The 98%: Proteomics reveals proteins made from ‘noncoding’ DNA

by Nancy Fliesler on December 1, 2014

Probing the genome's 'dark side' could change our view of biology.

Vast chunks of our DNA—fully 98 percent of our genome—are considered “non-coding,” meaning that they’re not thought to carry instructions to make proteins. Yet we already know that this “junk DNA” isn’t completely filler. For example, some sequences are known to code for bits of RNA that act as switches, turning genes on and off.

New research led by Judith Steen, PhD, and Gabriel Kreiman PhD, of Boston Children’s Hospital’s Proteomics Center and Neurobiology program, goes much further in mapping this “dark side” of the...
In a report published last month in *Nature Communications*, they describe a variety of proteins and peptides (smaller chains of amino acids) arising from presumed non-coding DNA sequences. Since they looked in just one type of cell—neurons—these molecules may only be the tip of a large, unexplored iceberg and could change our understanding of biology and disease.

**Overturning genomic dogma**

“The central dogma in textbooks rests upon the ‘canonical’ gene structure, which is highly annotated and well-characterized in public databases,” says Steen. “It has clearly denoted exons (coding sequences) interspersed with introns (noncoding sequences), all with clear beginning and end points.”

Some 22,000 proteins are known to arise from this canonical genome. But what proteins are cells actually making? Steen, Kreiman and colleagues set out to create a systematic tally.

As their test case, they turned to mouse neurons grown in a dish. They looked at both the proteome (the proteins being made by the cell) and the cell’s transcriptome—the complete set of RNA molecules it was making, transcribed from the DNA sequence. Because a cell’s transcriptome and proteome are always changing in response to its environment, they sampled the neurons at different time points (0, 1, 2, 3 and 6 hours) after stimulating the cells chemically.

Starting with the proteome, the researchers used the technique of mass spectrometry to digest the proteins into peptides and blow them apart into fragments. By sequencing the amino acids in those fragments, they could identify and quantify the peptides the cell was making, yielding a list of some 1.1 million peptide sequences, or spectra.

**Uncharted territory**

Eighty percent of these spectra were completely off the map—falling outside the known, annotated proteome. “Nobody had sequenced and quantified these non-canonical peptides using mass spectrometry showing their regulation,” Steen says.
An analysis of 250 previously unrecognized peptides being made by mouse neurons. Each row represents one peptide at four different time points after stimulation of the neuron. The colors indicate how frequently the peptide was detected.

Steen, Kreiman and colleagues then made a second list: Using transcriptome data from each time point, they calculated all proteins or peptides that the neurons could potentially make using these RNA sequences.

“No one had systematically looked at these RNAs to see if they matched the proteome,” Steen says. “Now, because deep sequencing is so cheap, you can have a customized transcriptome for anything.”

Next came a massive data crunch to look for matches, using a computational algorithm developed at Boston Children’s. After filtering out “noise” in the system, 1,584 peptides came up consistently enough that they were likely to have some biological relevance.

**New players in biology?**

Steen, Kreiman and colleagues then did a deep dive to validate 250 of these peptides. In so doing, they uncovered a variety of interesting, formerly unknown peptides and proteins, some made only at certain time points.

Among the garden of oddities:

- proteins that resemble other known proteins, but with different “stop” and “start” points
- 79 “intron inclusion events”—alternate forms of known proteins that incorporated a
“noncoding” sequence, or intron, changing their structure
- 66 antisense proteins, made from the chain of double-stranded DNA that normally isn’t used to make proteins
- proteins made from “pseudogenes”—stretches of DNA that resemble genes but are frame-shifted, causing a cascade of transcription errors
- very small peptides unrelated to any known gene in the genome

While some of these proteins and peptides may simply be mistakes, perhaps contributing to disease, others could have legitimate biological roles. In the case of neurons, Steen thinks that some might be important in fine-tuning brain function—helping to strengthen certain synapses, or perhaps secreted at the synapse to alter signaling. “A lot of these proteins go up in response to stimulation,” she notes.

The study's workflow for finding novel peptides

A similar dive into other tissue types could uncover whole other sets of novel proteins. The DNA sequences underlying those in this study seem to be selected for in evolution, so the researchers speculate that collectively, these proteins may help organisms adapt to changing conditions. “If, under stress, you can translate a proteome that can allow you to survive that stress, you have a lot more of the genome to pick from,” says Steen.
Interestingly, this work—which also involved first authors Sudhakaran Prabakaran, PhD, in the Department of Systems Biology at Harvard Medical School, Martin Hemberg, PhD, in Kreiman’s Neurobiology lab at Boston Children’s and Ruchi Chauhan, MSc, of Boston Children’s Proteomics Center—began as a side project. The researchers plan to continue their investigations and make their computational algorithms publicly available.

“We’re at a hospital, so we often study specific proteins involved in disease,” Steen says. “But this research gives us the opportunity to really explore something different—basic science that lays the foundation for new areas of research. That opportunity doesn’t come often.”
A garden of proteins made by "noncoding" DNA

@nansonaham

vectorblog.org/2014/12/the-98-proteomics-reveals-proteins-made-by-noncoding-dna

#proteomics #transcriptome #genomics @NatureComms

Spleen-on-a-chip for cleansing inflammation-triggering pathogens and toxins from the blood of patients with sepsis.

vectorblog.org/2014/11/a-bios

Retweeted by oriano perata

Expand

Animalusa

@Animilusa

Spleen-on-a-chip for cleansing inflammation-triggering pathogens and toxins from the blood of patients with sepsis.

vectorblog.org/2014/11/a-bios

Expand

Compose new Tweet...

Thriving blog posts

- “The right doctors and the right game plan” help teen with IBD
- Healthy Thanksgiving Recipes
- Lammily’s message: Average is beautiful
- Caring for Christopher’s Crohn’s disease close to home
- Looking for the flu shot? We have you covered
'Junk' DNA does code for proteins, overturning biological dog...

http://vectorblog.org/2014/12/the-98-proteomics-reveals-protei...
'Junk' DNA does code for proteins, overturning biological dog...

http://vectorblog.org/2014/12/the-98-proteomics-reveals-protei...