Introduction to Proteomics

Åsa Wheelock, Ph.D.
Division of Respiratory Medicine &
Karolinska Biomics Center
asa.wheelock@ki.se

In: *Systems Biology and the Omics Cascade*,
Karolinska Institutet, June 9-13, 2008

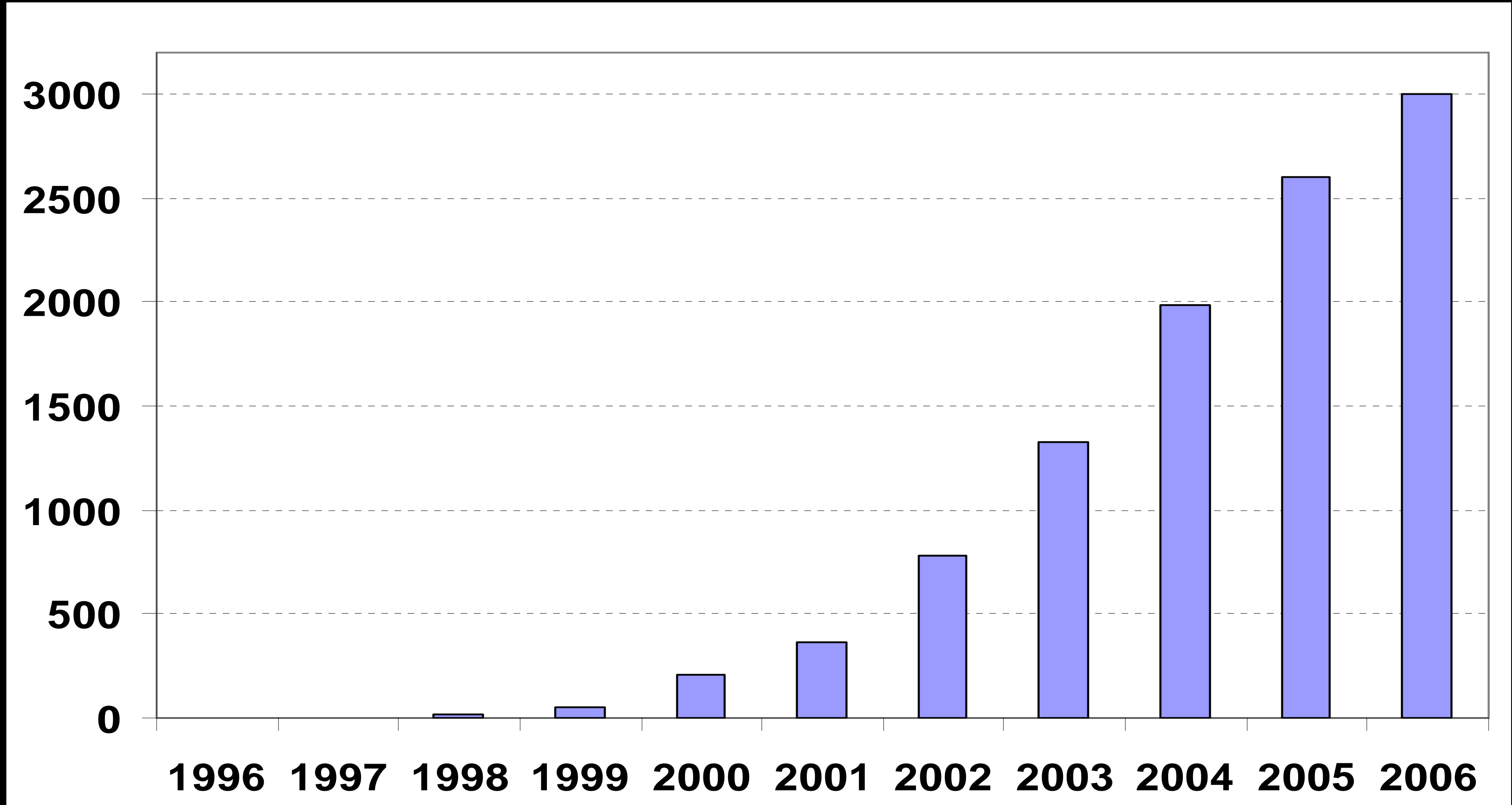
Focus of course: Tools for data analysis

Your analysis is no better than data you have collected...

The goals of this proteomics overview:

- Understand possibilities & limitations
- Pros and cons of different method
- Sources of variance in proteomics
- Take advantage of proteomics core facilities
- Perform proteomics collaborations
- Write a short research proposal in proteomics

Proteomics publications in Pubmed

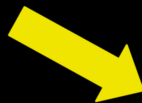


First ”proteomics” publication

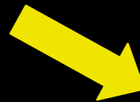
Why Proteins?!?

- Business end of the cell
- Detailed information with limited efforts
  - – As compared to metabolomics
- Relatively robust methods available

TRANSCRIPTOMICS Limited info from mRNA



PROTEOME Detailed info
Robust technology



METABOLOME Techincally challenging

Proteomics Methodology

- No ”protein PCR”
  - 4 nucleotides vs 20+ amino acids
  - Post-translational modifications (PTM)

3 MAIN PROTEOMICS PLATFORMS

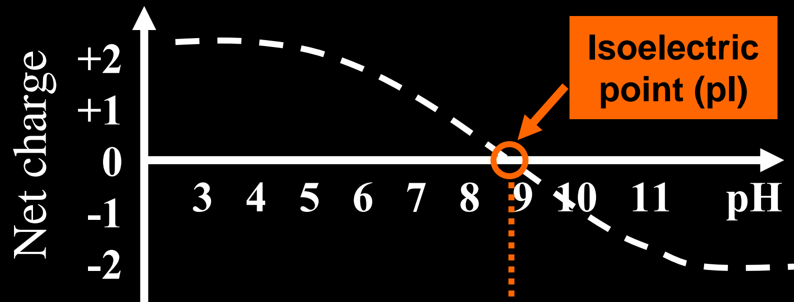
- Gel based methods
- Shotgun methods (mass spec-based)
  ”chromatography-based”, ”gel-free”
- Array based (antibody based)

Gel based: 2-Dimensional Electrophoresis

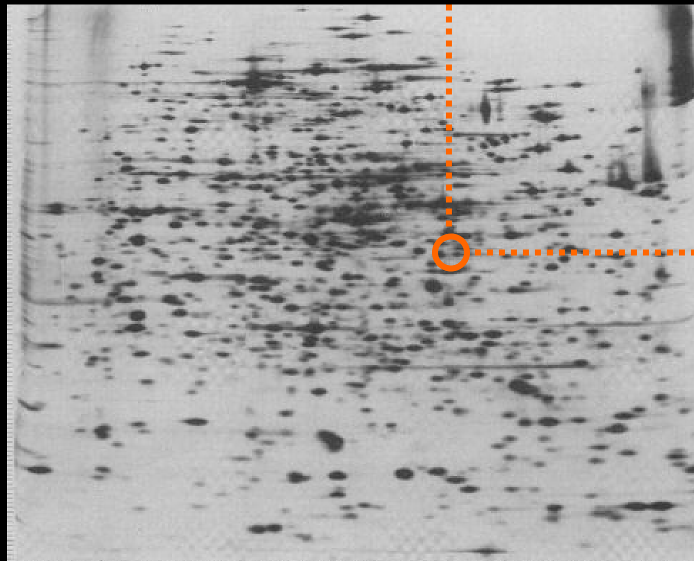
SEPARATION

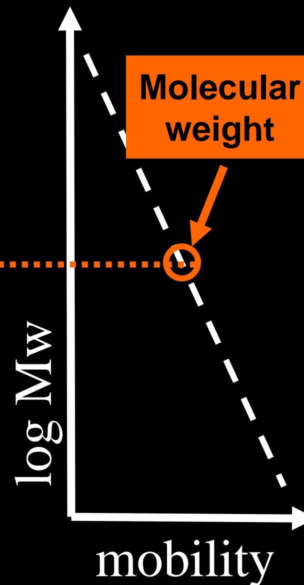
VISUALISATION

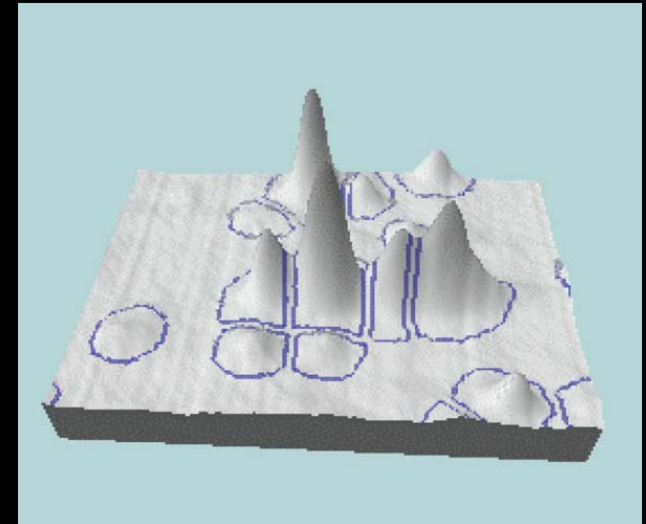
QUANTIFICATION



Stoichiometric protein stain => 3rd dimension







Klose, J. 1975. Humangenetic 26, 231-43
O'Farrel, P. 1975. J. Biol.Chem, 250, 4007-21

Image acquisition

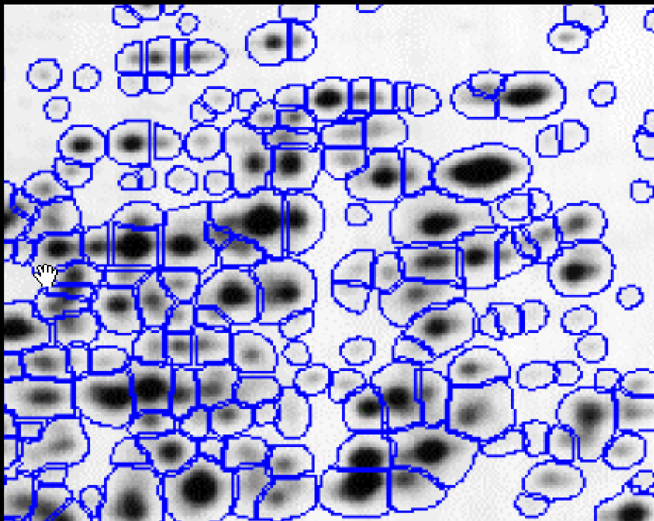
Sample Tray
Gel
Emitted Light
Laser
Laser Light
(488 nm)
Mirror
Fiber Optic Collector
530 nm
610 nm
PMT
Filters

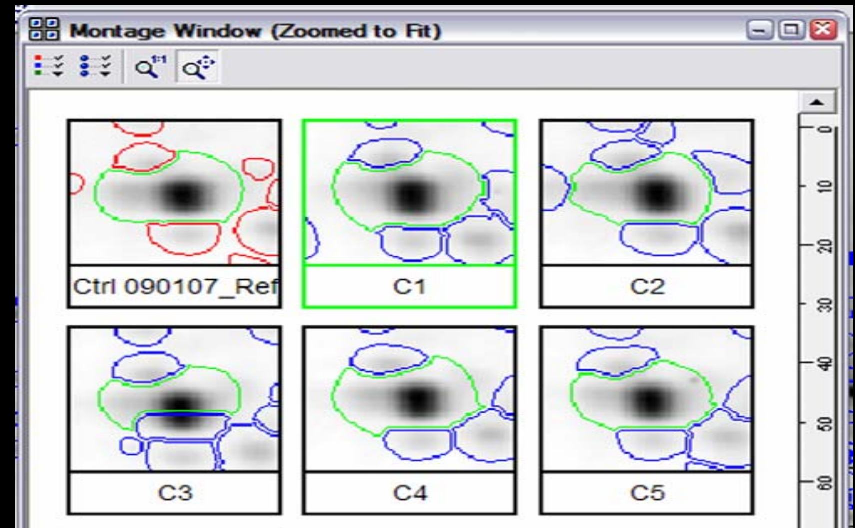
Image acquisition using fluorescent scanner

Quantitative image analysis

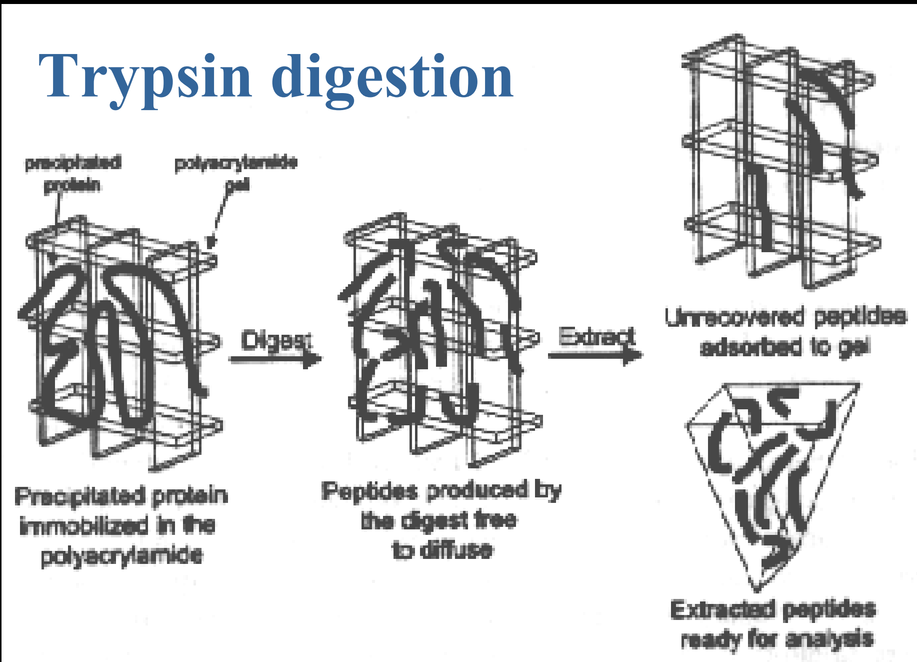
Pixel intensity => 3rd dimension
Spot volume = protein quantity

1. Detect spots
2. Match spots across gels
3. Quantify spot volumes





Protein identification

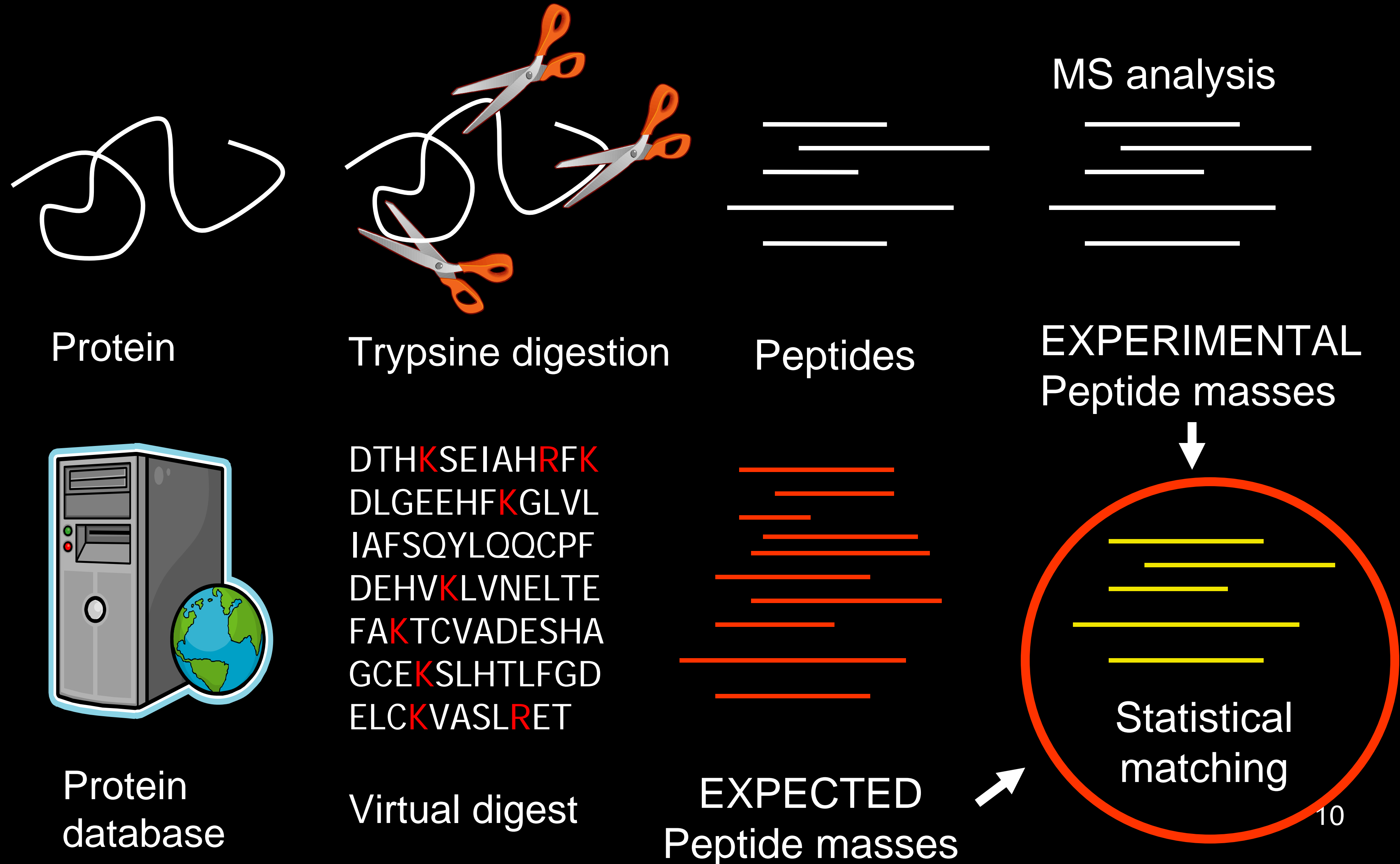


⇒ **Mass spectrometry (MALDI-TOF/TOF)**

⇒ **DATABASE SEARCH => IDENTIFICATION**

(Swiss-prot, EnSemble) **(statistical probability)**

Protein Identification



Peptide mass mapping

MS analysis=>
peptide masses

DLGEEHFK

LHTLFGDR

MS/MS analysis=>
sequence information

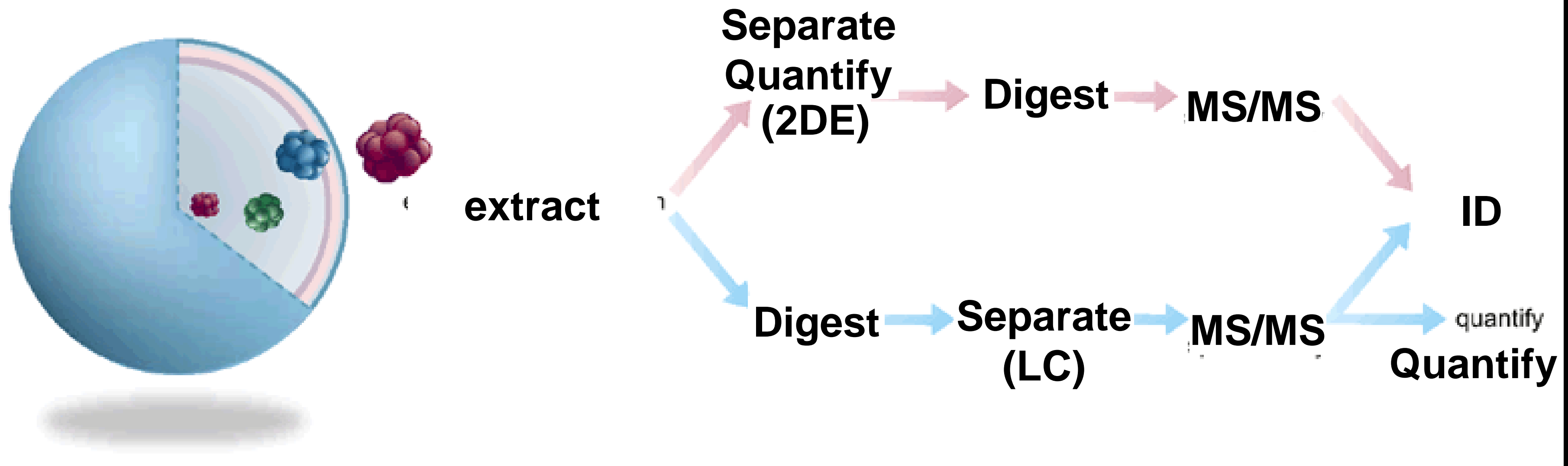
database
search

Statistical
matching

DTHKSEIAHRFK
DLGEEHFKGLVL
IAFSQYLQQCPF
DEHVKLVNELTE
FAKTCVADESHA
GEKSLHTLFGRE
LCKVASLRET

Homology search
Validate statistical hit

Shotgun vs. Gel-based proteomics



Adapted from Patterson and Aebersold, Nature Genetics 2003, 33:311-23. Fig. 3

Semi-quantitative proteomics

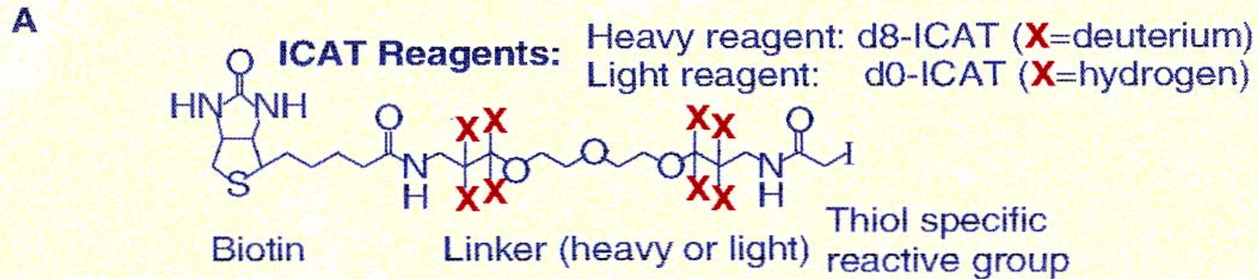
Both 2DE and MS-based methods

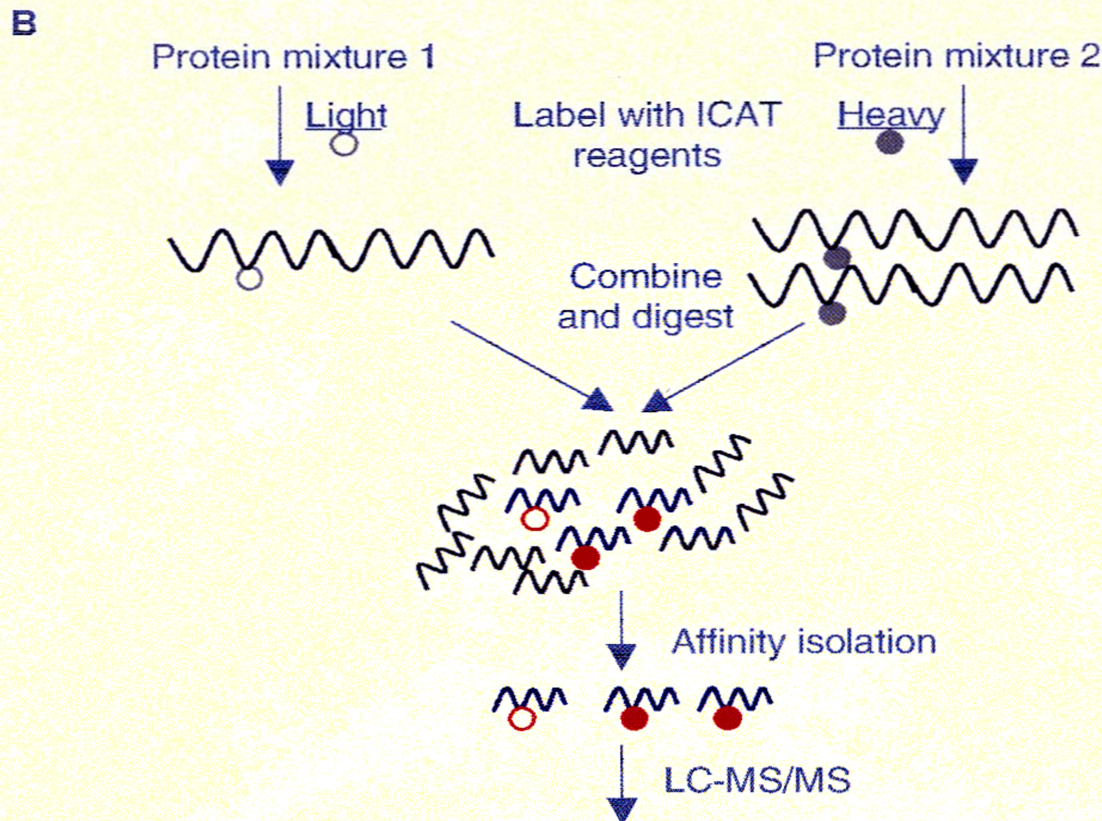
NOT quantitative by nature

Co-separation: 2 samples => ratios

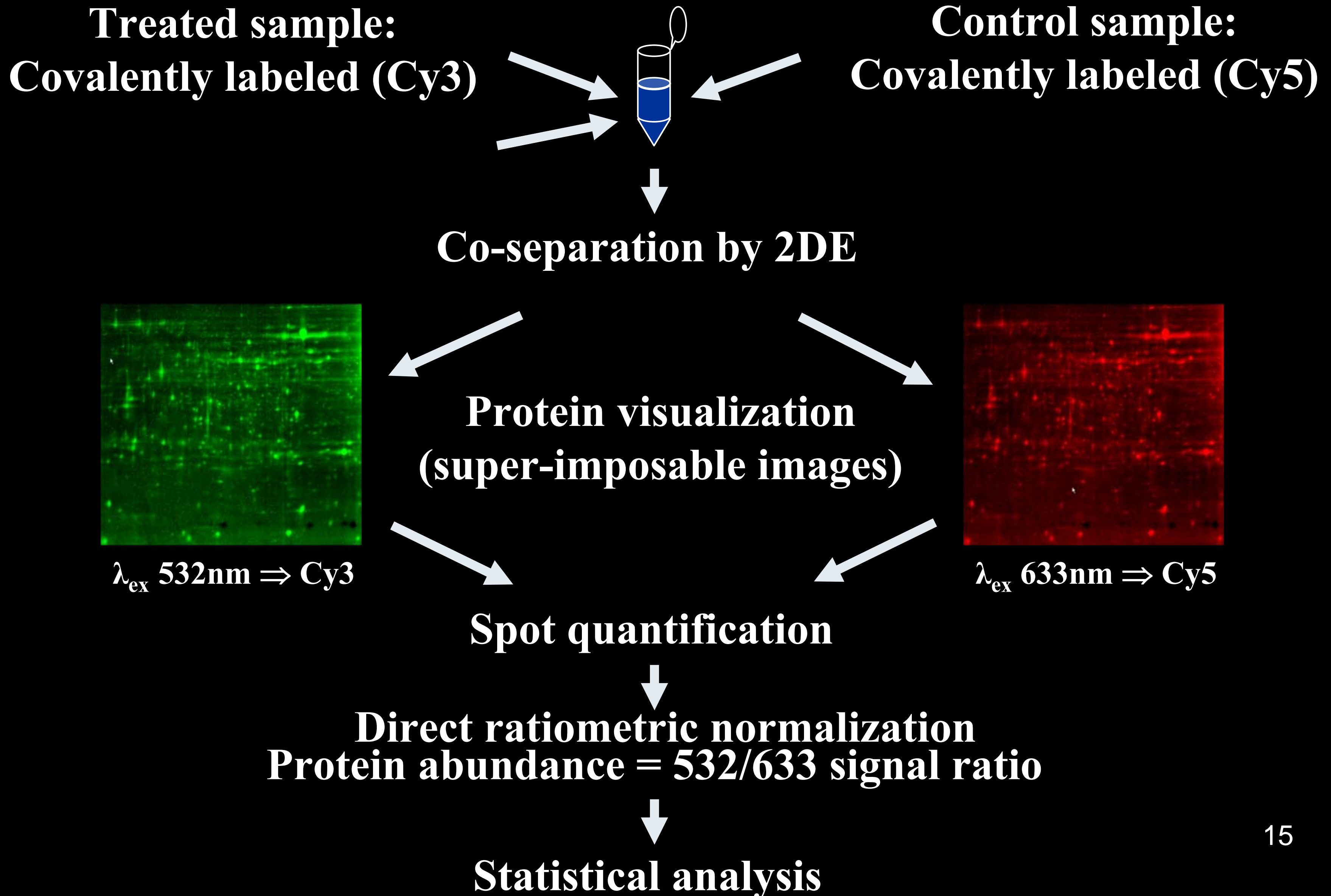
Tags => Semi-quantitative proteomics

Shotgun isotope-tagging : Isotope coded affinity tag (ICAT)





Multiplexing in 2DE: DIGE
- Differential Gel Electrophoresis



Semi-quantitative proteomics

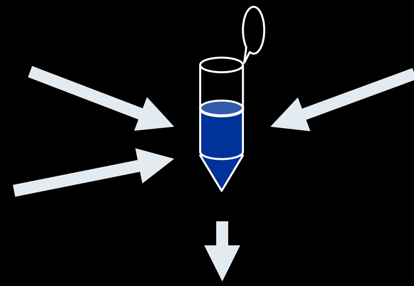
Both 2DE and MS-based methods
NOT quantitative by nature

Co-separation: 2 samples => ratios
Tags => Semi-quantitative proteomics

Pooled internal standard + 2-3 samples =>
Relative quantification

Internal Standards in 2DE: DIGE

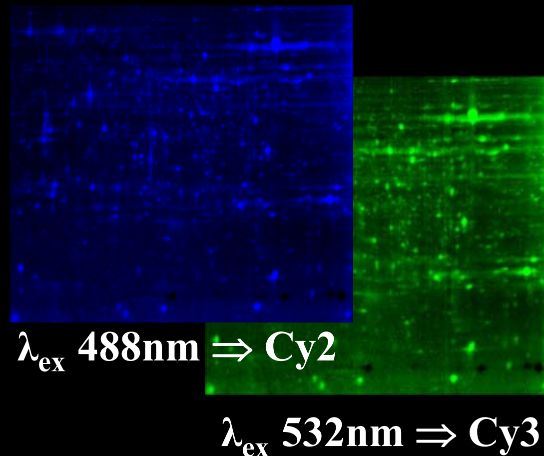
Treated Sample:
Covalently labeled (Cy3)



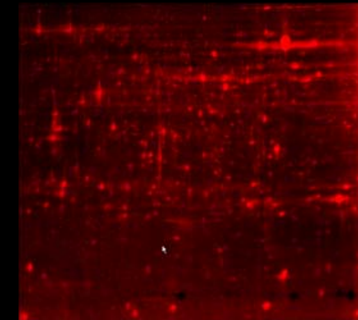
Control Sample:
Covalently labeled (Cy5)

Pooled Internal
standard (Cy2)

Co-separation by 2DE



Protein visualization
(super-imposable images)

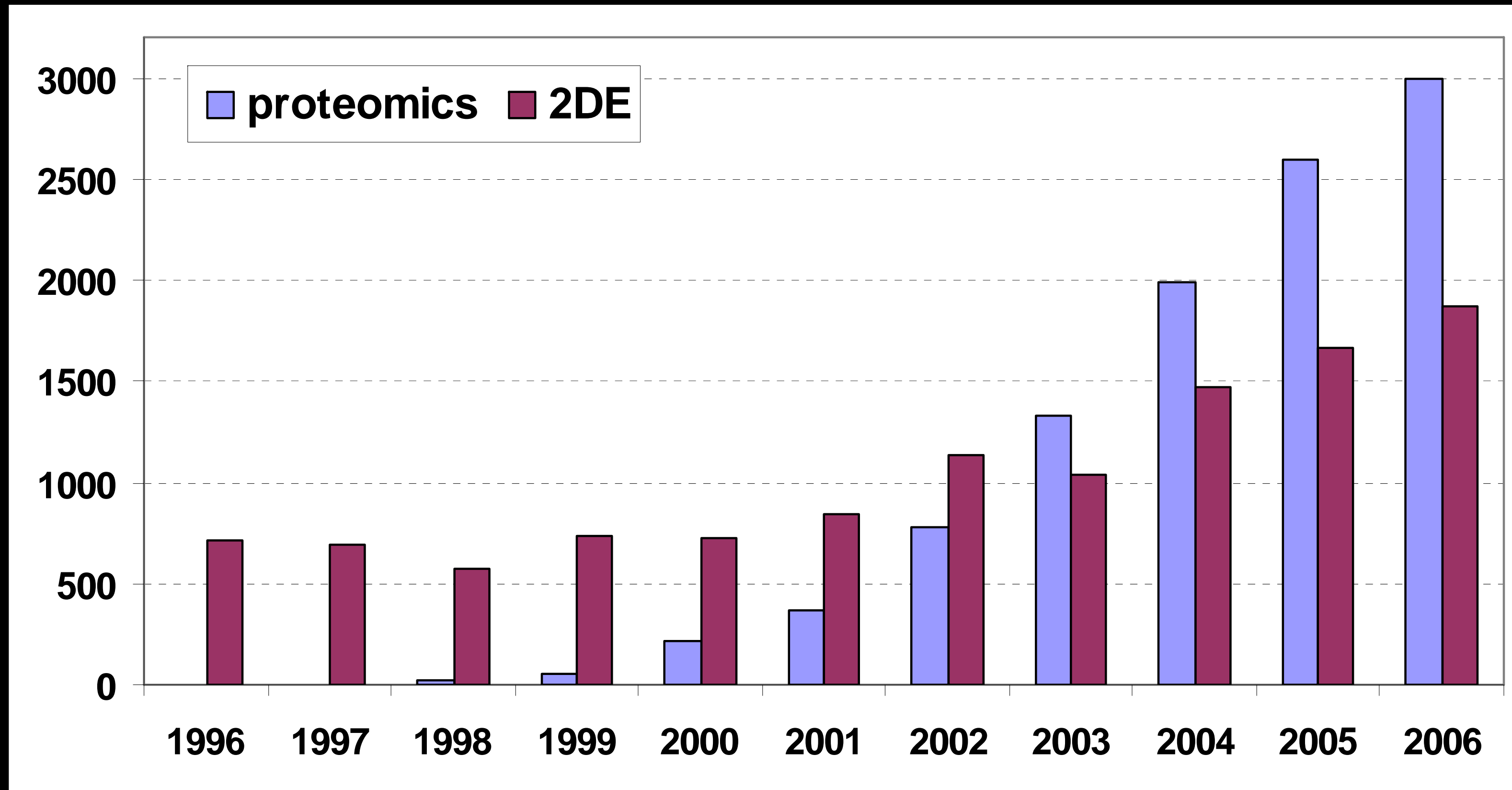


Spot quantification

Direct ratiometric normalization
Protein abundance = relative to IS

Statistical analysis

Proteomics in Pubmed

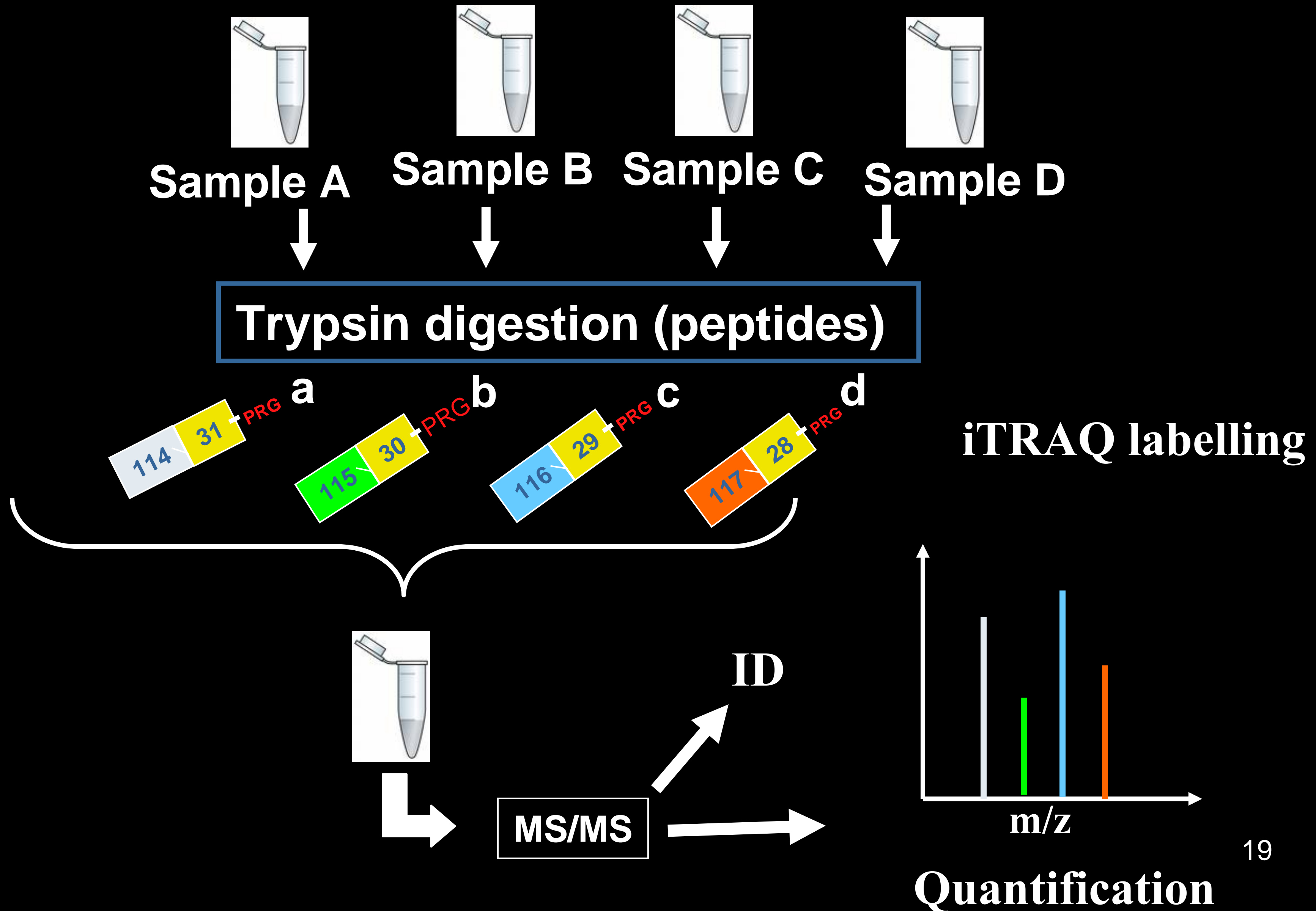


First
”proteomics”
publication

First
DIGE-method
published

Multiplexing in MS: iTRAQ

- isobaric Tag for Relative and Absolute Quantitation



Differential labelling opens up new possibilities

- Cysteine oxidative states
- Identify peptides on plasma membrane surface
- Cellular re-localization

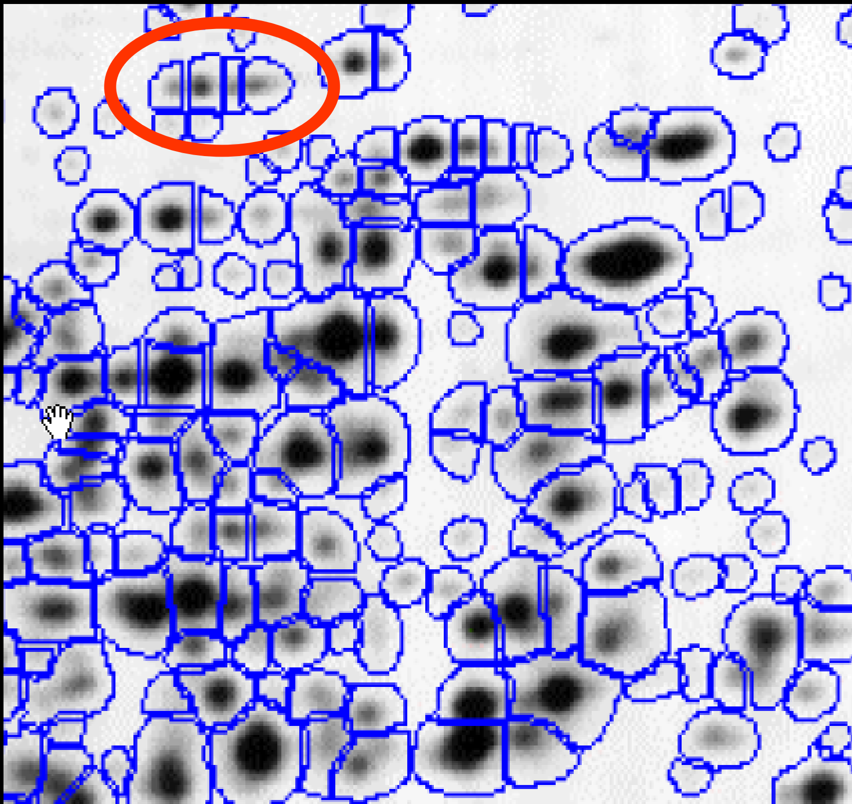
2D or not 2D?

Gel-based methods: 2-D electrophoresis

+ soluble proteins

+ post-translational modifications

Post-translational modifications
”Spot trains”
Intact proteins



2D or not 2D?

Gel-based methods: 2-D electrophoresis

+ soluble proteins

+ post-translational modifications

- technical variance, time consuming

MS-based (Gel-free) methods: ICAT, iTRAQ

+ membrane proteins

+ low abundance proteins

Extremes of physiochemical properties: Peptides

- Charge
  - pI range from 3-12
- Size
  - Mw range of 5 – 500,000 kDa
- Hydrophobicity
  - membrane proteins

2D or not 2D?

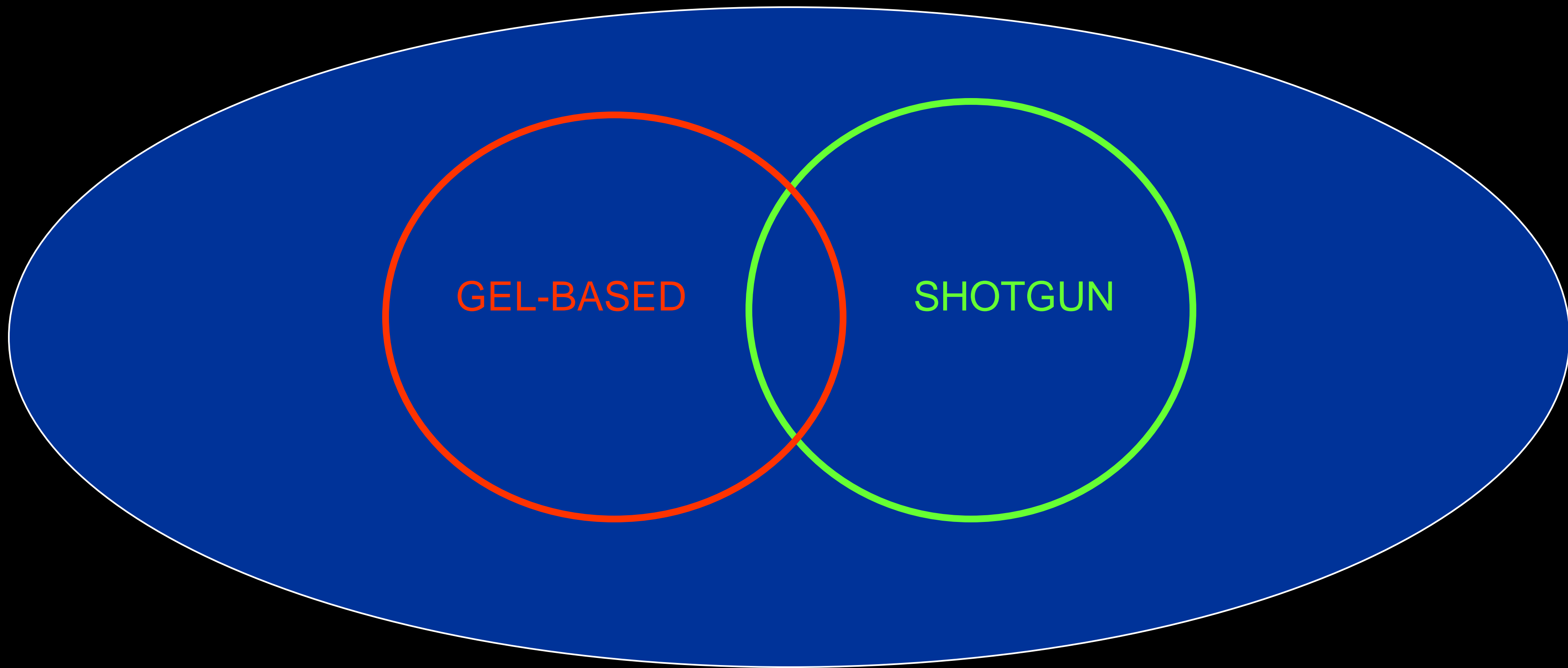
Gel-based methods: 2-D electrophoresis

- **+ soluble proteins**
- **+ post-translational modifications**
- **- technical variance, time consuming**

MS-based (Gel-free) methods: ICAT, iTRAQ

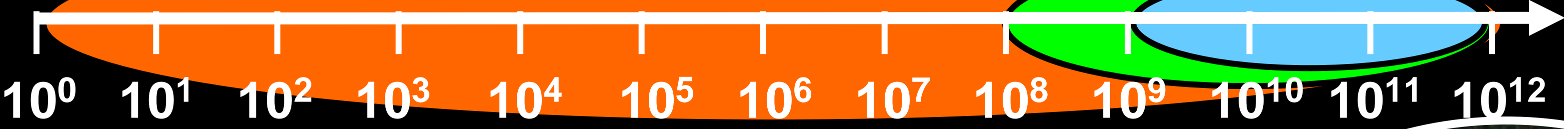
- **+ membrane proteins**
- **+ low abundance proteins**
- **- expensive, data intense**

Shotgun approcahes and gel-based approaches complementary



No ”true” proteomics technique yet

DYNAMIC RANGE



COPIES of each PROTEIN

10^{12}

10^4

Kyoto
京都

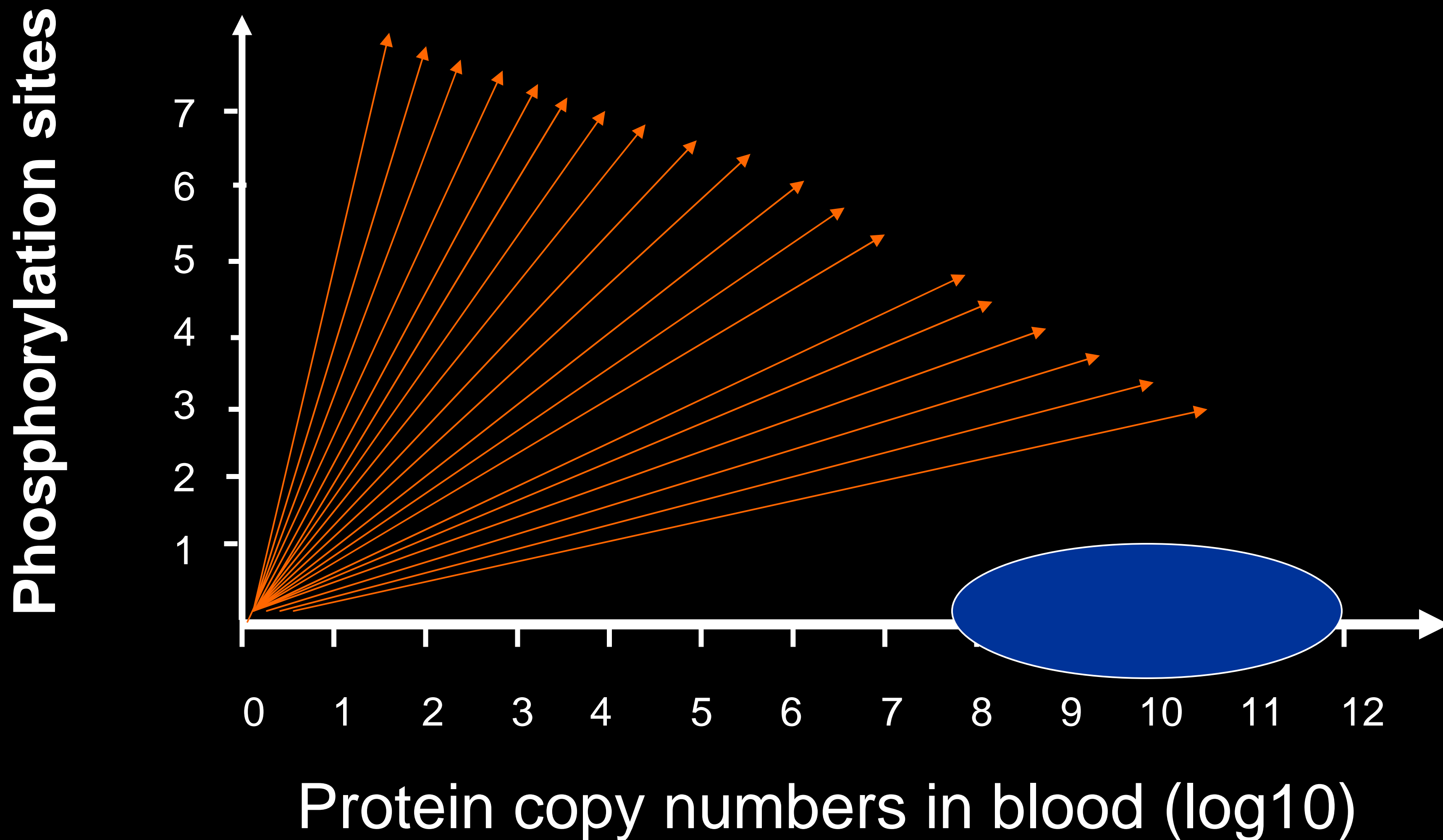
Kobe
神戸

Osaka
大阪

日本
Japan

10^3

Post-translational Modifications (PTMs) - 400 reported PTMs



Variance in 2DE

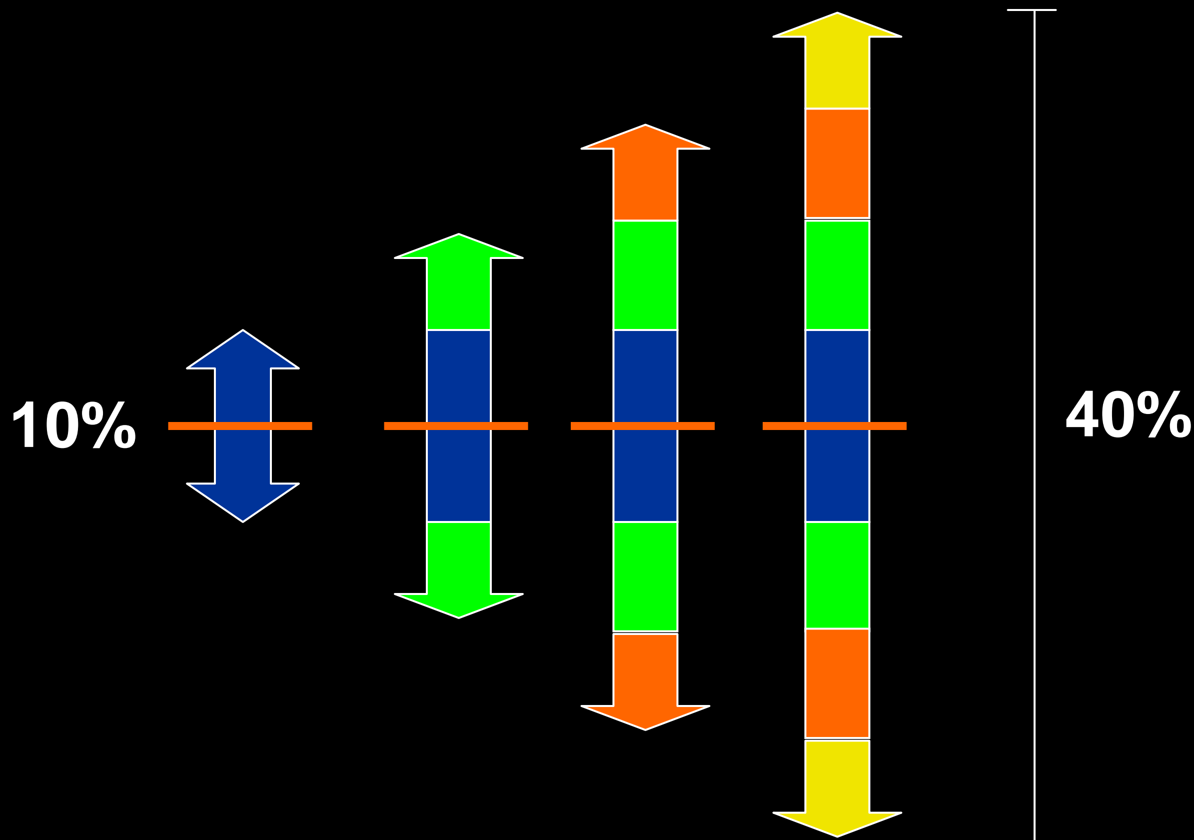
- BIOLOGICAL VARIANCE
- Experimental variance
  - Pre-fractionation, isolation & labelling of proteins
  - Protein staining
- Technical variance
  - Gel-to-gel variation in 2DE
  - Image acquisition (scanner)
- Post-experimental variance
  - Software-induced variance
  - User dependant variance

Variance in 2DE

- BIOLOGICAL VARIANCE
- **Experimental variance**
  - Pre-fractionation, isolation & labelling of proteins
  - Protein staining
- Technical variance
  - Gel-to-gel variation in 2DE
  - Image acquisition (scanner)
- Post-experimental variance
  - Software-induced variance
  - User dependant variance

Remember that variance adds up:

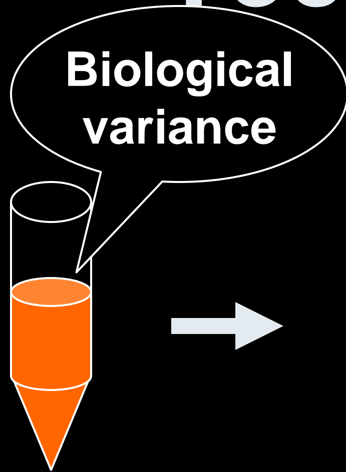
Multiple-step method is not your friend...

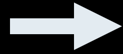


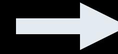
Variance in 2DE

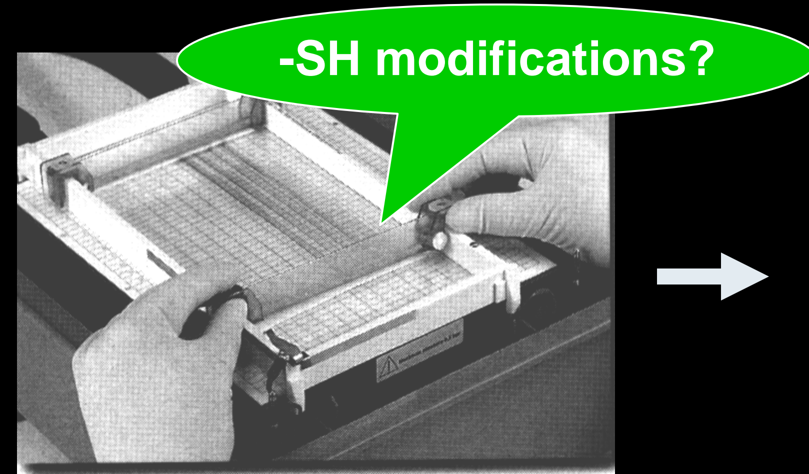
- BIOLOGICAL VARIANCE
- Experimental variance
  - Pre-fractionation, isolation & labelling of proteins
  - Protein staining
- **Technical variance**
  - **Gel-to-gel variation in 2DE**
  - **Image acquisition (scanner)**
- Post-experimental variance
  - Software-induced variance
  - User dependant variance

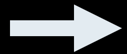
Technical variance in 2DE







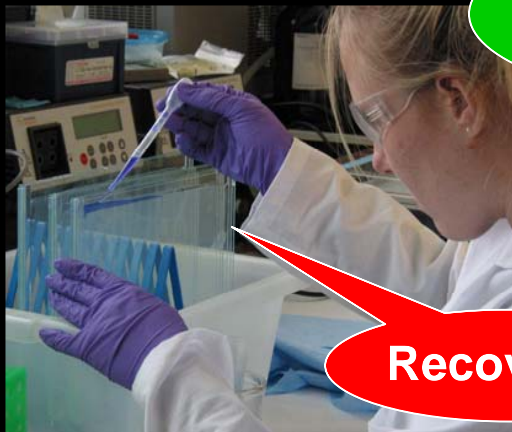


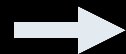


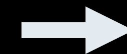
Solubilization

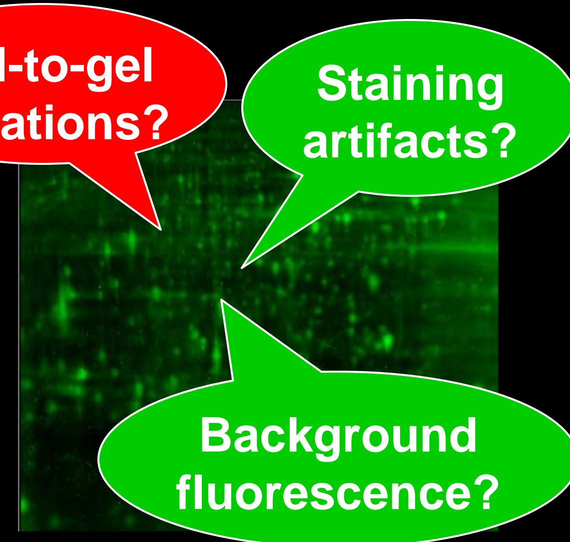
Rehydration

Isoelectric focusing









Load IPG strip

SDS-PAGE

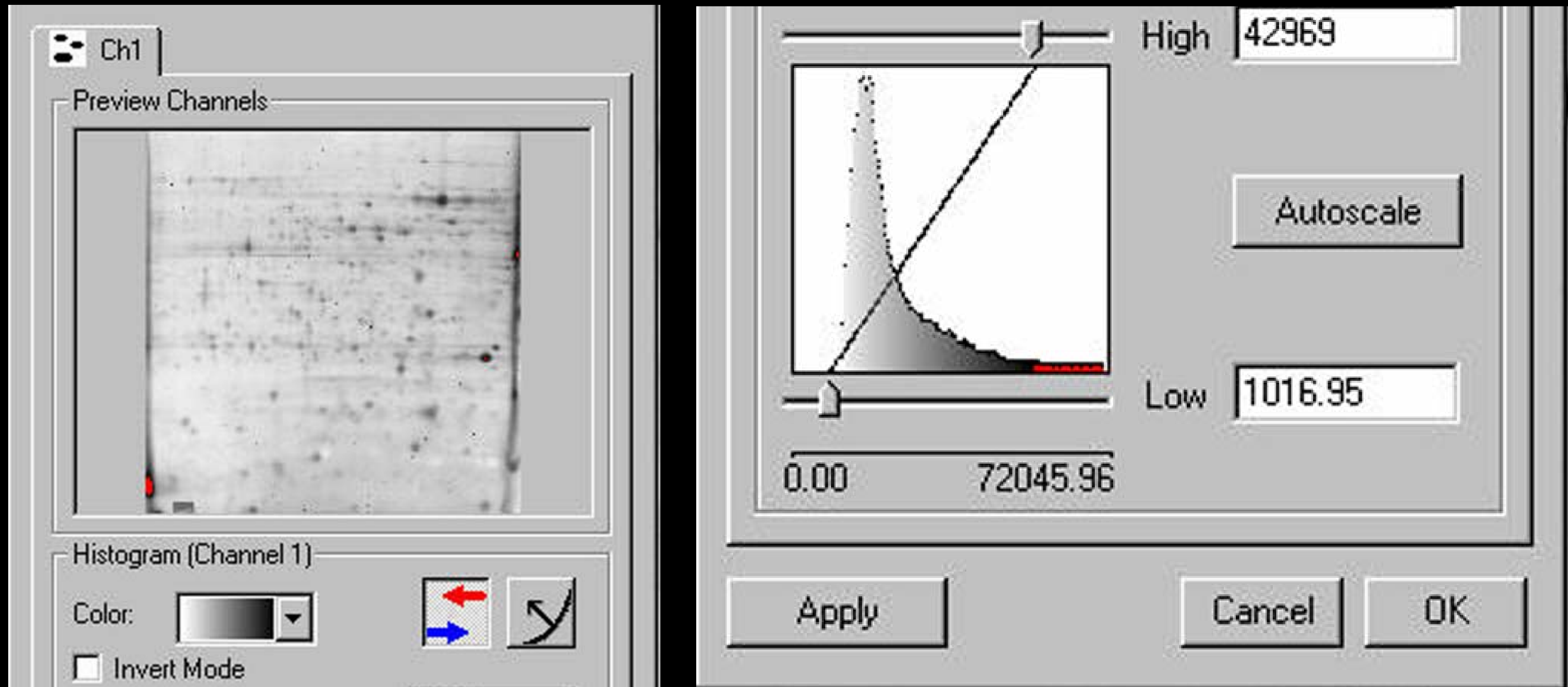
Protein visualization

Tools to reduce variance

Technical variance

- Internal standard:
  - DIGE
- Software algorithms:
  - Background subtraction
  - Normalization

Dynamic range of scanner



16 bit pixel resolution (2^{16} ~ 65,000 ~ 10^5)

Make sure you are using the entire range!

Variance in 2DE

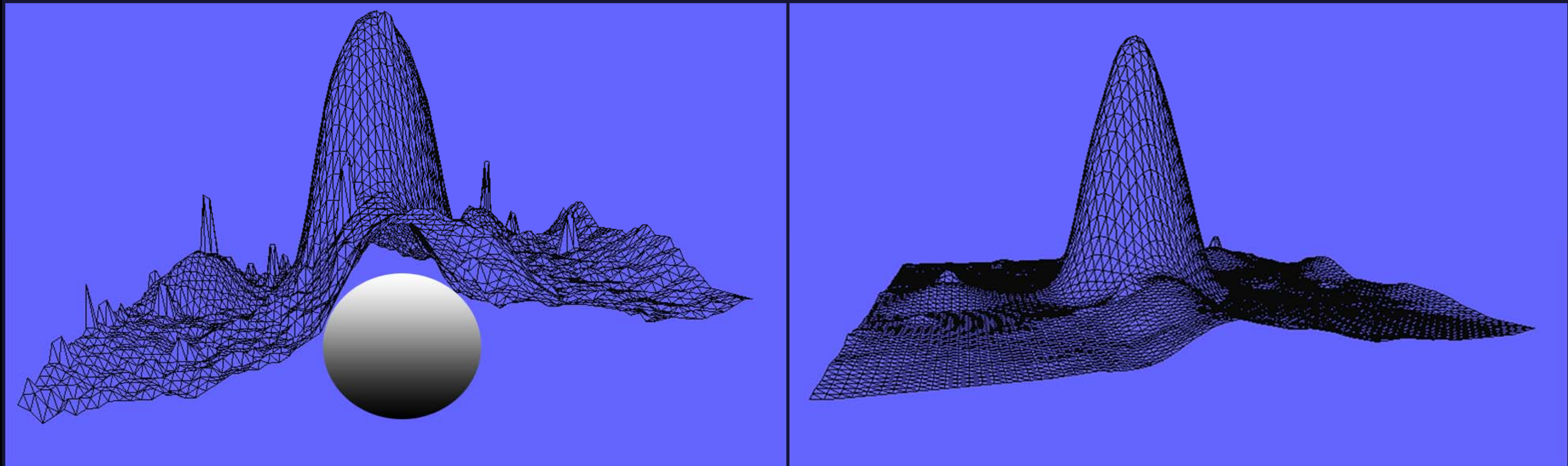
- BIOLOGICAL VARIANCE
- Experimental variance
  - Pre-fractionation, isolation & labelling of proteins
  - Protein staining
- Technical variance
  - Gel-to-gel variation in 2DE
  - Image acquisition (scanner)
- **Post-experimental variance**
  - **Software-induced variance**
  - **User dependant variance**

2DE analysis software

- Main purpose: match and quantify spots
- Normalization: reduce gel-to-gel variation
- Background subtraction:
  - Reduce background noise
  - Increase signal/noise ratio
  - Increase sensitivity

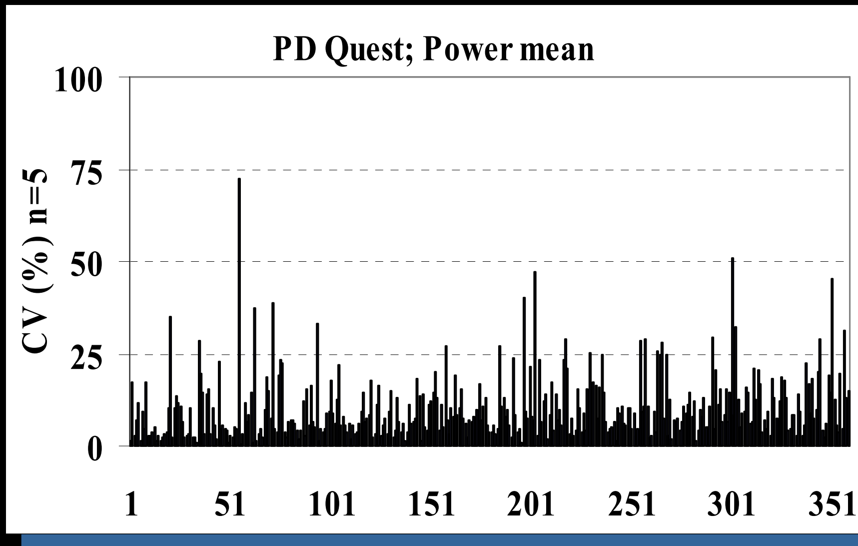
Global Background Subtraction

PDQuest: Floating/Rolling Ball

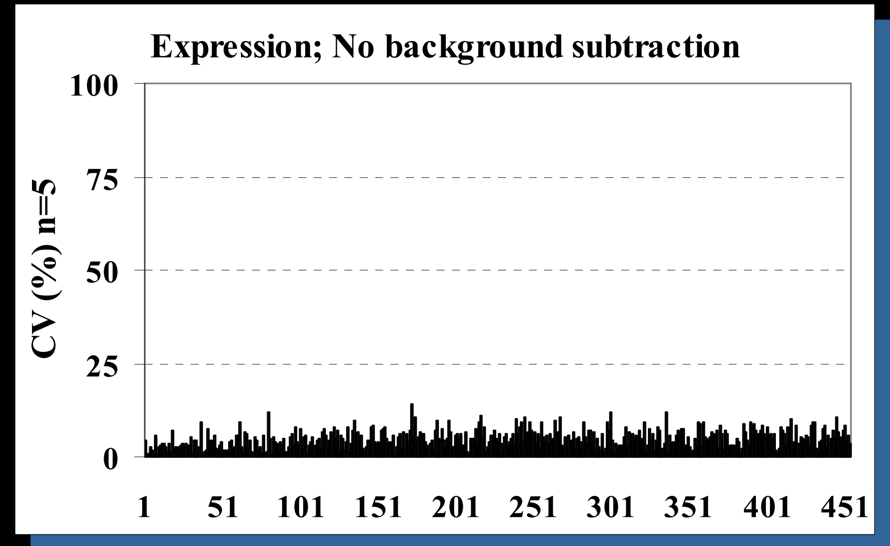


Software induced variance

PDQuest



PG200



Average CV=10%

Average CV=4.6%

Software variance up to 30% of technical variance

Applications of proteomics

BIOMARKER DISCOVERY

- Biomarker of disease & susceptibility

CLINICAL APPLICATIONS

- Pharmaceutical target identification
- Improved diagnostics

MECHANISTIC STUDIES

- Protein-protein interactions
- Protein adduction /Altered protein expression
- Hypothesis generation: avoid local ”maxima”
- Systems Biology

Proteomics in the future

- Improved sensitivity
  - Currently: scratching the surface
  - laser capture microdissection
- Protein microarrays
  - Antibody arrays (e.g. for cytokines)
  - Tissue microarrays (Peter Nilsson, Friday)
- In vivo subcellular localization assays
- Protein amplicifation method?
  - i.e. ”protein-PCR”

Proteomics in the **NEAR** future...

Focus on **INTERPRETING** data,
not on **ACQUIRING** data.

Pathway Analysis

- Integrate data from omics cascade
- Integrate heatmap with biological pathways

Preview of coming attractions… KEGG & KegArray

KegArray Array Image

File Edit View Tools Configure Help

ARACHIDONIC ACID METABOLISM

00590 2/2/06

Take home messages...

...keep your variance down
and your dynamic range up!

...keep your false positives down,
and your power up!