

A man in a yellow and white striped shirt is holding a large red ball. He is looking towards a woman in a white lab coat who is smiling. They are in a museum exhibit with various scientific models, including a DNA double helix and a cell diagram. The background is dark blue with glowing purple and red spheres.

The house guest
that stayed



Lars Juhl Jensen, Fabiana Perocchi,
Lars Steinmetz and Julien Gagneur

Foreground: A schematic diagram of the 876 proteins found in the mitochondria and the 9780 connections building up the mitochondrial network. Background: an electron microscope image of a mitochondrion.

ABOUT TWO BILLION YEARS AGO, ancient bacteria were engulfed by free-living, single cells and formed symbiotic relationships with them, living as bacterial colonies inside those cells. The cell came to depend on its mitochondria, as those colonies became known, mainly to produce its energy but also to perform other functions such as regulating programmed cell death – a natural part of the cell's life cycle. Because of that dependency, mitochondria now contribute to a large burden of human disease, which is why Lars Steinmetz would like to understand how they work.

He studies mitochondria in yeast, because yeast is a eukaryotic organism, its cells have a nucleus, placing it in the same branch of life as humans, and its mitochondria are similar to those of other eukaryotes, including humans. His goal is to construct a map of all the interactions that take place between mitochondrial proteins in a cell, and to understand why disruptions in those networks lead to disease. He takes a systems biology approach to the problem, drawing on information generated by different molecular biological techniques, across a wide range of organisms.

“The mitochondria are particularly attractive for that purpose, because their function has been highly conserved through evolution,” says Lars. “Not only can we test technologies in yeast, but also we can transfer some of what we learn directly to humans. The prediction of human disease genes is one example of that.”

But mapping mitochondrial interactions is not easy, for one simple reason: not all mitochondrial proteins are encoded by mitochondrial DNA (mtDNA). Over the course of evolution, the mitochondria transferred many of their genes to the cell's nucleus, or simply lost them. The result is that although the yeast *Saccharomyces cerevisiae*

has around 800 mitochondrial proteins, only eight of these are encoded by mtDNA – the others are encoded by nuclear DNA. In other words, the mitochondria must recruit proteins from other parts of the cell to perform their functions. Just over half of the mitochondrial proteins in yeast are known, but that's not enough to build a picture of mitochondrial function.

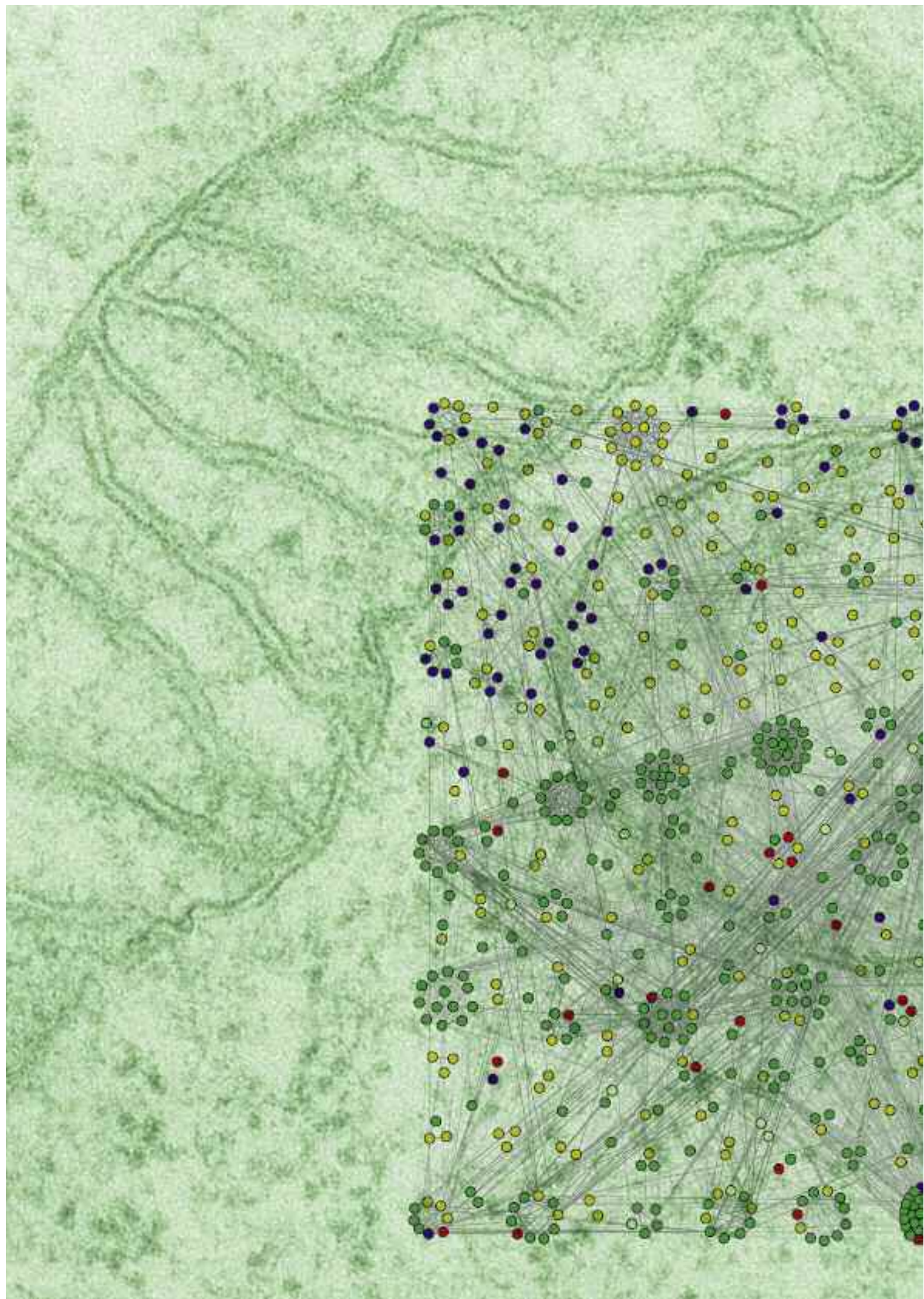
Lars wants the global view. He made a first stab at getting it in 2002, while still at Stanford University in the USA, where he was involved in constructing a set of deletion strains of *S. cerevisiae*. In each strain, one gene was deleted and replaced by an identical ‘dummy’ cassette of DNA and a molecular barcode – a tag that meant the strain could be identified by DNA microarray, or gene chip, analysis. By growing all the strains together and varying environmental parameters – the food source, for example – Lars and his colleagues could see how they fared compared with one another, and measure the contribution of each deleted gene to the strain's fitness, or ability to reproduce.

Using this method, they identified almost 500 genes whose deletion impaired mitochondrial function, around half of which were new. They then switched techniques, to proteomics, purifying

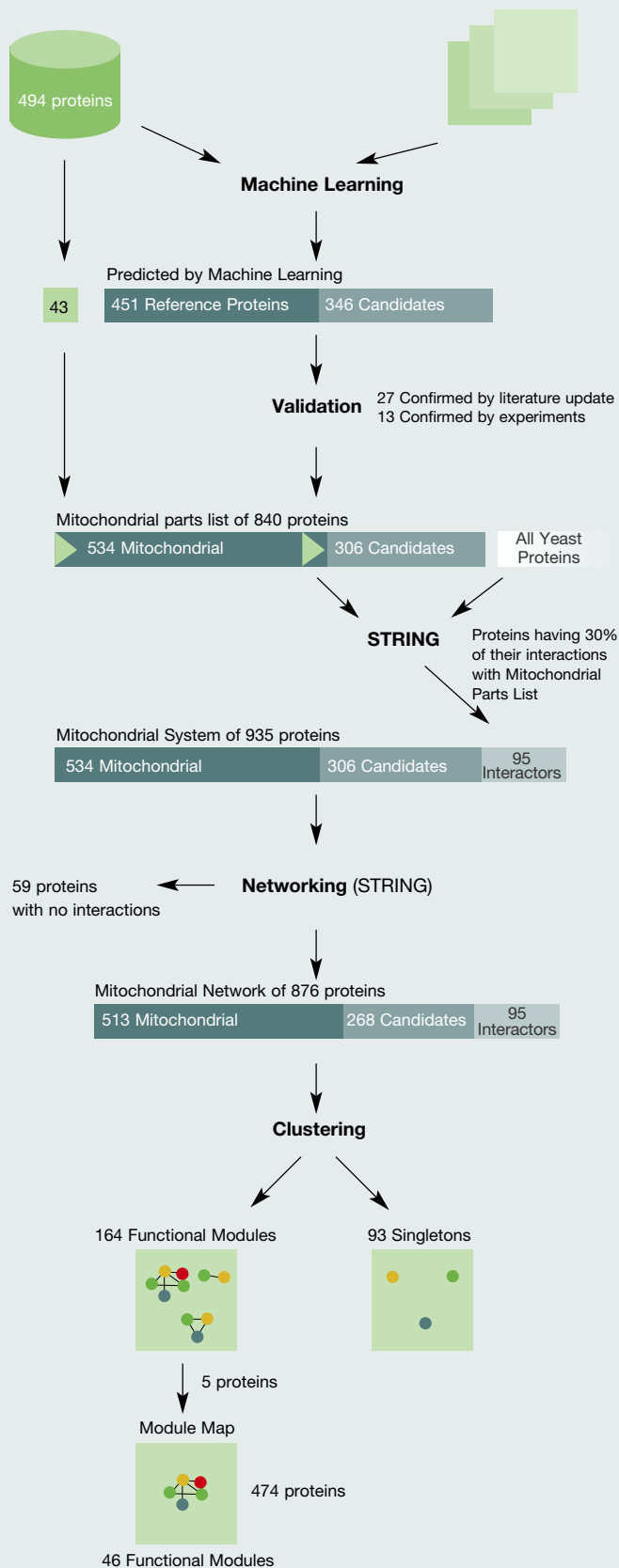
all the proteins in the organelle and using mass spectrometry to identify them by the mass of their components. Again, they came up with around 500 mitochondrial proteins. When they compared these with the 500 identified by the deletion screen, however, they were surprised. “Here were two global datasets in which we found significant enrichment of mitochondrial components, and yet they overlapped for less than 30 percent of the proteins,” Lars says.

This set him thinking. “The deletion screen identifies components which, when knocked out, produce a mildly defective mutant, but it might not show up redundant fac-

*Over the course of evolution, the mitochondria transferred many of their genes to the cell's nucleus, or simply lost them. Although the yeast *Saccharomyces cerevisiae* has around 800 mitochondrial proteins, only eight of these are encoded by mtDNA.*



Reference Dataset (MitoP2) 24 High-Throughput Datasets



Schematic summary of the integrated systems biology approach used to create a map of all mitochondrial interactions.

tors – factors whose loss can be compensated for by other enzymes, for example. The proteomics approach, on the other hand, identifies all the proteins which are found in the organelle, but it’s biased towards those that are expressed there in larger quantities. It occurred to us that these different types of information might be complementary, and that there might be an advantage in combining them.”

They generated one more dataset themselves, by identifying the fraction of yeast mitochondrial genes which are transcribed into messenger RNA – a predictor of those which eventually generate proteins – and then they went mining the literature for large-scale datasets that other groups had published. They came up with 24 in total, which had been derived using different high-throughput techniques, including computational models that predicted mitochondrial proteins from the signalling pathways associated with them, and phylogenetic studies that inferred yeast mitochondrial components from those known to exist in other species. “Some of these produced very powerful datasets, but none was able to capture all the mitochondrial proteins on their own,” he says.

At that point the problem became a computational one, and he called in Lars Jensen, a member of Peer Bork’s computational biology group in EMBL Heidelberg, to work with his PhD student Fabiana Perocchi in integrating the 24 datasets and extracting some meaningful information from them. Using computer programs which are capable of ‘learning’ from the information they process, they came up with a list in which every gene in the yeast genome was ranked according to the probability that its protein product was linked with the mitochondria.

To their delight, they found that almost all of the genes encoding the 500 mitochondrial proteins they already knew about were near the top of the list. Their computational method seemed to work. To test just how reliable it was, however, they took a handful of proteins that the computer had ranked highly, but which were new to them, and put them through some experiments. In collaboration with Uwe Ahting and Holger Prokisch of the Institute of Human Genetics at Munich’s Technical University in Germany, they tested the ability of live yeast

mitochondria to import these proteins when mixed with them in a dish. This screening method showed that 13 out of 16 of their candidate proteins were indeed imported, confirming their status as mitochondrial proteins – not bad considering the screen itself is not 100 % sensitive, says Lars Steinmetz.

Identifying mitochondrial proteins is one thing, but to know how they contribute to disease you need to know what they do in the healthy cell. Here the researchers had a helping hand in the form of a database that has been under development at EMBL since 2000. STRING, the Search Tool for the Retrieval of Interacting Genes/proteins, is a collection of known and predicted

“It is becoming increasingly clear that biology cannot be understood by studying individual molecules one at a time, nor by studying an entire biological system using a single experimental technique.”

protein-protein interactions across a large number of organisms.

Using this database, the two Lars, Fabiana and their colleagues were able to place their putative mitochondrial proteins in the context of known protein-protein interactions, and hence to start to trace out, step by step, the pathways that mitochondria use to perform their vital functions. At last they could begin to build their map.

A survey of the map reveals certain patterns that are likely to hold for humans as well as for yeast. For example, the older the gene, the more important its role, with genes of ancient bacterial origin being more likely to control core mitochondrial functions such as respiration, and to be implicated in disease. To date, says Lars Steinmetz, one in ten of the genes which are known to be involved in heri-

table diseases has been found to code for mitochondrial proteins. “The mitochondria are implicated in a whole range of disorders from muscular to neurological diseases,” he says. “Tissues which have high energy requirements tend to be affected, but the age of onset and the severity of the symptoms can be quite diverse.”

To try to disentangle some of this complexity, researchers will need to pay attention to those mitochondrial proteins that modulate disease pathways, as well as to the actual components of those pathways. Here the map will prove invaluable because of the broader connections it reveals. Proteins that interact with known disease-causing proteins may themselves be implicated in disease, at a stroke increasing the number of candidate disease genes which warrant investigation. And the map can be considered a work in progress. As more genome-wide datasets are collected, it can be updated, simply by integrating the new data with the existing datasets and running the computer learning algorithms on the combined whole.

“A global approach to mitochondrial function was the only way to proceed because of the large number of components, the complexity of their interactions and their involvement in so many crucial functions in the cell,” says Lars Steinmetz, adding that he thinks this systems approach will now be adopted for other organelles and organisms. Lars Jensen agrees: “It is becoming increasingly clear that biology cannot be understood by studying individual molecules one at a time, nor by studying an entire biological system using a single experimental technique,” he says. “Integration of different types of high-throughput data will play a key role in biology in the future.” ■

Prokisch, H., Scharfe, C., Camp D.G., 2nd, Xiao, W., David, L., Andreoli, C., Monroe, M.E., Moore, R.J., Gritsenko, M.A., Kozany, C., Hixson, K.K., Mottaz, H.M., Zischka, H., Ueffing, M., Herman, Z.S., Davis, R.W., Meitinger, T., Oefner, P.J., Smith, R.D. and Steinmetz, L.M. (2004). Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biology* 2(6): e160