Can we create easy-to-use software to capture the complex behaviors of proteins?

Proteomics is the large-scale study of proteins enabled by mass spectrometry technology. Much like a fingerprint can place someone at a scene despite the lack of an eye witness, a mass spectrometer records protein fingerprints that are introduced as peptides or pieces of the protein. The mass spectrometer itself does not identify the proteins from their peptides; instead, highly specialized software programs are required for this complex analysis.

Proteomics can monitor changes in protein abundances between samples such as normal versus diseased cells. One method that measures changes of a cell's proteome in response to a stress is called isobaric tandem mass tagging (IMT). IMT involves tagging all protein peptides from a sample with a unique chemical barcode. One can barcode up to six distinct proteomes for comparison. Barcodes provide two major advantages. Firstly, tagged proteomes can be combined and analyzed simultaneously by the mass spectrometer. If the samples are injected individually, this introduces unnecessary sources of technical error. A second advantage of the barcodes is that intensities are measured and directly correlate with the abundance of their respective proteins. IMT data thus provide both fingerprint and quantitative information to be analyzed by downstream software.

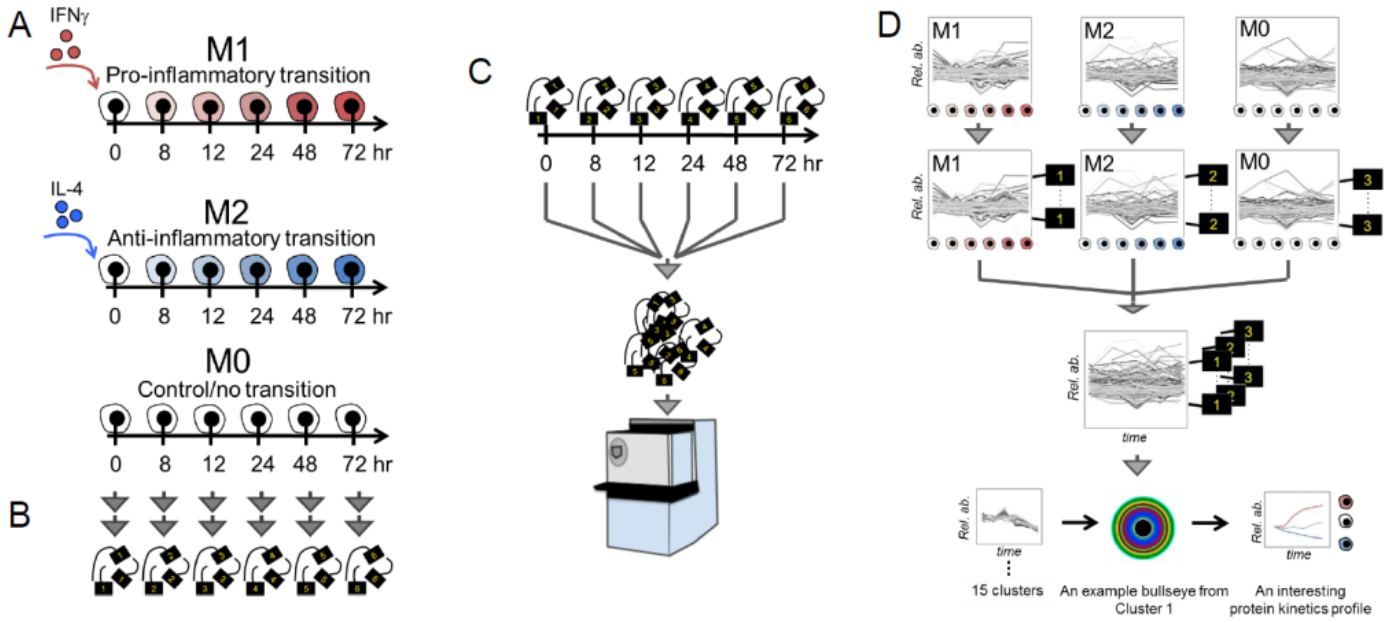

Fig. 1. The barcoding strategy for proteomes (A-C) and proteomics data (D). A, Macrophages were cultured in vitro for 72 hours to induce a pro-inflammatory phenotype (M1) with interferon gamma (IFN?); an anti-inflammatory phenotype (M2) with interleukin-4 (IL-); or a neutral phenotype (no stimulation, M0). B, The arrows indicate the multiple steps for sample collection, protein isolation and chemical barcoding (the tags labeled 1 to 6) pertaining to the IMT strategy. C, The barcoded peptide samples are pooled and analyzed by a mass spectrometer. D, Overview of the steps

involved in barcoding the protein datasets acquired from the three macrophage experiments. The barcoded datasets are combined and analyzed simultaneously by the mIMT-visHTS software.

It is thought that a dominance of pro-inflammatory (M1) over anti-inflammatory (M2) macrophages promotes the chronic inflammation associated with atherosclerosis and metabolic disorders. We measured proteomic changes in macrophages cultured in vitro to mimic these two phenotypes. We also performed a third no stimulation (M0) experiment that served as a noise control. We monitored the proteomes at six time points over three days in the three culture experiments, resulting in three IMT 6-plex experiments (Fig. 1A-C).

Once the IMT data were collected, protein behaviors were clustered by similarity. This step produced 15 clusters for each M1, M2 and M0 dataset. We had the daunting task to screen the clusters of ~4,000 proteins representing ~12,000 kinetic profiles from the three datasets. One can imagine that important trends may be missed due to the large sizes of the datasets. At first glance, we observed that some proteins behaved the same in M0, M1 and M2; others behaved similarly in two of three; whereas others exhibited different kinetics in all three states. However we struggled to represent these various scenarios in a high throughput manner. We then realized that our problem could be fixed if we took a similar approach to that of IMT itself: assign a barcode to each dataset, combine the datasets, and have the software analyze them simultaneously. We envisioned that a single output would facilitate the throughput analysis (Fig. 1D).

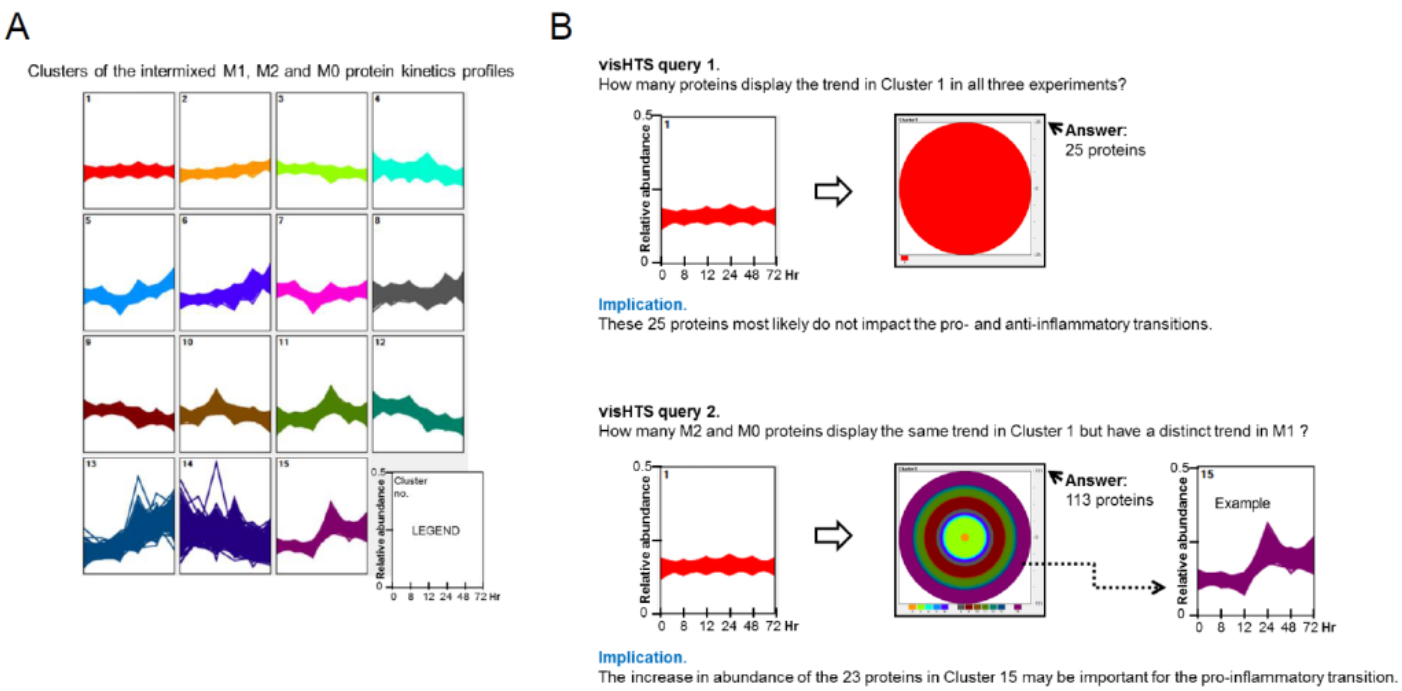

Fig. 2. A closer look at the mIMT-visHTS software. A, The software groups protein abundance

(kinetics) profiles according to similarity. The clusters represent observed protein kinetics in all three macrophage datasets. B, Since the data are barcoded the software can easily assist the user in exploring protein trends between datasets. Two questions a biologist is likely to ask are highlighted. Query 1 pertains to proteins whose abundances remain steady in all three experiments. The single color of the bullseye reflects the unperturbed behavior. Query 2 pertains to proteins whose abundances remain steady in M0 and M2, but are perturbed as a consequence of a pro-inflammatory transition. The multicolored bullseye indicates 113 proteins altered their kinetics to one of the 12 other clusters.

Our resulting software workflow, mIMT-visHTS, refers to the process of multiplexing various IMT datasets and then implementing a visualization graphical user interface for high throughput screening (Fig. 1D and 2). Each bullseye represents one type of macrophage protein behavior. The more colorful the bullseye, the more proteins deviate from a behavior as a consequence of the culture condition. The less colorful the bullseye, the less responsive the proteins are to the different conditions. The protein names can be extracted for follow-up studies. visHTS thus provided a simple way to evaluate the various perturbations to the proteome as a consequence of becoming pro- or anti-inflammatory macrophages. It operates as a screening system, without the need to understand the complicated mathematical operations behind the analysis.

Proteomics is becoming more commonplace. As a consequence, data is increasing rapidly and becoming more complicated to analyze. visHTS is the first step towards generating creative and intuitive output programs necessary for the challenges of today's biologists.

Publication

mIMT-visHTS: A novel method for multiplexing isobaric mass tagged datasets with an accompanying visualization high throughput screening tool for protein profiling.
Ricchiuto P, Iwata H, Yabusaki K, Yamada I, Pieper B, Sharma A, Aikawa M, Singh SA
J Proteomics. 2015 Oct 14