PrISM - A novel method for single-molecule protein identification and quantification at scale

Introduction

There have been major advances in the field of proteomics and particularly in the use of mass spectrometry to identify and quantify thousands of proteins in the proteome. Despite these advances, researchers can still only routinely measure 8 to 35% of the proteome. Current technologies still have difficulty quantifying proteins across the wide dynamic range of the whole proteome which can span more than 10 orders of magnitude.

In our pre-print titled "A theoretical framework for proteome-scale single-molecule protein identification using multi-affinity protein binding reagents," we describe a method to measure substantively all of the proteome using a technique called Protein Identification by Short Epitope Mapping (PrISM). Here we share some of the key findings from the preprint and invite you to dig deeper in the full manuscript.

Methodology

We modeled an experimental/computational workflow consisting of three phases. In the first phase, individual protein molecules are immobilized onto an array. In the second phase, those molecules are iteratively and repeatedly probed with fluorescently labeled multi-affinity probes (e.g. antibodies or aptamers that bind 2-4 amino acid long epitopes). Unlike traditional affinity reagents that are designed to bind to individual proteins, multi-affinity probes are designed to be promiscuous so that each multi-affinity probe binds to many proteins. In the third phase, the probe binding data, listing which probes bound to each individual position on the array, are fed into a machine learning algorithm, which translates these binding data into protein identities and quantities.

We used simulation to examine the performance characteristics of such a method, and how variations in the method might alter those characteristics. In particular, we used model samples in which proteins were represented at the concentrations they had been observed in experimental studies and varied our methodology to investigate questions like:

- How many cycles of probing are required to comprehensively measure the cellular and circulating human proteomes?
- How many molecules need to be measured to achieve a given dynamic range?

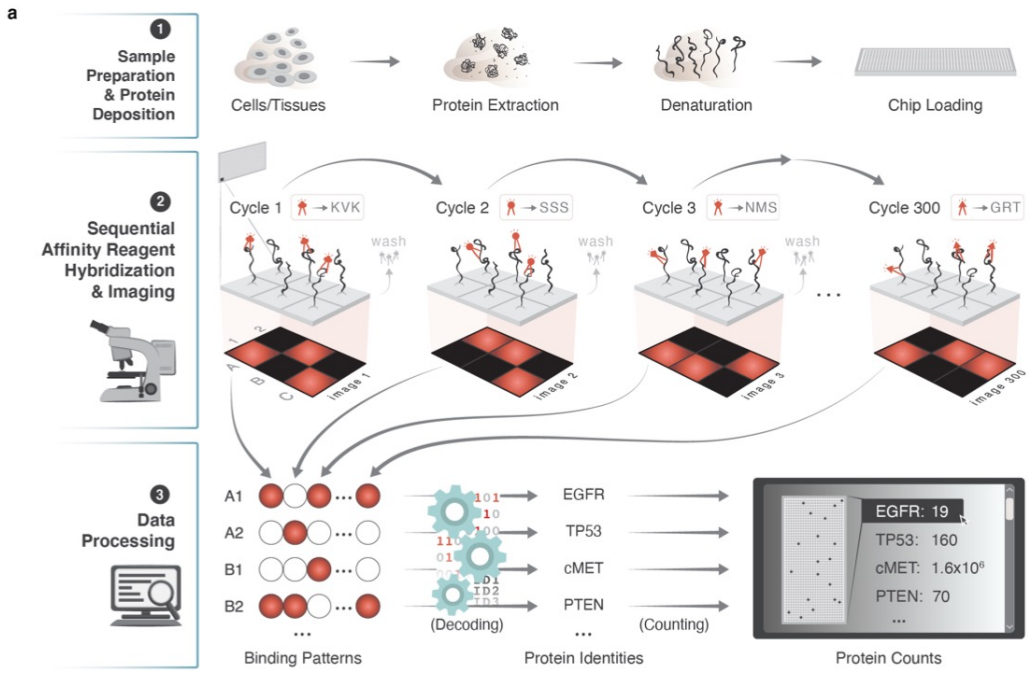

Figure 1. The processes underlying Protein Identification by Short epitope Mapping (PrISM).

Results

Our modeling showed that, using multi-affinity probes with 3 amino acid epitopes, it is possible to identify substantively all of the proteins in the proteome after 300 cycles of probing. This was true for a diverse range of organisms.

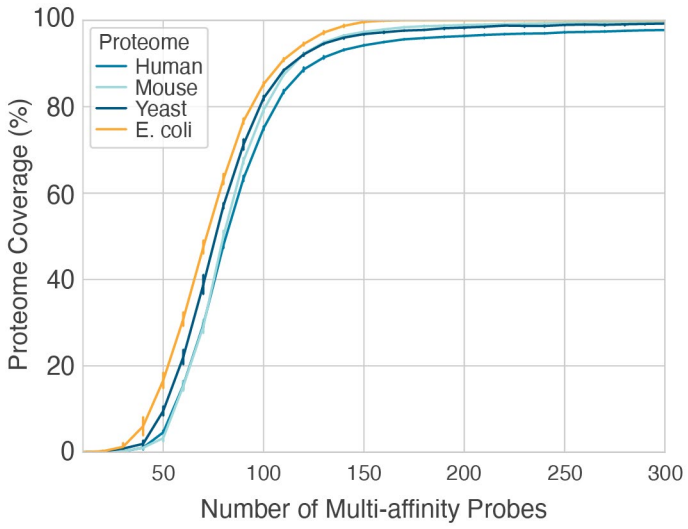

Figure 2. We modeled what percentage of various proteomes could be identified using an increasing number of 3AA-binding multi-affinity probes. With 300 multi-affinity probes, it is estimated we can cover substantively all of the proteome across a variety of organisms.

We additionally determined the relationship between number of molecules measured, assay dynamic range (the ability to measure both high and low abundance proteins in the same assay) and proteome coverage by simulating assays using arrays containing between 1 million and 10 billion molecules. We estimate that, interrogating 10 billion protein molecules would enable the measurement of 92.6% of the proteins in the HeLa cell proteome or 66% of the blood plasma proteome. With these parameters, the dynamic range of detected proteins is 9 orders of magnitude in HeLa cell lysates and 11.5 orders of magnitude in plasma. If 99% of the 20 most abundant proteins are depleted from plasma prior to analysis, coverage increases to 73% with interrogations of 10 billion protein molecules.

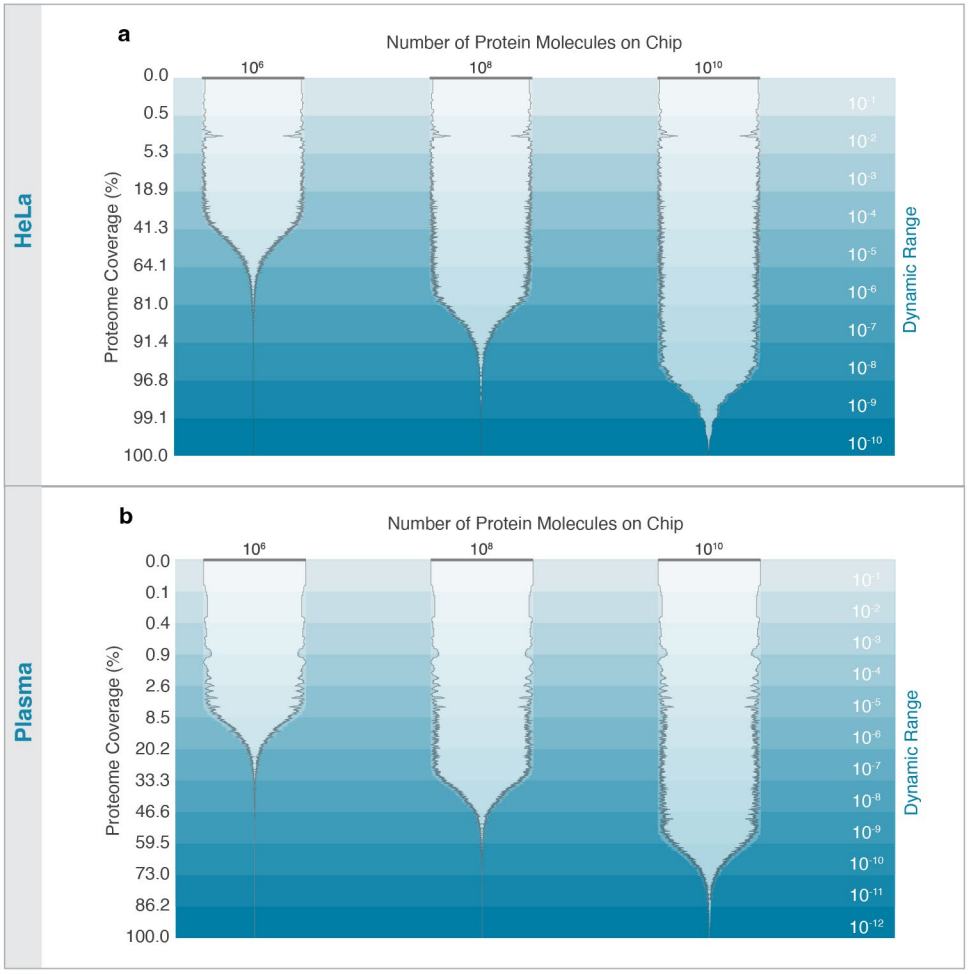

Figure 3. We modeled the dynamic range of protein abundance that we could cover with 300 multi-affinity probes used to bind proteins on arrays with an increasing number of single-molecule (single-protein) landing pads. With 10 billion landing pads, we estimate that detected proteins cover a dynamic range of 9 orders of magnitude in HeLa cells, 11.5 orders of magnitude in plasma. This represents the majority of the proteins in both sample types.

Conclusions

These modeling studies show that it is possible to measure substantively all of the proteome from diverse species with a practical number of cycles of multi-affinity probe binding. We also show that 10 billion landing pads enable coverage of the majority of the proteome in diverse sample types. In other work, we demonstrate that it is possible to create such arrays using nano-scale manufacturing techniques.

By designing our platform with these modeling studies in mind, we believe we can better enable researchers to access the proteome. We hope this will spur a revolution in proteomics and biology that will improve basic research, drug discovery, and life as we know it.

Learn more about PrISM in our preprint.

Learn more at:
nautilus.bio/technology

NAUTILUS™
BIOTECHNOLOGY