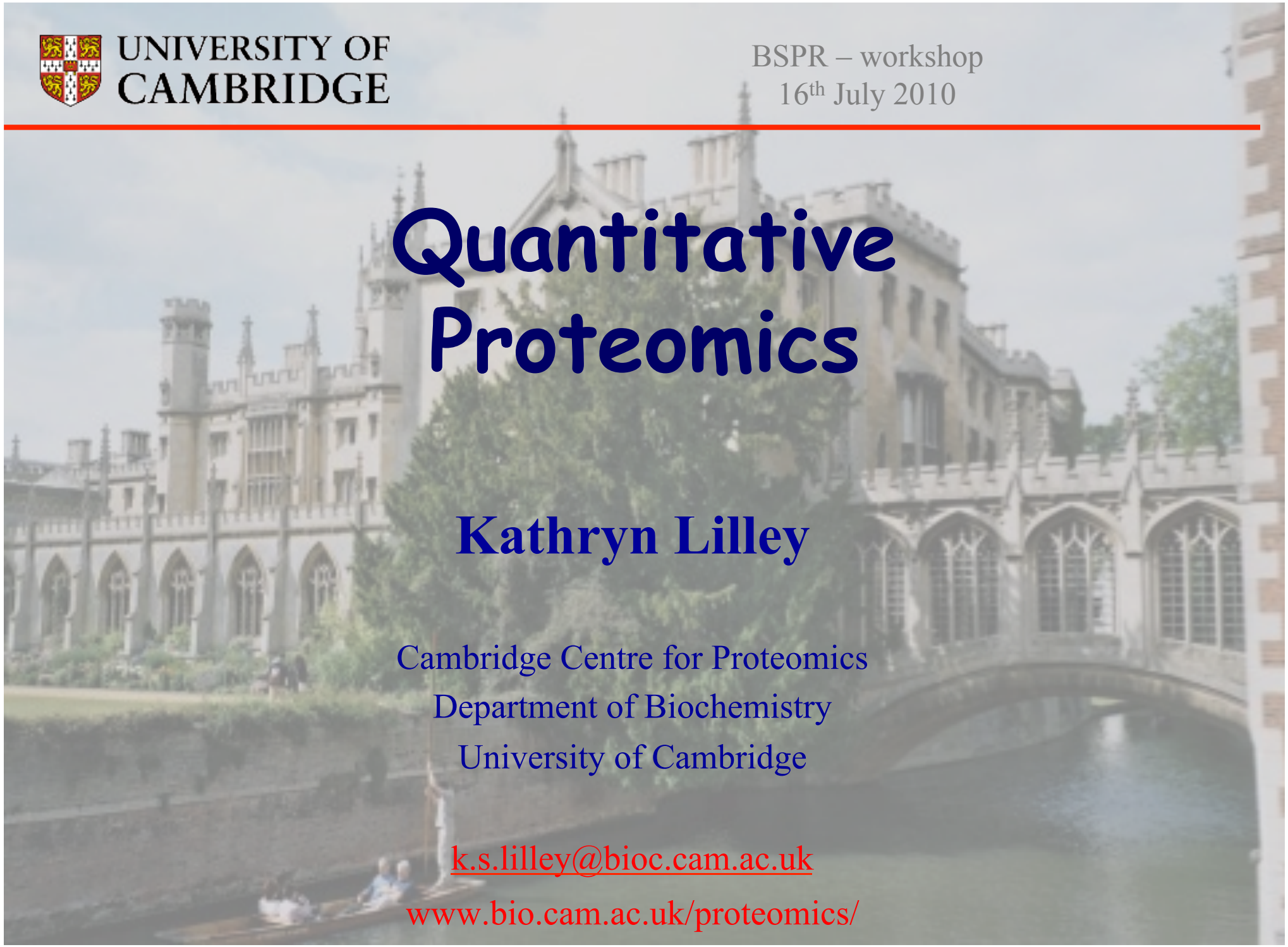


UNIVERSITY OF CAMBRIDGE

BSPR – workshop
16th July 2010

Quantitative Proteomics

Kathryn Lilley

Cambridge Centre for Proteomics
Department of Biochemistry
University of Cambridge

k.s.lilley@bioc.cam.ac.uk

www.bio.cam.ac.uk/proteomics/

Outline

Quantitation in proteomics

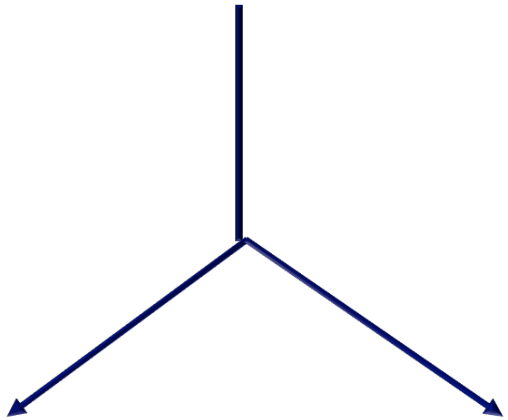
Relative Quantitation

Absolute Quantitation

Importance of Experimental Design

Importance of Suitable Data Analysis

Quantitative Proteomics



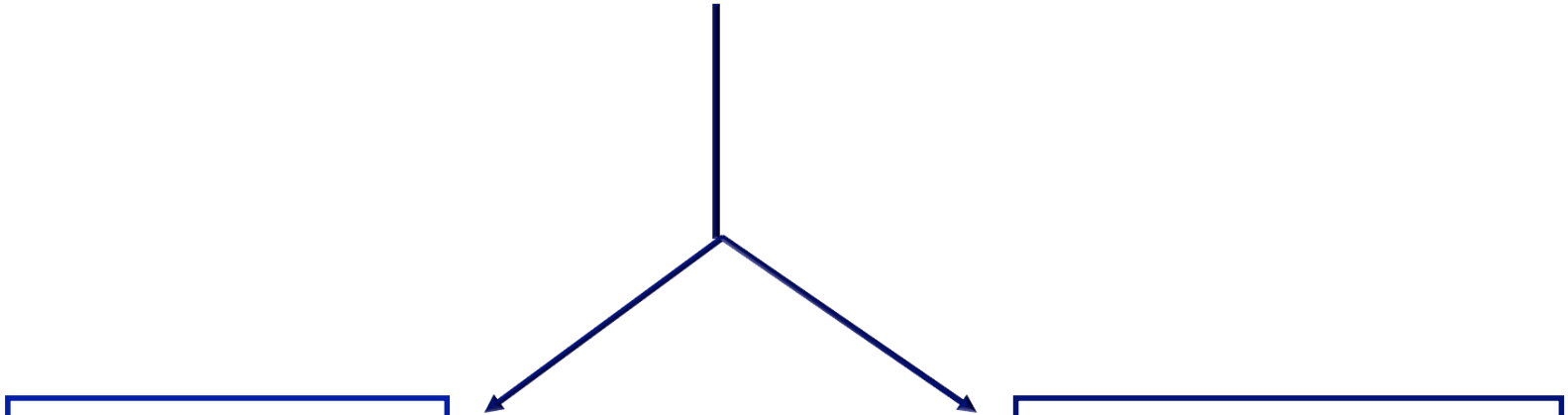
Relative

fold change

Absolute

absolute amount

Quantitative Proteomics



-Comparative levels of proteins between two or more samples

- 2D gel/ DIGE
- Isobaric labelling iTRAQ/TMT
- Metabolic labelling/ SILAC
- Label Free

- Rank order of protein abundance
- Assessment of stoichiometry
- Facilitates targeted analysis
- Transferable data sets

- Internal standards (usually peptide surrogates)

Outline

Quantitation in proteomics

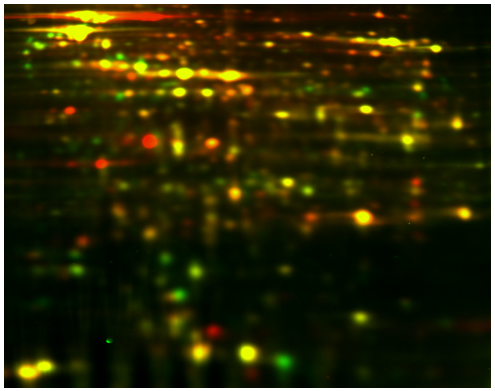
Relative Quantitation

Absolute Quantitation

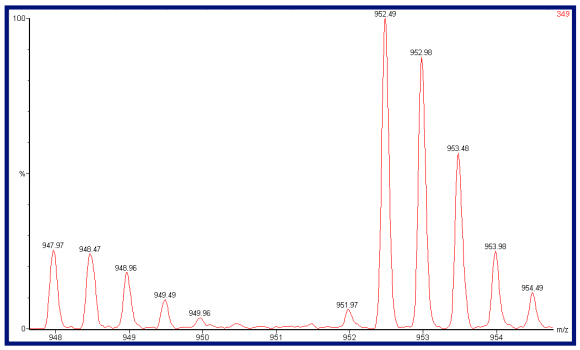
Importance of Experimental Design

Importance of Suitable Data Analysis

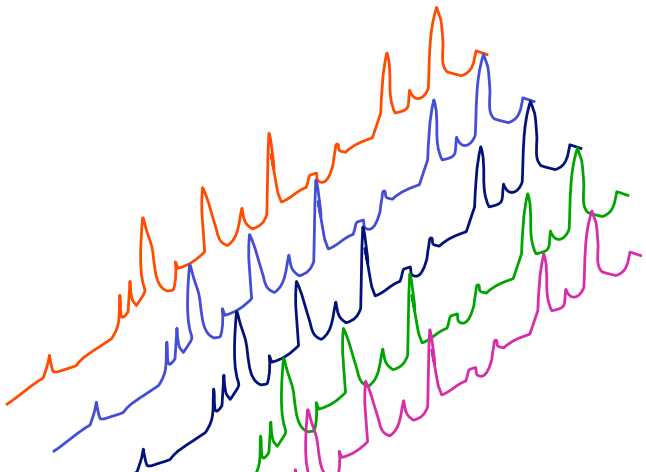
Quantitative proteomics methodologies



Gel based



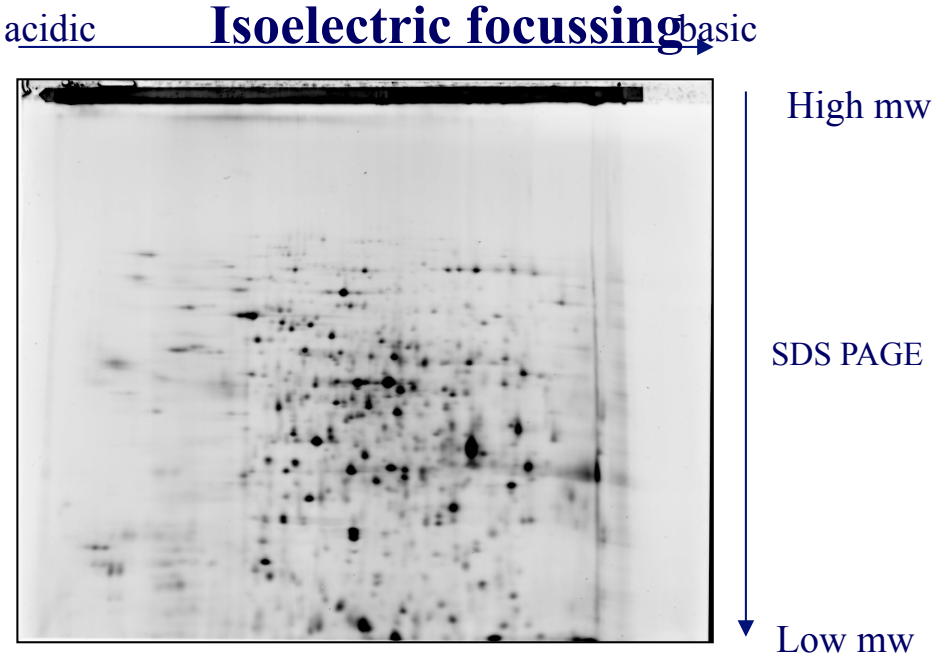
Stable isotope labelling



Label free

2D PAGE

- Visualize many proteins at once
- Relatively quick
- Great way of storing samples
- Detect isoforms if pI shift
- Relatively inexpensive
- Can use with functional stains
- Poor gel to gel reproducibility
- Many stains not linear along dynamic range
- No good for membrane proteins



1st dimension = pI

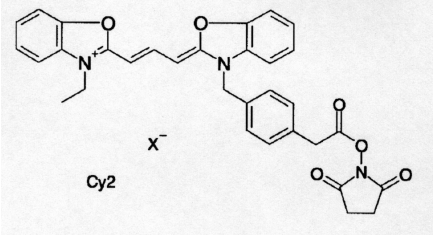
2nd dimension = MW

Difference Gel Electrophoresis

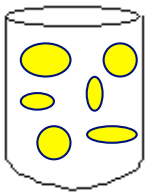
DIGE

- **First described by Jon Minden (Carnegie Mellon University. Pittsburg, USA**
  - **Ünlü M. *et al* (1997). *Electrophoresis*,18, 2071-2077**

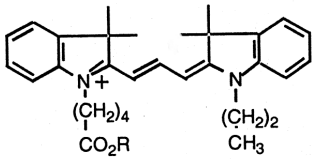
Sample 1



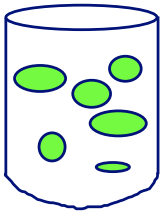
label with cy2
in dark 30mins @ 4^{O}C



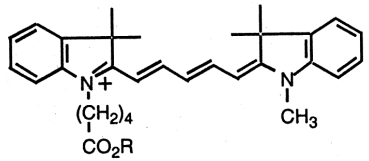
Sample 2



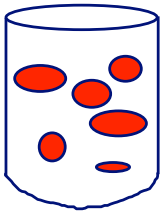
label with cy3
in dark 30mins @ 4^{O}C

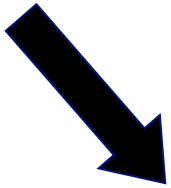


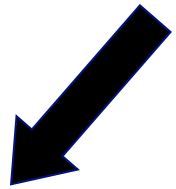
Sample 3



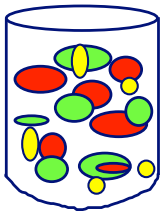
label with cy5
in dark 30mins @ 4^{O}C







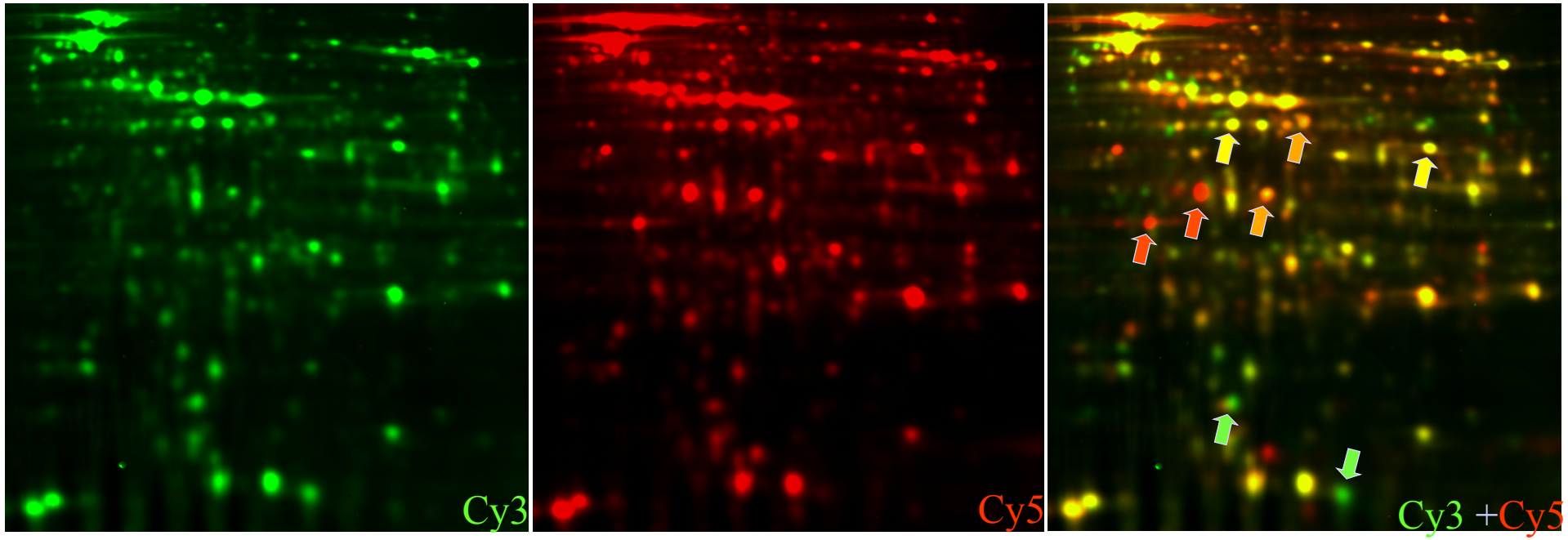
quench un-reacted dye
by adding 1mM lysine
in dark 10mins @ 4^{O}C



Difference Gel Electrophoresis

2D gel electrophoresis

•Ünlü M. *et al* (1997). ***Electrophoresis,***18, 2071-2077

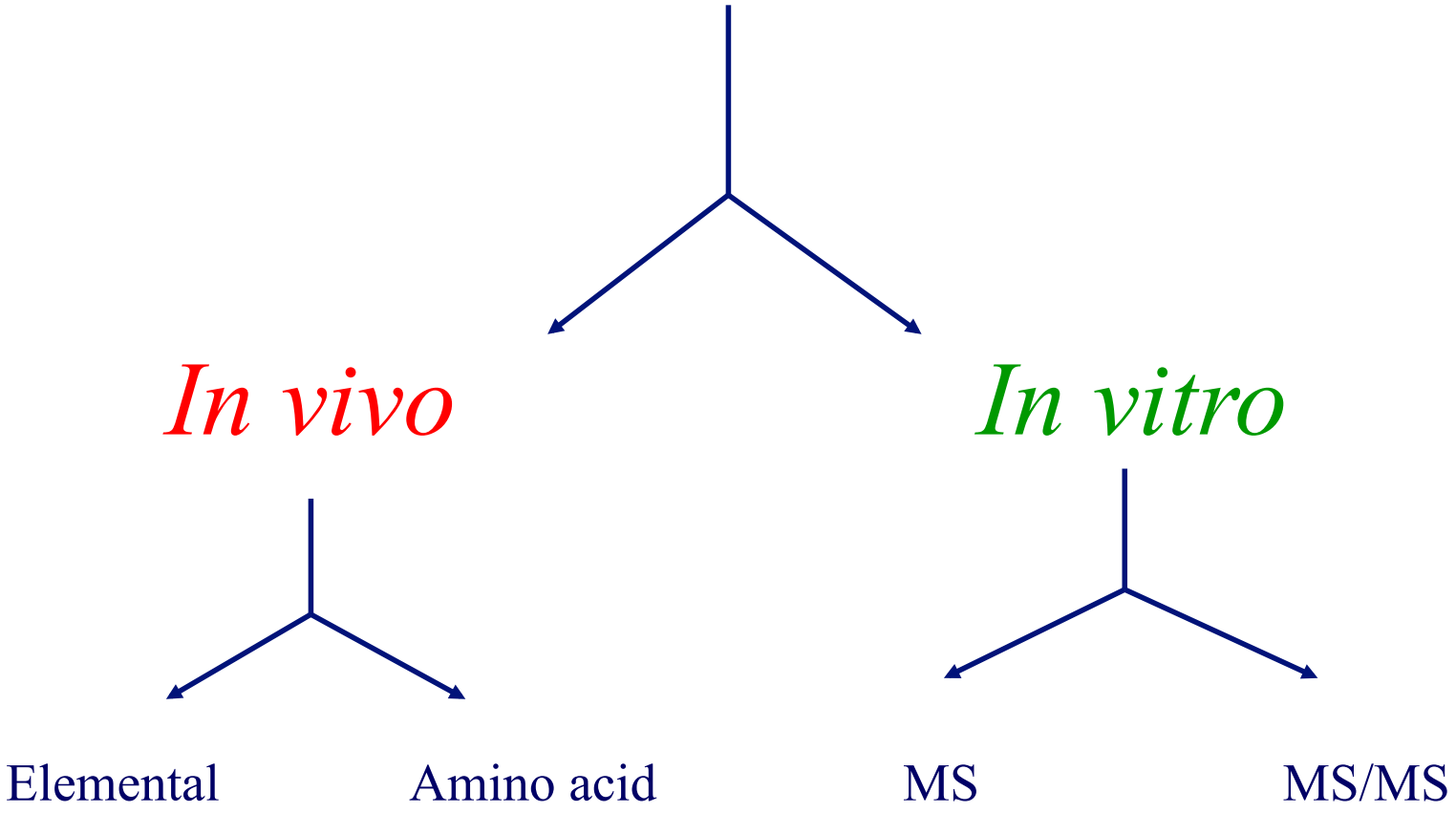


no difference

presence / absence

up / down-regulation

Quantification using stable isotope labelling



Stable Isotope Labelling – *in vivo*

Sample 1
incorporates
natural isotope

Sample 2
incorporates
heavier isotope

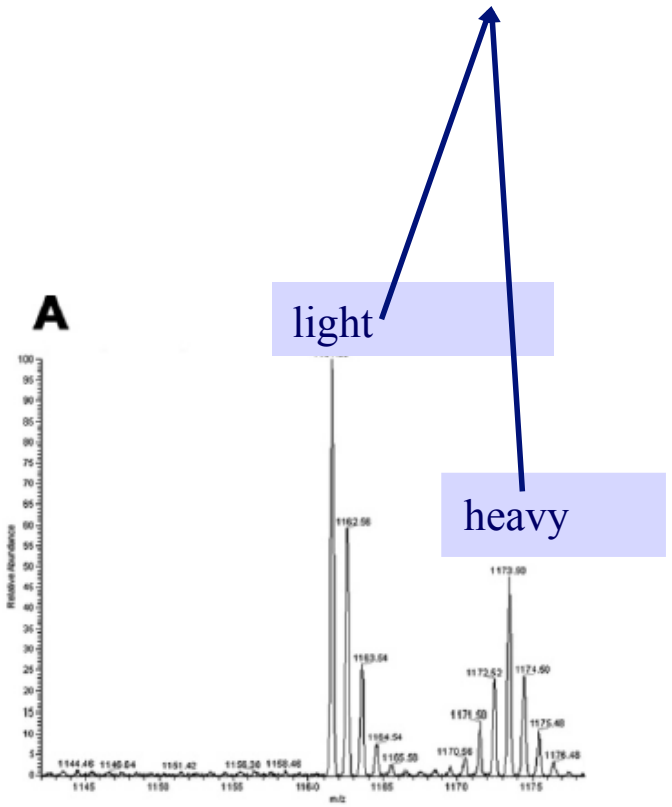
Digest with protease

Mixture of light/heavy peptides

LC separation – usually multi dimensional

Quantitation in MS

MS/MS to identify

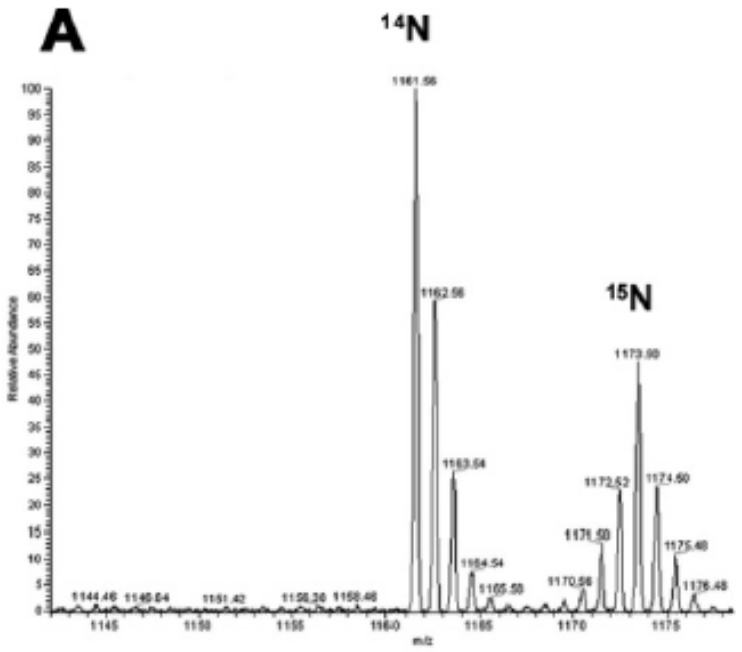


Stable Isotope Labelling – *in vivo*

1. Elemental

- Samples grown in medium where there is replacement of an element with a stable isotope
- Typically ^{15}N instead of ^{14}N, or ^{13}C instead of ^{12}C
- ^{13}C not often used as more carbon in proteins than nitrogen and therefore big mass shifts
- Do not known mass difference between light and heavy pairs unless sequence is deduced (retention times)

?

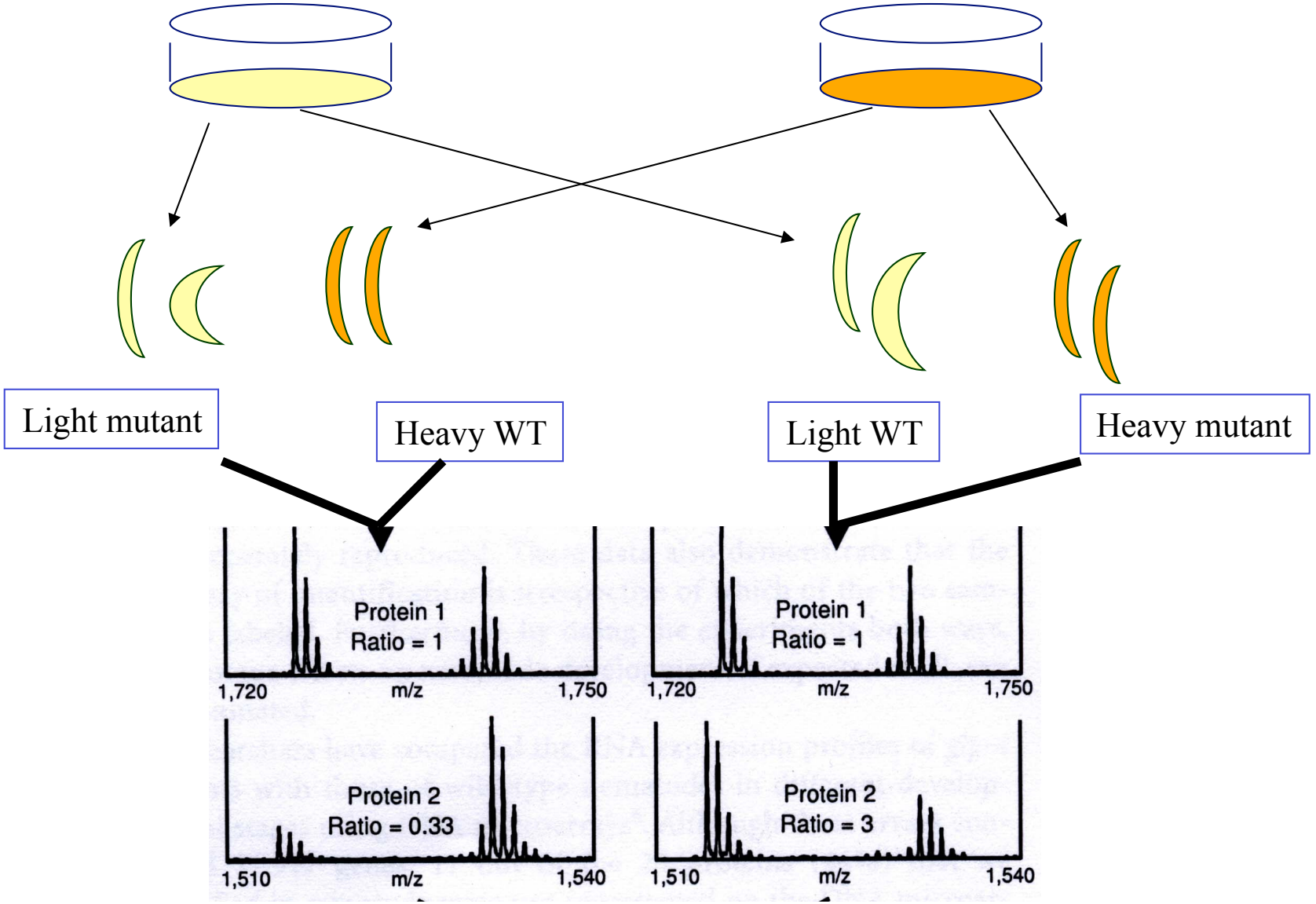


Types of samples suitable?

Bacterial / Cell culture

Examples

E coli grown on ^{15}N sole nitrogen source and then fed to *C.elegans*

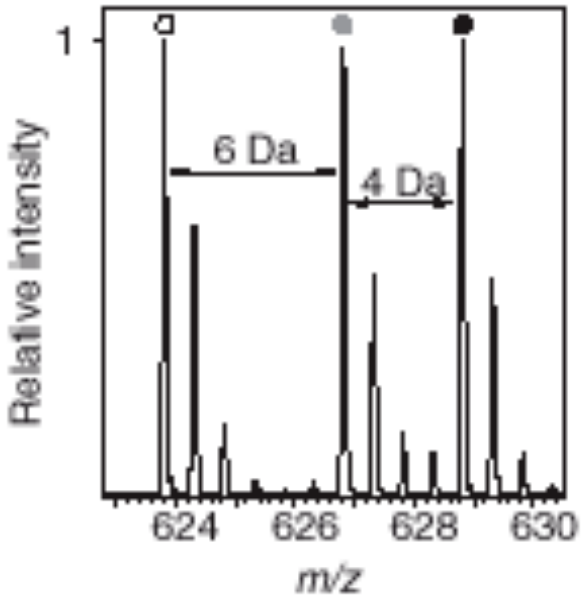


Krijgsveld *et al* (2003) Nat. Biotech.21:927

Stable Isotope Labelling – *in vivo*

1. Amino acid - SILAC (Stable Isotope Labeling with Amino acids in Cell culture)

- Samples grown in medium where there is replacement of an amino acid with heavier stable isotopic form of the amino acid
- Typically ^{13}C instead of ^{12}C – labelled lysine, arginine or leucine
- Know the mass difference between light and heavy pairs
- Need to check for extent of incorporation
- £££ as need also to buy depleted medium



Types of samples suitable?

Bacterial / Cell culture

Stable Isotope Labelling – *in vivo*

SILAC Mouse

Krüger et al (2008) Cell 134(2):353-64

SILAC Drosophila

Sury et al (2010) Mol. Cell Prot. On-line

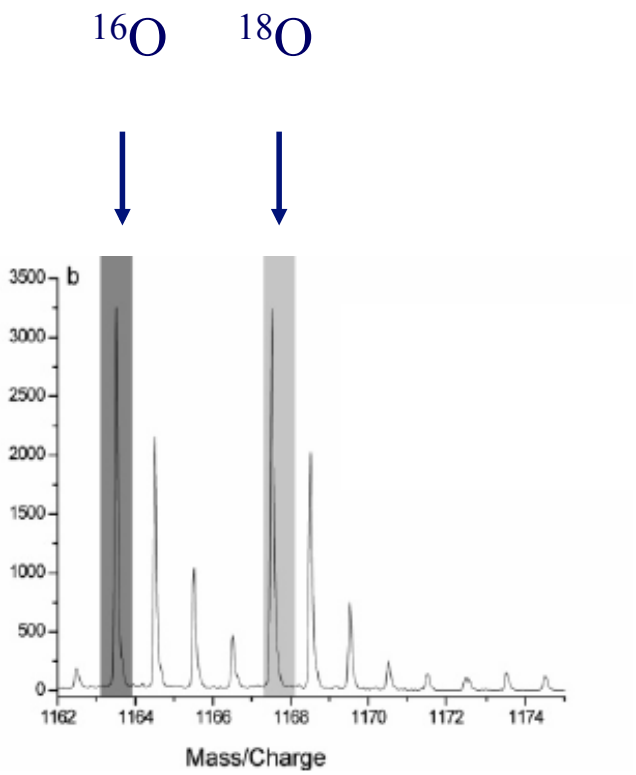
Problem is the conversion of Arg to Pro

many in higher organisms only use labeled lysine and digestion with LysC, this gives rise to longer peptides for analysis

Stable Isotope Labelling – *in vitro*

1. Analysis at MS stage

- Many variants including
  - Isotopes introduced during proteolysis ^{18}O – labelled water, C-termini
  - Guanidation of lysine using isotopes of O-methyl isourea – lysine residues
  - Dimethyl labelling – lysine residues
  - Mostly the above lead to small mass differences
  - Back exchange can be a problem with trypsin

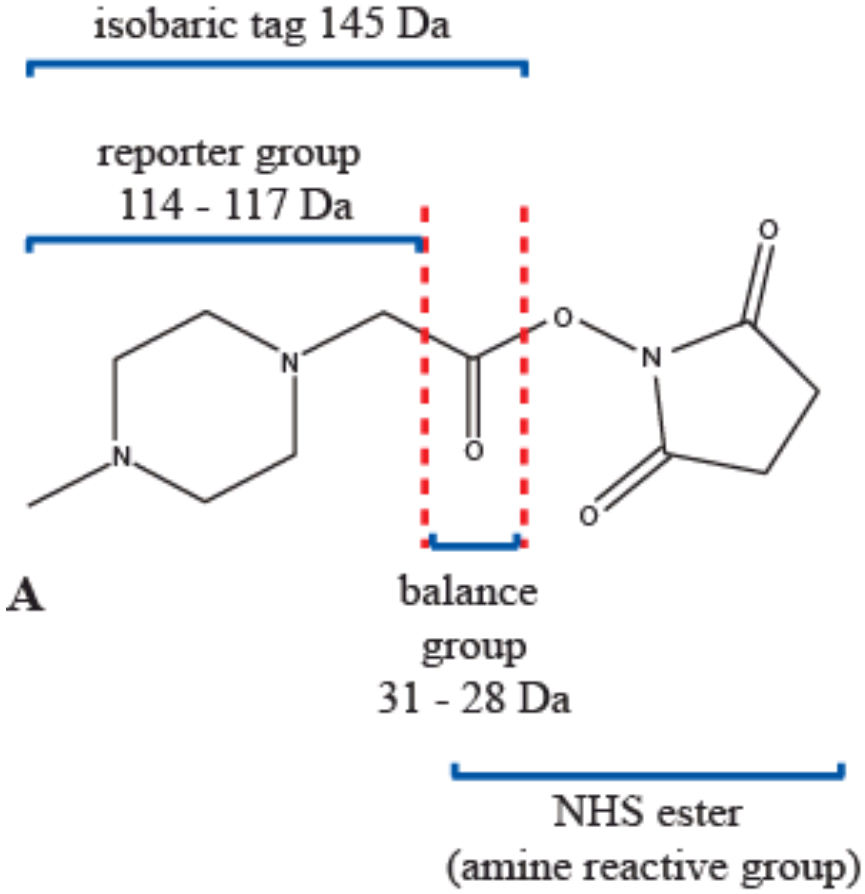


Stable Isotope Labelling – *in vitro*

2. Analysis at MS/MS stage

iTRAQ reagents (Amine Modifying Labeling Reagents for Multiplexed Relative and Absolute Protein Quantitation)

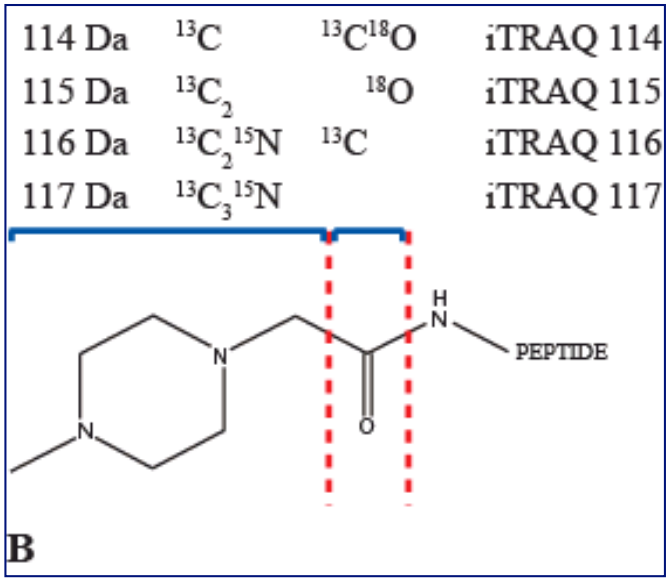
- 4 x isobaric tags - all 145 Da
- React with primary amines
- Label at peptide level
- Fragment during MSMS to produce characteristic reporter ion for each tag

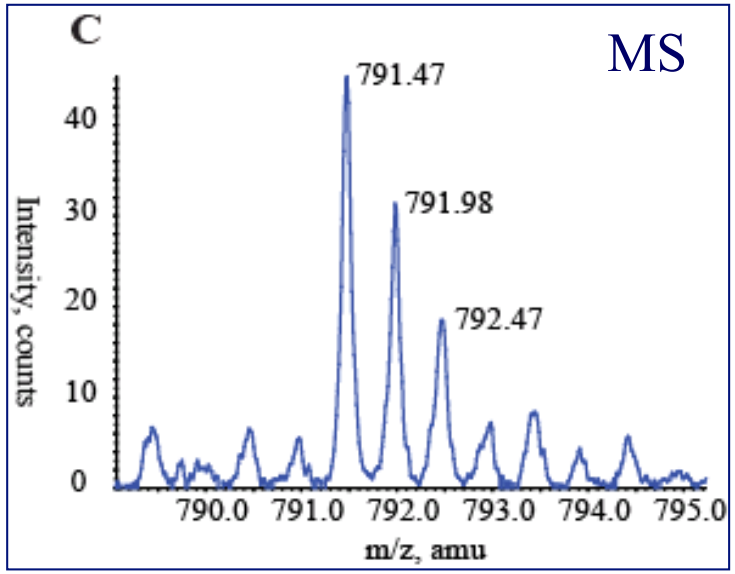


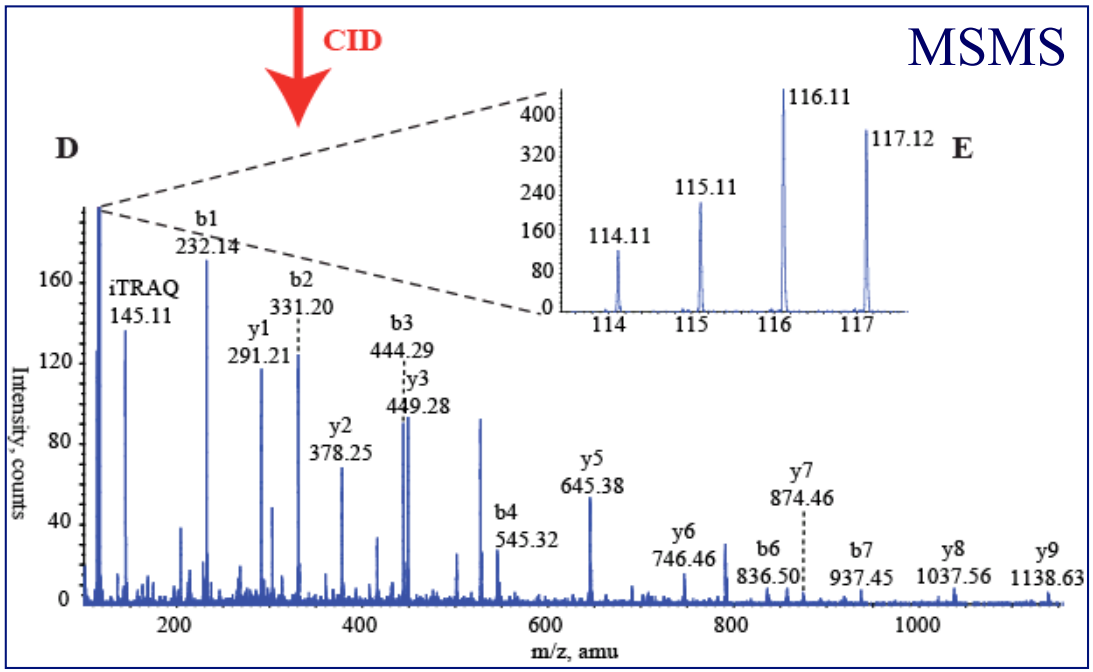
Ross et al (2004) Mol. Cell. Prot. 3:1154

Stable Isotope Labelling -*iTRAQ*

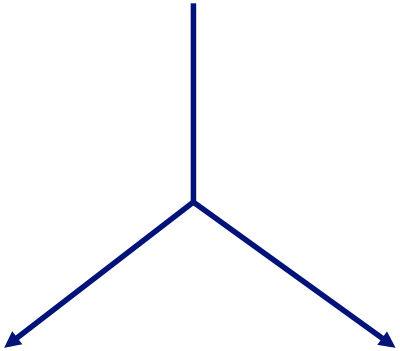
Isotopic Variation







Quantitation using a label free approach



Peak measurements

Spectral counting

Label Free Proteomics –Peaks

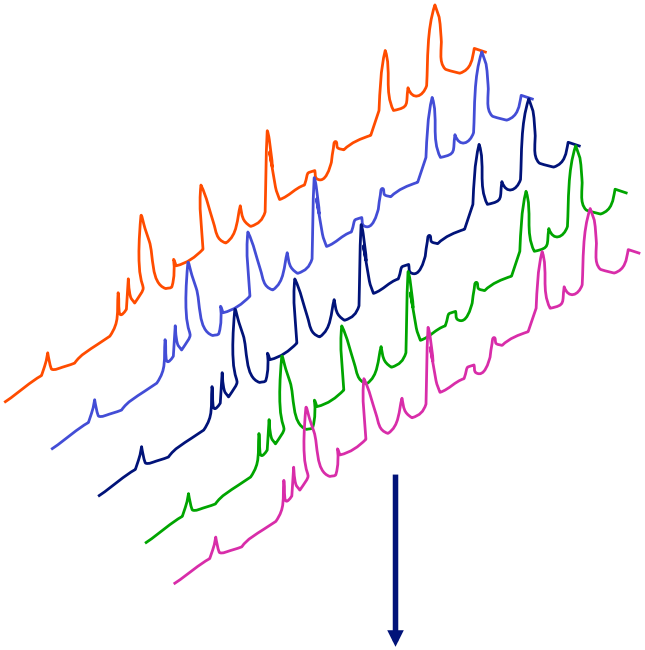
Ion intensity measurements

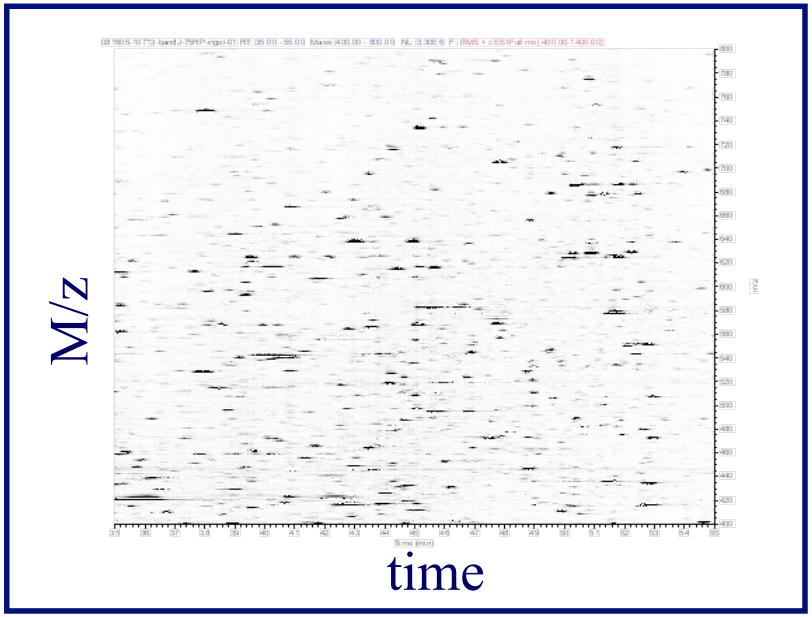
Compare peak intensities of the
same ion in consecutive LCMS runs

Need to match retention times with m/z values

Can be targeted approach collecting MSMS information in a separate run only fragmenting ions showing change in abundance

Essential to have good mass accuracy and reproducible retention times





Label Free Proteomics - Spectral counting

Spectral counts

Number of non-redundant spectra matching the same proteins

The number of redundant peptides observed correlates with abundance

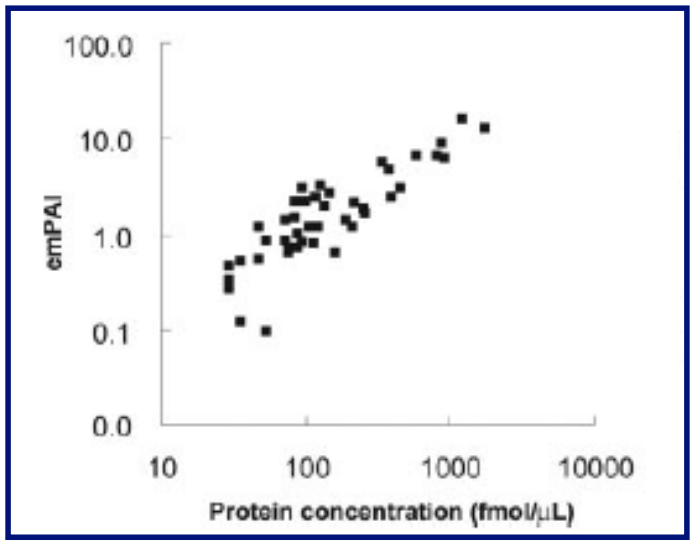
Must take length of protein into account
emPAI software available for analysis
(Exponentially modified protein abundance index)

See: Ishihama Y, *et al* Mol Cell Proteomics. (2005) 4(9):1265-72

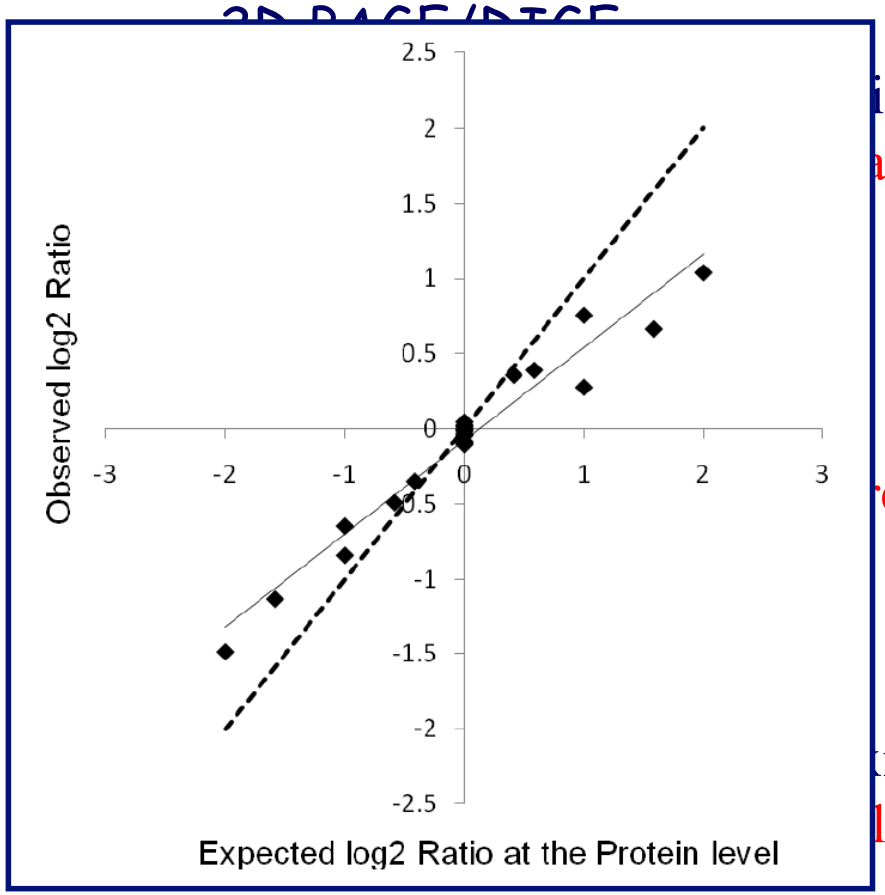
PAI = protein abudance index

number of observed peptdies /number of observable peptides

$emPAI = 10^{PAI} - 1$



Summary



in separation

age as no membrane proteins

potentially

e growth conditions can be specified

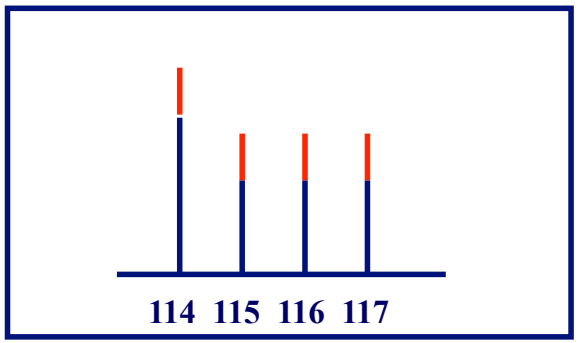
potentially

ting

leads to unestimation of large fold changes

Label free

- Cheap
- Complex data analysis
- Greatest variance?



Outline

Quantitation in proteomics

Relative Quantitation

Absolute Quantitation

Importance of Experimental Design

Importance of Suitable Data Analysis

Absolute Quantitation

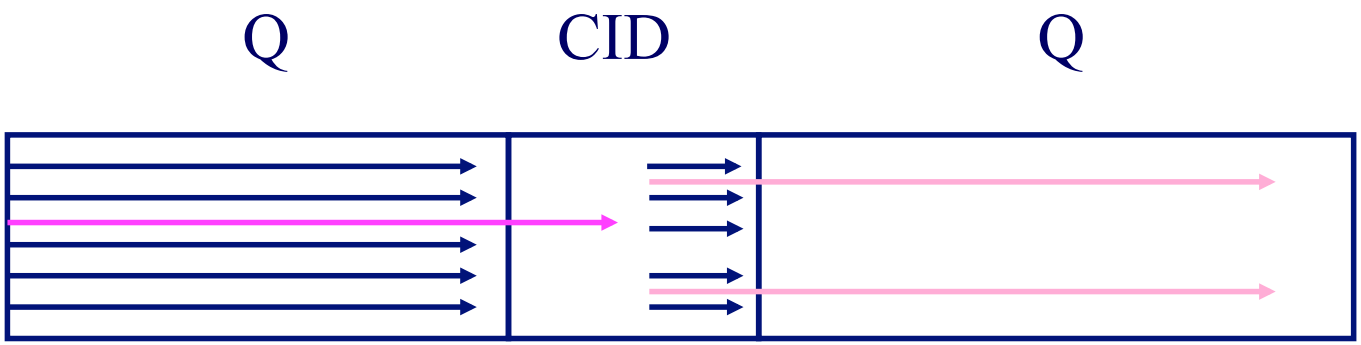
Assay proteins of interest

MS based absolute quantitation works by measuring peptide ‘surrogate’ simultaneously against quantified internal standard.

Surrogate = peptides

The ions that are used for measurement are generally MS/MS fragment ions which are discriminatory for the peptide of choice

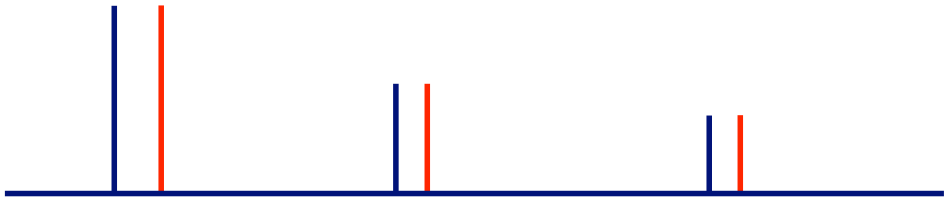
Multiple Reaction Monitoring



Precursor ion selected

Collision Induced Fragmentation

Diagnostic fragment ions selected = transitions



Natural peptide

Synthetic version of peptide containing stable isotope

How to create good peptide internal standard?

AQUA

- Gerber *et al* (2003) *PNAS* 100(12):6940-5

QconCAT

- Beynon *et al* (2005) *Nat. Methods* 2(8):587-9.

Labelled proteins ‘mass Western’

- Lehmann *et al* (2008) *The Plant Journal* 55:1039–1046

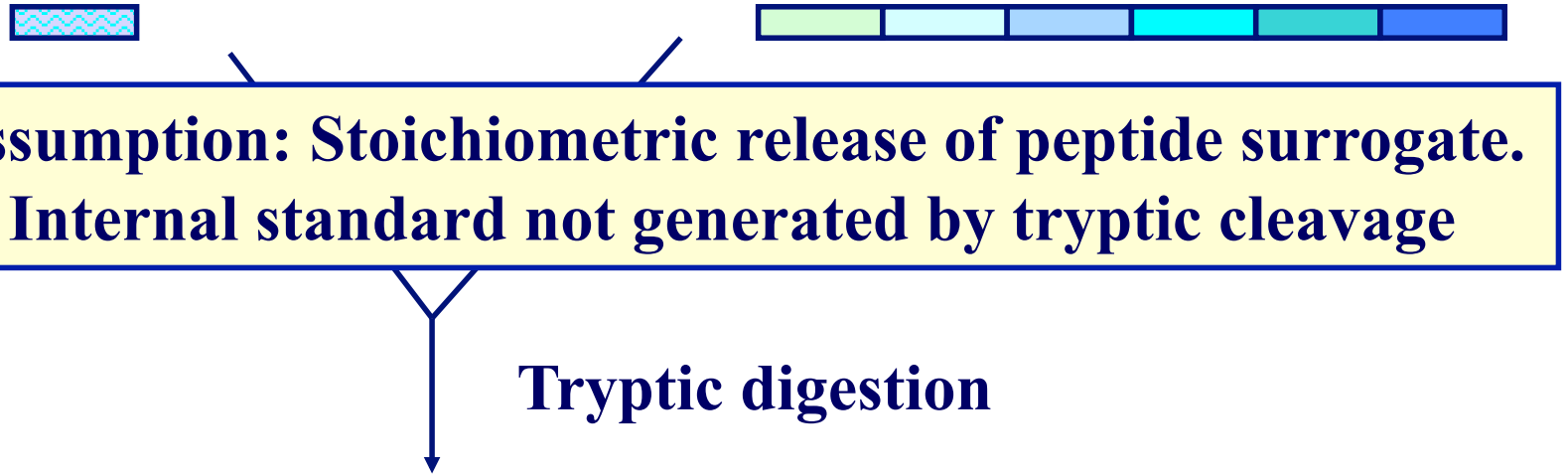
Good Example

- Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Cell. 2009 138(4):795-806

AQUA

Stable isotope tagged synthetic peptide

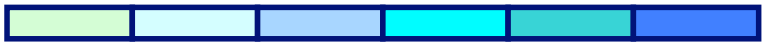
protein of interest



QconCAT

Stable isotope labelled synthetic protein
Constructed from concatenated peptides(Qprotein)

Protein of interest



Assumption: Stoichiometric release of peptide surrogate.
Internal standard not generated by identical tryptic cleavage

Tryptic digestion

Recombinant labelled protein Mass Western

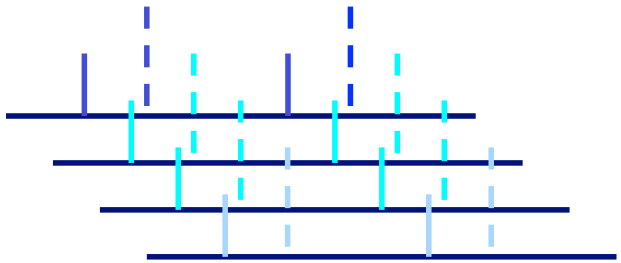
Stable

Assumption:

Identical tryptic cleavage for internal standard and surrogate.

Complete set of internal standards

Tryptic digestion



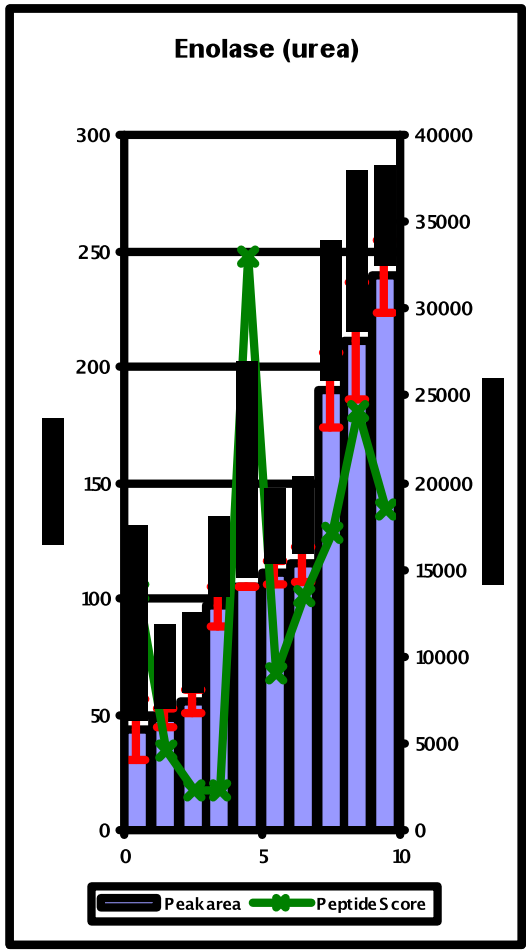
LC-MS^E
Multiplexed data acquisition

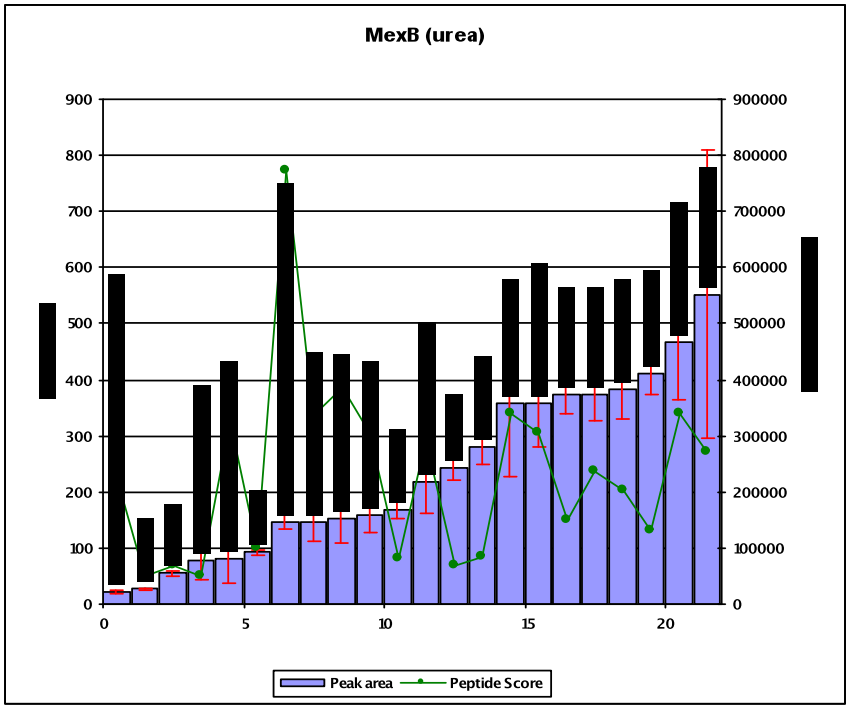
Add known amo
calibrate absolute
the performance o
pe

- Collision Energy in gas cell alternated between
  - Low energy (5eV)
  - Elevated energy (linear 15 eV - 42 eV)

Silva et. al., Anal Chem. (2005)
Liu et. al. Proteomics (2009)

MS^E Absolute and estimated Quantitation





Outline

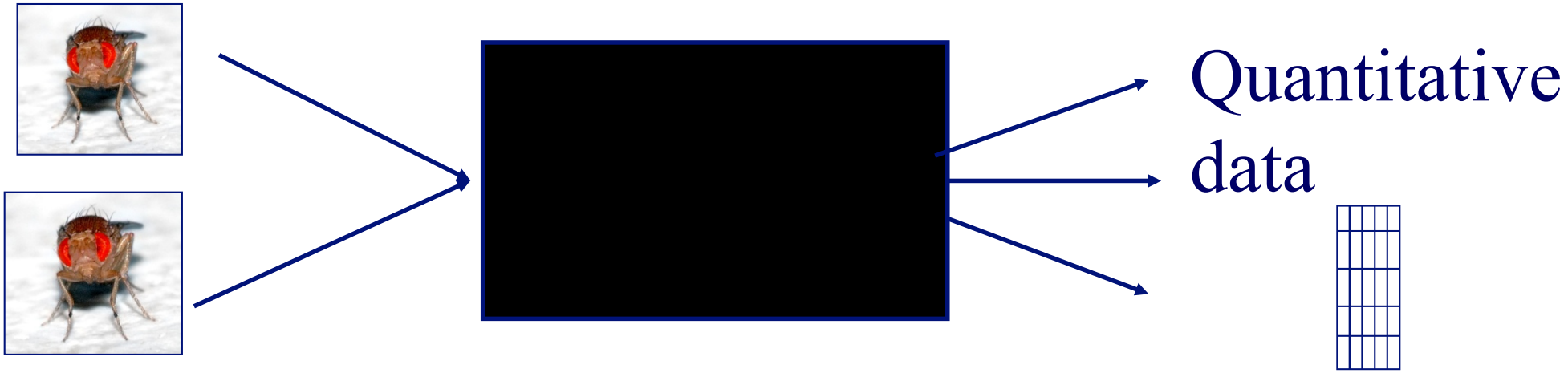
Quantitation in proteomics

Relative Quantitation

Absolute Quantitation

Importance of Experimental Design

Importance of Suitable Data Analysis



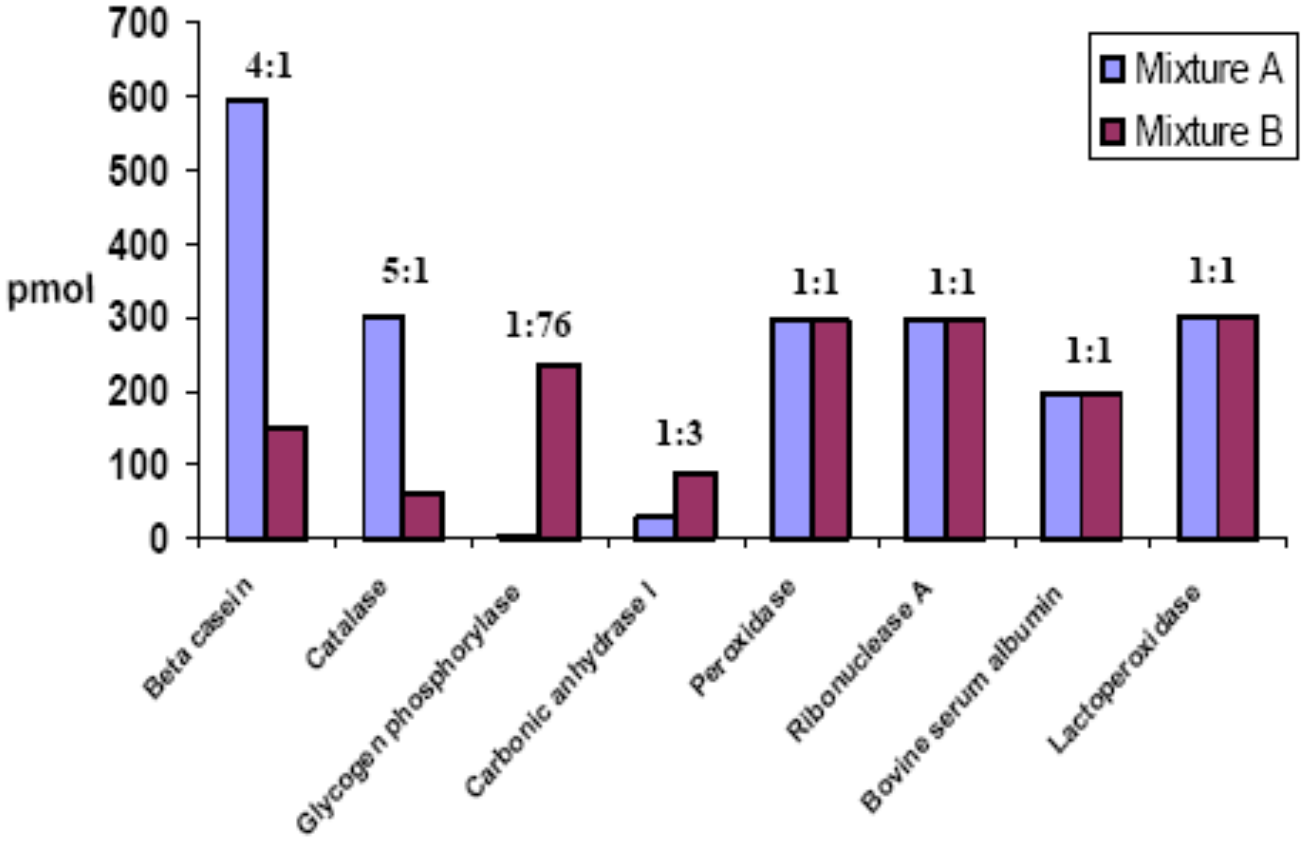
Knowledge of these facts influences

1. Design of experiment
2. Number of replicates utilised
3. Application of normalisation methods

The Association of Biomolecular Resource Facilities

www.abrf.org

ABRF Proteomics Research Group Study 2006



8 proteins

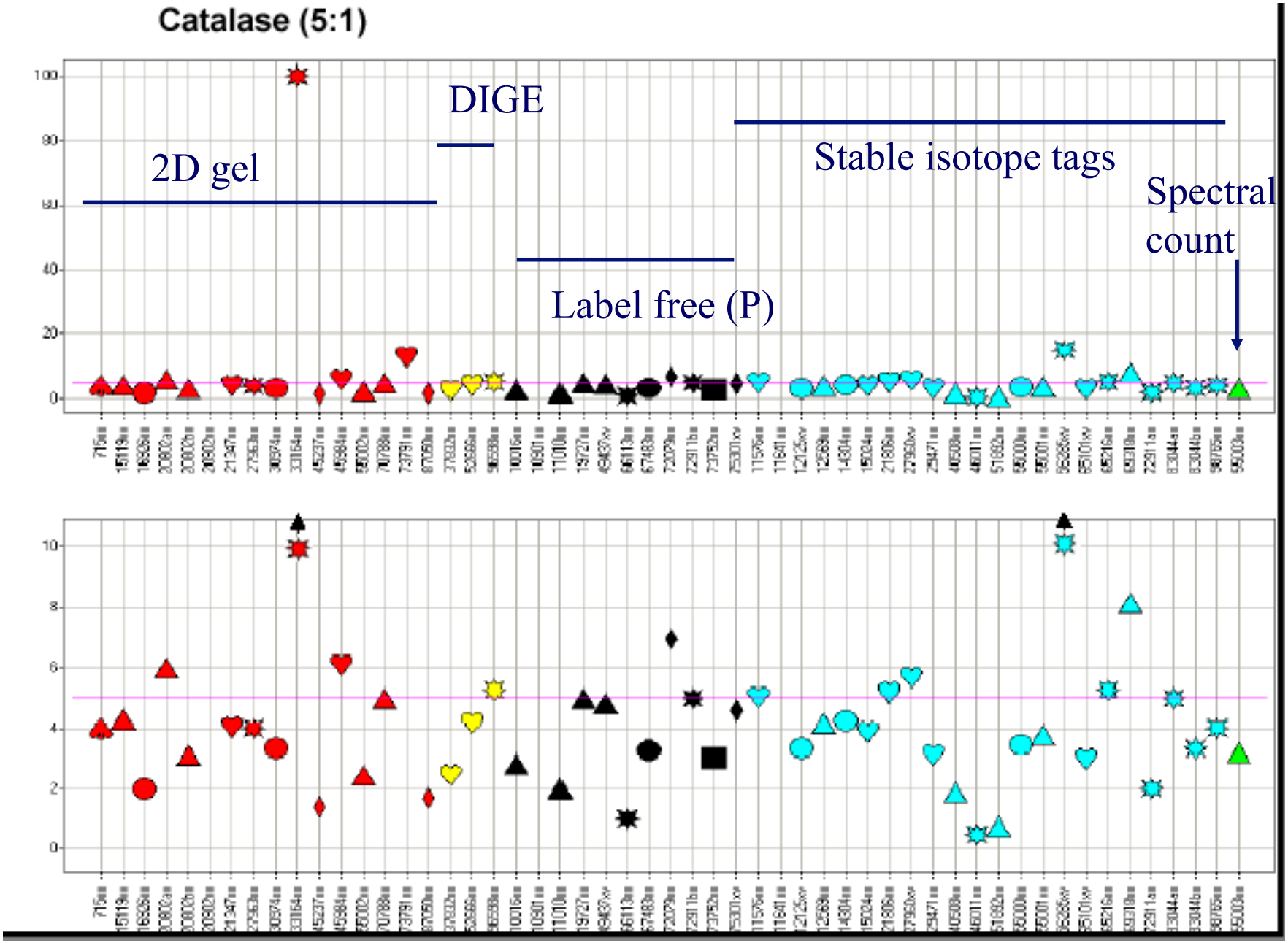
Same total amount of protein in each sample

52 responses

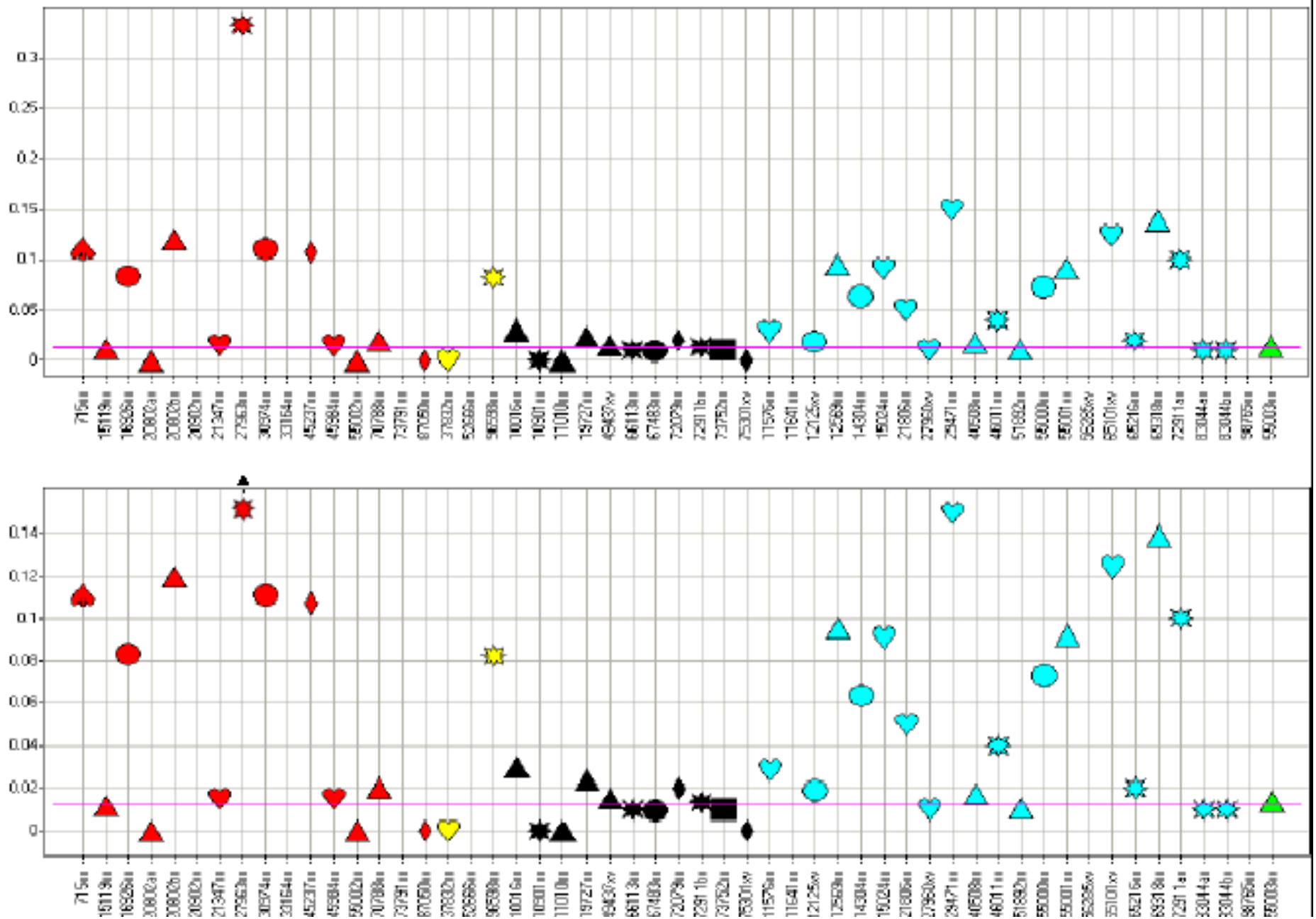
Turck *et al* (2007) Mol. Cell. Prot. 6(8):1291-8

Do they give the same results?

Cambridge Centre for Proteomics



Glycogen phosphorylase (1:76)



Why you don't get the same answer?

Variability in starting material
- Biological variation

Variability in experimental protocol (influences technical variance)
- Point at which you combine samples to be compared

Inappropriate experimental design
- Not enough replicates

Inaccuracy of measurement
- The wrong answer all the time
- The wrong answer some of the time

Inappropriate statistical testing
- Using a test that does not fit the data

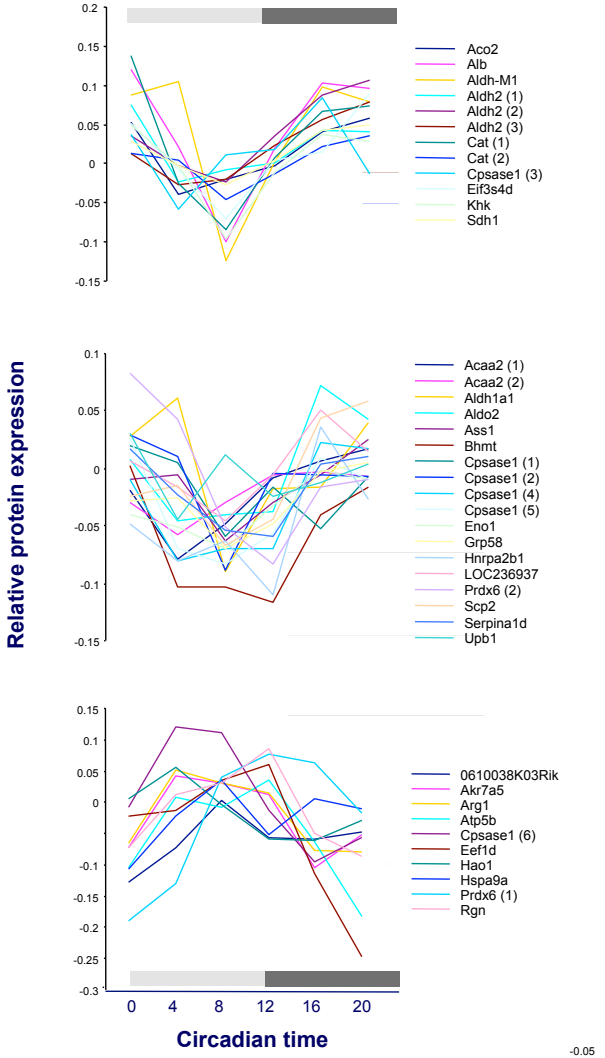
Biological Variance

Try to control as much of variance as possible

Standardised collection protocols

Appropriate samples (matched controls)

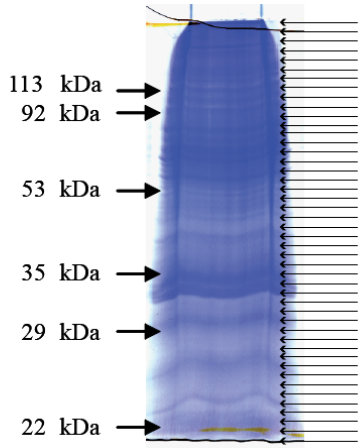
Time of harvest



Reddy et al (2006) Current Biology 16(11):1107-15.

Differential variance in a protocol

Extract proteins



in-gel tryptic digest

45 slices

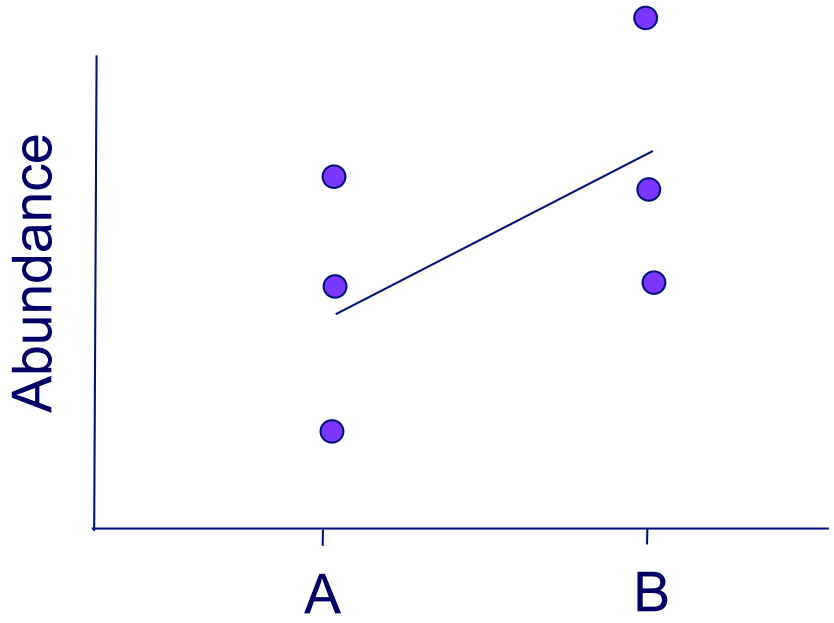
LC-MSMS

Points of variance

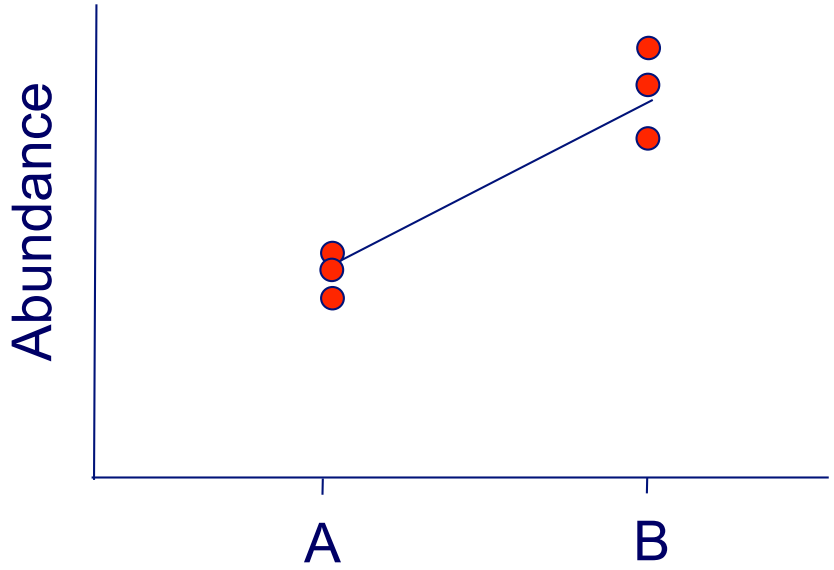
Extraction of proteins
1D gel
In gel digestion
LC
MS

Russell and Lilley in preparation

Types of Replicates



Biological Replicates



Technical Replicates

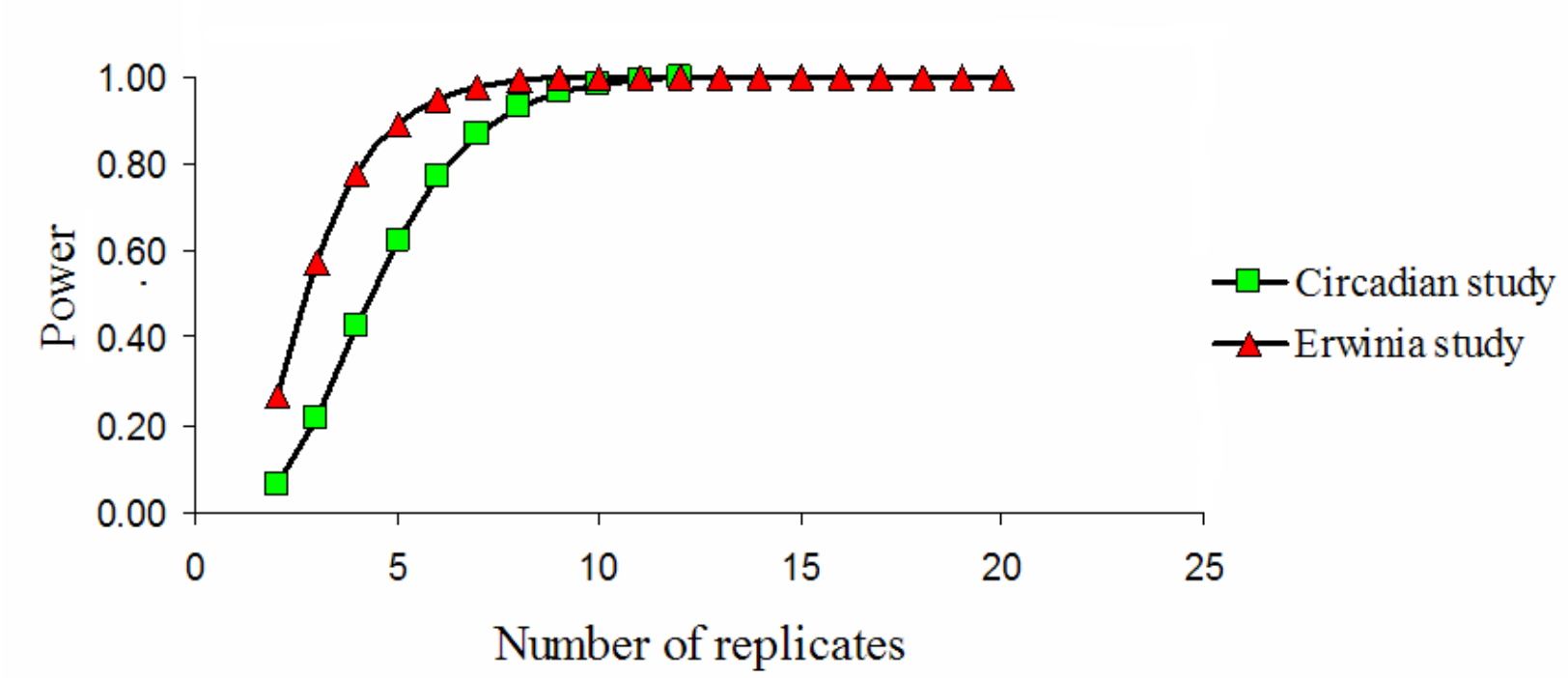
Technical replicates give an illusion of more power (sensitivity)

Karp *et al*, *J.Prot Res*, 2005, 4, 1867-71.

Power comparison

The power of a is the probability that the test will reject a false null hypothesis (i.e. that it will not make a Type II error). As power increases, the chances of a Type II error decrease. The probability of a Type II error is referred to as the false negative rate (β). Therefore power is equal to 1 − β.

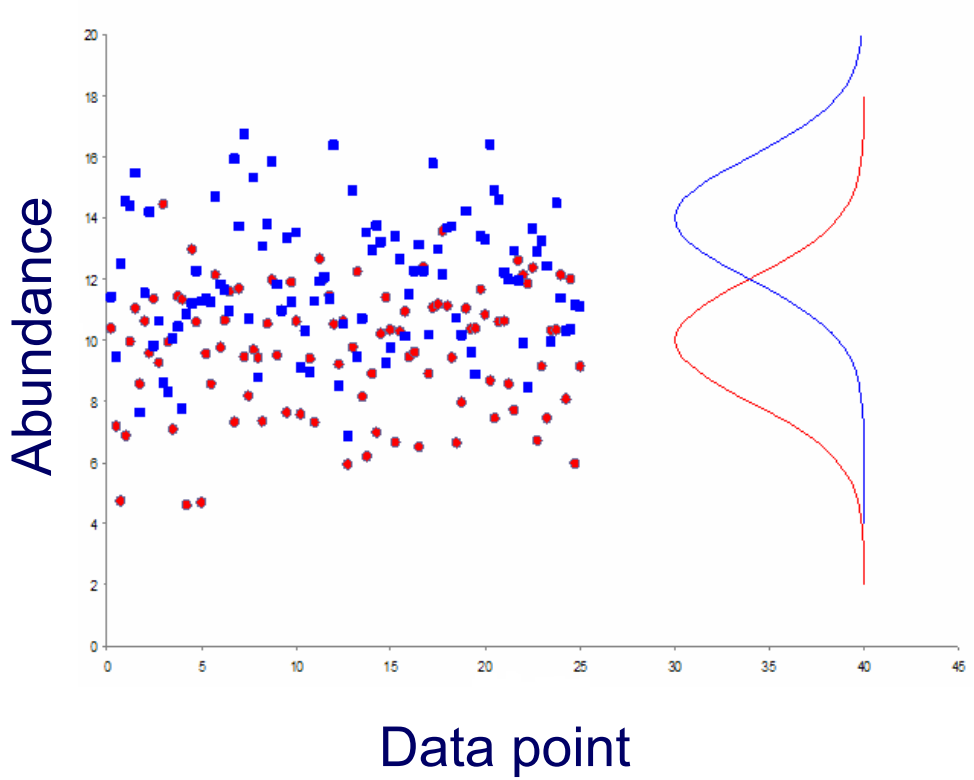
Depends on noise of system (variance), effect size (i.e. 2 fold), significance demanded by researcher (error you're prepared to live with), number of replicates.



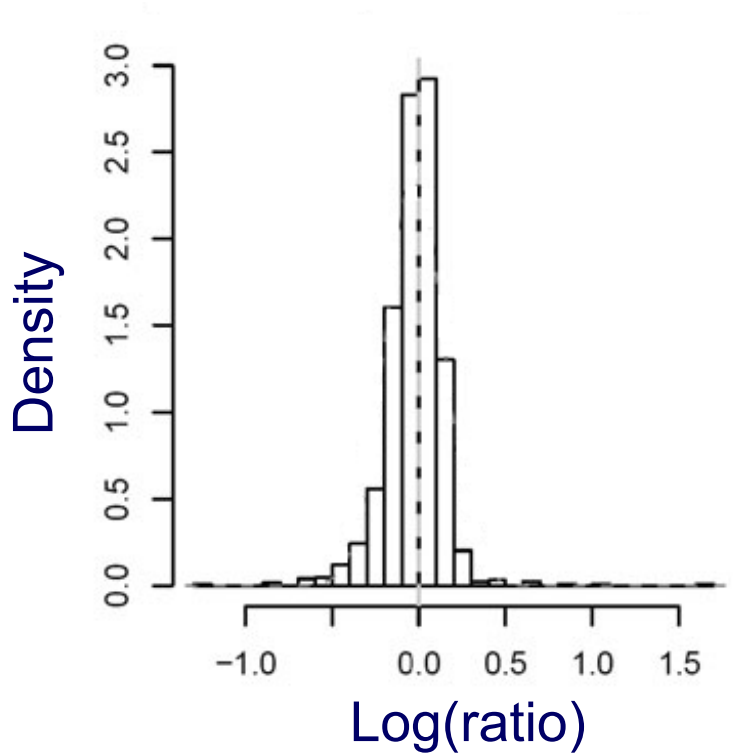
Calculated in detecting a 2 fold change with a noise measure that encompasses 75% of the species studied for a confidence of 0.01.

Karp *et al, Mol.Cell Prot*, 2007, 6, 1354-64.

Is the sample representative?



What threshold should you use?

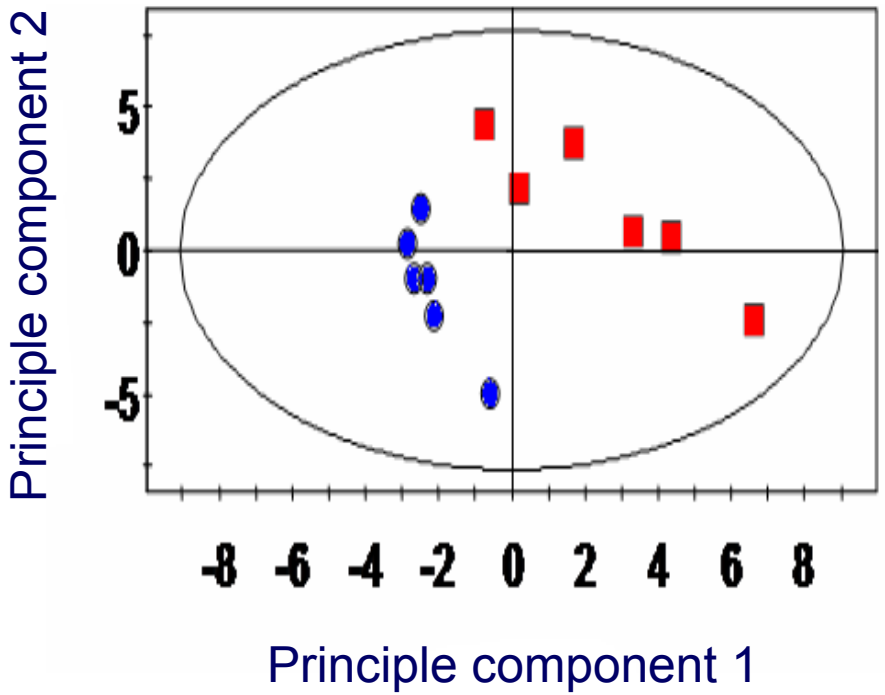


Karp *et al*, *Proteomics*, 2004, 4, 1421-32.

Randomisation in design

Cy3	Cy5	Cy5
control	treated	Internal standard
treated	control	Internal standard
control	treated	Internal standard
treated	control	Internal standard

Cy3	Cy5	Cy5
control	treated	Internal standard
control	treated	Internal standard
control	treated	Internal standard
control	treated	Internal standard



batch effects seen in same-same study.

Karp *et al*, *Proteomics*, 2005, 5, 81-90.

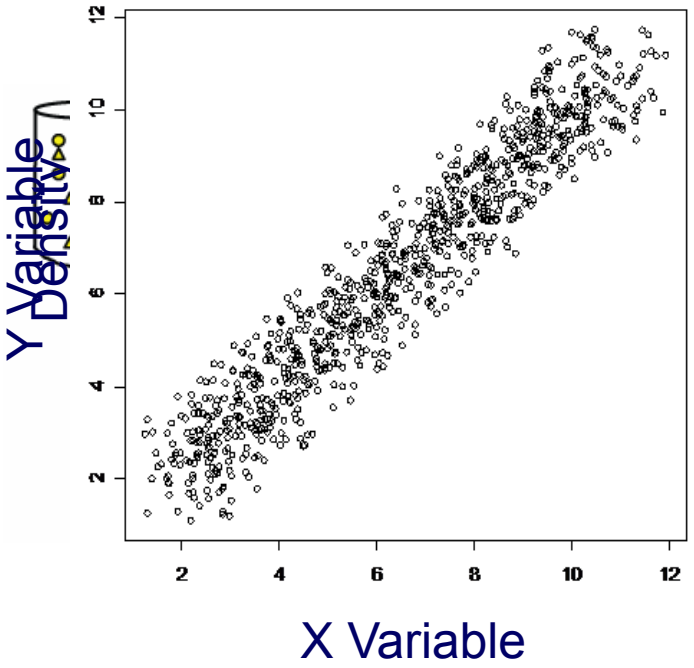
Are you using the correct statistical test?

Assumptions:

Normality

Homogeneity of variance

Independent sampling

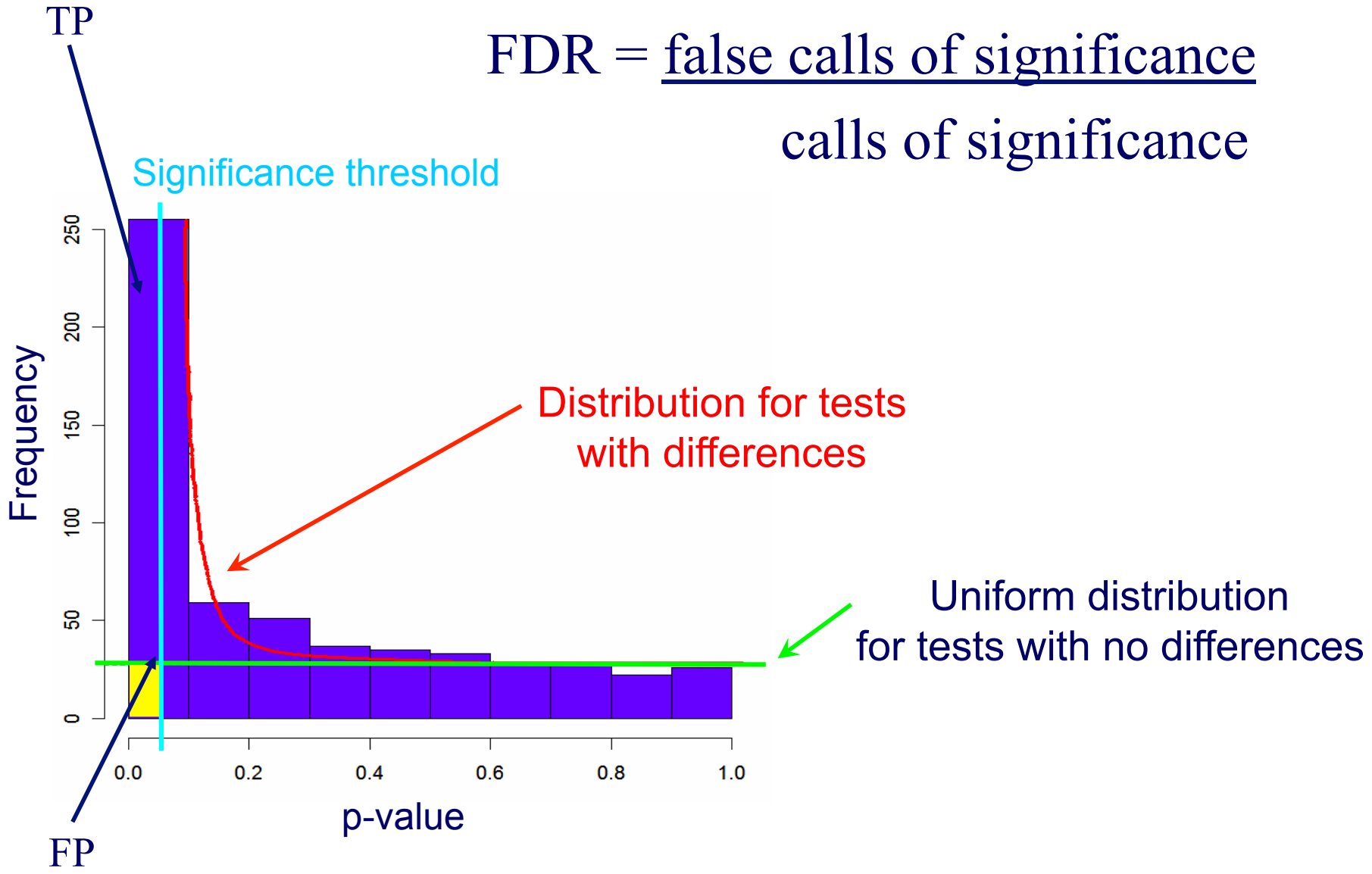


Karp *et al*, *MCP*, 2007, 6, 1354-64.

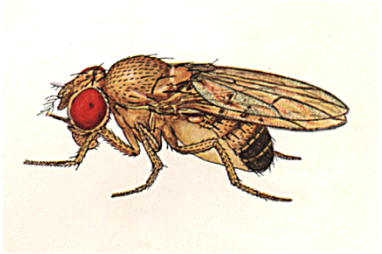
Karp & Lilley, *Proteomics*, 2005, 5, 3105-15.

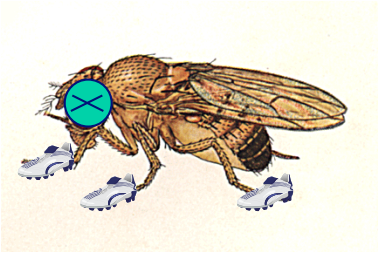
False Discovery Rate

$$FDR = \frac{\text{false calls of significance}}{\text{calls of significance}}$$



Importance of communication and design





Thank you for listening

Kathryn Lilley
Cambridge Centre for Proteomics

k.s.lilley@bioc.cam.ac.uk

www.bio.cam.ac.uk/proteomics