

PROTEOMICS

Charting the subcellular proteome

Human protein subcellular locations light up in the Cell Atlas resource.

A cell is not just a bag of proteins and other molecules; organelles provide niches with specific chemical environments in which certain members-only proteins are allowed. “Knowing the exact subcellular locations provides good clues for protein function and can also be used as boundaries for protein–protein interactions in systems biology applications,” says Emma Lundberg of KTH Royal Institute of Technology in Stockholm. “Knowing the whole set of proteins in an organelle also facilitates the understanding of its morphology and function.” But despite the availability of numerous methods, identifying proteins’ subcellular addresses has turned out to be challenging to do on a proteomic scale. As a result, most human proteins have not been mapped to specific organelles.

Lundberg, her KTH colleague Matthias Uhlén, and a large team of their colleagues now present the Cell Atlas, a comprehensive map of subcellular addresses of human proteins. This project, a part of the Human Protein Atlas (HPA) effort, complements the previously reported Tissue Atlas, a map of protein distribution across human tissues and organs. The data are freely available to the community and should be of great interest to the already hundred thousand or so monthly HPA users. The work also nicely complements other emerging initiatives to build whole human cell maps.

In building the Tissue Atlas, the HPA team generated a large number of antibodies against human proteins, which were used for immunohistochemistry-based mapping. “When we started the Cell Atlas effort 10 years ago we realized that in order to resolve not only the major organelles in the cell but also the fine intricate substructures we needed to work with high resolution at a large scale,” recalls Lundberg. The Cell Atlas team therefore applied a more sensitive approach combining antibody-based immunofluorescence and high-resolution confocal microscopy. Once they had established a robust, automated pipeline, they were able to analyze

BIOCHEMISTRY

CRISPR KINETICS

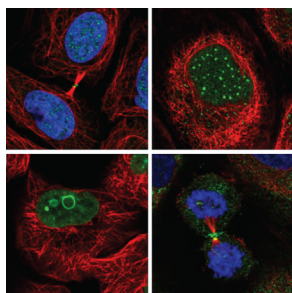
In vitro Cas9-binding assays show the effects of mutations in the target sequence on binding kinetics.

Despite the popularity of the CRISPR–Cas9 genome editing system, our understanding of its biophysical underpinnings remains vague. William Greenleaf from Stanford University aims to change that. Working with Jennifer Doudna’s group at the University of California, Berkeley, his goal was to “move towards a predictive, biophysical model for how Cas9 finds its target.”

A rough outline of the molecular underpinnings of the CRISPR system is known; first, Cas9 binds genomic DNA at a three-nucleotide motif, the protospacer adjacent motif (PAM) most often comprised of the nucleotides NGG. Following PAM binding, the single guide RNA (sgRNA) associated with Cas9 starts to invade the double-stranded genomic DNA, and its twenty terminal nucleotides seek to pair with a complementary DNA strand. Successful DNA–RNA base pairing triggers the activation of the nuclease domains in Cas9 and leads to DNA cleavage. Greenleaf and his team wanted to know how this process works in detail.

They devised an *in vitro* high-throughput assay to profile the binding of Cas9 to its target sequences containing mutations among the twenty nucleotides that bind the sgRNA and the three nucleotides of the PAM. First, these target DNA molecules flanked by sequencing adapters were arrayed and sequenced on an Illumina flow cell. Not only did this process provide the exact sequence of each target, but it also amplified the sequences. The team then probed the target clusters with a fluorescently labeled complex of sgRNA and nuclease-dead Cas9 (dCas9), measured the association and dissociation rates of dCas9 and calculated on and off rates.

Greenleaf recalls being surprised when he saw the results. “Frankly, it was way more complicated than we thought it would be.” At a target with perfect complementarity to the sgRNA, the researchers found the off rates to be immeasurably slow, but effects of



Immunofluorescence maps subcellular protein localization in the Cell Atlas. From Thul *et al.* Reprinted with permission from AAAS.

about 1,000 samples per week, says Lundberg.

Altogether, the researchers mapped the subcellular locations of more than 12,000 proteins to 30 subcellular structures across 22 human cell lines. The cellular addresses of about 47% of the proteins that the team mapped were previously unknown. Approximately one-third of the proteins they charted are expressed in all cells, indicating housekeeping functions. Interestingly, the researchers also found that more than half of the detected proteins localize to multiple subcellular structures. “Multilocalizing proteins may have context-specific function and ‘moonlight’ in different parts of the cell, thus increasing the functionality of the proteome and the complexity of the cell from a systems perspective,” notes Lundberg.

The validation of their immunofluorescence results was a crucial part of the project. The researchers used a four-tier scoring system based on the availability of supporting validation data to determine the reliability of their antibodies. To confirm a portion of their subcellular localization results with a completely orthogonal method, they used a mass-spectrometry-based mapping technique, hyperLOPIT, which assigns proteins to organelles based on their density gradient fractionation profiles.

The HPA team plans to continue work on the Cell Atlas to map the complete subcellular human proteome. “This means that we will need to work with more specialized cell types, such as primary cells and stem cells,” Lundberg notes. The team’s ultimate goal is to develop a model of the entire proteome of a human cell over the course of one cell cycle and to understand how protein localization contributes to cell function.

Allison Doerr

RESEARCH PAPERS

Thul, P.J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).

mutations along the target sequence varied greatly. “Mutations in different parts of the target modulate off rates.” says Greenleaf.

When the PAM sequence, NGG, was varied to represent the noncanonical PAMs NGA and NAG, the researchers still saw binding over background levels, indicating that dCas9 binding, at least *in vitro*, is not confined to the NGG PAM.

The team observed that mutations and base composition in the bases proximal to the PAM had the highest effect on association rates of dCas9; but variants distal to the PAM also had an effect, particularly in conjunction with a mutation in the proximal region. Off rates were affected by sequences distal to the PAM; for example, a single mutation sixteen nucleotides away from the PAM led to a rapid dissociation of dCas9 from its otherwise complementary target. The team confirmed some of the findings with an orthogonal assay that probed the kinetics of dCas9 binding in bulk.

“It’s a complicated energy landscape,” explained Greenleaf, “but you can tune different parts of the kinetics, and this is the first step to rationally tune the behavior of this programmable, biochemical agent.”

One of the next steps will be to test more sgRNAs against more targets and vary the incubation times with Cas9. The researchers want to expand the investigation beyond Cas9’s on and off rates and study the kinetics of sgRNA strand invasion in more detail.

“We want to build a generalizable model for how the energetics of base pairing and between DNA, RNA and protein all come together,” says Greenleaf. He also hopes to build a platform to address these questions which can then be applied to other CRISPR nucleases.

How Cas9 binding relates to target cleavage will be the next intriguing question to tackle and will result in rationally designed sgRNAs with optimal specificity and activity.

Nicole Rusk

RESEARCH PAPERS

Boyle, E.A. *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Nat. Acad. Sci. U.S.A.* **114**, 5461–5466 (2017).