analyses using the now-free software to interrogate aberrant cells in neurodegenerative disorders, cancers, and other illnesses.

"It's very exciting for us, and I think for the research community in general, among other reasons, because we can now go back through the vast amount of existing transcriptome data, knowing that new and important things will emerge," said senior author Yoseph Barash, PhD, an assistant professor of Genetics. Barash is also a senior fellow at the Penn Institute for Biomedical Informatics.

The report by Barash and his team, which included co-lead authors Jorge Vaquero-Garcia, Alejandro Barrera, and Matthew R. Gazzara, all research staff at Penn, was published online in *eLife* this week.

An Incomplete Picture

Barash's laboratory has been devoted principally to the study of RNA transcripts and their variations by using machine learning and computational modeling. One major mechanism of variation, called alternative splicing, has been known to scientists since the 1970s. When a protein-coding gene is first transcribed into RNA, cellular machinery slices the fresh RNA transcript into segments. It then discards non-protein-coding segments (introns) and splices back together protein-coding segments (exons) into a finished transcript of messenger-RNA, which is later translated into a protein.

Sometimes, depending on circumstances in the cell, the [splicing machinery](#) deliberately omits splicing in one or more exons, and the

result is a shorter messenger-RNA transcript, which in turn codes for a different form of the protein. In this way, a single gene may code for multiple forms of the same protein, each of which has its own distinct biological role, for example working only in one set of cell types or only during fetal development. Splicing patterns that deviate from normal are known to contribute to many diseases.

A long-standing problem for biologists has been that they have no easy, error-free way to identify and quantify all the distinct messenger-RNA splice variants in a sample. Modern RNA-sequencing technology (RNA-seq) is a powerful scientific tool but mostly yields the sequences only of fragments of messenger RNAs. Those fragment sequences essentially have to be stitched back together, with the aid of sophisticated software and existing RNA databases, to get a complete picture of the transcriptome. But that picture isn't necessarily a complete one.

"The reads from RNA-seq are sparse and also short compared to actual messenger-RNA transcripts, so you don't directly know what transcripts those reads came from," Barash said. "Therefore you also don't directly know the abundance of those transcripts."

A New View of the Transcriptome

The new approach devised by Barash and his team begins with the mapping of what they call local splice variations (LSVs)—essentially the variable junctions between exons, which are detectable sequences that span more than one exon.

"These are places where the splicing machinery of a cell makes a choice about which exon is spliced to another," Barash said.

The team developed software to generate LSV maps from RNA-seq data and combine those data with existing RNA databases to yield pictures

that include ordinary, known splice variants, as well as complex splice variants that other methods fail to detect.

To gauge the importance of the hitherto-unseen part of the transcriptome, the team used the new MAJIQ software (Modeling Alternative Junction Inclusion Quantification) to analyze RNA-seq data from a variety of species including lizards, mice, and humans. The analysis revealed that complex splice variants are much more common than previously thought—comprising, for example, about 37 percent of the transcriptome variations in human samples.

"These variations are a bit like the dark side of the moon," said Barash. "They were known to exist, yet we lacked the ability to shine a light on them—and now they turn out to make up a third of the variations in human messenger RNAs."

The complex splicing variants detected with MAJIQ included a highly conserved, yet previously unreported variant from the gene *Ptbp1*, which is known to be critical for proper brain development. Further analysis suggested that the newly discovered variant is involved in controlling the expression of *Ptbp1* after birth by introducing a "poison exon" which marks the transcript to subsequent degradation.

Another complex variant detected with MAJIQ, from the human, synapse-related gene *CAMK2D*, turned out to be expressed about 40 percent less in brain tissue from Alzheimer's patients, compared to controls. The team later found a similar drop in a second, larger RNA-seq dataset, also from Alzheimer's disease cases. Overall, the team identified approximately 200 cases of altered splicing in Alzheimer patients that were reproducible in the two independent studies.

"We think that findings like those are just the tip of the iceberg," Barash said. He and his colleagues now plan to do further MAJIQ-based

investigations of complex splice variants in other disorders. However, Barash emphasizes that the MAJIQ software package will now be freely available for other academic researchers, who will all be able to use it in their own ways. In addition to MAJIQ, the team also has produced a complementary software package, VOILA, which enables researchers to visualize the complex splice variants detected by MAJIQ.

Provided by University of Pennsylvania School of Medicine

Citation: Researchers illuminate 'dark side' of the transcriptome (2016, February 9) retrieved 27 April 2024 from <https://phys.org/news/2016-02-illuminate-dark-side-transcriptome.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.