Mapping proteins with spatial proteomics

Vivien Marx

A number of techniques address the location of proteins within cells.

To maintain a cell's bustling activity, proteins handle specific tasks at particular locations. Researchers can detect these subcellular locations with spatial proteomics techniques and thereby obtain clues about protein function.

Localization also forms the basis of cellular maps. The team building the subcellular protein atlas[1], part of the Human Protein Atlas, is localizing human proteins to organelles and substructures; MitoCarta[2] positions proteins in mouse mitochondria; and a map based on a technique called BioID[3] is in the making in the lab of Anne-Claude Gingras, a biochemist at Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital in Toronto.

Gingras says that techniques to globally position proteins within a cell help researchers better understand poorly characterized proteins through a 'guilt-by-association' approach. Maps compiled with spatial proteomics techniques also help to characterize disease states.

Among the proteomic address–finding techniques are fractionation and mass spectrometry, live-cell tagging combined with mass spectrometry, and immunofluorescence approaches. The choice of localization method involves trade-offs, for instance, between throughput and specificity. And there are challenges in data interpretation and disagreements about which organelle is a protein's true home. In Gingras's view, the approaches are complementary: the union of these tools provides more solid and detailed data than a single approach can.

Spatial proteomics helps scientists journey from lists to function, says chemist Alice Ting from the Massachusetts Institute of Technology (MIT). For these travels they need maps of the proteome at different scales: in a complex environment such as in brain tissue, in an organelle or in organellar subcompartments.

Gamers will help to classify immunofluorescence images for the subcellular protein atlas. Here, a scenario involving analysis of a 'Jovian' sample.

Fractionation and mass spectrometry

Common practice in spatial proteomics involves cell fractionation and purification followed by mass spectrometry. John Bergeron, who works at the interface of proteomics and cell biology at McGill University, used this approach to see which proteins localize to the Golgi apparatus[4]. It was once an atypical approach that gained credence because it helped him find Golgi phosphoprotein 3, an uncharacterized protein later shown to help regulate Golgi function.

More recently, Bergeron has used this approach to find the most abundant Golgi protein in germ cells of the testis, which was also a protein that had not been previously characterized. The key to performing fractionation, he says, is a gentle protocol that isolates but does not fragment an organelle. He also performs electron microscopy (EM) of the isolated fractions to verify findings.

There has often been disagreement about whether a protein localizes to one organelle or another[5,6], and fractionation itself has been divisive, says Bergeron. He uses an approach based on work by Nobel laureate George Palade, which involves attempting to purify an organelle to homogeneity, followed by protein characterization and localization *in situ* of the abundant proteins. Generally, Bergeron recommends using quantitative proteomics; it also provides a way to see contaminants and to balance out the perceived weaknesses of purification.

As MIT's Ting explains, purification can work well, such as in the case of intact mitochondria. This technique, followed by mass spectrometry, helped to create

K. Girard/Cookiephoto Studio

The key to good fractionation is a gentle protocol, says John Bergeron.

MitoCarta, a map of the mitochondrial proteome[2]. Yet too many organelles and the contact sites between them do not purify well. For example, existing protocols to purify the contact site between mitochondria and the endoplasmic reticulum (ER) are crude, and they eliminate many integral membrane proteins at that junction, says Ting. That is a pity given that the interface is important in a variety of cellular processes.

The ER itself is challenging to purify in that cell lysis and ER shredding generate ER vesicles called microsomes, which need to be fished out after purification. "In the process you are going to pick up a lot of contaminants from other vesicular structures in the cell," says Ting. Some organellar substructures entirely defy purification. Ting and her team worked on the proteome of the mitochondrial intermembrane space[7], which can be neither purified nor accessed by traditional proteomics, she says. Another such area is the synaptic cleft in neurons.

The excitatory synapse has a postsynaptic density, a highly cross-linked and detergent-resistant meshwork of proteins that can be purified. But the inhibitory synapse is much more of a "black box," says Ting. Lacking a postsynaptic density, it cannot be purified, which means that only a handful of proteins are known at inhibitory synapses. She and her team have developed a reporter called engineered ascorbate peroxidase (APEX) for live-cell subcellular proteomic mapping and have applied it to the inhibitory synapse. "Now that we've mapped it with APEX, we see dozens more," she says.

APEX

The APEX reporter can be genetically targeted to an organelle or a cellular subcompartment of interest. When biotin-phenol and hydrogen peroxide are added, APEX biotinylates proteins in its immediate neighborhood. Biotinylated proteins can then be isolated and analyzed by mass spectrometry.

APEX can be fused to any protein and can be used in any cell type and in any subcellular compartment, says Ting, unlike the more traditional horseradish peroxidase tag, which is inactive in the mammalian cytosol. She and her team recently created APEX2, devised to have higher sensitivity[8] than its sibling APEX, now called APEX1. In some situations, APEX1 can create artifactual cellular perturbations when it is expressed at high enough levels to generate a detectable signal; the higher-sensitivity APEX2 can resolve these experimental impasses.

Once Ting's group had tested APEX2, they offered it prepublication to the 300–400 labs who had APEX1 "just to save themselves headaches," says Ting. APEX2 is now likely in the hands of more than 1,000 labs.

Ting recommends using APEX methods with quantitative proteomics approaches such as stable-isotope labeling by amino acids in cell culture (SILAC), isobaric tags for relative and absolute quantitation (iTRAQ) or tandem mass tags. "Otherwise you cannot get high-quality data and the kind of specificity that you need from this method," she says.

Quantitative proteomics is not easily accessible, so she recommends that researchers seek out collaborators or core facilities. Traditional proteomics using spectral counting will deliver "true hits," she says, but those will likely be the most enriched proteins, making it hard to distinguish true and false positives in a more middle zone.

APEX can also be used as an EM tag to verify the proper localization of a construct, such as prior to a proteomics experiment. When fixed cells are treated with hydrogen peroxide and diaminobenzidine, APEX polymerizes diaminobenzidine into a precipitate that attracts osmium, a widely used EM stain. For work on the mitochondrial intermembrane space, immunofluorescence did not clearly reveal whether APEX targeted the right mitochondrial subcompartment. "In EM it was very unambiguous," says Ting.

BioID

Gingras and her team are using BioID to localize proteins to their subcellular homes, a method originally developed by Kyle Roux of Sanford Children's Health Research Center and his colleagues to study the proximal and interacting neighbors of a protein of interest[3]. In this technique, the bait of interest is expressed in a cell as a fusion to BirA*—a promiscuous

bacterial biotin ligase. As a result, proteins near the bait are biotinylated. Once this labeling is complete, the sample is lysed, and the biotinylated proteins are purified using streptavidin agarose and then analyzed by mass spectrometry.

When used to study protein complexes, standard affinity purification coupled to mass spectrometry relies on keeping interactions intact through cell lysis and purification. But both APEX and BioID are based on biotinylation of proximity partners. So cell lysis and purification do not need to maintain protein interactions and can be done under fairly harsh conditions in which cellular compartments are solubilized.

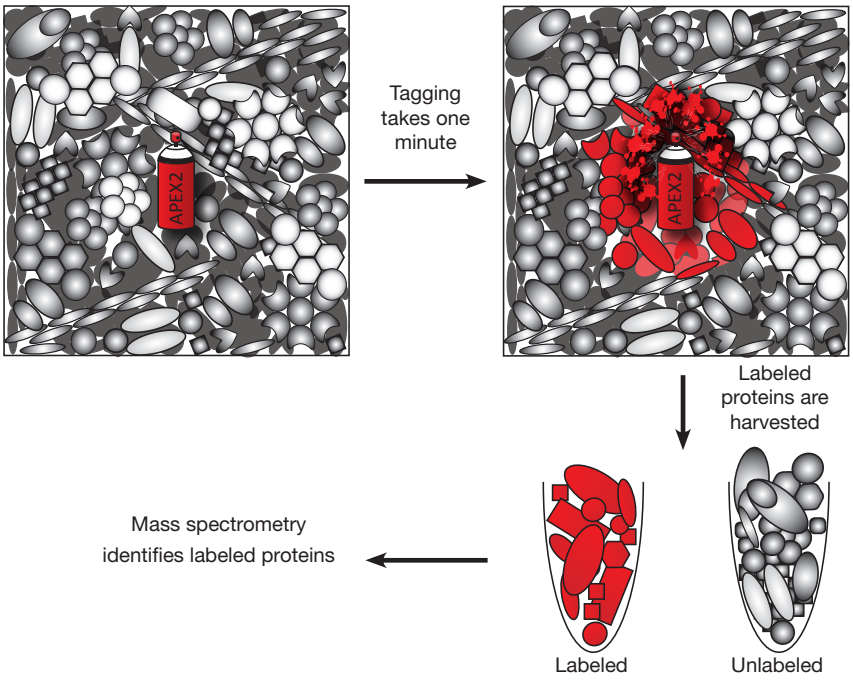
Gingras acknowledges that the slow reaction of BioID—labeling takes anywhere from 6 to 24 hours—could be seen as a drawback. It is less of a localization snapshot and more like a proteomic travel journal, she says. It allowed her team to 'amplify' some weaker signals from dynamically regulated proteins. But signal to noise is an issue. There needs to be biotinylation of the bait partners that is detectable above the background of endogenously biotinylated proteins and other contaminants.

Gingras has a BioID-based proteomic survey of key organelles and subcellular structures in the works. Each BioID experiment is high throughput and recovers hundreds to a thousand proteins in the mass spectrometer. "We do use the CRAPome to filter out the obvious contaminants, but the interaction profiles for some baits are still massive," she says.

A well-annotated resource of BioID-based subcellular localization would make it possible to confirm localization by comparing similar interactions. BioID seems to scale well, says Gingras. The more experiments she and her group perform, the more definition they obtain in terms of each organelle and the organization of the structures themselves. Together with colleagues Brian Raught and Laurence Pelletier, Gingras has used this approach to create a high-resolution map of the centrosome-cilia interface.

Immunofluorescence

In their work building the first draft of the subcellular protein atlas, planned for late 2016, Emma Lundberg and her team at KTH Royal Institute of Technology use immunofluorescence. This atlas is part of the Human Protein Atlas (HPA), a project devoted to the exploring the human proteome, based at Uppsala University and the Royal Institute of Technology. To date, the subcellular protein atlas has localized 10,000 human proteins in each of three cell lines in up to 18 organelles and substruc-



A. Ting, K. Cox, MIT; E. Dewalt, Nature Publishing Group

APEX can be genetically targeted to an organelle or cellular subcompartment. Next, biotinylation 'paints' proteins in the tag's immediate vicinity.

Emma Lundberg wonders whether some proteins moonlight, performing a different function in a different location.

tures. In about a year, Lundberg and her team plan to have more than 80% of nonsecreted human proteins mapped to their subcellular homes, which will include more than 25 subcellular locations such as nuclear bodies and most types of vesicles.

It appears, says Lundberg, that there is a core proteome, which is likely present to assure organelle function, and an associated proteome, which is present in cells only under certain conditions. And there seems to be a dual-localized proteome. One well-known dual-localized protein is human epidermal growth factor receptor 2 (HER2), a cell surface receptor often overexpressed in breast cancer. Two highly validated antibodies both show HER2's nuclear localization, reflecting its transcriptional activity in the nucleus, she says. This approach with highly validated antibodies makes it possible to confirm unexpected locations for other, less characterized proteins. Serine hydroxymethyltransferase (SHMT2) stains both in mitochondria, a known address, and in microtubules, which is a new finding. Many human proteins also show spatial variation related to the cell cycle, a phenomenon Lundberg and colleagues are studying. The more refined the location, the more biological relevance can be made of it, she says.

To additionally help classify the thousands of images generated in this project, Lundberg and her team are tapping into the crowd, in collaboration with Swiss start-up Massively Multiplayer Online Science (MMOS) and the Icelandic game company CCP. The idea is to create a subplot in Eve Online, the online science-fiction role-playing game that tens of thousands of people around the world play at any given time.

Immunofluorescence images match this sci-fi game well, says Attila Szantner, who cofounded MMOS. He is exploring ways to keep gamers interested in proteomic challenges for the long term. The narrative and the game mechanics are still a work in progress, but he shares one possible scenario: using immunofluorescence images from the subcellular protein atlas, a scientist in the game analyzes a 'Jovian' sample to see if it matches existing data. The protein localization that the gamers discover will also become part of the atlas, says Lundberg. An immunofluorescence image might link to another with a note that this similarity was discovered by 5,000 gamers.

Comparative views

Bergeron says he regularly consults the subcellular protein atlas and that it already has high value as a community resource. As Lundberg explains, the major drawback with immunofluorescence, however, is its need for reliable affinity reagents or tagged genes. The results need to be validated, as does the antibody specificity in each case, which is part of the pipeline at the HPA. "Also, the throughput is lower than for mass spec experiments," she says.

Mass spectrometry and immunofluorescence approaches complement one another, says Lundberg, each balancing the other's weaknesses. Mass spectrometry–based approaches are potentially quantitative, can deliver protein identity and are high-throughput methods in which a single measurement detects many proteins. But mass spectrometry is typically a batch method, delivering an average measurement from a cell population, she says. Additionally, there are the challenges of purifying organelles and the need to identify contaminants.

Imaging with fluorescently labeled antibodies, in contrast, gives a protein's spatial distribution *in situ* on a single-cell level. A cell surrounded by others is in its "natural surroundings," says Lundberg. Immunofluorescence can identify variations in protein distribution from one cell to its neighbor, such as cell cycle–based variation or differences due to cell density in tissue. "It is also easy to identify multiple localizations in the same cells," she says, or to see colocalization of specific proteins in the same cell and sample.

Proteins likely have more than one subcellular home, says Lundberg. According to her team's preliminary findings, approximately 50% of human proteins investigated thus far localize to multiple cellular compartments. She wonders whether the same proteins might be performing the same function in different locations, or they could be "moonlighting" in that

different milieu, performing a different function.

Some proteins show clear evidence of moonlighting activity, says Gingras, referring to her emerging BioID-based subcellular proteome map, with proteins popping up in unexpected locations. "However, at the moment, this is far too preliminary for me to comment on a given percentage," she says. A deeper understanding of moonlighting activity could help explore possible adverse effects of drugs.

Many proteins have multiple locations, says Ting, in part because they are shuttling from one place to another. Most mitochondrial proteins are translated in the cytosol and then shuttled into mitochondria, so a given spatial proteomics method might capture them anywhere along their trafficking route.

Many scientists, says Gingras, will map a protein of interest with fluorescence microscopy, using an antibody or a recombinant tagged protein. These are relatively low-cost options with widely available equipment, and they let researchers rapidly test changes in a protein's location or reactions to a treatment or perturbation. But used on its own, immunofluorescence microscopy in spatial proteomics faces some challenges. In addition to the requirement for specific antibodies, Gingras explains, mapping an organellar proteome using immunofluorescence requires high-resolution microscopes, as implemented by the HPA teams. Fractionation coupled to mass spectrometry, when possible for the structure in question, or proximity-based approaches are much more appropriate in such cases, she says.

"It is also in general fairly difficult using immunofluorescence alone to reveal the organization of proteins within a given structure, which is a task that proximity approaches seem particularly good at doing," says Gingras. Yet, fluorescence approaches coupled with super-resolution microscopy, and techniques such as fluorescence resonance energy transfer, the proximity ligation assay or 'split-GFP', can all provide more information regarding the relative organization of proteins or the distance between them within a given structure. But, she says, these latter approaches are hardly proteome scale at the moment.

Ting finds BioID potentially powerful and looks forward to more data assessing its spatial specificity. Effective BioID labeling takes longer than labeling with APEX, meaning that scientists cannot study a biological process shorter than the labeling time, which might be 12 hours. That would exclude, for example, looking at a mitochondrial protein in one phase of the cell cycle. With APEX, the labeling takes one minute, which offers a temporal snapshot of the proteome, says Ting, such as looking at the mitochondrial proteome after a drug treatment.

Both BioID and APEX need to be combined with quantitative proteomics, in Ting's view. A researcher will want to distinguish between more strongly and less strongly biotinylated proteins, which is revealed only with quantitative proteomics.

There is likely to be much synergy between spatial proteomic methods. And as larger resources—the fluorescence-based HPA, subcellular fractionation studies and, eventually, proximity-labeling studies—become widely available and when these projects all cross-reference each other, Gingras says, "we will be able to refine or redefine the compositions of subcellular locations for most human proteins." The journey to map proteins and to go from list to function may be full of methods challenges, but it is a trip worth taking, says Lundberg, because "this is where the interesting biology begins."

1. Lundberg, E. & Uhlén, M. *Proteomics* **10**, 3984–3996 (2010).
2. Pagliarini, D.J. *et al. Cell* **134**, 112–123 (2008).
3. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. *J. Cell Bio* **196**, 801–810 (2012).
4. Bell, A.W. *et al. J. Biol. Chem.* **276**, 5152–5165 (2001).
5. Touret, N. *et al. Cell* **123**, 157–170 (2005).
6. Gilchrist, A. *et al. Cell* **127**, 1265–1281 (2006).
7. Rhee, H.-W. *Science* **339**, 1328–1331 (2013).
8. Lam, S.S. *et al. Nat. Methods* **12**, 51–54 (2015).

Vivien Marx is technology editor for *Nature* and *Nature Methods* (v.marx@us.nature.com).