

# 'Deep proteome' project provides atlas for human complexity

March 23 2023

---



Joshua Coon, professor of biomolecular chemistry, UW-Madison and investigator, Morgridge Institute for Research. Credit: Morgridge Institute for Research

A major puzzle of biology is that while the human genome contains

roughly 20,000 genes, many comparatively primitive organisms—including the universally-studied worm *C. elegans*—have almost the same number of genes.

If not genes alone, what accounts for that quantum leap in complexity between the two species?

One answer may lie in the field of proteomics, which focuses on identifying and defining the [protein](#) building blocks that make up an individual cell. Rather than one gene coding for one protein with one purpose, human genes act like powerful compressed files, where a single gene can code for hundreds of distinct proteins that each perform precise functions in the body.

As many as 95% of human [genes](#) have this capability, known as [alternative splicing](#).

A [new study released today](#) (March 24) in the journal *Nature Biotechnology* outlines a meta-scale approach to quantifying the human proteome and the massive number of protein variants produced by the human body. Proteomics is a cornerstone of biology and a precursor to understanding how protein dysfunction contributes to disease.

Led by Joshua Coon, professor of biomolecular chemistry at the University of Wisconsin-Madison and investigator at the Morgridge Institute for Research, the research team developed a method called "deep proteome sequencing" that offers unprecedented characterization of the proteins that show up in standard proteomics experiments.

The project used six different human cell types and six proteases—enzymes that break down proteins into smaller fragments (peptides) that serve as the raw material for detection in the experiment. The team then analyzed the peptides by employing different methods of

mass spectrometry, the leading technology for identifying proteins.

The researchers identified more than 1 million peptides from 17,717 different protein groups. From these data, they were able to detect approximately 80% of the sequences of all individual proteins within those samples—a vast increase over standard approaches that sequence only ~20% of proteins.

Achieving this more complete picture is the Holy Grail of proteomics.

"In the field of mass spectrometry and proteomics, there has always been a goal of detecting all proteins that are present in a sample, then fully sequencing all the individual proteins present," Coon says. "But we really haven't been detecting the whole protein, just small pieces of it."

"Data generated from this study represent the deepest proteomics map collected to date," Coon adds. "These methods and resources lay the foundation for comprehensive mapping of protein diversity and are expected to catalyze future research efforts."

The research team created an online, publicly available resource called [deep-sequencing.app](#), in which scientists can query any gene and examine the corresponding peptides and protein modifications that are associated with that gene.

The project, primarily sponsored by the National Institutes of Health, received major input from research groups at the Max Planck Institute of Biochemistry in Germany, the University of Toronto in Canada, and the Garvin Institute in Australia. Pavel Sinitcyn, a scientist at Max Planck Institute and now a postdoc in the Coon Lab and Morgridge Interdisciplinary Postdoctoral Fellow, led the massive data analysis work for a project that generated more than five terabytes of data over 10 years. At Toronto, investigator Benjamin Blencowe provided expertise

on alternative splicing.

Scientists have disagreed about how much alternative splicing contributes to protein diversity, primarily because the process is very hard to detect at the protein level. The Coon Lab project is the first to specifically target evidence of splicing events in the actual proteins. They found that most of the alternative splicing detected at the RNA stage of gene expression is also present in the proteins.

"I think this knowledge tells us that, yes, these ideas about splicing—allowing the cell to have this repertoire of proteins for distinct purposes—are now validated. This is the first time we've been able to measure it and prove it," Coon says.

While at Max Planck Institute, Sinitcyn worked in the lab of Jurgen Cox, a world-leading bioinformatics group in the field of computational mass spectrometry. Sinitcyn developed software solutions to be able to detect evidence of single amino acid variants and alternative splicing in the mass spec data.

"We are dealing with more than five terabytes of data from heterogeneous sources, so our first problem was to find a way to account for the high probability of generating false positives," says Sinitcyn. "But the second problem, the exciting one, was actually to demonstrate how relevant this dataset could be for important biological questions."

**More information:** Jürgen Cox, Global detection of human variants and isoforms by deep proteome sequencing, *Nature Biotechnology* (2023). [DOI: 10.1038/s41587-023-01714-x](https://doi.org/10.1038/s41587-023-01714-x).  
[www.nature.com/articles/s41587-023-01714-x](https://www.nature.com/articles/s41587-023-01714-x)

Provided by Morgridge Institute for Research

Citation: 'Deep proteome' project provides atlas for human complexity (2023, March 23)  
retrieved 10 April 2024 from

<https://phys.org/news/2023-03-deep-proteome-atlas-human-complexity.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>
--