illumina®

The Power of Proteomics Via NGS: Study Proteins and RNA in a Single Sample

Proteins form the physical building blocks for cells. They are key players in metabolism, cell structure dynamics, and in the processing of genetic information that allows organisms to survive. Studying the protein profile of cells is central to helping researchers elucidate the functions of these proteins at the cellular level and their impact on health and disease.

Flow Cytometry and mass cytometry (CyTOF) have been the go-to tools for studying and characterizing proteins at the cellular level and on millions of cells at once. **To expand scientific discovery, over the last decade, many researchers began to move from protein detection to RNA analysis because it allows for the simultaneous analysis of more parameters**.

Thus, over the last decade, scientists have used bulk mRNA sequencing (RNA-Seq) to study the transcriptomic profile of cell populations. Furthermore, **the inherent unbiased nature of RNA-Seq makes it a powerful tool for discovery-type research**.

The above-mentioned modalities for studying proteins and RNA have all provided useful insights that advance biomedical knowledge and improve human health. However, standalone studies of the proteome or the transcriptome only present a one-sided picture of complex protein-gene interactions. This presents unique limitations when scientists are studying rare or niche cell populations or when unbiased protein discovery and understanding the genes that express or control that protein is the goal.

This is why researchers from early-stage discovery to translational research are **migrating to an alternative approach that applies well-established Illumina next-generation sequencing (NGS) tools to the study of proteins (proteomics via NGS). Using this combination, scientists can study both the proteins and RNA in a single sample, while greatly increasing the breadth of targets that can be investigated in each experiment**.

1960s	1970s	1990s	2000s	2010s
Flow cytometry	Immunoassays (e.g., ELISA/ Western Blot)	Mass cytometry	Protein-based microarrays	Proteomics via NGS

Figure 1: Major proteomics methods that have emerged over the last fifty years.

The Discovery Challenge

Studying the Proteome

For high throughput profiling of proteins in single cells, scientists require methods that are both sensitive and fast. Flow cytometry supports the study of cellular activity and the expression of multiple proteins simultaneously. This can be applied to several million cells at once. However, even the most powerful flow cytometer only allows around 20-30 antibodies in panel design. While using a mass cytometer may increase that number, researchers are still limited in how many targets they can profile. Thus, limitations arise either because of the flow cytometer or mass cytometer they are using, or by the types of antibodies researchers can optimize for their experiment. **Flow and mass cytometry are limited when it comes to the discovery of new proteins or when samples are rare or limited.**

Studying the Transcriptome

Before RNA-Seq became the tool of choice for transcriptome analysis, researchers depended largely on microarrays. Microarrays allowed researchers to look at a few hundred genes at a time whereas RNA-Seq provides global and unbiased study of the transcriptome. Since mRNA molecules have poly-A tails, a feature of mRNA particularly leveraged by RNA-Seq, **researchers can now easily assay all 20,000+ genes in the human genome in one experiment[1].**

Traditional or bulk RNA-Seq tends to look at the average gene expression of a population of cells. However, since cells within the same tissue are heterogenous[2] , RNA-Seq at the resolution of a single cell (scRNA-Seq) adds important context. Single cell RNA sequencing reveals information about an individual cell's transcriptome – though it can be difficult with cells that have low levels of RNA due to inactivity or senescence. Additionally, while RNA-seq provides invaluable information on gene expression, transcriptomics alone is often incapable of distinguishing between molecularly similar yet functionally distinct cells. Thus, there is differential importance in studying both RNA and protein in these cells and along the cell developmental continuum[3].

The Solution

Scientists need high throughput tools to further understand protein expression, while studying the intricacies of cellular transcriptomes – whether it is at the bulk level or at the resolution of a single cell. **The combined proteomics/ transcriptomics approach allows researcher to discover and study little-known yet crucial proteins.**
Proteomics via NGS allows researchers to discover and study a greater breadth of antibody targets and removes the need for researchers to limit their studies to just a handful of proteins. While most traditional approaches to studying proteins are reductionist and relatively insufficient to understand how proteins interact or function in a biological system, proteomics via NGS provides researchers with deeper insights as they discover and study proteins.

Protein Detection with an NGS Workflow

An example of a proteomics via NGS workflow is Cellular Indexing of Transcriptomes and Epitopes (CITE-Seq)[4]. CITE-Seq incorporates the simultaneous detection of proteins with single-cell RNA sequencing. Antibodies are carefully conjugated to designed oligonucleotides. These are commercially available from vendors such as Biolegend, 10X Genomics and BD Biosciences. See Figure 2 and 3 below. Each antibody-oligonucleotide conjugate contains a PCR handle, an RNA capture sequence and a 15-nucleotide barcode that identifies a unique cell. After the antibody-oligonucleotide conjugate binds to a cell target, the antibody-oligonucleotide-cell unit is encapsulated in single droplets. Using Illumina's NGS systems (Figure 3), the proteo-transcriptomic profiles of individual cells can be measured and analyzed.

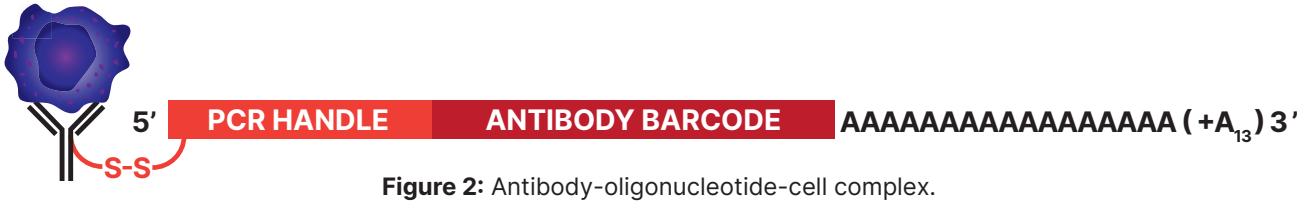

Figure 2: Antibody-oligonucleotide-cell complex.

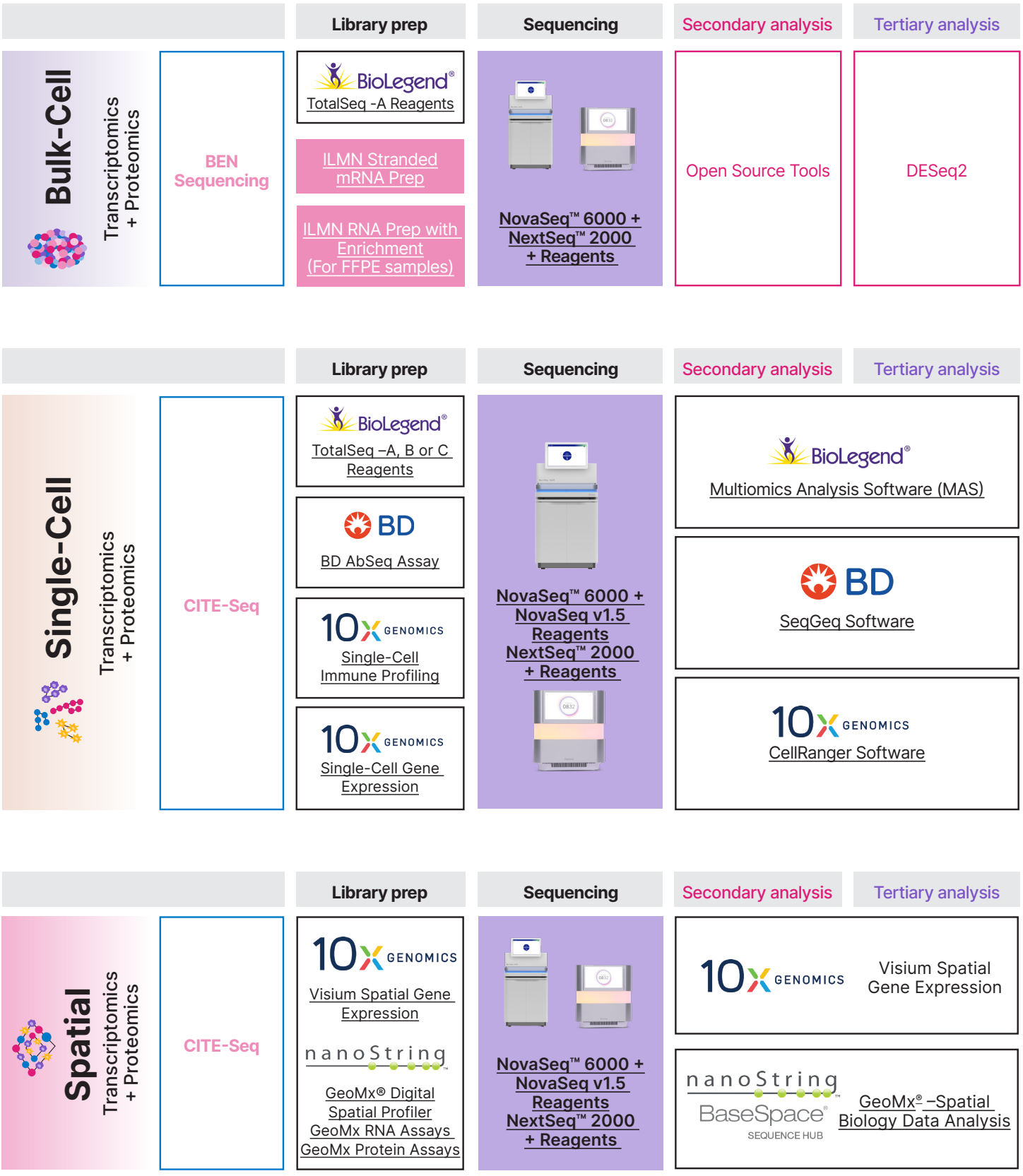

Figure 3: BEN-Seq and CITE-Seq workflows with reagents, instruments, and bioinformatics analysis software.

Experimental Applications

With its system-wide view of biological processes, proteomics via NGS can be used to study key biological processes, including:

- **Tumor biology**[5] – Understanding the protein and transcriptional profile of tumors can help illuminate the underlying causes of cancer. It may also provide insights into responses to specific medications or treatments[6].
- **Biomarker discovery research**[7] – Protein biomarkers continue to provide new insights into disease biology and help clinicians predict patient outcomes. The discovery of these life-saving biomarkers can be expedited using proteomics via NGS, which can profile as many as 280 proteins on a single cell, in a single experiment.
- **Host-pathogen interactions**[8] – Elucidating the interactions between the immune system and pathogens that cause deadly diseases is key to developing vaccines and treatments.

Furthermore, studying **proteomics via NGS can offer the following advantages** to your research:

- **Cost-effective and time-saving flow/mass cytometry panel design** that outperforms current approaches to panel design.
- **Accurate measurement of complex cellular topologies.**
- For researchers interested in enriching a specific cell population, it permits **purification of precisely defined cell states.**

Bringing NGS into a protein detection workflow allows scientists to generate large and novel data sets that break new ground in their respective fields.

Tapping into Existing Infrastructure

You may have everything you need within your reach.

Proteomics via NGS is more accessible than you think. It's highly likely that a core lab on your campus or at a neighboring research institution is already furnished with the Illumina sequencer you need for your experiments.

NGS-Based Proteomics Enters Its Prime

Combining proteomics and NGS is an exciting and increasingly accessible research approach that is now entering the mainstream. The additional data, context, and breadth of targets brings new value to countless applications – either as a standalone tool or alongside traditional methods.

Despite only being available for less than five years, papers that use study proteomics using an NGS approach are rapidly picking up steam.

To date (September 2021), a little **over one hundred publications have used proteomics via NGS** in answering complex scientific questions. Amongst these are seminal papers that have been published in *Nature Methods, Cell, Immunity, Nature Medicine and Current Protocols in Immunology*. A study published in *Cell* demonstrated how profiling proteins from 139 COVID-19 patients using NGS helped to resolve the sharp difference that is observed between mild and moderate disease[9].

Studying the protein profile of cells using NGS puts your research on the cutting edge, and now is the time to get started.

Partnering with Illumina

We're here to support you as you embark on the journey to studying and discovering proteins using Illumina's proteomics via NGS technology.

Our field applications scientists are happy to assist with the planning and design of experiments. When you hit a roadblock, it's not the end. Our field team comes to you, helping you troubleshoot and resolve any problems you run into. We provide guidance on everything from appropriate controls to bioinformatic analysis.

Contact our Illumina team:
https://emea.illumina.com/forms/contact_request.html

[1] Hebenstreit D. (2012). Methods, Challenges and Potentials of Single Cell RNA-seq. *Biology*, 1(3), 658–667. https://doi.org/10.3390/biology1030658

[2] Gutierrez-Arcelus M, et al. (2019). Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nature communications*, 10(1), 687. https://doi.org/10.1038/s41467-019-08604-4

[3] Hao Y, et al. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

[4] Stoeckius M et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9), 865–868. https://doi.org/10.1038/nmeth.4380

[5] Tirosh I, et al. (2019) Deciphering Human Tumor Biology by Single-Cell Expression Profiling. *Annual Review of Cancer Biology*, 3:151-166

[6] Helmink BA, et al. (2020) B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577, 549–555 https://doi.org/10.1038/s41586-019-1922-8

[7] Zugazagoitia J, et al. (2020). Biomarkers Associated with Beneficial PD-1 Checkpoint Blockade in Non-Small Cell Lung Cancer (NSCLC) Identified Using High-Plex Digital Spatial Profiling. *Clinical cancer research* 26(16), 4360–4368. https://doi.org/10.1158/1078-0432.CCR-20-0175

[8] Chattopadhyay PK, et al. (2018). A deadly dance: the choreography of host-pathogen interactions, as revealed by single-cell technologies. *Nature communications*, 9(1), 4638. https://doi.org/10.1038/s41467-018-06214-0

[9] Su, Y, et al (2020). Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell*, 183(6), 1479–1495.e20. https://doi.org/10.1016/j.cell.2020.10.037