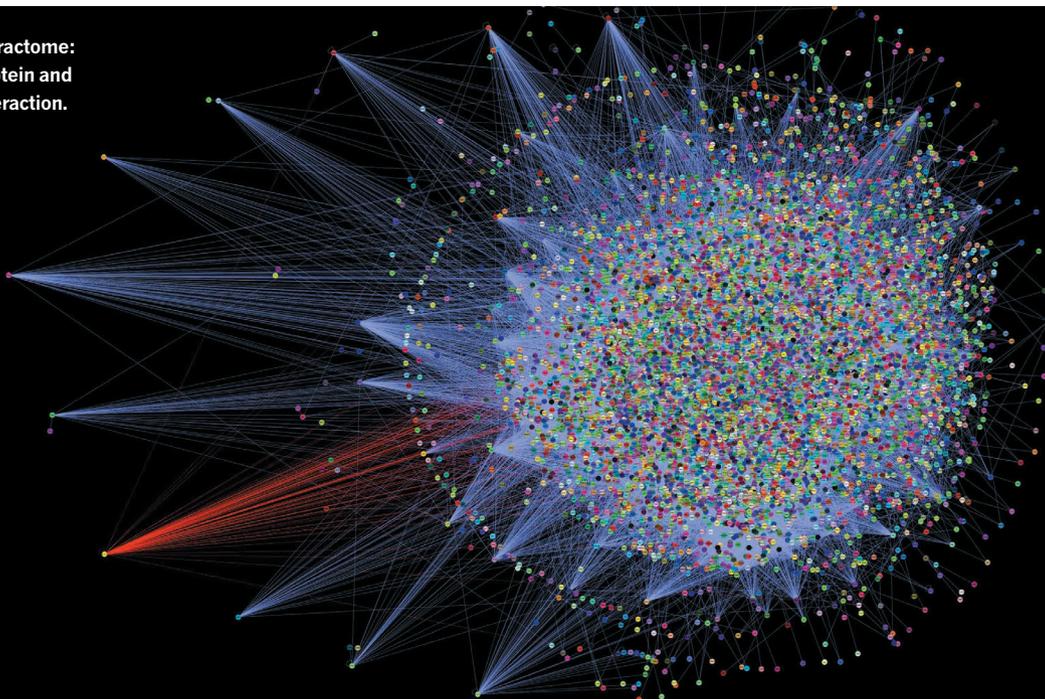


## TECHNOLOGY FEATURE

# PROTEIN MAPS CHART THE CAUSES OF DISEASE

*Improvements in mapping protein–protein interactions are allowing researchers to deconstruct the delicate mechanics of cells.*

The human interactome: each dot is a protein and each line an interaction.



ANDREW GARROW

BY MARISSA FESSENDEN

In 1987, researchers in Switzerland described two sisters who were born separately but had similar abnormalities. A curl of tissue in their cerebellums was missing. Their hearts contained holes and clefts. One died aged three following cardiac surgery; her sister had a similar operation at age four, but survived. Because neither of the girls' parents had these abnormalities, the researchers concluded that their daughters had inherited two copies of an atypical gene, leading to a previously unknown syndrome<sup>1</sup>.

The scrambled nucleotides responsible for the girls' condition may reside in a single gene. Yet several other genes have subsequently also been associated with what has been dubbed Ritscher–Schinzel syndrome. The functions of those genes, and how they related to the syndrome, remained a mystery for years.

Today, those molecular underpinnings are

coming into focus thanks to the systematic study of protein–protein interactions, a discipline called interactomics. By mapping the network of connections between proteins, three research teams independently discovered a complex called Commander that's made up of proteins produced by the mutated genes<sup>2</sup>. Commander is an essential cell component that sorts and delivers proteins, and its malfunction causes the devastating defects of Ritscher–Schinzel syndrome.

Proteins and other biological molecules rarely work alone; they brush up against one another in fleeting interactions or band together to form complex cellular machines. Only through such partnerships can proteins perform their many functions. Breakdowns in those interactions can affect human health.

“If you break a gene coding a protein that goes into a complex, then that complex is dysfunctional in some way and that gives rise to a condition or disease,” says Edward Marcotte,

a systems biologist at the University of Texas at Austin.

Biochemists have long studied the ways in which one or a few proteins interact with others. But now they are developing tools to chart more comprehensive sets of protein–protein interactions at levels from organellar to organismal. These interactomes typically look like dense starbursts, with protein dots or nodes joined by the interactions between them. Self-contained clusters of interconnected proteins that emerge from these webs may represent key complexes and communal functions or, as in the case of Ritscher–Schinzel syndrome, provide clues to the roots of disease.

In the past three years, research groups have published the first high-quality maps of the human interactome<sup>3–6</sup>. Together, the most recent iterations of those maps have identified around 93,000 unique protein–protein interactions.

The technologies underlying these maps ►

## Pick A or B

The two high-throughput methods for protein-interaction mapping have their supporters, but they are complementary.

Determining whether two proteins physically interact relies on yeast two-hybrid systems. The assay involves fusing the genes encoding two putative interaction partners to the two halves of a yeast DNA-binding protein. The strain carrying these hybrids can grow only when the target proteins interact and unite the halves of the yeast protein, which activates crucial genes. Sequencing the DNA from growing yeast colonies reveals the proteins involved in the interaction.

Yeast two-hybrid systems allow the quick screening of many protein pairs at once, although validating the interactions through further assays is essential: just because two proteins interact in the yeast nucleus does not mean that they partner in their native cell.

Researchers can also run protein complexes through a mass spectrometer.

These instruments convert the complexes to a cloud of charged particles and identify the pieces by their mass. In one common approach, affinity purification followed by mass spectrometry, researchers label protein ‘baits’ with peptide or protein ‘handles’. Those handles provide a way to recover the bait proteins from cell slurries, along with their interaction partners or ‘preys’, which are identified by mass spectrometry. Alternatively, researchers can take the full mixture of proteins from cells and run it through a series of biochemical separation steps. The proteins that tend to co-purify (or ‘cofractionate’) in this approach are interaction partners.

Mass-spectrometry-based approaches allow researchers to work directly in the cells where proteins occur, rather than in yeast, but not all complexes can survive the extraction steps. Also, these approaches cannot distinguish between direct, physical interactions and looser associations. **M.F.**

► are not new; protein-interaction mapping dates back to the 1990s. And researchers have been producing interactome maps since at least the early 2000s. But methodological refinements as well as advances in protein purification, mass spectrometry and gene-editing techniques have empowered researchers to explore the interactome — and the insights it promises into development and disease — with ever-finer precision.

It isn't easy: capturing all interactions is a challenge, as the set of protein partners varies across different tissues, cells and even time. The interactome is dynamic, breaking and forming connections as the cell responds to its environment. Mapping it to completion may require fresh methods and ways of thinking about systems biology.

Still, the field is yielding results. “New machines that are ubiquitous but deeply understudied — that’s fundamental biology coming out of the maps,” Marcotte says. “We’ve clearly passed a critical threshold.”

### NUMBERS GAME

There essentially are two approaches to building interactome maps. The yeast two-hybrid assay tests for direct interactions between protein pairs by coupling gene expression to protein interactions in the cell. The second approach maps both direct and indirect protein contacts by isolating complexes with antibodies and identifying their component parts with mass spectrometry (see ‘Pick A or B’ and ‘Mapping tools’).

Marcotte’s laboratory uses a variation on the second approach that involves biochemically

separating proteins — for instance, using sucrose density gradients — to see which molecules tend to stay together.

The resulting maps allowed Marcotte and Anna Mallam, a postdoctoral researcher in his lab, to draw inferences about the Commander complex’s cellular role<sup>2</sup>. Previous studies revealed that two components were structurally similar to proteins that build and maintain eukaryotic hair-like structures called cilia and flagella; other components seem to move proteins across membranes. Those data and other findings suggest that Commander moves specific proteins from the cell membrane to a compartment called the Golgi apparatus, where they are recycled.

The largest maps encompass thousands of proteins, resembling tangled hairballs more than starbursts. But by unravelling them, researchers have identified signatures that distinguish cancer-causing genes from ‘normal’ ones, and that define key biological processes, such as chromosome segregation during cell division.

Even with multiple approaches, interactome maps are “still largely incomplete”, says computational biologist Katja Luck at the Dana-Farber Cancer Institute in Boston, Massachusetts. It’s a question of numbers. The human genome contains roughly 20,000 protein-coding genes. If one assumes that each protein has only one form — a massive oversimplification — there are approximately 200 million possible interactions. The real number is likely to be much smaller because many interactions are indirect; estimates for one-to-one interactions range from 120,000 to 1 million.

Proteins are incredibly diverse, biochemically speaking, and thus their interactions cannot be captured by every assay equally. Membrane-protein interactions, for instance, are difficult to study because when the membrane is stripped away, their shape and behaviour changes; they may not link with their typical partners. But, the extent to which this incompleteness alters the current maps isn’t yet clear. “We are just at the beginning of understanding the biases of different methods,” Luck says.

As a postdoctoral researcher in the lab of geneticist Marc Vidal, Luck has helped to implement protocols to eliminate errors in their two-hybrid approach. The core method dates back to the 1980s. “We are just doing some tweaks to make it better,” she says. By tagging the protein genes with barcodes, the team can test more than one interaction at a time in a large range of growing yeast. Rigorous attention to detail, automation of key steps and sequencing in quadruplicate has allowed them to identify more than 60,000 interactions, the majority of which were previously unknown.

That data set forms the bulk of the interactions reported in the collaborative Human Reference Protein Interactome Mapping Project, and it is still growing. “By 2020 we want something that people will be able to refer to as a reference map for the human interactome,” Vidal says. The work hasn’t always gone smoothly. The early days of interactomics generated error-prone networks. Only about 3% of identified interactions had support from more than one method, according to one 2006 review<sup>7</sup>. “People were extremely cautious about using those data sets,” Vidal says. “But in ten years we have made really incredible progress.”

### BETTER MAPPING WITH CRISPR

The eventual reference map Vidal envisions is likely to contain only a subset of all possible interactions. Cell and tissue variation as well as shifting cellular responses add up to many possible versions of the full interactome. For Matthias Mann, a biochemist at the Max Planck Institute of Biochemistry in Martinsried, Germany, those variations are daunting. But he is optimistic about the power of gene-editing technology, such as CRISPR–Cas9, to address them.

Mann’s mapping method involves libraries of cell lines expressing hundreds of proteins, which are tested for interactions using an ultra-high-resolution mass spectrometer called Orbitrap. The bait proteins are fused to a green fluorescent protein, producing a luminosity profile that allows the researchers to quantify interactions through live-cell imaging. In the late 2000s, creating the cell-line library was “quite laborious”, he says. “Now our method gets wings due to the CRISPR engineering that can be brought to bear.”

Since introducing the quantitative approach in 2010, Mann’s team has mapped and

quantified the strength of more than 28,000 interactions. Interactions in which the partners exist in one-to-one ratios are considered 'strong' and are likely to exist in stable and abundant complexes. Without such information, "it is very hard to say something about the structure of the network", Mann explains. Analysis of his team's map showed that the human interactome is dominated by weak associations, which may reflect low-abundance regulatory proteins acting on more stable protein machines.

**SUBTLE TWEAKS**

A common trend across the field is the adoption of relatively gentle protocols for sample preparation that aim to faithfully capture all protein-protein interactions in the cell.

"We are trying to find less disruptive methods," says Rosa Viner, a biochemist at Thermo Fisher Scientific, a life sciences company in San Jose, California. The firm's focus on improving sample preparation, workflow and mass-spectrometry technology aims to help researchers identify interactions as they exist in cells. "This is the hardest challenge: finding the methods that will give us the best picture without any artefacts," she adds.

Artefacts can include protein complexes that fall apart before their interactions are detected. To hold complexes together, Viner has worked with researchers at the University of California, Irvine, to chemically fuse complexes, an approach called crosslinking, before mass-spectrometric analysis. A strategy called QMIX (quantitation of multiplexed, isobaric-labelled crosslinked peptides) has been developed that integrates crosslinking compounds with chemical labels to allow researchers to stabilize as well as track protein complexes<sup>8</sup>.

Good analysis also takes into account the

blind spots of any given method. "There are still classes of proteins that are very challenging," says Wade Harper, a cell biologist at Harvard Medical School in Boston. "When you do high-throughput analysis, you are limited in how much care you can take with individual protein." That's because such analyses tend to treat every reaction the same, leaving little room for customization.

Harper and his colleague Steven Gygi, also at Harvard, created a lab group to fine-tune their approach. "With a relatively small team of four to six people we can create four or five hundred cell lines a month," he says. That dedication has yielded the largest collection of human-protein-complex data yet achieved from a single pipeline. Their map, called BioPlex, includes around 120,000 interactions.

**BIG PICTURE**

But to get a closer look at interactions, researchers must dive into the crowded landscape of the cell itself.

Anne-Claude Gingras, a biochemist at the University of Toronto in Canada, uses a technique called BioID, which tags proteins on the basis of their proximity to one another. The tagged protein of interest adds a chemical tag to nearby proteins, leaving evidence of its interactions like a crayon-wielding toddler's trail through the house.

The result is a map of the physical neighbourhood surrounding the initial protein. Identifying a protein's larger community is

likely to reveal details about its cellular function, Gingras explains.

Proximity mapping also allows researchers to track proteins that cannot be picked up by other assays, such as difficult-to-isolate membrane-embedded proteins. "We and others have looked at proteins on chromatin, mapped the organization of the centrosome and detected interactions that span all kinds of membranes," Gingras says. Using BioID, the group found new components in a signalling pathway that regulates organ size during development<sup>9</sup>.

Harper's lab uses a similar method called APEX. In it, an engineered plant enzyme called ascorbate peroxidase chemically restricts the time window during which the protein of interest can tag others, resulting in a fainter but more spatially precise signal.

Having multiple approaches in interactomics means that when interactions appear on more than one map, they carry more weight. That is where the insights will come, says Jennifer Lippincott-Schwartz, a cell biologist at the Howard Hughes Medical Institute Janelia Research Campus in Ashburn, Virginia. "If we are going to understand how cells are working, it is critical to connect all the protein-protein interaction maps with spatial maps within the cell," she says.

Cells are packed with large structures or organelles, all floating in the protein-rich soup of the cytoplasm. Understanding which proteins are interacting and why will require researchers to actually see what this world looks like.

Lippincott-Schwartz's lab has developed an arsenal of tools for visualizing proteins inside living cells using fluorescent labels. These tools have revealed six organelles — the endoplasmic reticulum, the Golgi apparatus, lysosomes, peroxisomes and lipid droplets — moving and interacting in 3D. The team calls it the organelle interactome<sup>10</sup>.

The interactome, Lippincott-Schwartz says, is "a hypothesis generator" for cell biologists. "You go in and start testing once you see a protein you know interacting with a whole bunch of other proteins that have functions you didn't know."

With interactome maps finally becoming fleshed out with high-quality, abundant interactions, researchers can start putting those hypotheses to the test. ■

*Marissa Fessenden is a freelance science writer in Bozeman, Montana.*

1. Ritscher, D. et al. *Am. J. Med. Genet.* **26**, 481–491 (1987).
2. Mallam, A. L. & Marcotte, E. M. *Cell Syst.* **4**, 483–494 (2017).
3. Rolland, T. et al. *Cell* **159**, 1212–1226 (2014).
4. Huttlin, E. L. et al. *Cell* **162**, 425–440 (2015).
5. Wan, C. et al. *Nature* **525**, 339–344 (2015).
6. Hein, M. Y. et al. *Cell* **163**, 712–723 (2015).
7. Qi, Y., Bar-Joseph, Z. & Klein-Seetharaman, J. *Proteins* **63**, 490–500 (2006).
8. Yu, C. et al. *Anal. Chem.* **88**, 10301–10308 (2016).
9. Couzens, A. L. et al. *Mol. Cell Proteomics* **16**, 1098–1110 (2017).
10. Valm, A. M. et al. *Nature* **546**, 162–170 (2017).

