

# Illumina SomaSeq™ Discovery

NGS-based solution for  
deeper insights into the  
proteome



Analyze 11K protein  
measurements using a scalable,  
NGS-based proteomics workflow



Go from sample to results in  
< 2.5 days with just 4 hours  
hands-on time with a streamlined,  
automated workflow



Interpret proteomic data with  
integrated analysis via  
DRAGEN™ Protein Quantification and  
Illumina Connected Multiomics

## Introduction

Proteins play a key functional role in human biology, reflecting a real-time snapshot of health and disease states. Insights from proteomics serve as a critical link between genotype and phenotype, enabling a deeper understanding of disease mechanisms and supporting the prediction, monitoring, and prevention of disease progression. High-throughput proteomic assays with next-generation sequencing (NGS)–based readouts enable large-scale proteomic studies from discovery to translation and have the ability to connect genomic and proteomic data sets, accelerating multiomics research.

Illumina SomaSeq Discovery is a comprehensive high-throughput proteomics solution that integrates trusted Illumina sequencing by synthesis (SBS) chemistry with the high sensitivity of SOMAmer™ (slow off-rate modified aptamer) proteomics. This innovative assay uses SOMAmer Reagents for protein capture to achieve high specificity for target proteins.<sup>1</sup> Combining SOMAmer Reagent detection with an NGS-based readout and the bioinformatics capabilities of Illumina data analysis software provides researchers with a streamlined workflow for evaluating 11K protein measurements from a single sample (Figure 1).

## Comprehensive content

SomaSeq Discovery enables 11,000 protein measurements for comprehensive protein identification and quantification across serum, plasma, cell lysate, and tissue homogenate samples. This comprehensive content targets human proteins associated with a broad range of biological pathways and disease areas, including cancer, inflammation, immune biology, cardiometabolic health, and neurological function. The content covers major molecular targets, including receptors, kinases, growth factors, and hormones, spanning secreted, intracellular, and extracellular proteins across more than 200 biological pathways.

All SOMAmer Reagents undergo a rigorous, multi-step qualification process before inclusion in the SomaSeq Discovery assay. During qualification, key performance characteristics such as sensitivity, specificity, cross-reactivity, affinity, and dynamic range are thoroughly evaluated.<sup>1</sup> With the content included in the SomaSeq solution, researchers can gain deep insights into the proteome, enabling the discovery of novel biomarkers, drug targets, and disease mechanisms.

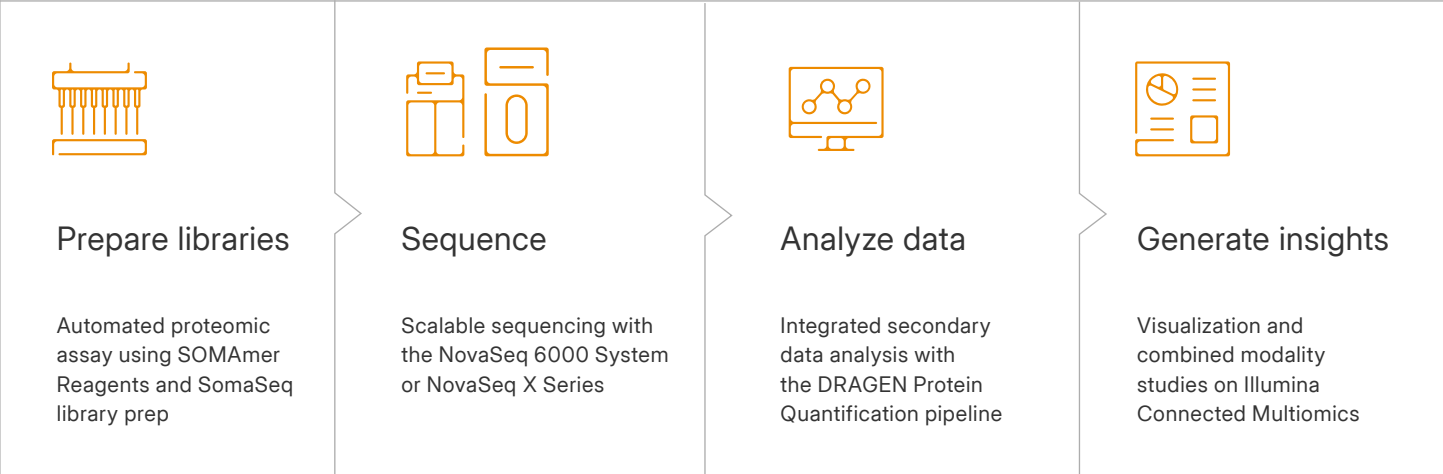


**Figure 1: Overview of the Illumina SomaSeq solution**  
The high-throughput proteomics solution uses SOMAmer Reagents for high-plex protein detection in serum, plasma, cell lysate, or tissue homogenate samples, followed by Illumina library preparation and sequencing on the NovaSeq 6000 System or NovaSeq X Series. Secondary analysis is performed using integrated DRAGEN Protein Quantification and Illumina Connected Analytics, with optional tertiary analysis using Illumina Connected Multiomics for deeper biological insights. The proteomic assay and library preparation workflow is automated on the Illumina SomaSeq Automation System, a custom Tecan Fluent 780 platform, to deliver consistent and reproducible results.

## Expanded sample types and study flexibility

In addition to standard plasma and serum workflows, Illumina offers specialized SomaSeq Discovery kit options for a growing number of biological sample types, enabling researchers to extend proteomic investigations beyond blood-based studies. The additional kit options support proteomic profiling in a broad range of cell- and tissue-derived samples.

For cell lysate and tissue homogenate samples, which often have limited protein abundance or sample quantity, the Illumina SomaSeq Discovery Multi Matrix kit is designed to enable high-content proteomic analysis, helping researchers generate meaningful biological insights from valuable or challenging specimens.

## Sensitive and precise protein detection

SomaSeq Discovery enables reliable detection of low-abundance regulatory proteins, including cytokines, growth factors, and signaling molecules that can be difficult to detect using other proteomic approaches without enrichment. Consistent relative quantification supports comparisons across treatments, timepoints, and cohorts. This is enabled by innovative slow off-rate modified aptamers, or SOMAmer Reagents, which are synthetic monoclonal affinity reagents that provide exceptional sensitivity and specificity for protein detection across a broad dynamic range.<sup>2</sup> These aptamers are chemically modified single-stranded oligonucleotides with hydrophobic functional groups that adopt unique three-dimensional structures and bind specific protein epitopes through complementary shape recognition. Unlike polyclonal antibodies, which have variable performance and specificity across batches, SOMAmer Reagents exhibit favorable binding kinetics, providing femtomolar sensitivity<sup>2</sup> and excellent reproducibility. In serum and plasma samples, the assay demonstrates low median total-run coefficients of variation (CVs) of ~5%–6% (Table 1).

Table 1: Summary of expected performance metrics<sup>a</sup>

| Sample | Median intra-run CV | Median inter-run CV | Median total-run CV | 90 <sup>th</sup> %ile total-run CV |
|--------|---------------------|---------------------|---------------------|------------------------------------|
| Plasma | 4.7%                | 5.3%                | 5.8%                | 8.7%                               |
| Serum  | 4.7%                | 5.0%                | 5.0%                | 7.4%                               |

a. Expected median CV calculated using samples from healthy donors. CV, coefficient of variation.

## Hybridization-based NGS assay

The first step in the SomaSeq Discovery workflow is the SOMAmer proteomic assay, which quantitatively transforms the protein epitope availability in a biological sample into a specific SOMAmer Reagent-based DNA signal.<sup>3</sup> The initial SOMAmer Reagent-protein binding step is followed by a series of bead capture and wash steps that translate relative protein concentrations into corresponding SOMAmer Reagent abundances (Figure 2). Next, SOMAmer Reagents are converted into barcoded sequencing libraries using a hybridization-based approach (Figure 3). Each SOMAmer Reagent has a unique pair of probes, with one enabling capture on magnetic beads and the other carrying a barcode specific to that SOMAmer Reagent. Probe pairs are hybridized to SOMAmer Reagents in an overnight incubation step before being captured on magnetic beads. Unbound probes are washed away, ensuring that the abundance of the barcoded probes is proportional to the relative abundance of the SOMAmer Reagents. Index PCR primers are then added to append sample indexes for sequencing, followed by amplification to generate individually indexed, barcoded libraries. The SomaSeq Discovery assay is fully automated using a single platform, the Illumina SomaSeq Automation System. Samples are pooled and sequenced on the NovaSeq™ 6000 System or NovaSeq X Series.

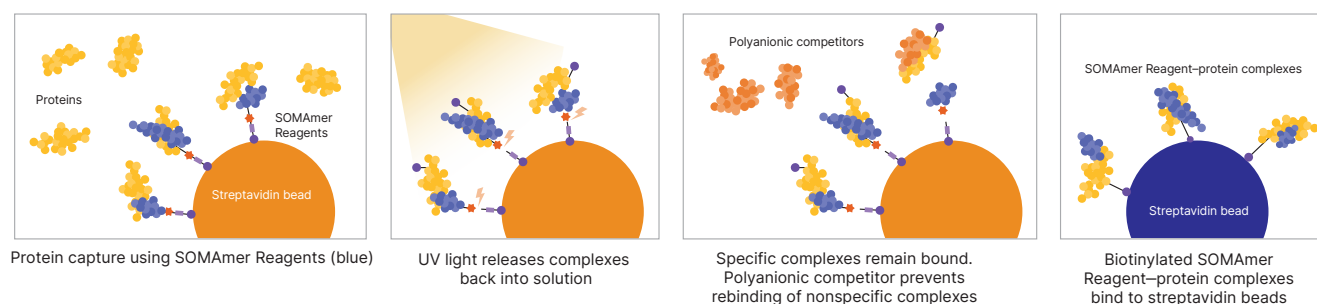


**Figure 2: Protein capture using SOMAmer proteomics**

SOMAmer Reagents (blue), which contain a biotin moiety and a photocleavable linker, are immobilized on streptavidin beads and used to capture target proteins (yellow) from complex biological samples, such as serum, plasma, cell lysate, or tissue homogenates. Unbound proteins are washed away and bound proteins are tagged with biotin. Ultraviolet (UV) light is used to cleave the linker, releasing SOMAmer Reagent-protein complexes into solution. During a subsequent incubation step, nonspecific complexes dissociate while specific complexes remain bound. The inclusion of polyanionic competitors during this step prevents nonspecific rebinding of dissociated proteins. Specific SOMAmer Reagent-protein complexes are then captured on fresh streptavidin beads and eluted for relative quantification by next generation sequencing (NGS).

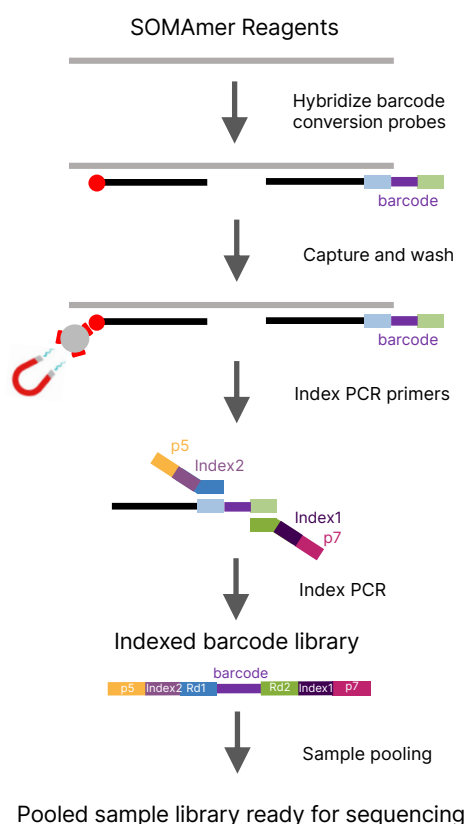


**Figure 3: SomaSeq Discovery NGS conversion chemistry**

Following the SOMAmer proteomic assay, SOMAmer Reagent abundances are prepared as barcoded, sequencing-ready libraries using a hybridization-based conversion process for downstream next-generation

## Scalable, streamlined workflow

The SomaSeq solution uses a streamlined workflow that starts with a range of supported sample types (Figure 1), followed by highly sensitive protein capture using innovative SOMAmer Reagents and Illumina library preparation. The entire workflow is automated on the Illumina SomaSeq Automation System, enabling labs to go from samples to results in less than 2.5 days with ~4 hr hands-on time (Table 2). SomaSeq Discovery kits require input volumes of 35–55  $\mu$ L, depending on the sample type, and demonstrate median intra-assay CVs of ~5.5% for plasma and serum (Table 3).

Libraries are sequenced on the NovaSeq 6000 System (170 samples and 22 controls per run using the S4 flow cell), the NovaSeq X Series (85 samples and 11 controls per run using the 5B flow cell or 170 samples and 22 controls per run using the 10B flow cell), or the NovaSeq X Plus System (340 samples and 44 controls per run using the 25B flow cell).

After sequencing is complete, the DRAGEN Protein Quantification pipeline is autolaunched, generating normalized protein expression counts and quality control reports. The resulting data are then ready for downstream visualization and interpretation.

Table 2: SomaSeq Discovery specifications

| Parameter                  | Specification                 |                              |                               |                               |
|----------------------------|-------------------------------|------------------------------|-------------------------------|-------------------------------|
| Total no. SOMAmer Reagents | 11,000 protein measurements   |                              |                               |                               |
| Dynamic Range              | > 10-log (fM to $\mu$ M)      |                              |                               |                               |
| Total workflow time        | 2.5 days                      |                              |                               |                               |
| Hands-on time              | ~4 hr                         |                              |                               |                               |
| Throughput per flow cell   | NovaSeq 6000 System           | NovaSeq X Series             |                               |                               |
|                            | S4 flow cell                  | 5B flow cell                 | 10B flow cell                 | 25B flow cell                 |
|                            | 170 samples + 22 controls/run | 85 samples + 11 controls/run | 170 samples + 22 controls/run | 340 samples + 44 controls/run |

Table 3: SomaSeq Discovery kit specifications

| Parameter                                                                                                                                                                                                                                                                               | SomaSeq Discovery Plasma or Serum kit | SomaSeq Discovery Multi Matrix kit                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Sample type                                                                                                                                                                                                                                                                             | Plasma or serum                       | Cell lysate; tissue homogenate                                                                                 |
| Input volume                                                                                                                                                                                                                                                                            | 55 $\mu$ L                            | 35 $\mu$ L at 400 $\mu$ g/mL                                                                                   |
| Reproducibility                                                                                                                                                                                                                                                                         | ~5.5% median total CV                 | Cell lysate <sup>a,c</sup> : ~12.4 median total CV<br>Tissue homogenate <sup>b,c</sup> : ~13.6 median total CV |
| <p>a. Determined by HeLa cell line cell lysate analysis.</p> <p>b. Determined by human liver tissue homogenate analysis.</p> <p>c. Performance and reproducibility results will vary depending on the sample type and pre-analysis processing.</p> <p>CV, coefficient of variation.</p> |                                       |                                                                                                                |

## Integrated data analysis

The DRAGEN Protein Quantification pipeline is autolaunched in Illumina Connected Analytics, a secure, cloud-based platform for managing data and running scalable bioinformatics workflows. The pipeline converts sequencing reads into normalized protein expression counts and generates quality control metrics using proteomic assay-specific normalization methods, providing a streamlined and reproducible approach to secondary analysis.

For additional details, see the [Illumina SomaSeq Discovery data normalization technical note](#).

Following secondary analysis, researchers can progress through an interactive data analysis workflow that includes quality assessment, exploratory analysis, differential expression analysis, and biological interpretation (Figure 4).

Illumina Connected Multiomics provides an easy-to-use bioinformatics platform for exploring and interpreting proteomic data. The platform streamlines the analysis of NGS and large-scale multiomics data through an intuitive visual user interface and integrated statistical tools. Researchers can perform quality control review, differential expression analysis, and biological interpretation while generating publication-ready visualizations (Figure 5).

Illumina Connected Multiomics integrates seamlessly with the broader Illumina informatics ecosystem for data management. This workflow simplifies proteomic data analysis and supports integration of proteomic, genomic, and transcriptomic data sets, enabling deeper biological insights from multiomic studies.

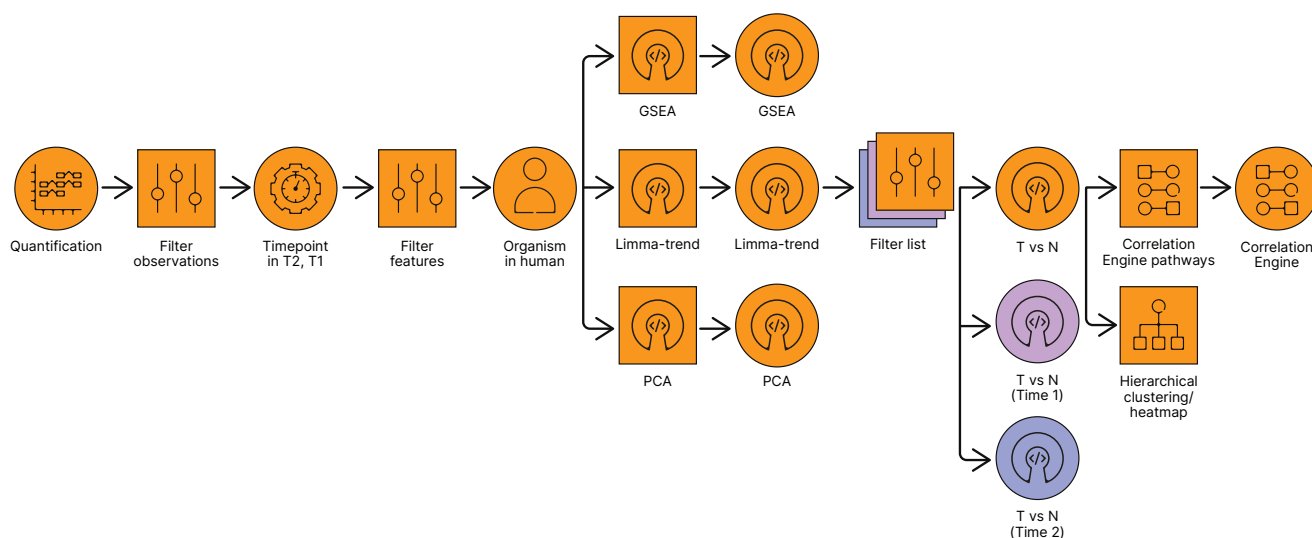
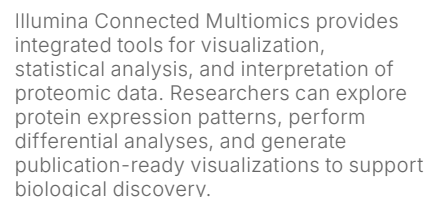


**Figure 4: Protein expression analysis and biological interpretation workflow**

Illumina Connected Multiomics supports protein expression analysis from quality assessment and exploratory analysis through differential expression analysis and biological interpretation. Correlation Engine provides access to curated biological knowledge to help interpret proteins and gene sets in the context of pathways, biological processes, diseases, and prior studies.

GSEA, gene set enrichment analysis; ORA, over-representation analysis; PCA, principal component analysis; T vs N, tumor versus normal.



[Illumina Connected Analytics](#)

For Research Use Only. Not for use in diagnostic procedures.

References

1. SomaLogic. [SomaScan platform. somalogic.com/somascanplatform](https://somalogic.com/somascanplatform). Updated December 2023. Accessed January 2, 2024.

2. Kraemer S, Schneider DJ, Paterson C, et al. [Crossing the halfway point: Aptamer-based, highly multiplexed assay for the assessment of the proteome](#). *J Proteome Res*. 2024;23(11):4771-4788. doi:10.1021/acs.jproteome.4c00411.

3. Gold L, Ayers D, Bertino J, et al. [Aptamer-based multiplexed proteomic technology for biomarker discovery](#). *PLoS One*. 2010;5(12):e15004. doi:10.1371/journal.pone.0015004.

Ordering information

| Product                                                | Catalog no. |
|--------------------------------------------------------|-------------|
| Illumina SomaSeq Discovery v2 Plasma (96 reactions)    | 20167042    |
| Illumina SomaSeq Discovery v2 Serum (96 reactions)     | 20167043    |
| Illumina SomaSeq Discovery Plasma (96 reactions)       | 20137827    |
| Illumina SomaSeq Discovery Serum (96 reactions)        | 20137828    |
| Illumina SomaSeq Discovery Multi Matrix (96 reactions) | 20167045    |
| Illumina SomaSeq Automation System                     | 20116818    |



1.800.809.4566 toll-free (US) | +1.858.202.4566 tel  
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see [www.illumina.com/company/legal.html](https://www.illumina.com/company/legal.html).  
M-GL-02533 v5.0