PROTEOMICS

Exploring how the organelles are organized

New work from several laboratories describes ongoing efforts to explore a new proteomic frontier—mapping and indexing the protein content of the cellular organelles.

The cell is more than just a sack of proteins, even if it can sometimes be convenient to think of it that way. Every cell contains dozens of functionally distinct compartments, and it is well established that the Golgi, mitochondria and other organelles each have unique proteomic profiles. But until recently, efforts to cleanly isolate individual organelle types and analyze their protein content have been met with limited success.

To improve the quality of their organelle analyses, Max Planck Institute researcher Matthias Mann and his colleagues developed a technique called protein correlation profiling (PCP), in which mass-spectrometric data obtained from gradient-fractionated cell extracts are compared against proteins known to localize to specific organelles, allowing researchers to confidently map organelles to particular fractions. Using PCP with extracts obtained from mouse liver, Mann and his colleagues obtained data that allowed them to confidently assign nearly 1,500 proteins to 10 cellular organelles (Foster *et al.*, 2006). These data not only took Mann's multinational team a step closer toward assembling a better cellular proteomic map, but it also proved useful for the preliminary identification of organelle-specific *cis*-regulatory elements and the tentative assembly of networks of coregulated genes.

Sometimes dividing the cellular proteome into compartments can also be a pragmatic decision. "The fractionation that we did to enrich our organelles, to some degree, was to get around technical limitations associated with mass spectrometry," explains Andrew Emili of the University of Toronto. "[With] a crude extract, we'd probably identify far fewer

BIOINFORMATICS

MORE THAN JUST 'DOING THE MATH'

Two new articles show how computational tools continue to move beyond mere sequence-based bioinformatic analysis into more advanced arenas of prediction, deduction and network building.

As interest grows in the still-young field of organelle proteomics, inventive *in silico* strategies are essential if researchers are to construct accurate hypotheses from mountains of raw data. Computational and experimental approaches have a symbiotic relationship, explains Vamsi Mootha of the Broad Institute of MIT and Harvard University: "They complement each other—you can't tease them apart. In order to support high-quality computational approaches, you need to begin with high-quality datasets."

Mootha recently illustrated this relationship, describing a 'smarter' *in silico* approach for identifying mitochondrial proteins (Calvo *et al.*, 2006). Earlier strategies have largely emphasized motif-based predictors, but the Mootha group's 'Maestro' program takes a more holistic approach, integrating eight different 'predictors', based on both structural and experimental data, to generate scores predicting the likelihood of mitochondrial localization. After training Maestro with a 'gold standard' set of known positive and negative controls, Mootha's team confirmed hundreds of known mitochondrial proteins and confidently identified nearly 500 that were previously unidentified. Notably, Maestro also proved capable of tentatively identifying genes associated with several human mitochondrial diseases, including at least one that had not been previously recognized as mitochondrial.

Søren Brunak, of the Technical University of Denmark, and his colleagues recently described an alternative computational tool for organelle proteomics and used *in silico* methods to predict protein complexes in the nucleolus (Hinsby *et al.*, 2006). They began by constructing an interaction atlas for a collection of known human nucleolar proteins based on publicly available interaction data and then subjected each putative complex to component-by-component computational analysis based on dozens of protein features, to predict the likelihood of nucleolar localization. Using conservative parameters, Brunak's team confidently predicted 15 nucleolar complexes; several of them were expected, but many were rather surprising from a functional standpoint (for example, proteins involved in DNA repair). This work also revealed 11 new nucleolar proteins, which were confirmed by experimental data from Brunak's collaborator, Matthias Mann, in a process the two call 'reverse proteomics'.

Both groups benefited from smart use of existing data sets, and Mootha suggests that more data should mean more options for future computational efforts. "More generally," he says, "if we get different types of really good functional genomics data sets, it might be possible to reconstruct all organelles *in silico*." Both approaches, however, also illustrate the value of using conservative cutoffs to eliminate 'junk' data and to ensure confidence in one's analysis. "Mapping something often means to throw a lot of information away, and this is, I think, what we try to do with our work," says Brunak. "We would rather not waste the precious time of the experimentalists!"

Michael Eisenstein

RESEARCH PAPERS

Calvo, S. *et al.* Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat. Genet.* **38**, 576–582 (2006).

Hinsby, A.M. *et al.* A wiring of the human nucleolus. *Mol. Cell* **22**, 285–295 (2006).