PROTEOMICS

The human proteome takes the spotlight

Two papers report mass spectrometry–based draft maps of the human proteome and provide broadly accessible resources.

For years, members of the proteomics community have been trying to garner support for a large-scale project to exhaustively map the normal human proteome, including identifying all post-translational modifications and protein-protein interactions and providing targeted mass spectrometry assays and antibodies for all human proteins. But a lack of consensus on how to exactly define the proteome, how to carry out such a mission and whether the technology is ready has not so far convinced any funding agencies to fund on such an ambitious project.

"I have been very impatient about this," says Akhilesh Pandey of Johns Hopkins University School of Medicine in Baltimore, the principal investigator of one of two recent reports describing mass spectrometry–based draft maps of the human proteome. While listening to the debate during a session about a possible Human Proteome Project at a conference of the Human Proteome Organization several years ago, he realized that his lab was in a good position to perform one of the human proteome project's goals: finding evidence for all human protein-coding genes using mass spectrometry. As the founder of the Institute of Bioinformatics in Bangalore, India, his team also had the resources and bioinformatics know-how to carry out such a study.
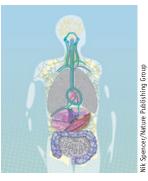
Around the same time, at the Technische Universität München in Germany, Bernhard Kuster was realizing that existing tools for managing proteomics data were not meeting his laboratory's needs for mining and cross-referencing. After reading an article about in-memory database computing, he contacted the German business operations software company SAP to see whether this might provide a solution for handling proteomics data. This resulted in the creation of ProteomicsDB, which contains useful, computationally efficient tools for analyzing big data. "We then thought, 'What is a potentially good illustration for the utility of such a database?'" says Kuster. "We very quickly got to the idea, 'Why don't we try to put together the human proteome?'"

The two groups took slightly different strategies towards this common goal. Pandey's lab examined 30 normal tissues, including adult and fetal tissues, as well as primary hematopoietic cells, subjecting them to comprehensive, label-free quantitative mass spectrometry analysis and analyzing the data with a stringent bioinformatics pipeline (Kim *et al.*, 2014). In total, Pandey and colleagues detected about 84% of the roughly 20,000 annotated human protein-coding genes. They also created a biologist-friendly resource called the Human Proteome Map, which allows users to explore protein expression across tissues.

Kuster's team took a different tack (Wilhelm *et al.*, 2014). "In talking to colleagues over the last few years, most people agreed that we had collectively probably seen the human proteome already, just that we hadn't put it together," says Kuster. His lab amalgamated publicly available raw data sets and those from colleagues, which together made up about 60% of the data that they analyzed using their own rigorous bioinformatics pipeline. The other 40% of the data came from new quantitative mass spectrometry experiments, in which they profiled 60 tissues, 13 body fluids and 147 cell lines. In total they obtained mass spectrometry evidence for about 92% of predicted protein-coding genes.

Both groups found evidence for many proteins that had not been previously observed by mass spectrometry, complementing genome annotation efforts. The data also

Two groups provide mass spectrometry evidence for ~90% of the human proteome.

include some surprising findings. For example, Kuster's team found protein evidence for 430 long intergenic noncoding RNAs, which have been thought not to be translated into protein. Pandey's team refined the annotations of 808 genes and also found evidence for the translation of many noncoding RNAs and pseudogenes.

Obtaining evidence for the last roughly 10% of proteins not detected in these studies, will not be easy. Pandey suggests that proteomics researchers will need to look at highly specialized tissues, such as parts of the eye or the nasal epithelium, to find the missing proteins. Kuster believes that the lack of mass spectrometry evidence for some predicted proteins suggests that these proteins may have been recently "retired" by evolution; determining this with certainty will require focused efforts. To facilitate this, ProteomicsDB has an "adopt-a-protein" feature to entice researchers to follow up on predicted proteins that still elude detection.

The Human Proteome Map and ProteomicsDB will be valuable resources for the entire biological community. The fact that two independent data sets are available also increases the impact of the work. "The community, they have cross-validation," says Pandey. "This is beautiful." The two groups were unaware of how far each other's efforts had progressed but were happily surprised to see that the papers were submitted to *Nature* at roughly the same time and published together. Now they are collaborating to analyze the collective data.

Allison Doerr

RESEARCH PAPERS
Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).