Introduction to Proteomics

Workshop written by Julia Broadbent and Emma Gail

Content and images in this presentation sourced from the Galaxy Training Network's
Introduction to proteomics, protein identification, quantification and statistical modelling

Melbourne Bioinformatics
BIOINFORMATICS + DATA SERVICES + INFRASTRUCTURE, FOR LIFE SCIENCES TODAY

Acknowledgement of Country

The University of Melbourne acknowledges the Traditional Owners of the unceded land on which we work, learn and live: the Wurundjeri Woi-wurrung and Bunurong peoples (Burnley, Fishermans Bend, Parkville, Southbank and Werribee campuses), the Yorta Yorta Nation (Dookie and Shepparton campuses), and the Dja Dja Wurrung people (Creswick campus).

The University also acknowledges and is grateful to the Traditional Owners, Elders and Knowledge Holders of all Indigenous nations and clans who have been instrumental in our reconciliation journey.

We recognise the unique place held by Aboriginal and Torres Strait Islander peoples as the original owners and custodians of the lands and waterways across the Australian continent, with histories of continuous connection dating back more than 60,000 years. We also acknowledge their enduring cultural practices of caring for Country.

We pay respect to Elders past, present and future, and acknowledge the importance of Indigenous knowledge in the Academy. As a community of researchers, teachers, professional staff and students we are privileged to work and learn every day with Indigenous colleagues and partners.

In making this Acknowledgment of Country we commit to respectful and responsible conduct towards all others according to the Traditional lores of this land, particularly at times of formal ceremony.

Housekeeping

- Facilitator introductions
- Building access
- Emergency exits
- Refreshments
- Setup
  - WiFi
  - Installing & loading R packages
  - Downloading data (parquet file + Sample Annotation)
  - Slack

Workshop overview

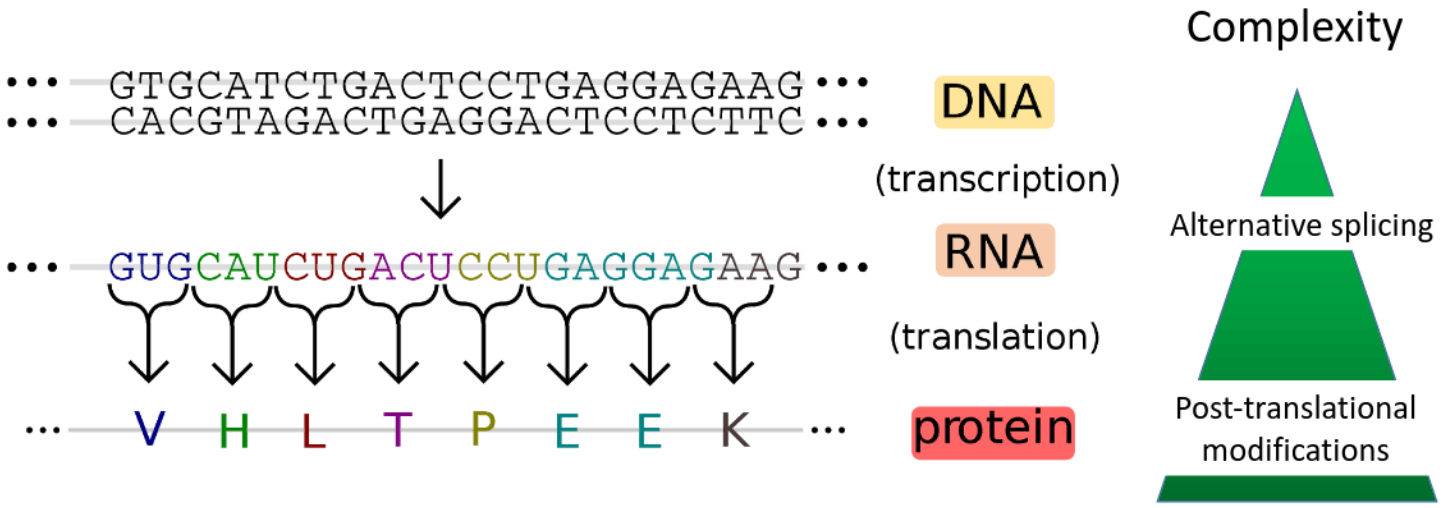
- Introduction to proteomics
- Timing:

Time	Workshop section	Activity
~ 30 minutes	Proteomics workflow overview	Lecture
~ 30 minutes	Proteomics data processing	Tutorial – follow along
~ 1 hour	Cleaning proteomics data	Tutorial – work individually, discuss as a group
BREAK		
~ 1 hour 30 mins	Statistical analyses	Tutorial – work individually, discuss as a group

- Sticky-note system

What is the proteome?

- "*A **Proteome** is the entire complement of proteins that is or can be expressed by a cell, tissue, or organism at a given time*." - Marc Wilkins 1996
- Proteins complete biological functions
- **Proteomics** = study of the proteome (protein structure and function)



This workshop

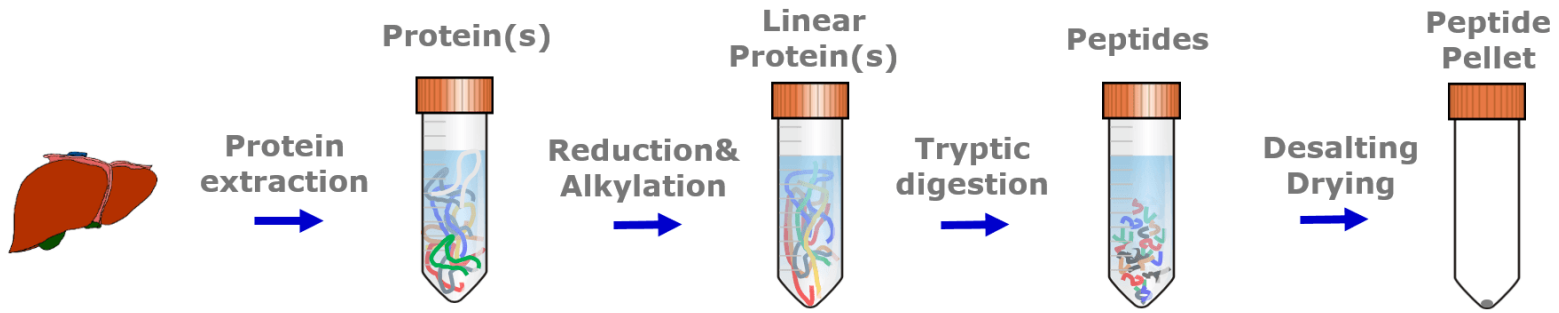
- **Quantitative expression proteomics** = identifying and quantifying amounts of proteins in a sample

  - Bottom-up
  - Liquid chromatography tandem mass spectrometry (LC-MS/MS)
  - Label-free
  - DIA

Top-down vs Bottom-up proteomics

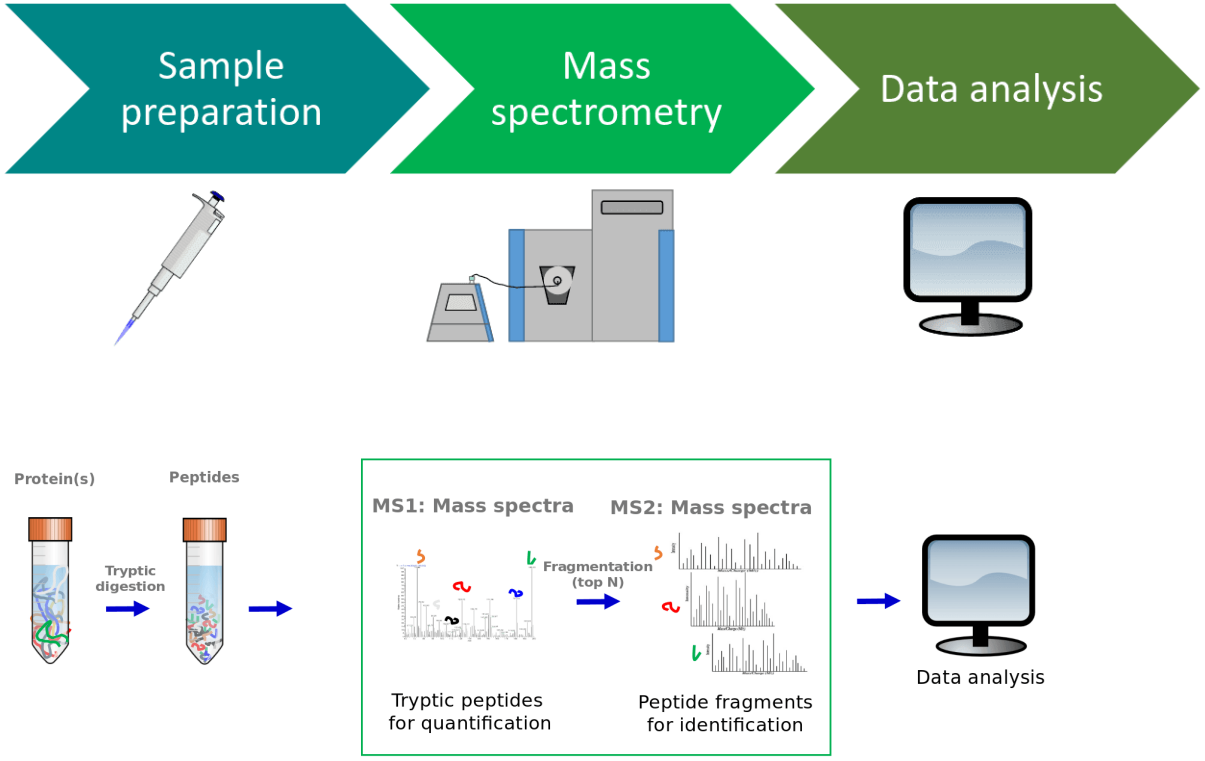
- **Top-down** = Analysing fully intact proteins
  - **Pros:** proteoforms, post-translational modifications

- **Bottom-up** = Proteins → peptides/precursors (MS1) → ions (MS2) →
  Infer presence of proteins from detected peptides & ions
  - **Pros:** higher coverage, higher sensitivity, higher throughput, faster, not limited by protein size
  - Shotgun proteomics = bottom-up LC-MS/MS proteomics



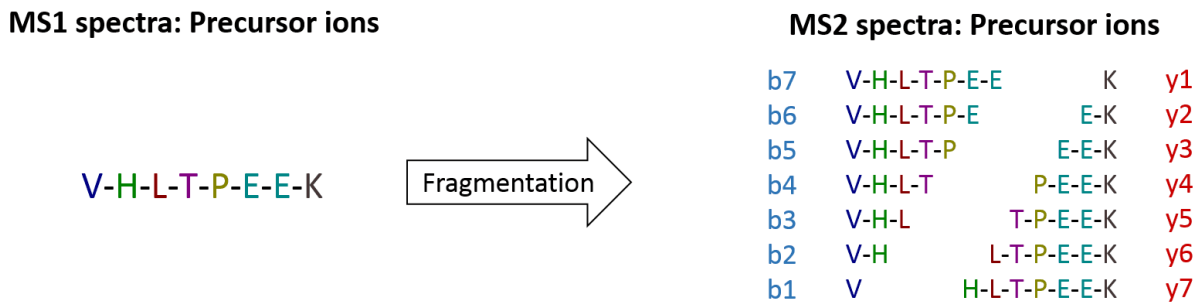
Liquid chromatography tandem mass spectrometry (LC-MS/MS)

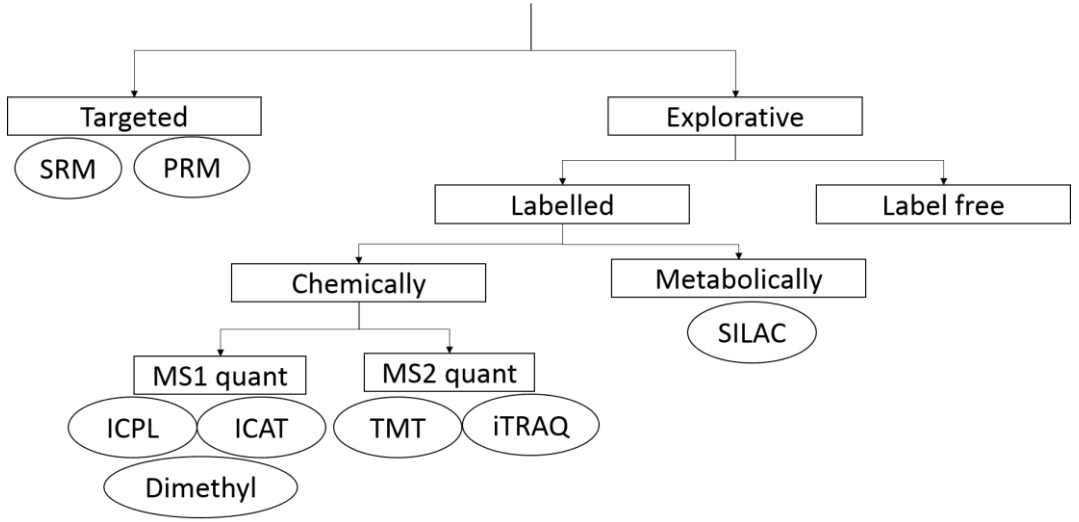
- **Mass spectrometry**
  - High throughput
  - High sensitivity
- **Workflow**
  - **Sample prep**: sample → proteins extracted → digested into peptides
  - **LC**: separates peptides
  - **MS/MS**: Measuring mass-to-charge ratio (m/z) of peptides
  - **Data analysis**: mass spectra (m/z-intensity pairs)



Label-free quantitative proteomics

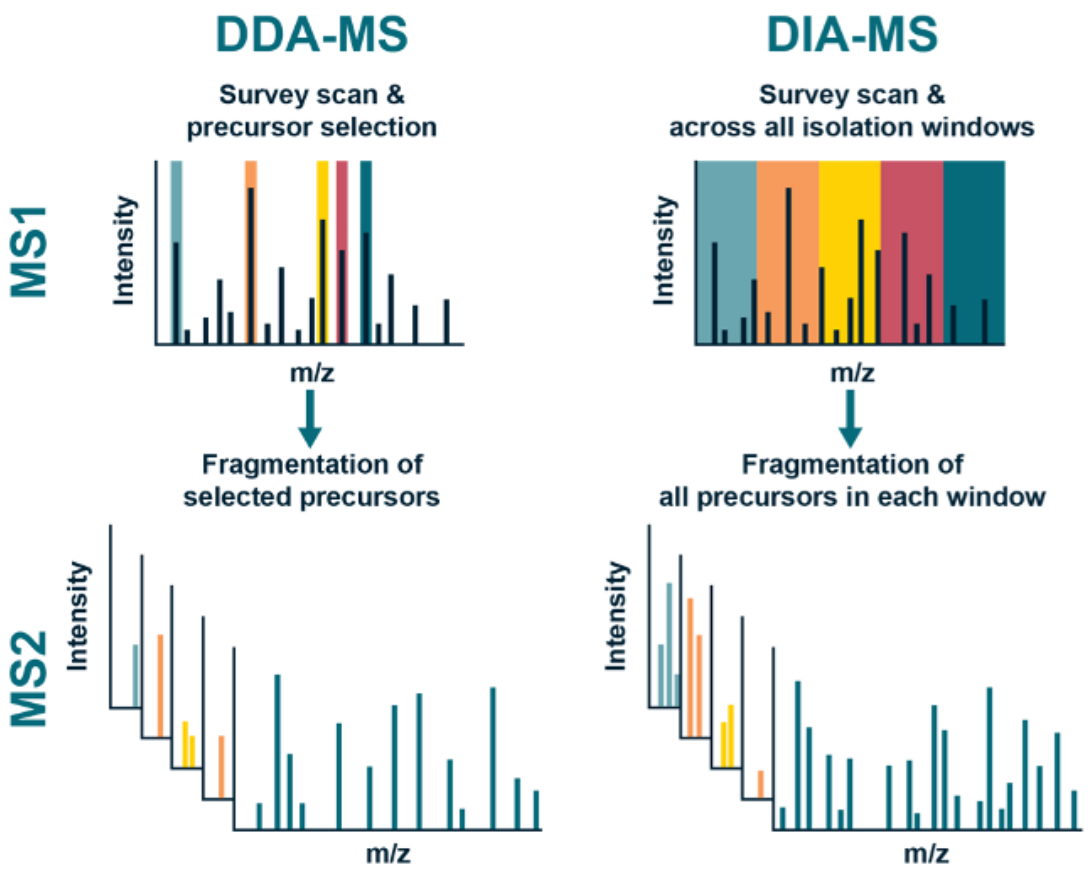
- **Label-free** quantification methods
  - Based on spectral counts and intensities
  - Relative abundance
- **MS/MS spectra**
  - MS1 precursor ions (peptide quantification)
  - MS2 fragment ions (peptide identification and quantification)
- **Quantification strategies:**
  - Labeling (TMT, iTRAQ, SILAC)
  - Label-free (spectral counts, intensities)





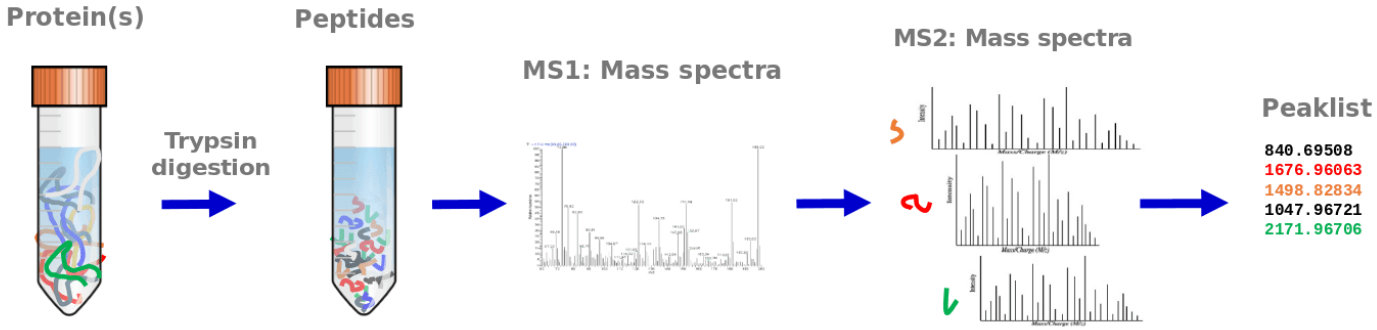
DIA vs DDA

- **Data-dependent acquisition (DDA)**
  - Most abundant precursors selected for further fragmentation
  - Biased, results in more missing data
- **Data-independent acquisition (DIA)**
  - All precursors fragmented within a predetermined m/z range
  - Higher coverage, precision, reproducibility
  - **Some* bias towards highly abundant and easily ionised proteins/peptides
  - **SWATH-MS** = a DIA method



Workflow overview

- Quantitative expression proteomics
  - Bottom-up
  - Liquid chromatography tandem mass spectrometry (LC-MS/MS)
  - Label-free
  - DIA
- Output
  - Mass spectra (m/z vs intensity) per sample



Introduction to the dataset we are using today

- The study
  - Exploring a non-invasive diagnostic approach for Inflammatory Bowel Disease (IBD) using mass spectrometry-based proteomics
  - Stool samples
- Our data
  - Subset of the main dataset
  - 11 controls + 9 aCD (active Crohn's disease)
  - Two batches (B1 & B3)

Workshop

Time	Workshop section	Activity
10:00am	Proteomics workflow overview	Lecture
10:30am	Proteomics data processing	Tutorial – follow along
11:00am	Cleaning proteomics data	Tutorial – work individually, discuss as a group
BREAK		
12:30pm	Statistical analyses	Tutorial – work individually, discuss as a group