

# Extensive cataloging of human proteins uncovers 193 never known to exist

May 28 2014

---



An artistic interpretation of the human proteome. Credit: Corinne Sandone, MA, CMI and Jennifer Fairman, MA, CMI, Department of Art as Applied to Medicine, Johns Hopkins University

Striving for the protein equivalent of the Human Genome Project, an international team of researchers has created an initial catalog of the

human "proteome," or all of the proteins in the human body. In total, using 30 different human tissues, the team identified proteins encoded by 17,294 genes, which is about 84 percent of all of the genes in the human genome predicted to encode proteins.

In a summary of the effort, to be published May 29 in the journal *Nature*, the team also reports the identification of 193 novel proteins that came from regions of the genome not predicted to code for proteins, suggesting that the [human genome](#) is more complex than previously thought. The cataloging project, led by researchers at The Johns Hopkins University and the Institute of Bioinformatics in Bangalore, India, should prove an important resource for biological research and medical diagnostics, according to the team's leaders.

"You can think of the [human body](#) as a huge library where each [protein](#) is a book," says Akhilesh Pandey, M.D., Ph.D., a professor at the McKusick-Nathans Institute of Genetic Medicine and of biological chemistry, pathology and oncology at The Johns Hopkins University and the founder and director of the Institute of Bioinformatics. "The difficulty is that we don't have a comprehensive catalog that gives us the titles of the available books and where to find them. We think we now have a good first draft of that comprehensive catalog."

While genes determine many of the characteristics of an organism, they do so by providing instructions for making proteins, the building blocks and workhorses of cells, and therefore of tissues and organs. For this reason, many investigators consider a catalog of human proteins—and their location within the body—to be even more instructive and useful than the catalog of genes in the human genome.

Studying proteins is far more technically challenging than studying genes, Pandey notes, because the structures and functions of proteins are complex and diverse. And a mere list of existing proteins would not be

very helpful without accompanying information about where in the body those proteins are found. Therefore, most protein studies to date have focused on individual tissues, often in the context of specific diseases, he adds.

To achieve a more comprehensive survey of the proteome, the research team began by taking samples of 30 tissues, extracting their proteins and using enzymes like chemical scissors to cut them into smaller pieces, called peptides. They then ran the peptides through a series of instruments designed to deduce their identity and measure their relative abundance.

"By generating a comprehensive human protein dataset, we have made it easier for other researchers to identify the proteins in their experiments," says Pandey. "We believe our data will become the gold standard in the field, especially because they were all generated using uniform methods and analysis, and state-of-the-art machines."

Among the proteins whose data patterns have been characterized for the first time are many that were never predicted to exist. (Within the genome, in addition to the DNA sequences that encode proteins, there are stretches of DNA whose sequences do not follow a conventional protein-coding gene pattern and have therefore been labeled "noncoding.") The team's most unexpected finding was that 193 of the proteins they identified could be traced back to these supposedly noncoding regions of DNA.

"This was the most exciting part of this study, finding further complexities in the genome," says Pandey. "The fact that 193 of the proteins came from DNA sequences predicted to be noncoding means that we don't fully understand how cells read DNA, because clearly those sequences do code for proteins."

Pandey believes that the [human proteome](#) is so extensive and complex that researchers' catalog of it will never be fully complete, but this work provides a solid foundation that others can reliably build upon.

**More information:** Paper: [dx.doi.org/10.1038/nature13302](https://doi.org/10.1038/nature13302)

Provided by Johns Hopkins University School of Medicine

Citation: Extensive cataloging of human proteins uncovers 193 never known to exist (2014, May 28) retrieved 23 April 2024 from <https://phys.org/news/2014-05-extensive-human-proteins-uncovers.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.