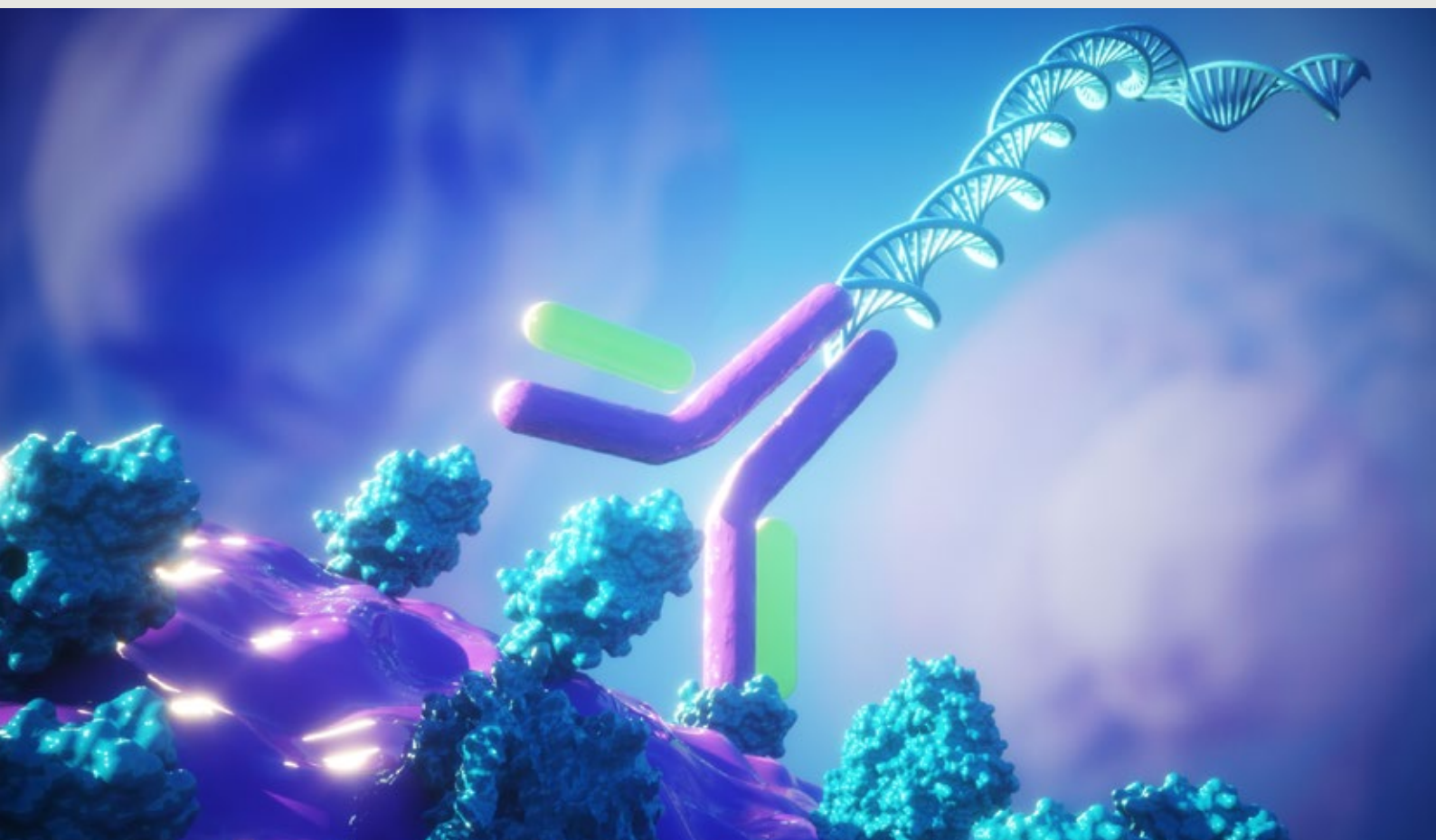


# PROTEIN DETECTION BY SEQUENCING: TOWARDS A DEFINITIVE CELLULAR PHENOTYPE

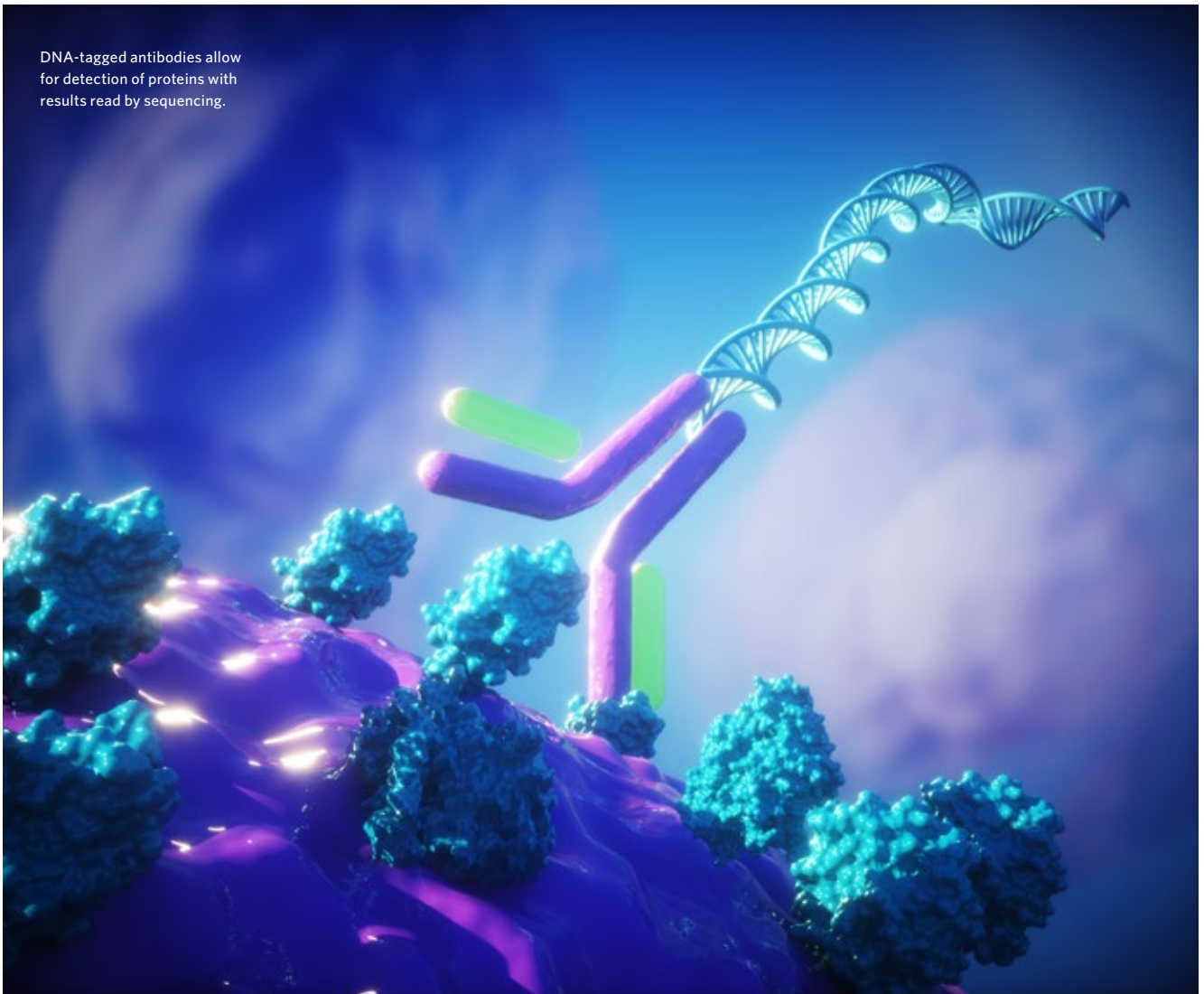
Methods that allow researchers to simultaneously sequence RNA and detect extracellular proteins in individual cells reveal new cell types and states associated with disease.



For **illumina**<sup>®</sup>

by **nature**research  
CUSTOM MEDIA

DNA-tagged antibodies allow for detection of proteins with results read by sequencing.



# PROTEIN DETECTION BY SEQUENCING: TOWARDS A DEFINITIVE CELLULAR PHENOTYPE

Methods that allow researchers to **SIMULTANEOUSLY SEQUENCE RNA AND DETECT EXTRACELLULAR PROTEINS** in individual cells reveal new cell types and states associated with disease.

Proteins are the workhorses of the cell, involved in almost every aspect of structure and function. Protein expression patterns help define a cell's identity and state. RNA transcripts are often used as a surrogate for protein expression, but it has long been understood that the relationship between abundance of proteins and mRNA is not one-to-one. There are differences caused by regulation of post-transcriptional, translational and protein degradation.

"It has been eye-opening to see how much we might be missing by using conventional sequencing," concedes Adeb Rahman, an immunologist at Icahn School of Medicine at Mount Sinai in New York. Rahman is developing innovative assays to monitor disease-relevant molecular profiles and discover new biomarkers and drug targets. "In some cases, we have populations of cells that are all clearly expressing a particular protein, but we may detect the corresponding transcript in only 5-10% of them," he explains.

Rahman and others are using techniques such as CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) and REAP-seq (RNA expression and protein sequencing assay) to investigate the relationship between RNA transcripts and proteins, and identify new cell types and cell states associated with disease. Both methods combine antibody-oligonucleotides with existing single-cell RNA-sequencing (scRNA-seq) approaches to measure the expression levels of genes and cell-surface proteins in individual cells.

A major advantage of CITE-seq and REAP-seq is that, with an appropriate antibody, they can detect a protein of interest even if the RNA that encodes it is in low abundance. "CITE-seq gives us more confidence in the presence of a marker of interest in a given cell type," Rahman explains.

For decades, researchers have used flow cytometry to detect proteins in single cells and characterize cell types based on protein expression patterns. Flow cytometry relies on antibodies directed against proteins of interest. This means that to obtain useful information researchers need to know what they are looking for.

**"IT HAS BEEN EYE-OPENING TO SEE HOW MUCH WE MIGHT BE MISSING BY USING CONVENTIONAL SEQUENCING."**

In recent years, single cell sequencing has become the method of choice for uncovering cell types and states based on gene expression patterns. One of the advantages of single cell RNA (or DNA) sequencing is that it is unbiased, it requires no prior knowledge about the sample — RNA (or DNA) from any cell type can be run through a sequencer to obtain its genomic or transcriptional profile. Advances in single-cell genomics are enabling large scale projects such as the Human Cell Atlas Project, which aims to create

comprehensive reference maps of all 37 trillion cells in the human body.

Now, by providing details about both protein and RNA expression, CITE-seq and REAP-seq allow researchers to fill in the gaps left by flow cytometry or single cell RNA-seq.

"Advances in multimodal single cell sequencing methods are providing an unprecedented opportunity to look at biological mechanisms of disease," says Tom Maniatis, a renowned biochemist and biophysicist leading Columbia University's Precision Medicine Initiative and the New York Genome Center. Technologies that combine single cell DNA or RNA sequencing with methods that provide details about the regulation of gene expression, protein expression, genetic mutations, cell lineage, and even the spatiotemporal order of molecular events not only allow the creation of a more refined human cell atlas, but offer new insights into cancer evolution, neurodegeneration and the immune system's response to disease.

#### **Multimodal single cell sequencing in action**

In CITE-seq and REAP-seq, cells are incubated with cell-surface antibodies linked to oligonucleotide barcodes which tag the protein's identity, and a stretch of adenine bases that serve as a starting point for RNA sequencing (FIG). The technologies differ in the way the DNA barcode binds to the antibody. In both cases, protein detection has been shown to be as robust as by flow cytometry.

To validate CITE-seq, the developers at the New York Genome Center and scientists at NYU Lagone Health used it to simultaneously profile the transcriptome and 13 cell-surface proteins of human and mouse immune cells. They showed that CITE-seq not only correctly identified different immune cell populations, it enhanced the characterization of known subtypes. Using CITE-seq they were able to identify subsets of natural killer cells with different roles in regulating the immune response in disease states that could not be detected using single-cell RNA-seq methods alone<sup>1</sup>.

REAP-seq was first employed to examine the response of a T cell subset to a drug that activates the immune system to attack and kill cancer cells in mice by targeting a cell surface protein. Differences between the proteins and genes expressed in the treated and untreated cells provided new insights into the drug's mechanism of action<sup>2</sup>.

Since these studies were published, more than 200 oligo-conjugated antibodies have become available for CITE-seq and REAP-seq. These antibodies have been used to assess immune cell activation in patients with advanced lung cancer after immunotherapy<sup>3</sup>, and to investigate immune dysregulations at atherosclerotic sites associated with clinical cerebrovascular events. When Rahman and colleagues used CITE-seq to examine the immune cell types in atherosclerotic

lesions in stroke patients, they found T cell subsets that were more activated, differentiated and exhausted compared to their blood counterparts. Characterizing immune cell profiles with these technologies could aid the design of tailored therapies and help predict a patient's treatment response<sup>4</sup>.

### Expanding the multimodal sequencing toolbox

Using the same concept of DNA barcodes, CITE-seq has been extended by Peter Smibert and Neville Sanjana at the New York Genome Center to reveal even more information from single cells. ECCITE-seq (expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing), can be used for simultaneous detection of the transcriptome, proteins, clonotypes and CRISPR perturbations<sup>5</sup>.

**"SINGLE CELL SEQUENCING METHODS ARE PROVIDING AN UNPRECEDENTED OPPORTUNITY TO LOOK AT BIOLOGICAL MECHANISMS OF DISEASE."**

"I love this technology," says Ya-Chi Ho, a molecular virologist at Yale School of Medicine, who is trying to understand how HIV persists in cells. HIV infected cells are very rare: less than 1% of T cells in peripheral

blood contain the virus. Moreover, HIV infected cells do not display any markers that distinguish them from uninfected cells, making them practically impossible to detect and analyse.

ECCITE-seq can capture the variability of T cell receptor sequences that confer specificity for particular antigens. These sequences are found at the 5' end of RNA molecules and are difficult to capture in scRNA-seq approaches that profile the 3' ends of RNAs. Furthermore, this method can be expanded to capture HIV RNA in the same cell. "ECCITE-seq allows us to tackle the heterogeneity, rarity and the lack of marker in these cells, which is very powerful," she explains.

Ho and her team are currently looking at cells in the blood of HIV-infected individuals during high-level viraemia and after antiretroviral therapy to determine how infected cells are able to continue dividing, leading to the clonal expansion of HIV-positive cells. Understanding the features of these resistant clones may offer clues for new therapies that target cells bearing the virus and, effectively, cure HIV.

Abbas Rizvi is a molecular neuroscientist working with Maniatis and colleagues at Columbia University to combine other single cell sequencing technologies to develop a comprehensive cell atlas of the human spinal cord. They are carrying out RNA-seq and ATAC-seq (assay for transposase-accessible chromatin using sequencing), which identifies

active transcriptional regulatory elements, in nuclei obtained from postmortem spinal cord cells to develop new treatments for spinal-cord disease and injury.

Rizvi is working closely with bioinformaticians and statisticians to develop mathematical approaches to harmonize

**"NOW WE CAN BEGIN TO REALLY UNDERSTAND THE COMPLEXITY OF NEURODEGENERATIVE DISEASES."**

the data. "Multimodal strategies enable us to make connections among simultaneously collected measurements in any cell type, and therefore decipher key molecular events associated with the development, establishment and maintenance of neural circuits," he explains.

When the team compares the transcriptional expression patterns from cells in cervical, thoracic and lumbar regions of the spinal cord, they see remarkable differences. "Ultimately, we want to carry out the same type of analysis on the spinal cord of patients with amyotrophic lateral sclerosis to understand the mechanisms underlying motor neuron degeneration, with the goal of identifying novel therapeutic targets," says Maniatis.

Computational tools are crucial to make biological sense out of all the data that are being generated by new sequencing technologies. Rizvi observed that these

tools are leading to a 'second wave' of excitement around single cell technologies. "Now we can begin to really understand the complexity of neurodegenerative diseases, and the contribution of multiple genes and variants to disease processes in a way that was not possible before," Maniatis adds.

### Future outlook

Barcoding cellular content is enabling researchers to obtain multiple readouts from single cells. As such, they can characterize cellular phenotypes in more detail than by using transcriptome measurements alone. These multi-omic readouts hold great potential to advance translational genomic science and precision medicine.

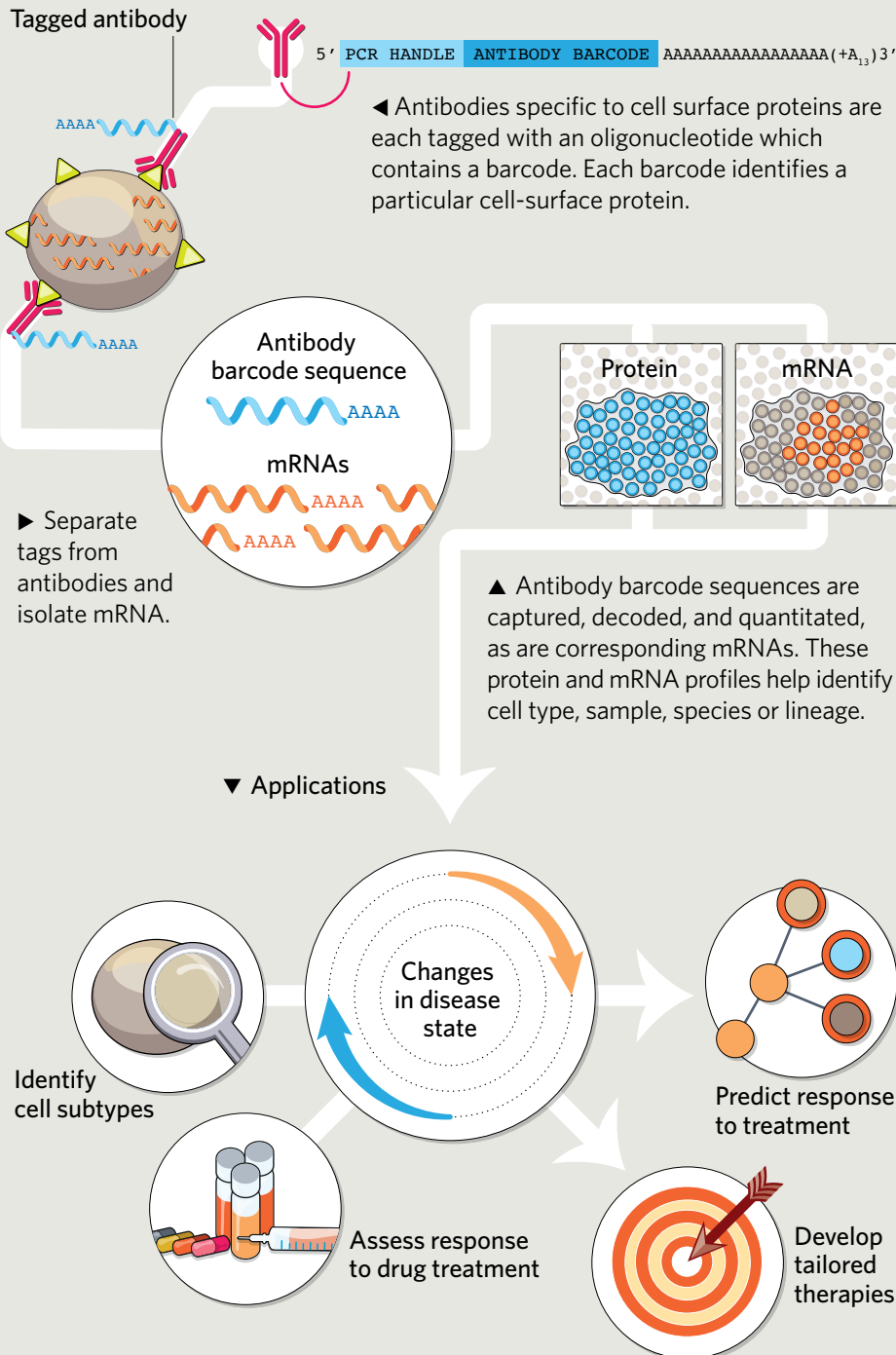
"We are also excited about the development of spatially resolved transcriptomic approaches to place single cell RNA sequencing data in its anatomical context," says Rizvi. This will allow the identification of cells that are vulnerable to neurodegeneration, as well as to define the features of cells that surround them, and which are likely to contribute to disease progression.

At present, single cell multi-omics technologies are primarily used to



## CITE-SEQ & REAP-SEQ

Barcoded cellular markers help simultaneously profile mRNAs and surface proteins from individual cells.



identify correlations among data types. But soon, by combining them with experimental perturbations, it will be possible to explore causal relationships.

Protein detection by sequencing requires large, pre-optimized panels of antibodies. At present there seems to be no technical limit to the number of antibodies that can be used to detect extracellular proteins via CITE-seq. Mild permeabilization methods that allow antibodies to get inside cells may enable researchers to sequence intracellular proteins in the future. "Such improvements would dramatically increase the power and applicability of this method to other systems," says Maniatis.

As CITE-seq and ECCITE-seq antibodies become increasingly available, Rahman expects these approaches will become rapidly and widely adopted: "If you are already doing single cell RNA-seq, adding in CITE-seq is a pretty minimal investment for the amount of extra data you get."

## REFERENCES

1. Stoeckius, M. *et al. Nat Methods* **14**, 865-868 (2017)
2. Peterson, V. *et al. Nat Biotechnol* **35**, 936-939 (2017)
3. Chin, V. T. *et al. J Clin Onc* **37**, 15\_suppl e20563 (2019).
4. Fernandez, D.M. *et al. Preprint at bioRxiv* <https://doi.org/10.1101/721688> (2019)
5. Mimitou, E. P. *et al. Nat Methods* **16**, 409-412 (2019)

illumina®