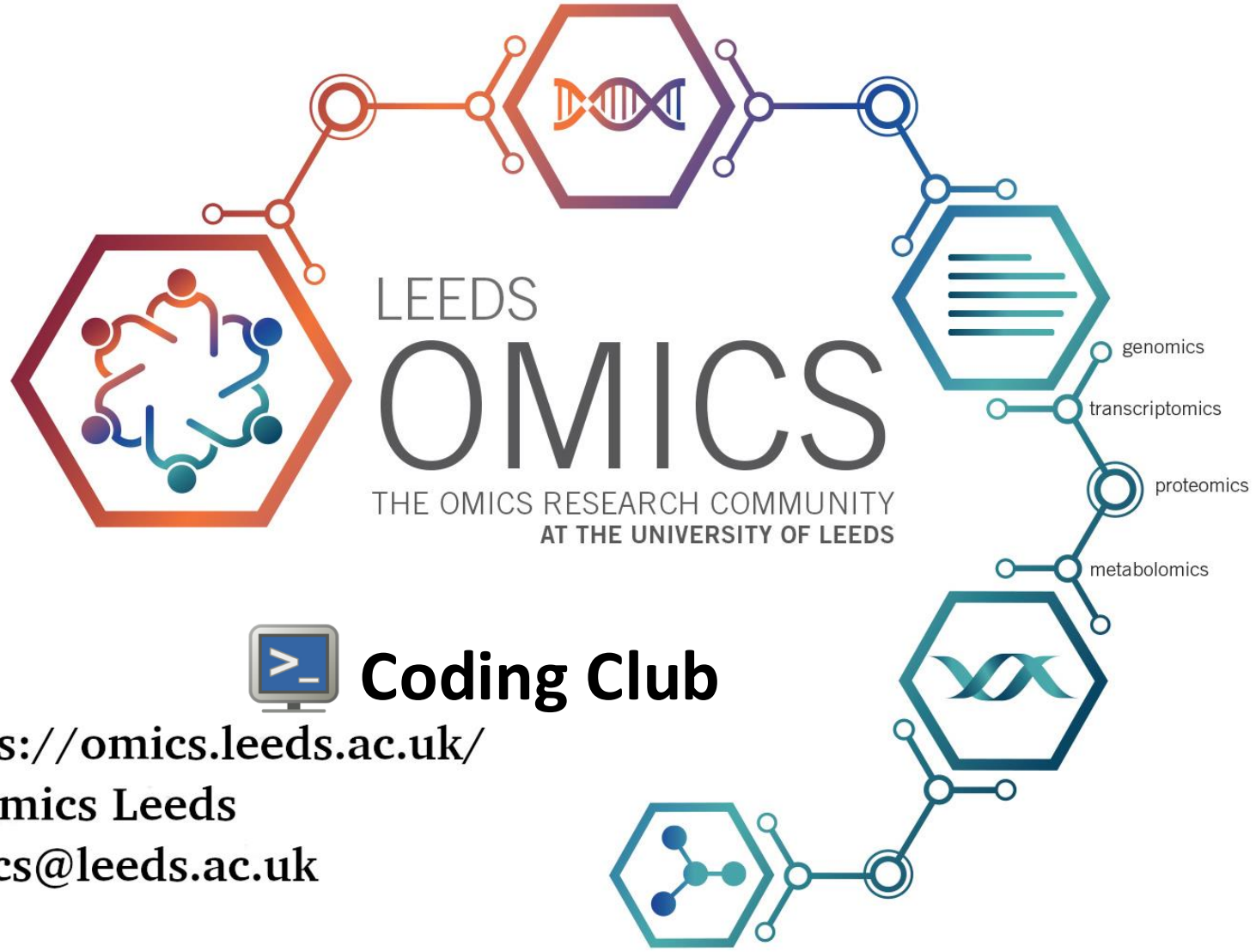


Coding Club

https://omics.leeds.ac.uk/

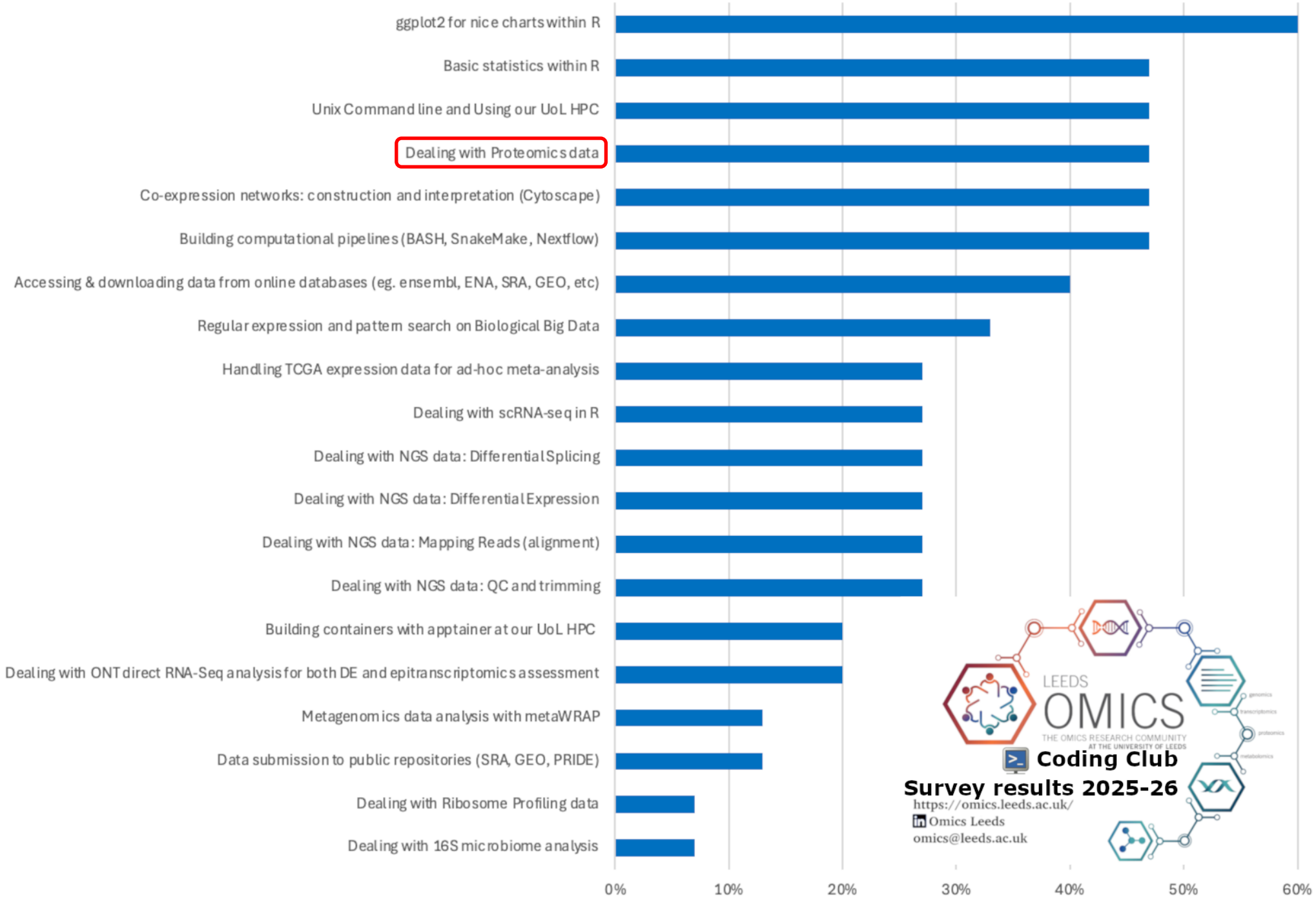
Omics Leeds

omics@leeds.ac.uk

Dealing with Proteomics Data

Club Moderators: Elton Vasconcelos, Eilidh Ward

Topics to be addressed on the **2025-26** season - **Survey Result**



LEEDS OMICS
THE OMICS RESEARCH COMMUNITY AT THE UNIVERSITY OF LEEDS

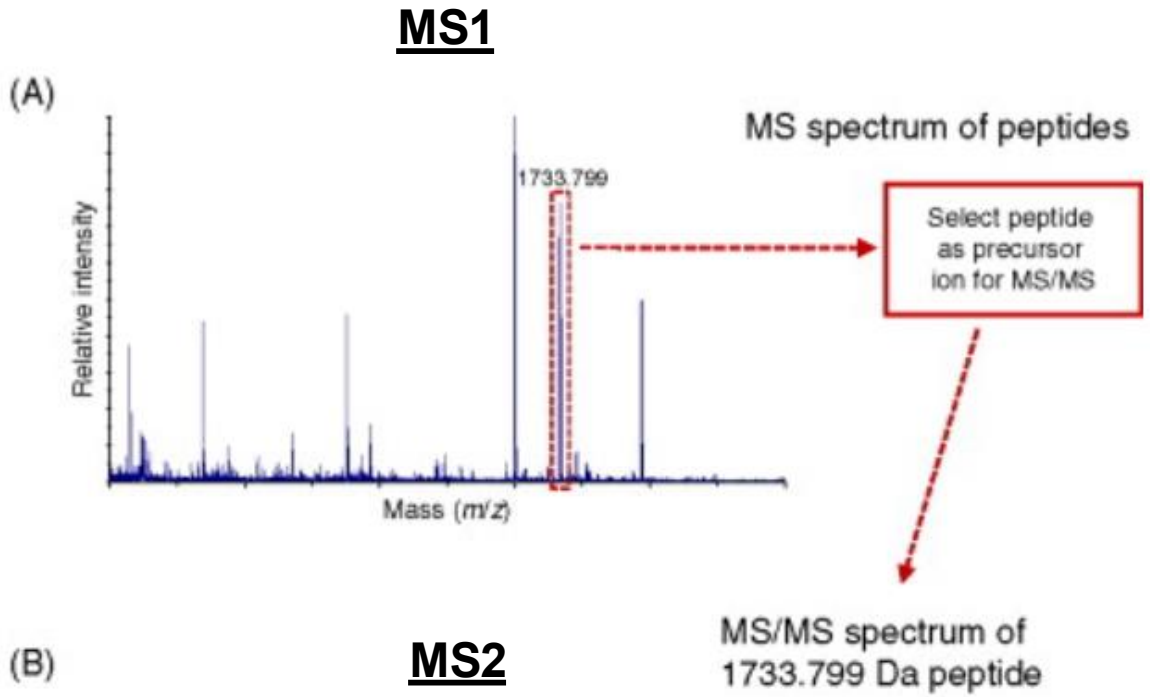
Coding Club

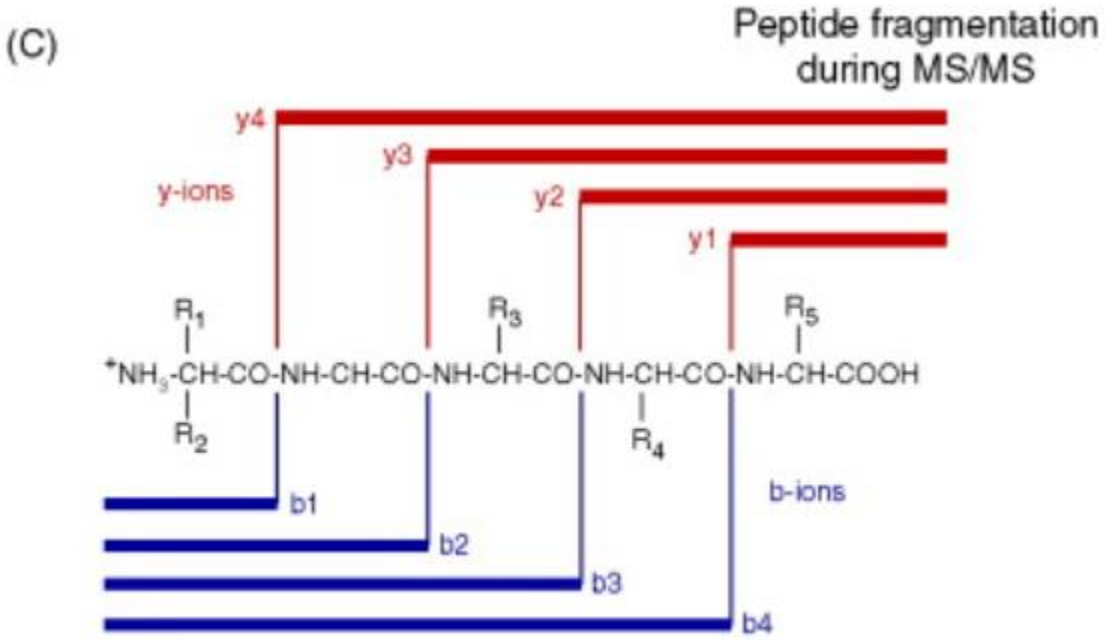
Survey results 2025-26

https://omics.leeds.ac.uk/
Omics Leeds
omics@leeds.ac.uk

MS/MS brief summary

Peptides are first detected by MS (panel A). In this example, a 1733.799-Da peptide was subjected to collision-induced dissociation (CID) to achieve peptide fragmentation and the MS/MS spectrum is shown in panel B. Sufficient energy is used to break the peptide bonds which are generally the weakest in the peptide. The trick uses a low enough energy to cause inefficient fragmentation so that peptide fragments are produced with differing numbers of amino acids present. Panel C is a diagrammatic representation of MS/MS fragmentation of peptides due to peptide bond cleavage.





https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/tandem-mass-spectrometry

DDA -vs- DIA

Fragments only a subset of peptides for **partial proteome coverage**

2,500-3,600 protein groups identifiable

Fragments and measures all detectable peptides in a sample to achieve **high proteome coverage**

>10,000 protein groups identifiable

- Label-free versus label-based proteomics approaches

Feature	Label-free	Label-based
Quantification	Spectral counting and ion intensity-based, relying on precursor ion abundances at the MS^1 level (during the course of protein fragmentation)	Measurement of reporter ions intensities, produced and detected at the MS^2 level (during the course of peptide fragmentation)
Sample preparation	Cost-efficient (no additional step)	Additional labelling steps introduced (expensive commercially available labelling reagents)
Multiplexed analysis	Not possible	3-plex (SILAC); 4- and 8-plex (iTRAQ); 6-, 10-, 11- and 16-plex (TMT)
Instrument time	Time-consuming (one sample per run)	Significantly decreased
Batch effect bias	Increased	Decreased
Proteome coverage	High	Mid

Proteomics Data Analysis Workflow

1. MS raw data conversion to plain text formats

ThermoRawFileParserGUI*

readMgfData ('RforProteomics' R package)

Quality control with plotMzDelta ('MSnbase' R package)

2. Protein Identification, Normalization and Quantification

Comet and CometUI*

Crux for label-free data only

SearchGUI*

Third party software coupled to MS/MS instruments

ProteomeDiscoverer ©

MaxQuant

IsobarQuant

3. Differential Expression Analysis

Regular t-Test, ANOVA and/or empirical Bayes on label-free data

Limma-adapted (eBayes) functions for iTRAQ/TMT strategies

MSqRob2 https://www.biostat.jhsph.edu/~kkammers/software/eupa/R_guide.html
https://www.bioconductor.org/packages/release/bioc/html/msqrob2.html

DEP R package https://bioconductor.org/packages/release/bioc/vignettes/DEP/inst/doc/DEP.html

Optional tools

- **PCA and batch effect adjustment**

  - prcomp function for PCA (‘ggfortify’ R package);

  - *removeBatchEffect* (‘limma’ R package) or ComBat (‘sva’ R package)**;**

- **Expression Correlation**

  Proteomics vs Transcriptomics

  - *ad hoc* ”Spearman” cor.test in R;
  - *plot_correlations* (**ReactomeGSA** R package)
  - fuzzy c-means-based clustering (**VSClust** R package**);**
  - Multi-Omics Factor Analysis (**MOFA2** R package or **mofapy2** in Python);

  NOTE: Large number of samples (>= 15) and high coverage depth (on both Omics approaches) are mandatory for achieving reliable results

Bring your issues on!