Exploring the word of metabolomics and metagenomics

Module 3: Basic Proteomics Approaches: From Theory to Practice

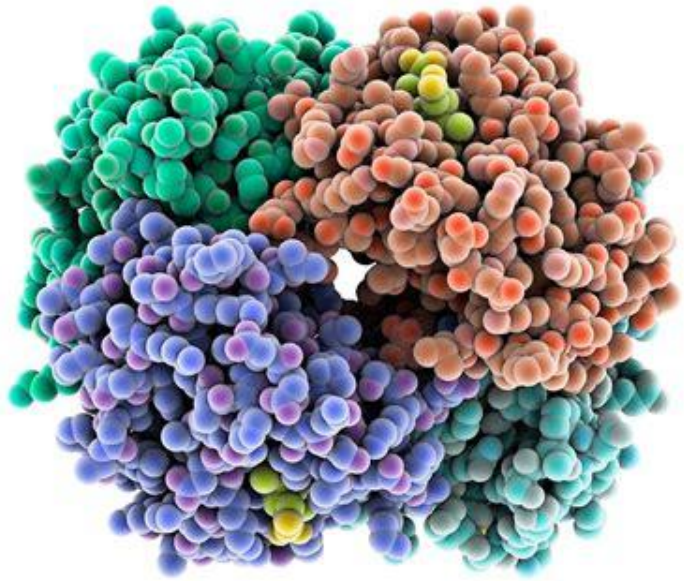
- Tutor: Prof. Flora Guerra

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

What is the Proteome?

- Proteins are biological molecules made up of building blocks called amino acids.
- Proteins are essential to life, serving a range of functions including structural, metabolic, immune, signaling, and regulatory roles, among others (*Tim and Tim, Science Advances, 2020, DOI: 10.1126/sciadv.aax8978)* (https://youtu.be/lijQ3a8yUYQ).



Laguna Design / Science Photo Library / Getty Images (https://www.thoughtco.com/protein-function-373550)

What is the Proteome?

- The term “proteome” was coined by an Australian PhD student, Marc Wilkins, in a 1994 symposium held in Siena, Italy (Wilkins, 2009,. Expert Review of Proteomics, doi.org/10.1586/epr.09.81).

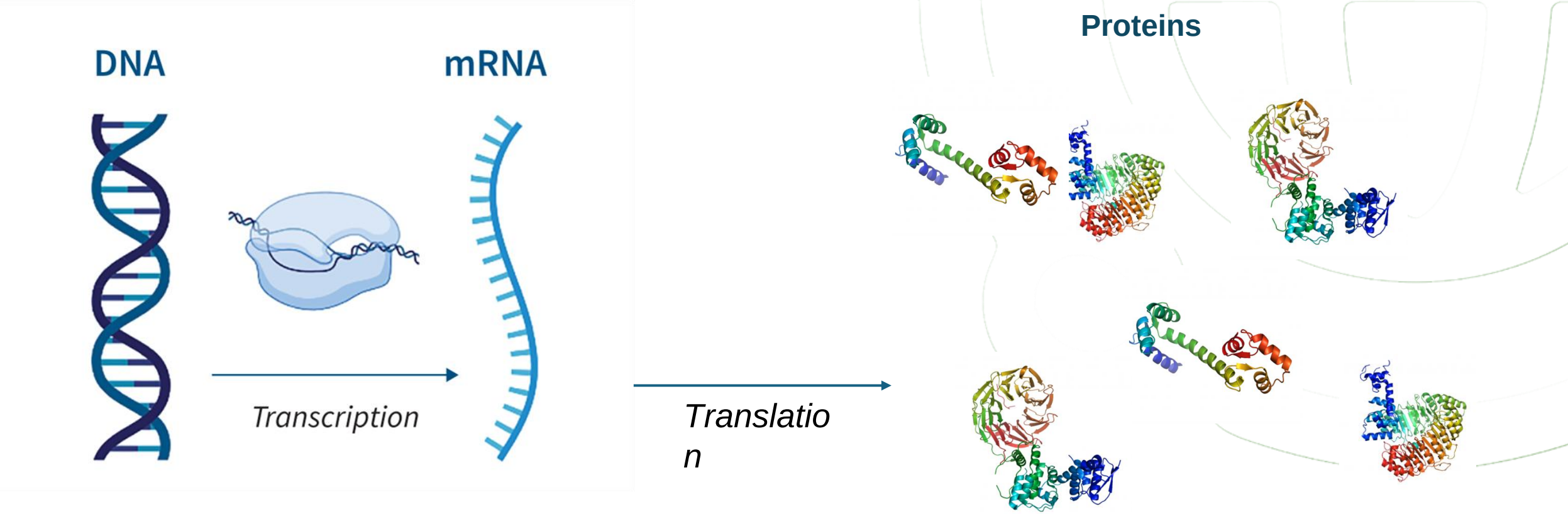
- It is a blanket term that refers to all of the proteins that an organism can express.
- Each species has its unique proteome.

What is the Proteome?

- Unlike the genome (the complete set of genes within each organism), the composition of the proteome is in a constant state of flux over time and throughout the organism *(Beynon, Briefings in Functional Genomics & Proteomics, 2005, https://doi.org/10.1093/bfgp/3.4.382)*

- Therefore, when scientists refer to the proteome, they are also sometimes referring to the proteome at a given point in time (such as the embryo versus the mature organism), or to the proteome of a particular cell type or tissue within the organism.

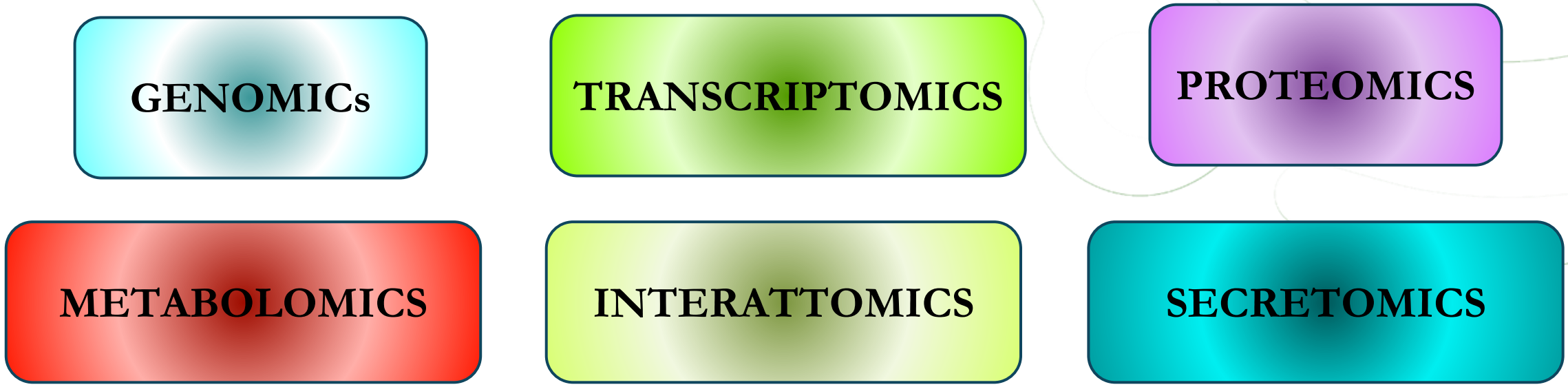
What is the Proteome?



- There are approximately 20,000 genes in the human genome, ~100,000 transcripts in the human transcriptome and over >1000000 proteoforms in the human proteome *(Credit: Technology Networks)*

The "*Omics*" Era

- The suffix «omics» (from the Greek «oma» to indicate all, each, or complete) to be traced back to the introduction of the term *genome*

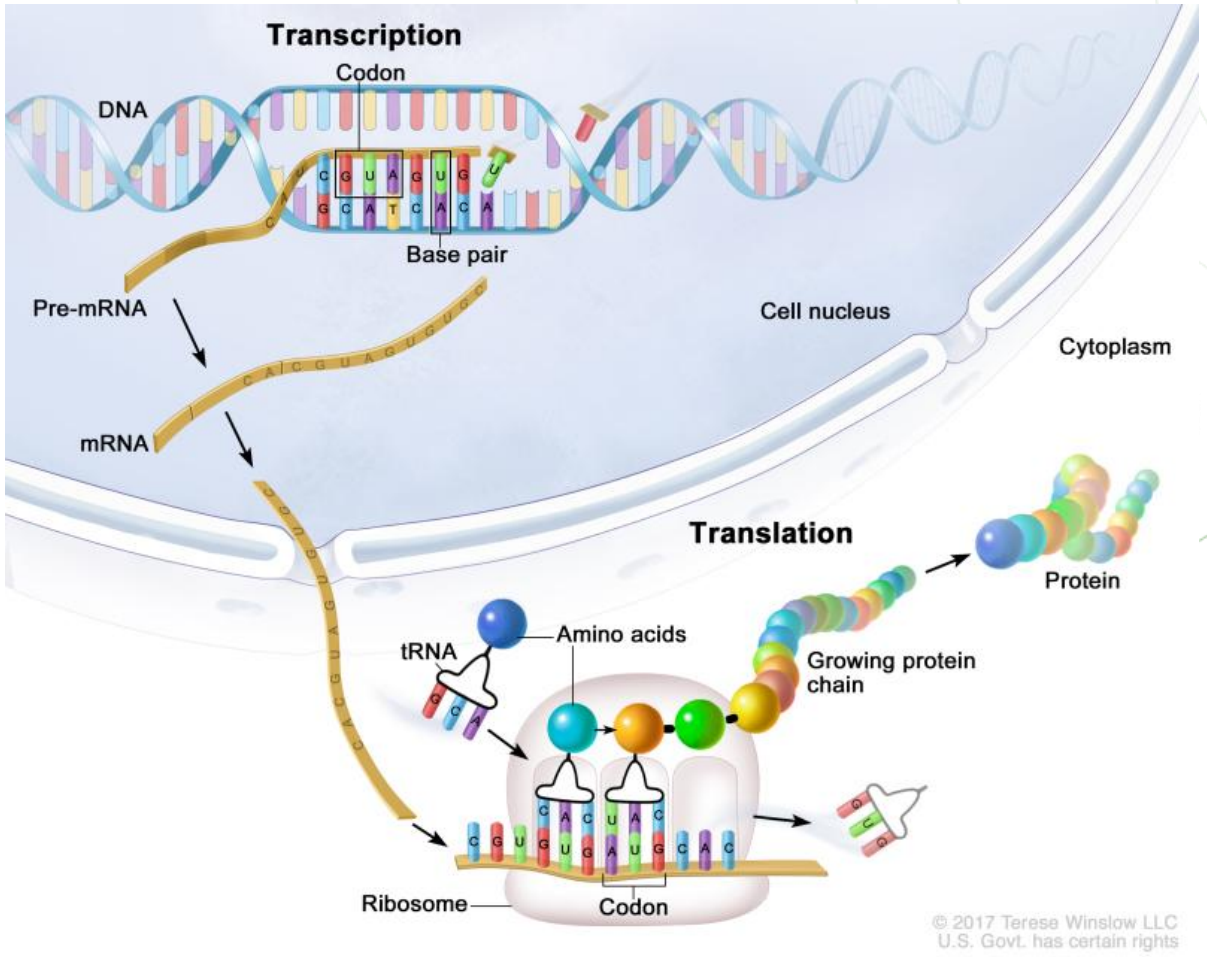


What is Proteomics?

- Proteomics is the study of the proteome, investigating how different proteins interact with each other and the roles they play within the organism

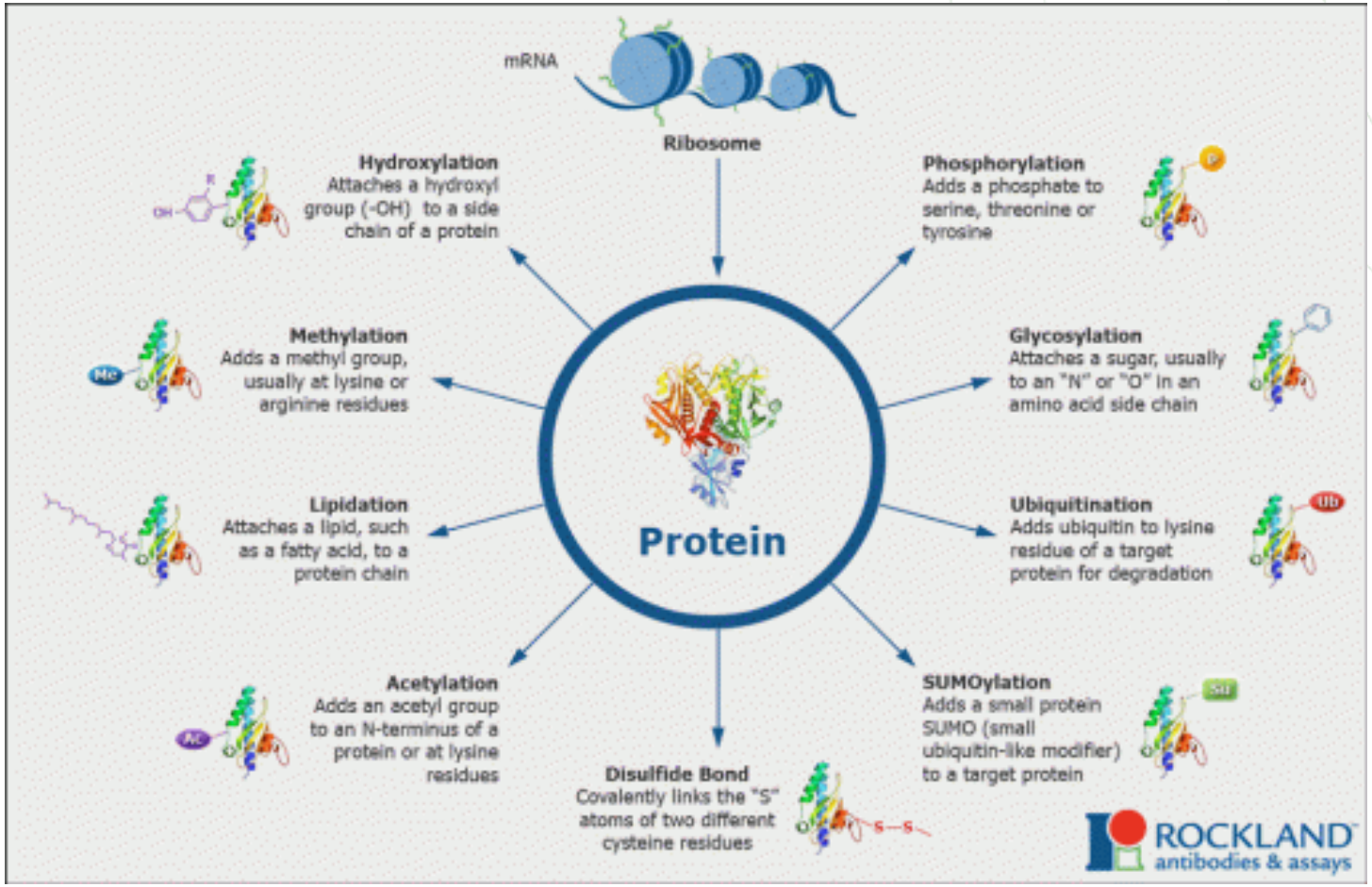
- Although protein expression can be inferred by studying mRNA expression, which serves as the intermediary between genes and proteins, mRNA expression levels do not always correlate well with protein expression levels. Furthermore, the study of mRNA does not consider protein post-translational modifications, cleavage, complex formation and localization, or the many variant mRNA transcripts that can be produced; all of which are key to protein function.

What is Proteomics?



https://www.ncbi.nlm.nih.gov/books/NBK65951/figure/CDR0000460135__5/

What is Proteomics?



https://tcohenlab.web.unc.edu/tau-post-translational-modification-puzzle/post-translational-modifications-ptm/

What is Proteomics?

- The first experiments that fit the label of “proteomic” studies were performed in 1975 with the development of 2D protein electrophoresis (*Graves, 2002, Microbiol Mol Biol Rev. doi.org/10.1128/mmbr.66.1.39-63.2002*)

- However, truly high-throughput identification of multiple proteins per sample only became possible with the development of mass spectrometry (MS) technology over 20 years later (*Andersen, 2000, FEBS Letter, https://doi.org/10.1016/S0014-5793(00)01773-7*)

- Since then, the sensitivity and accuracy of MS have advanced to the point where proteins can be reliably detected down to the attomolar range (1 target protein molecule per 10^{18} molecules), and various other proteomic techniques have been developed and optimized.

What are the key questions that proteomics can answer?

- Broadly speaking, proteomic research provides a global view of the processes underlying healthy and diseased cellular processes at the protein level.
- To do this, each proteomic study typically focuses on one or more of the following aspects of a target organism's proteome at a time to slowly build on existing knowledge:

What are the key questions that proteomics can answer?

- **Protein identification**:

Which proteins are normally expressed in a particular cell type, tissue or organism as a whole, or which proteins are differentially expressed?

- **Protein quantification:**

Measures total (“steady-state”) protein abundance, as well as investigating the rate of protein turnover (i.e., how quickly proteins cycle between being produced and undergoing degradation).

What are the key questions that proteomics can answer?

Protein localization:

Where a protein is expressed and/or accumulates is just as crucial to protein function as the timing of expression, as cellular localization controls which molecular interaction partners and targets are available.

- **Post-translational modifications:**

Post-translational modifications can affect protein activation, localization, stability, interactions and signal transduction among other protein characteristics, thereby adding a significant layer of biological complexity.

What are the key questions that proteomics can answer?

- **Functional proteomics**:

This area of proteomics is focused on identifying the biological functions of specific individual proteins, classes of proteins (e.g., kinases) or whole protein interaction networks.

- **Structural proteomics:**

Structural studies yield important insights into protein function, the “druggability” of protein targets for drug discovery, and drug design.

What are the key questions that proteomics can answer?

- **Protein-protein interactions**:

Investigates how proteins interact with each other, and which proteins interact, and when and where they interact.

Proteomics: Techniques

- **Low-throughput methods**
  - Antibody-based
  - Gel-based
  - Chromatography-based
- **High-throughput methods**
  - Analytical, functional and reverse-phase microarrays
  - Mass spectrometry-based proteomics

Proteomics: Techniques

- **Low-throughput methods**
  - **Antibody-based:**

Techniques such as ELISA (enzyme-linked immunosorbent assay) and western blotting rely on the availability of antibodies targeted toward specific proteins or epitopes to identify proteins and quantify their expression levels.

Proteomics: Techniques

- **Antibody-based methods:**
  - **ELISA:** An enzyme-linked immunosorbent assay (ELISA) is a qualitative or quantitative test that uses antibodies to bind and measure a molecule of interest. Similar to other immunoassays, both monoclonal and polyclonal antibodies can be utilized to identify the analyte (eg, peptides, proteins, antibodies, small molecules). The antibody provides the specificity for the analyte, and a moiety, such as horseradish peroxidase (HRP) is either directly or indirectly coupled to the antibody in order to provide the detection method and possible signal amplification.
  - Types of ELISA:
    - **Direct**
    - **Indirect**
    - **Sandwich**
    - **Competition/Inhibition**

Proteomics: Techniques

- **Direct ELISA**

•Enzyme conjugated to antibody that binds to antigen on a surface.

- **Indirect ELISA:**

•Similar to direct ELISA, but the antibody is not conjugated.

A second conjugated antibody

is used to detect the bound antibody.

Proteomics: Techniques

- **Sandwich ELISA**
  - The antigen is recognized by 2 antibodies, making a complex like a sandwich. One antibody is used for capture and one is used for detection. The detection can be direct or indirect..

Proteomics: Techniques

- **Competition/inhibition ELISA**
- Similar to a direct ELISA, but the quantitation is performed by competing or inhibiting the antibody binding with a measured amount of antigen.

Proteomics: Techniques

- **Antibody-based methods:**
  - **Western Blotting:** Western blot, also known as immunoblotting, is the process of separating proteins and identifying them in a complex biological sample
  - The use of polyacrylamide gel electrophoresis is a prerequisite for western blotting in order to separate proteins prior to their identification.

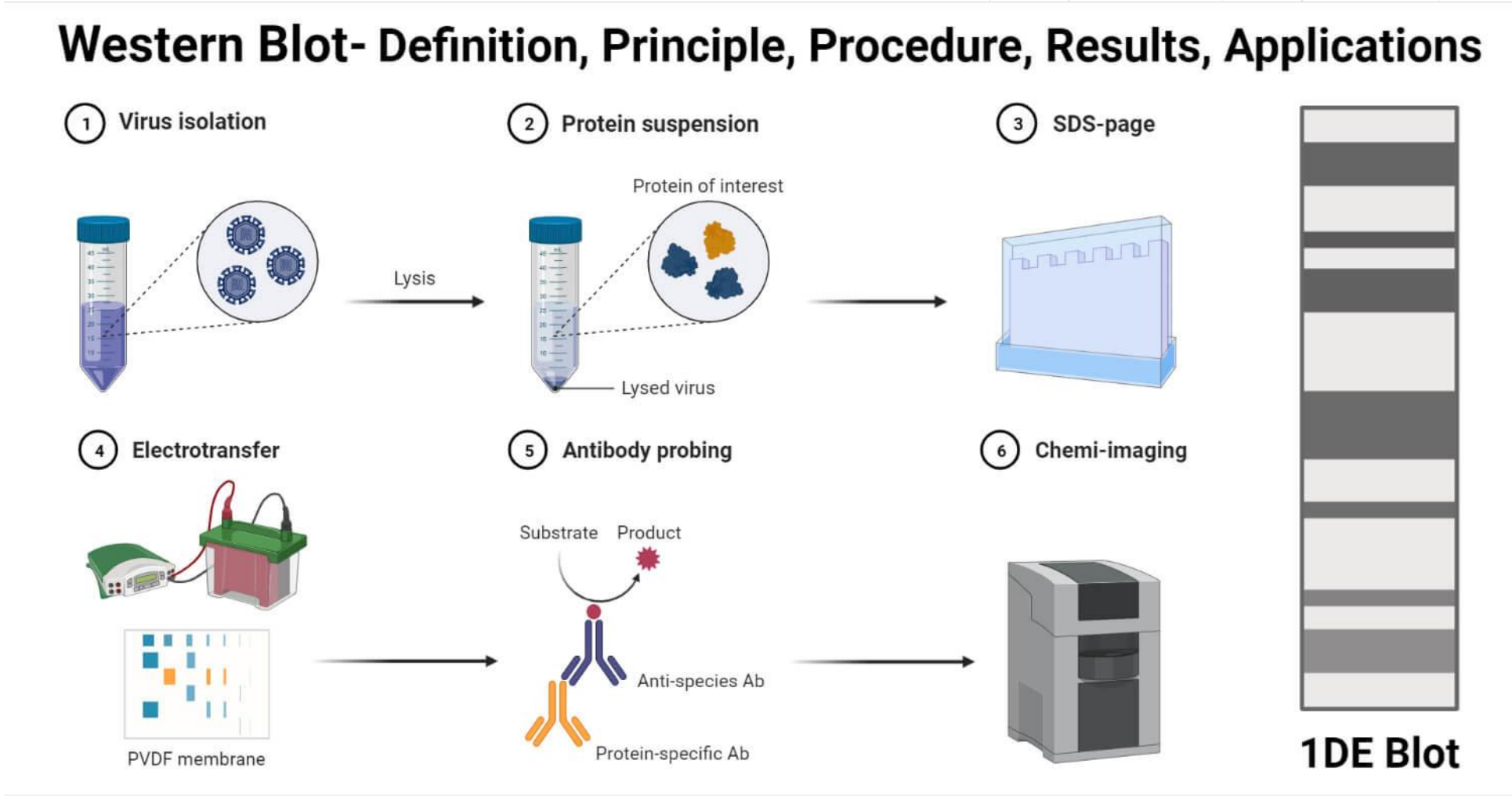
Proteomics: Techniques

- **Antibody-based methods:**
  - **Western Blotting:** The process of western blotting involves the transfer of proteins separated by SDS PAGE into an absorbent membrane. The proteins can then be identified on the membrane by different means.
  - The principle of western blotting is the interaction between the proteins and the probes used for the detection of the proteins.
  - The proteins used for western blotting are separated by gel electrophoresis to obtain them on a gel matrix.
  - The proteins are then transferred to a nitrocellulose or polyvinylidene fluoride (PVDF) membrane, where they are immobilized. The transfer of the protein is known as *blotting.*

Proteomics: Techniques

- **Antibody-based methods:**
  - **Western Blotting:**
  - The protein on the membrane can either be detected by the use of a reporter-labeled primary antibody directed against the protein or a reporter-labeled secondary antibody directed at the primary antibody.
  - The reporter or probe present on the antibody can be an enzyme that produces a color reaction or a luminescent signal at the antigen-antibody binding site that produces a fluorescent signal in the presence of a particular substrate.
  - The signal or color generated by the probe requires a detection system that is appropriate for the signal or intensity generated.

Proteomics: Techniques



https://microbenotes.com/western-blot/

Proteomics: Techniques

- **Low-throughput methods**
  - **Gel-based:**

Two-dimensional gel electrophoresis (2DE or 2D-PAGE), the first proteomic technique, utilizes an electric current to separate proteins in a gel based on their charge (first dimension) and mass (second dimension).

Differential gel electrophoresis (DIGE) is a modified form of 2DE that uses different fluorescent dyes to allow the simultaneous comparison of two to three protein samples on the same gel. These gel-based methods are used to separate proteins before further analysis by e.g., mass spectrometry (MS), as well as for relative expression profiling.

Proteomics: Techniques

- **Low-throughput methods**
  - **Gel-based:**
    - **Two-dimensional gel electrophoresis (2DE or 2D-PAGE):**
      - First dimension: Isoelectric Focusing
      - Second dimension: SDS-PAGE

Proteomics: Techniques

- **Gel-based:**
  - Two-dimensional gel electrophoresis (2DE or 2D-PAGE):
    - **First dimension: Isoelectric Focusing**

The fundamental principle behind IEF lies in the concept of the **isoelectric points (pI)** of proteins. The pI is the specific pH at which a protein has an overall net charge of zero. At this pH, the positive and negative charges on the protein are balanced, meaning that the protein will not move in an electric field because there is no net electrostatic force acting on it.

Proteins have varying numbers of acidic and basic residues, which impart different charge densities. As the pH of the environment changes, so does the net charge of the protein, leading to its migration in an electric field. This characteristic allows for the separation of proteins with different pIs when a pH gradient is established.

https://www.youtube.com/watch?v=mkMPx49QZtw

Proteomics: Techniques

- **Low-throughput methods**
  - **Chromatography-based:**

Chromatography-based methods can be used to separate and purify proteins from complex biological mixtures such as cell lysates. For example, ion-exchange chromatography separates proteins based on charge, size exclusion chromatography separates proteins based on their molecular size, and affinity chromatography employs reversible interactions between specific affinity ligands and their target proteins (e.g., the use of lectins for purifying IgM and IgA molecules). These methods can be used to purify and identify proteins of interest, as well as to prepare proteins for further analysis by e.g., downstream MS.

Proteomics: Techniques

- **Low-throughput methods**

Chromatography-based:

Affinity Chromatography, sometimes called bio-specific adsorption, is a separation method based on a specific binding interaction between an immobilized ligand and its binding partner.
Examples include antibody/antigen, enzyme/substrate, and enzyme/inhibitor interactions. This technique is well suited to large molecules and consequently is popular for protein purification.

Proteomics: Techniques

- **High-throughput methods**
  - **Analytical, functional and reverse-phase microarrays:**

Protein microarrays apply small amounts of sample to a “chip” for analysis (this is sometimes in the form of a glass slide with a chemically modified surface).

Specific antibodies can be immobilized on the chip surface and used to capture target proteins in a complex sample.

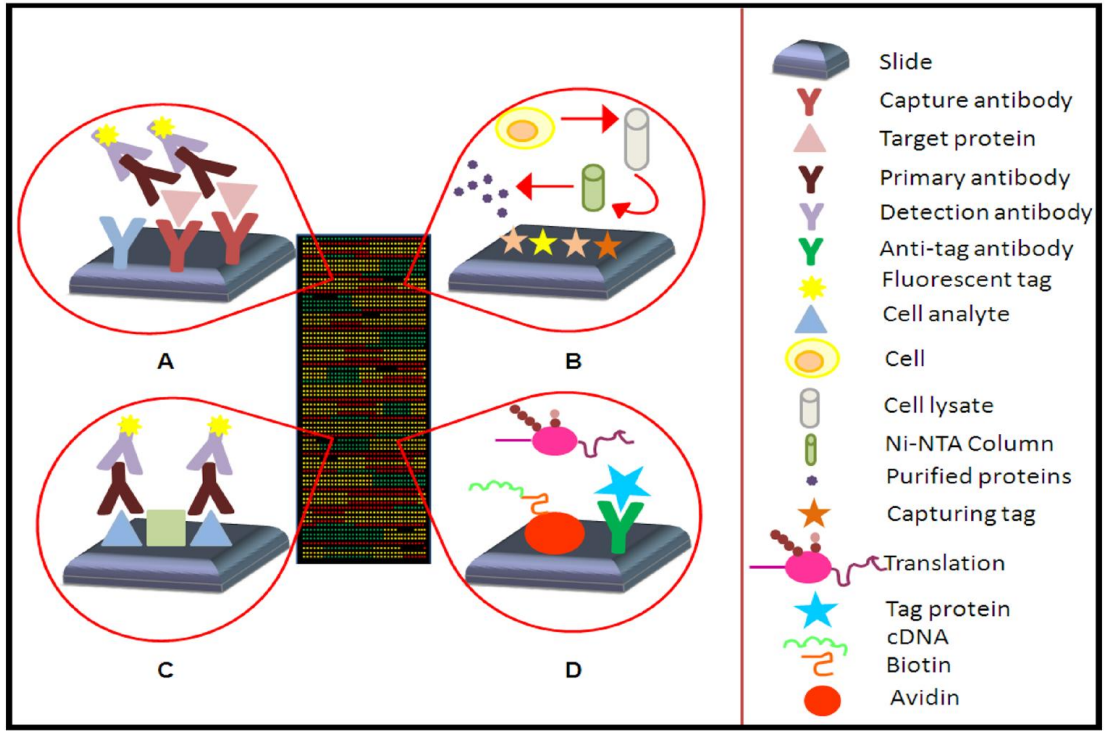
This is termed an analytical protein microarray, and these types of microarrays are used to measure the expression levels and binding affinities of proteins in a sample.

Functional protein microarrays are used to characterize protein functions such as protein–RNA interactions and enzyme-substrate turnover.

In a reverse-phase protein microarray, proteins from e.g., healthy vs. diseased tissues or untreated vs. treated cells are bound to the chip, and the chip is then probed with antibodies against the target proteins.

Proteomics: Techniques

- **High-throughput methods**
  - **Analytical, functional and reverse-phase microarrays:**



Diez et al., 2012 https://doi.org/10.3390/microarrays1020064

Types of different microarrays. (**a**) Capture arrays. (**b**) Cell-based protein microarrays. (**c**) Reverse phase arrays. (**d**) Cell-free nucleic acid programmable protein array.

Proteomics: Techniques

- **High-throughput methods**
  - **Mass spectrometry-based proteomics**:

There are several “gel-free” methods for separating proteins, including isotope- coded affinity tag (ICAT), stable isotope labeling with amino acids in cell culture (SILAC) and isobaric tags for relative and absolute quantitation (iTRAQ). These approaches allow for both quantitation and comparative/differential proteomics.

There are also other, less quantitative techniques such as multidimensional protein identification technology (MudPIT), which offer the advantages of being faster and simpler. Other gel-free, chromatographic techniques for protein separation include gas chromatography (GC) and liquid chromatography (LC) *(Bilal et al., 2017, Journal of Chromatographic Science, https://doi.org/10.1093/chromsci/bmw167)*

Mass spectrometry workflow (1)

- Regardless of how the protein sample is separated, the downstream MS workflow comprises three main steps:
  1. The proteins/peptides are ionized by the ion source of the mass spectrometer.
  2. The resulting ions are separated according to their mass-to-charge ratio by the mass analyzer.
  3. The ions are detected.

Mass spectrometry workflow (2)

- When using gel-free techniques upstream of MS such as iTRAQ or SILAC, the samples are used directly for input into the mass spectrometer.
- When using gel- based techniques, the protein spots are first cut out of the gel and digested before being either separated by LC or directly analyzed by MS.
- There are two main ionization sources, namely:

  - Matrix-assisted laser desorption/ionization (MALDI)•
  - Electrospray ionization (ESI)

Top-down proteomics vs. bottom-up proteomics

- In top-down proteomics, proteins in a sample of interest are first separated and then individually characterized.

- With bottom-up proteomics – also termed “shotgun” proteomics – all the proteins in the sample are first digested into a complex mixture of peptides, and these peptides are then analyzed to identify which proteins were present in the sample *(Zhang, Chem. Rev. 2013, https://doi.org/10.1021/cr3003533)*

Top-down proteomics

Proteins in a sample of interest are first separated before being individually characterized

- **Methods:**

Protein separation is performed based on mass and charge, using techniques such as 2D electrophoresis, DIGE, or mass spectrometry (MS). When employing 2D electrophoresis techniques, proteins are first resolved on a gel and then individually digested into peptides, which are analyzed by a mass spectrometer.

When using MS directly, the undigested sample containing whole proteins is injected into the mass spectrometer, where the proteins are separated. Individual proteins are then selected for digestion and a further round of MS for analysis of the peptides.

Bottom-up proteomics shotgun proteomics

All the proteins in the sample are first digested into a complex mixture of peptides, and these peptides are then analyzed to identify which proteins were present in the sample.

- **Methods:**

Proteins are first digested, and the digested peptide mixture is fractionated and subjected to MS, frequently in an LC-MS/MS configuration.

The resulting peptide sequences are compared to existing databases using automated search algorithms.

These search engines match the experimentally obtained peptide spectra to the predicted spectra of proteins produced by in silico digestion (this is called “peptide-spectrum matching”).

There are several different bottom-up workflows possible, including data-dependent and data-independent methods, as well as hybrids of these.

Data analysis in proteomics

- Proteomic studies, particularly those employing high-throughput technologies, can generate huge amounts of data. In addition to the sheer quantity of data produced, proteomic data analysis can also be relatively complex for certain techniques, such as shotgun MS.
- Adding to this complexity is the range of bioinformatics tools available for proteomic analyses.
- Proteomic studies often require multiple data processing and analysis steps that need to be performed in a specific sequence. To address this need, researchers are increasingly assembling the needed scripts, tools and software into customized proteomic analysis pipelines suited to their particular research questions.

Applications of proteomics

- Personalized Medicine
- Biomarker Discovery
- Drug discovery and development
- System biology
- Agriculture
- Food Science
- Paleoproteomics
- Astrobiology

Applications of proteomics

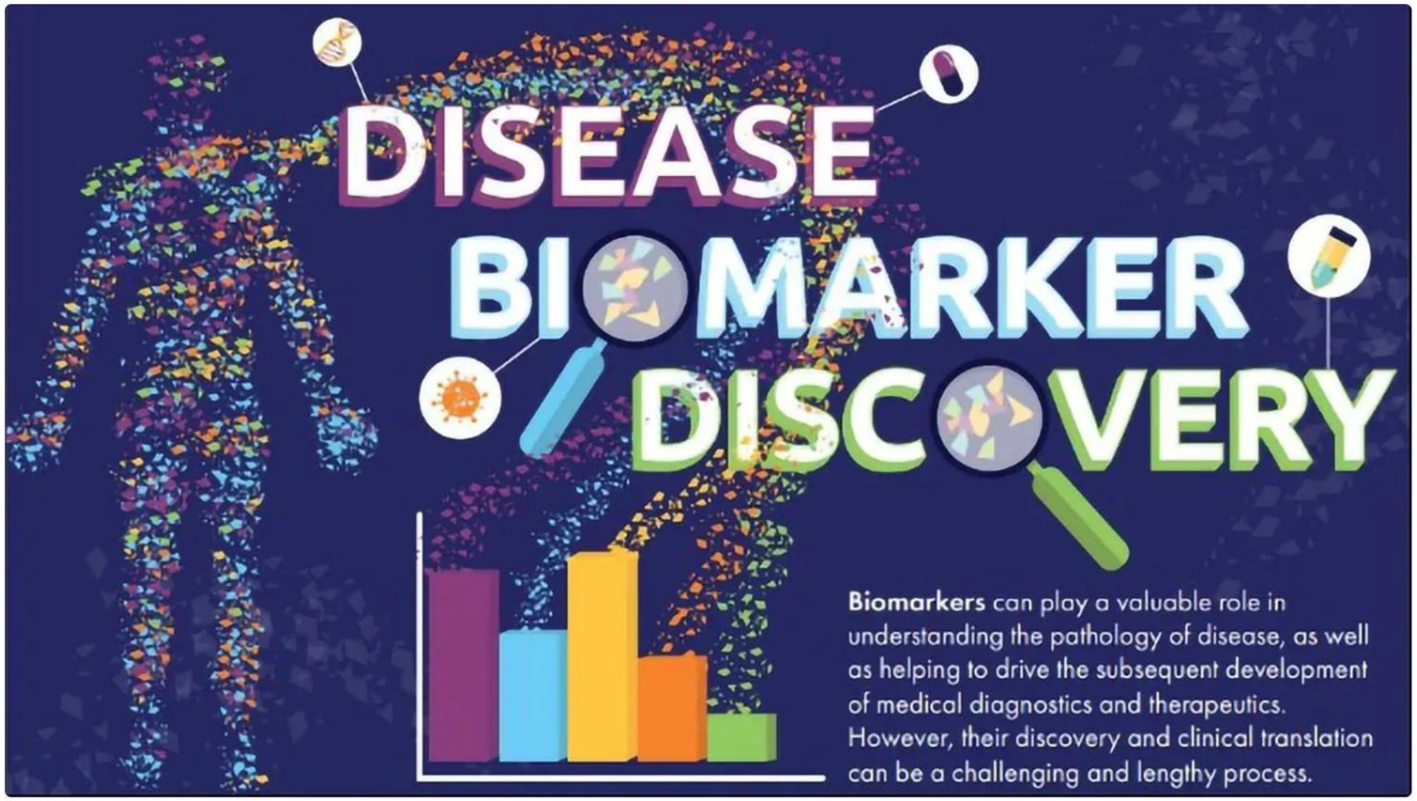
- Personalized Medicine

Tailoring disease treatment to each patient based on their genetic and epigenetic makeup, to improve efficacy and reduce adverse effects. While genomics and transcriptomics have been the main focus of such studies to date, proteomics data will likely add a further dimension for patient-specific management.

https://www.technologynetworks.com/genomics/lists/personalized-medicine-precision-patients-and-promises-in-2019-328583?

Applications of proteomics

- Biomarker Discovery
- Identification of protein markers e.g., for the diagnosis and prognosis of glioblastoma, and evaluating patients' response to therapeutic interventions such as stem cell transplantation.



https://www.technologynetworks.com/proteomics/infographics/disease-biomarker-discovery-340899

Applications of proteomics

- Biomarker Discovery

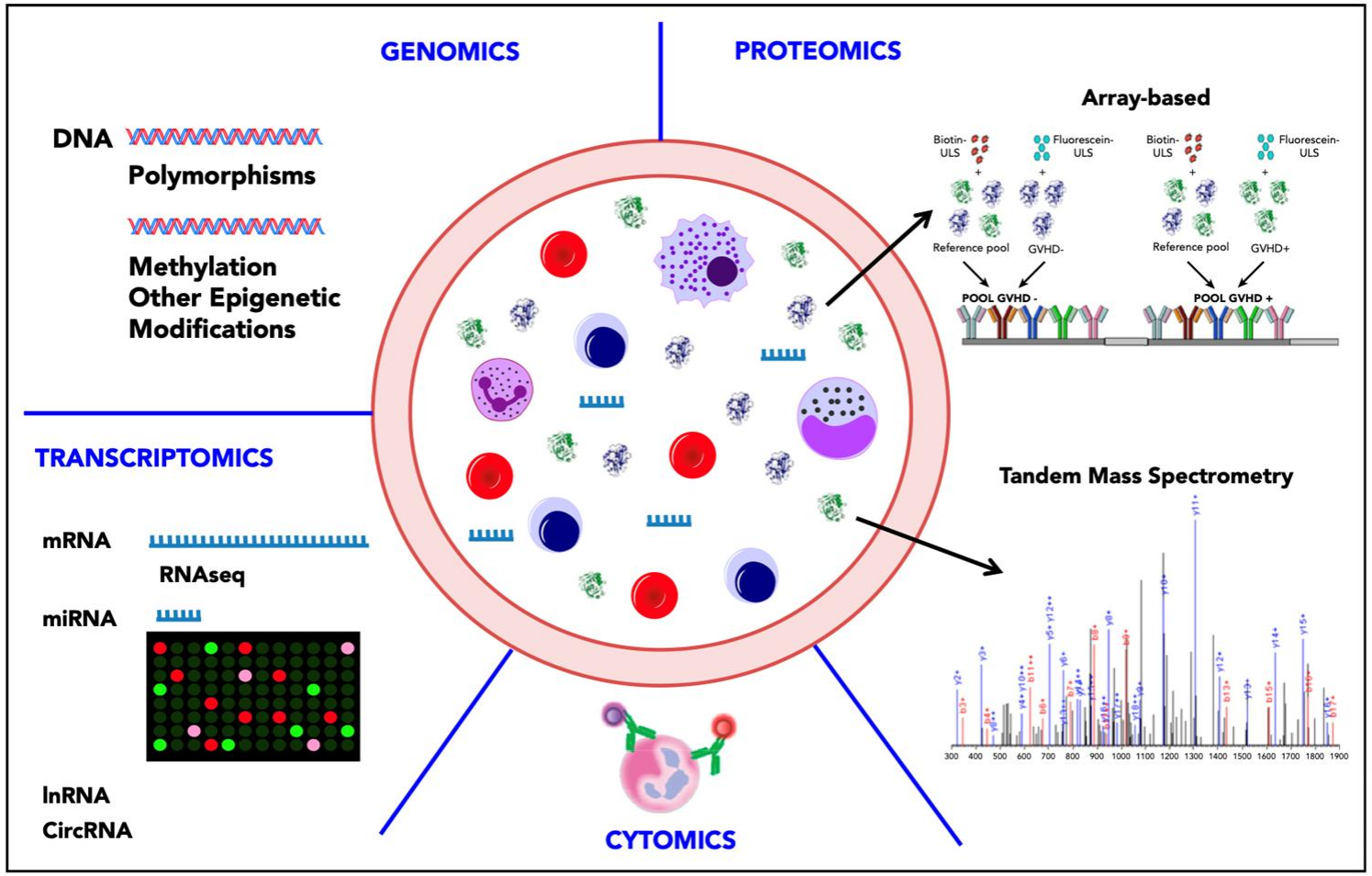
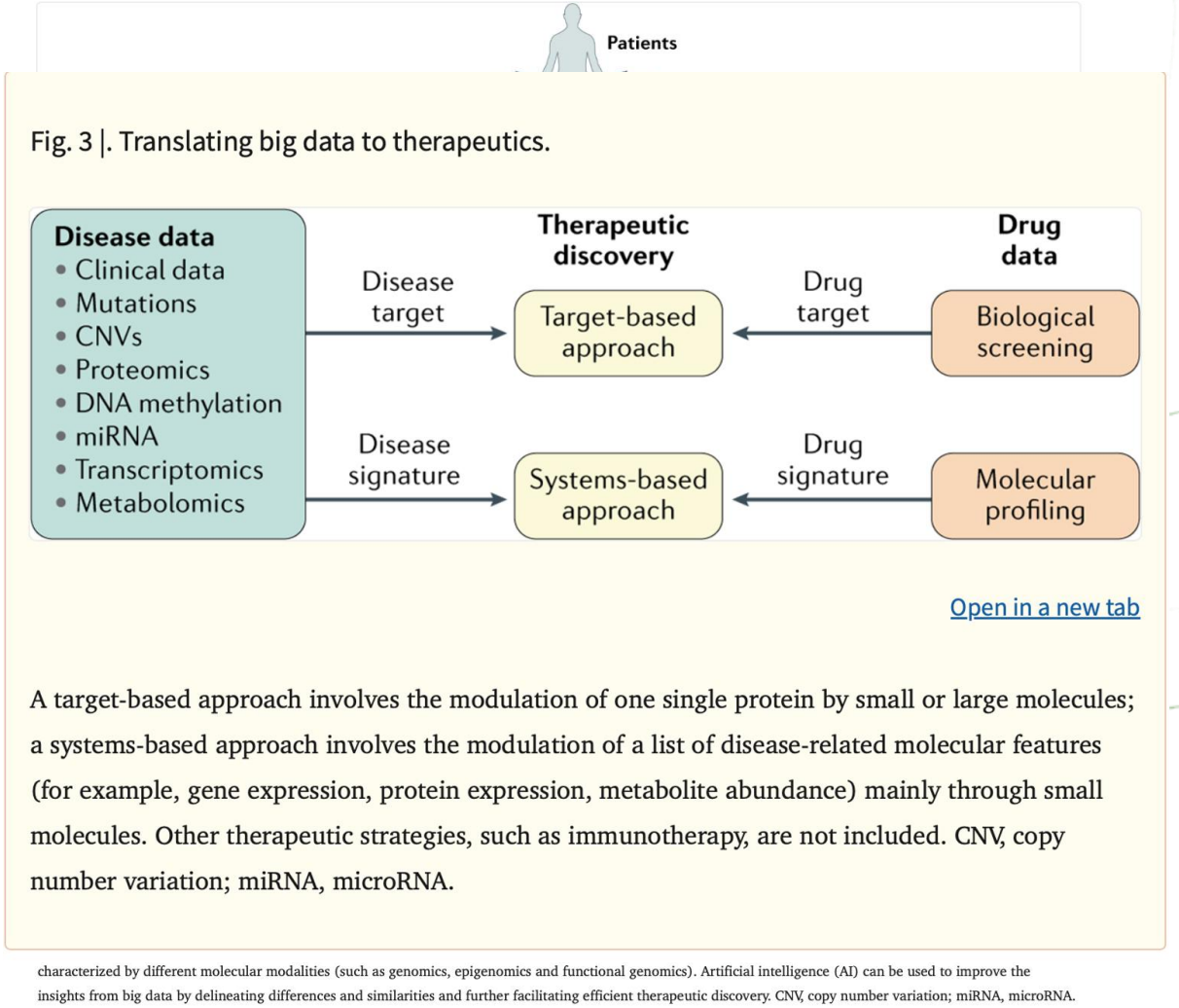


Figure 1. How can we assess GVHD through blood omics? The number of high-throughput omics tools that can be used to profile the human immune system in the blood is increasing rapidly. Genomic approaches consist of measuring abundance of cellular RNA and also microRNAs (miRNAs) that are present in cells or in the serum/plasma. Other genomic approaches consist of determining gene sequence and function (eg, genome-wide association studies, RNA-interference screens, exome sequencing, next-generation whole-genome sequencing). Proteomic assays are used to determine antibody specificity or measure changes in serum/plasma levels of cytokines or chemokines using multiplex assays or global profiles through tandem-mass spectrometry. Cellular profiling assays or cytomics are used to phenotype immune cells based on intracellular or extracellular markers using polychromatic flow cytometry or mass cytometry. CircRNA, circular RNA; lnRNA, long noncoding RNA; mRNA, messenger RNA; RNAseq, RNA sequencing; ULS, Universal Linkage System.

https://doi.org/10.1182/blood-2018-02-791509

Applications of proteomics

- Drug discovery and development
- Identifying potential drug targets, examining the druggability of selected. protein targets, and developing drugs aimed at candidate therapeutic protein targets (e.g., for hepatocellular carcinoma).



Chen et al., 2020, https://doi.org/10.1038/s41575-019-0240-9

Applications of proteomics

- System biology:
- System-wide investigations of disease pathways and host–pathogen interactions to identify potential biomarkers and therapeutic targets; system- wide investigations of drug action, toxicity, resistance and efficacy.
- Agriculture:
- Investigations of plant–pathogen interactions, crop engineering for increased resilience to e.g., flooding, drought and other environmental stresses.
- Food Science:
- Food safety and quality control, allergen detection and improving the nutritional value of foods.
- Astrobiology:
- Investigations of how mammals' immune systems may respond to exo-microbes found in space and studies of the prebiotic organic matter found on meteorites.

Protein Expression Database

- UniProt (Universal Protein Resource) is one of the most comprehensive protein sequence and annotation databases available. While not a true protein expression database, it provides detailed, high-quality information on the function of proteins, their structures, and their roles in various biological processes

- The PRIDE (Proteomics Identifications Database) is a prominent public repository for protein and peptide identifications and is part of the ProteomeXchange consortium. It contains data from a wide range of mass spectrometry-based experiments, making it a go-to resource for researchers needing raw data for re-analysis or integration.

- The Peptide Atlas aggregates mass spectrometry-based proteomics data to create a high-quality, publicly accessible resource for peptides. The Peptide Atlas collects raw mass spectrometry output files from human, mouse, yeast and several other organisms and reprocesses them through a unified analysis and validation pipeline. The results are loaded into a database and the information derived from the raw data is made available to the community through several exploration tools.

- The Human Protein Atlas (HPA) is a comprehensive resource focusing on the human proteins in cells, tissues, and organs using the integration of various omics technologies.

Protein Interaction Databases

- The STRING **(Search Tool for the Retrieval of Interacting Genes/Proteins)** is a widely used database of known and predicted protein-protein interactions. These interactions include direct (physical) and indirect (functional) associations which are derived from computational prediction, high-throughput laboratory experiments, conserved co-expression, text mining and other knowledge bases.

- IntAct is a high-quality, open-source database maintained by the European Molecular Biology Laboratory (EMBL) that focuses on protein-protein interactions.

- KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for: Metabolism, Cellullar Processes, human diseases etc…

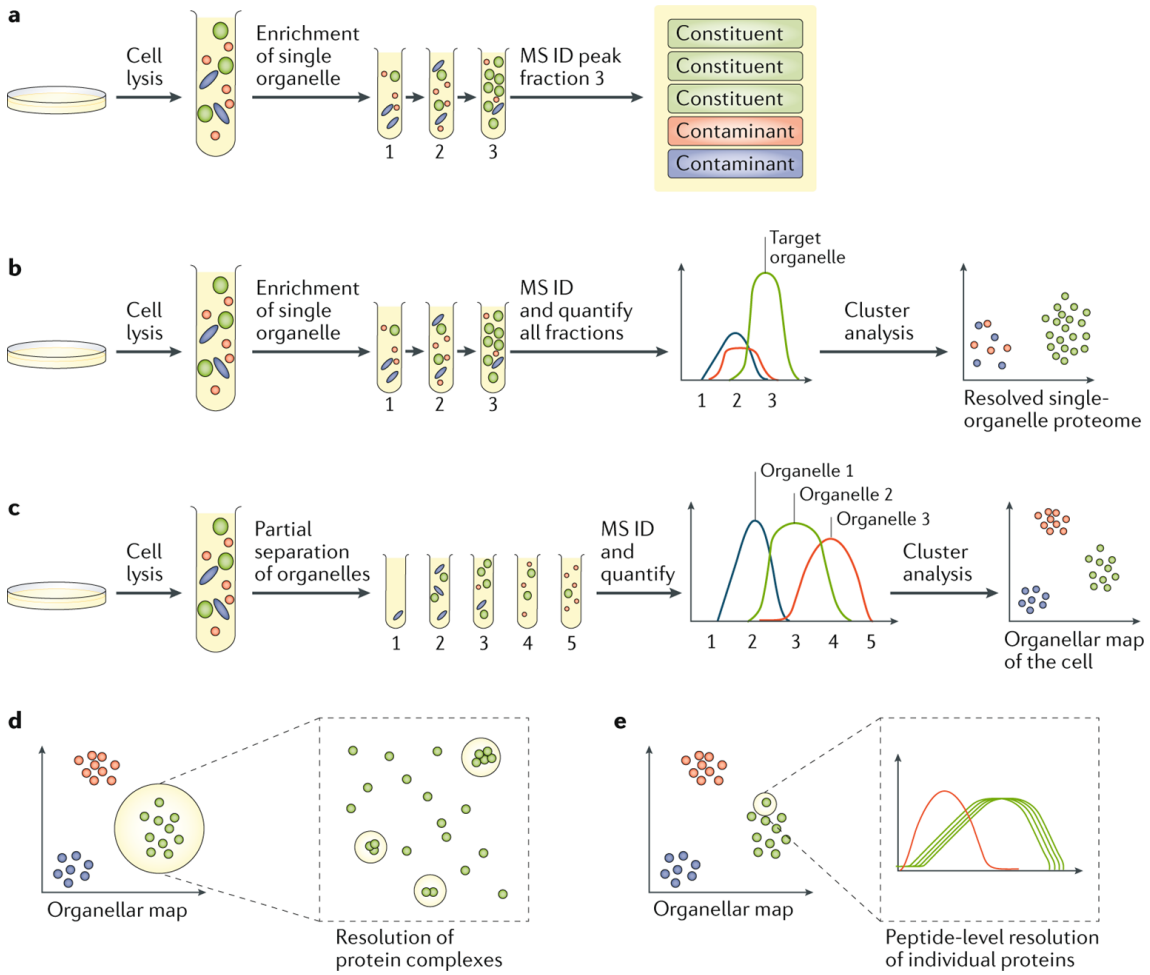
Spatial Proteomics

- Protein function is closely linked to subcellular localization, as different compartments provide different chemical environments (such as pH and redox conditions), potential interaction partners, or substrates.

- Recent substantial advances in high-throughput microscopy, quantitative mass spectrometry (MS) and interactomics mapping, as well as machine learning applications for data analysis, have enabled proteome-wide investigations of spatial cellular regulation.

- Several studies have successfully harnessed the power of global spatial proteomics to investigate diseases, including acute viral infection and liver disease, or to pinpoint the cellular defects that underlie monogenic disorders

Spatial Proteomics by MS analysis of fractionated organelles.

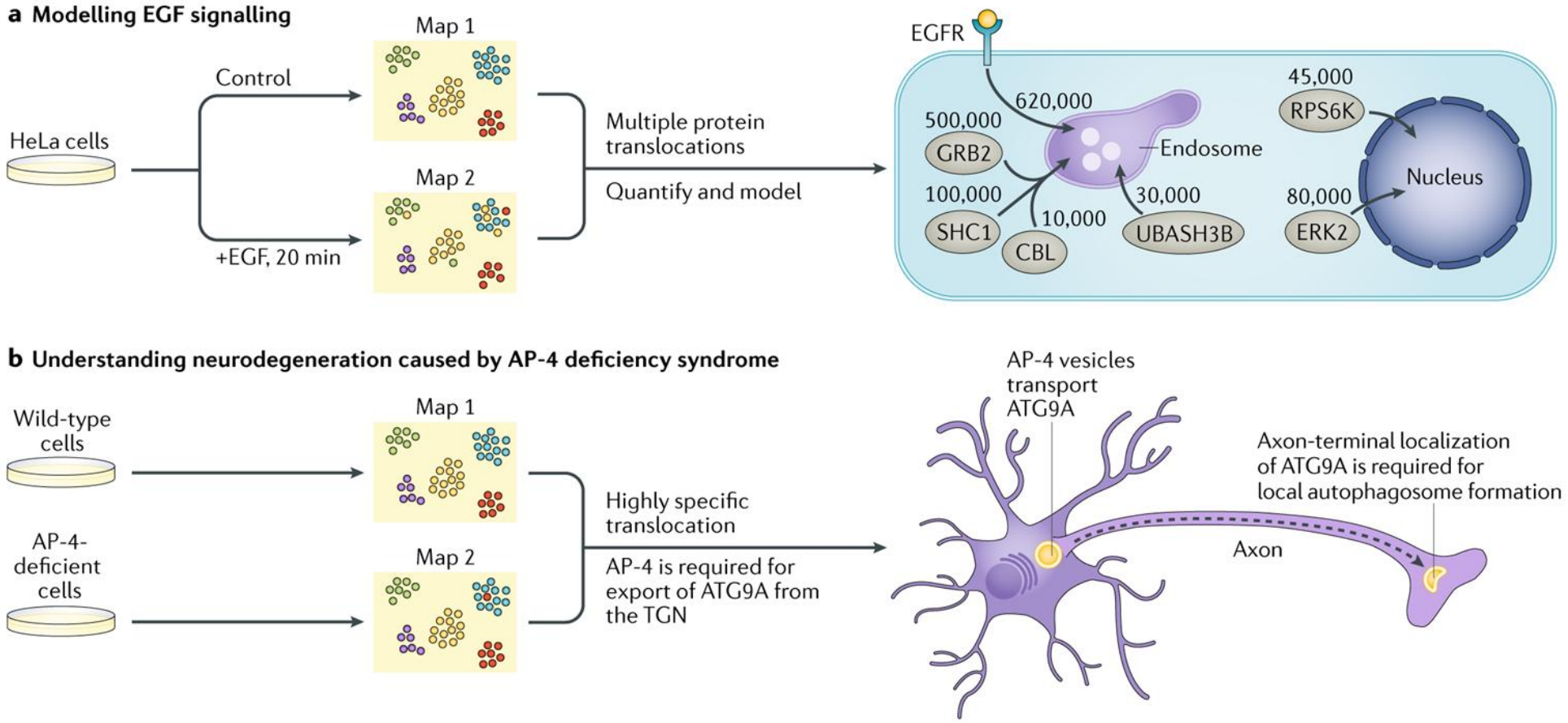


Lundberg et al., 2019, https://doi.org/10.1038/s41580-018-0094-y

d | Organellar maps have high local resolution. Proteins that are part of the same complex have tightly linked fractionation profiles that appear as microclusters within organellar maps, a feature that can be used to identify (ID) novel protein complexes

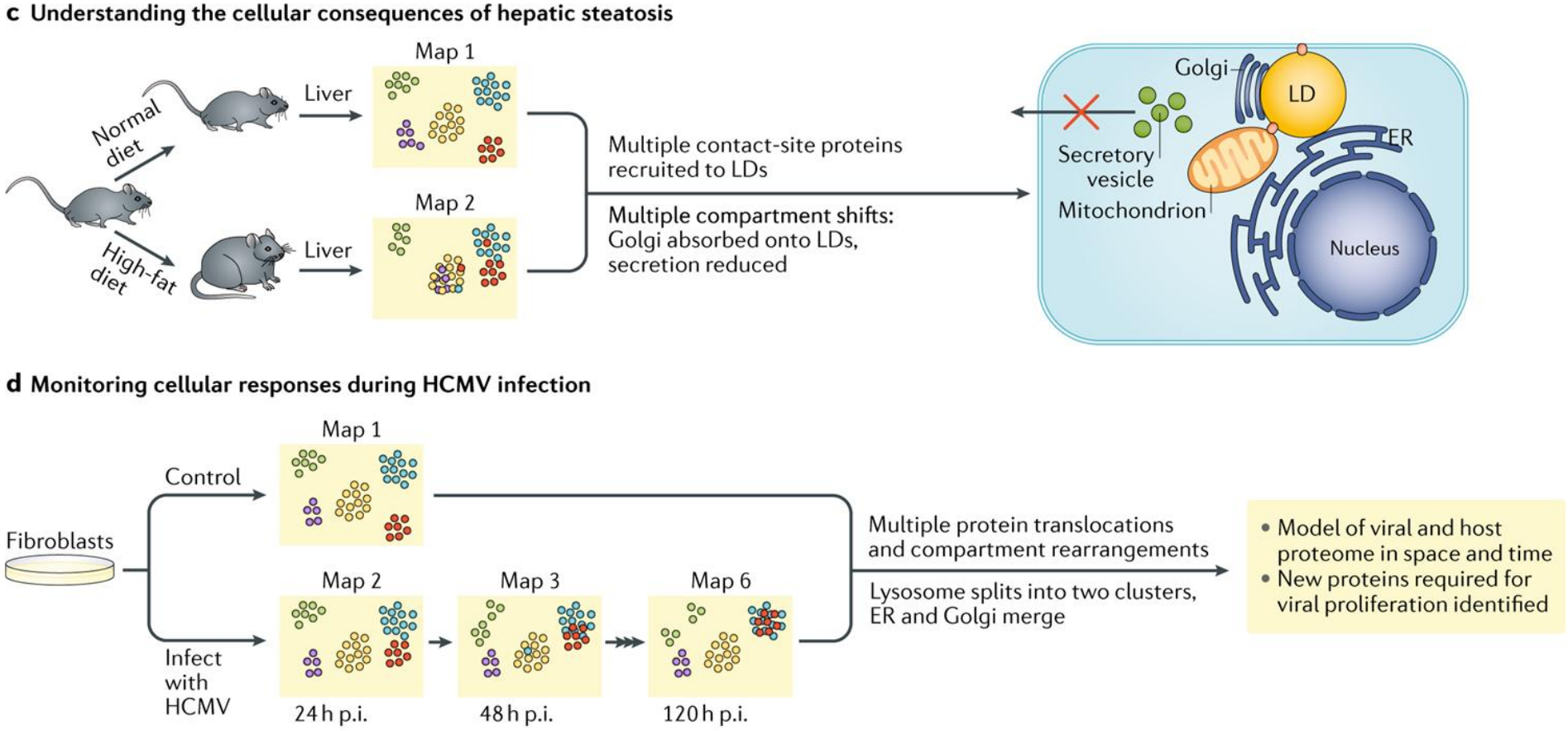
e | Organellar maps can provide peptide-level resolution. Quantification by MS works at the level of peptides. The fractionation profile of a protein (a dot on the organellar map, left panel) is an average (or median) of the profiles from all peptides matching the sequence of that protein. Peptide profiles from the same protein tend to be closely aligned (right panel, green). However, if a post-translational modification (such as phosphorylation) causes a shift in the subcellular localization of a protein, the peptide containing the modification will have a different profile (red). As usually only a fraction of the copies of a protein present within a cell are modified, the profile of the modified peptide is different from that of the bulk of the other peptides.

MS-based comparative spatial proteomics: example applications.



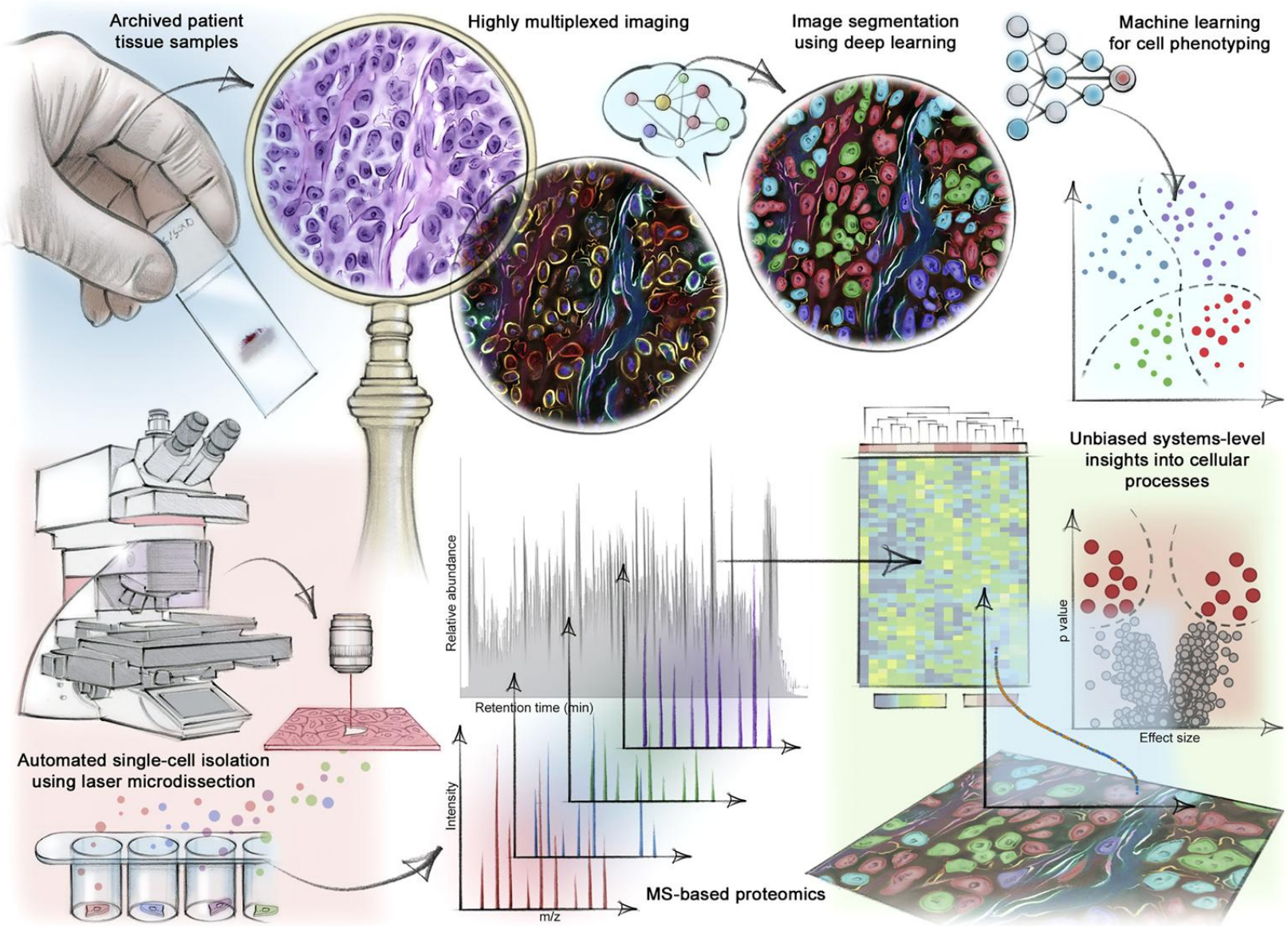
Lundberg et al., 2019, https://doi.org/10.1038/s41580-018-0094-y

MS-based comparative spatial proteomics: example applications.



Lundberg et al., 2019, https://doi.org/10.1038/s41580-018-0094-y

Single cell-resolution in tissues



10.1016/j.molcel.2022.05.022 External Link

https://www.youtube.com/watch?v=mRZfZrviKIQ

THANKS!

https://www.youtube.com/watch?v=cx7l9ZGFZkw

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani