Proteomics for Biomarker Discovery

UWC BIOMARKERS MODULE

13TH AUGUST 2019

DR. CAROLINE BELTRAN

Biomarkers Recap

Measurable entity providing diagnostic, prognostic or treatment-oriented care

Ideal marker

- Easily attainable
- Adequate sensitivity
- Adequate specificity
- Clinical applicability

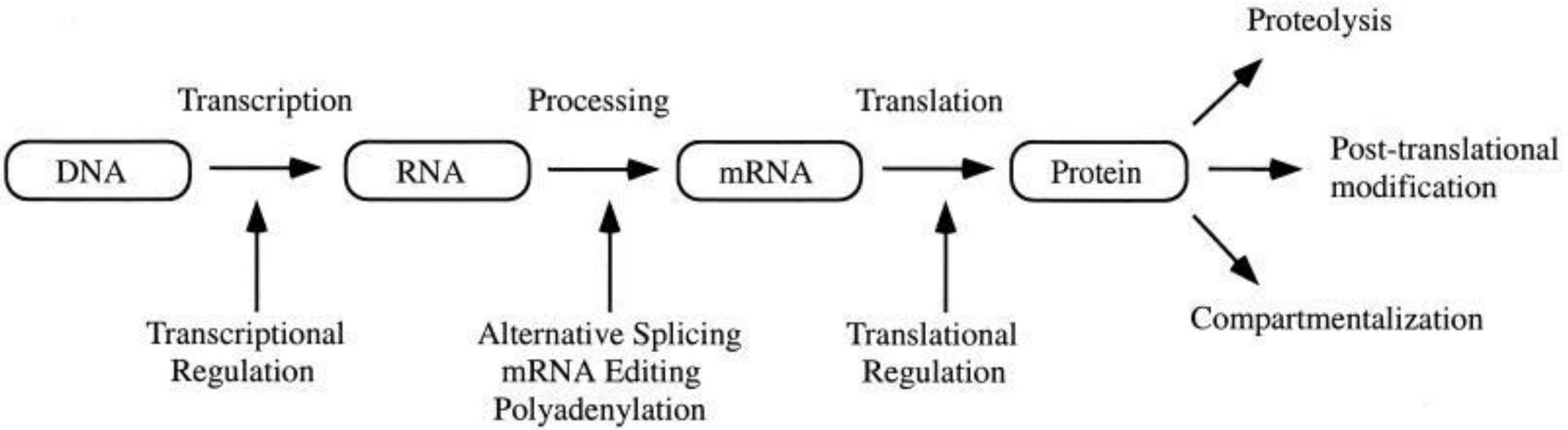
Many technologies for biomarker discovery

- Proteomics
- Transcriptomics (gene expression assays)
- Genomics

Why proteins?

Studying genes and mRNA does not provide all information

Proteins are responsible for the **phenotype**



PR Graves and TAJ Haystead. *Microbio. Molec. Bio. Rev.* (2002) 10.1128/MMBR.66.1.39-63.2002

Proteomics as a biomarker discovery tool

"Post-genomic" technologies: study of total protein alterations

Proteome responds dynamically (in localization and time), with different protein-protein interactions forming.

Each gene may yield many ***proteoforms***:

- Genetic variation and RNA splicing give different mRNA sequences.
- Proteolysis and clipped signals truncate mature protein sequence.
- Post-translational modification may dramatically alter activity.

Phenotype may be specific to a particular proteoform!

LM Smith, NL Kelleher. *Nature Methods* (2013) 10: 186-187.

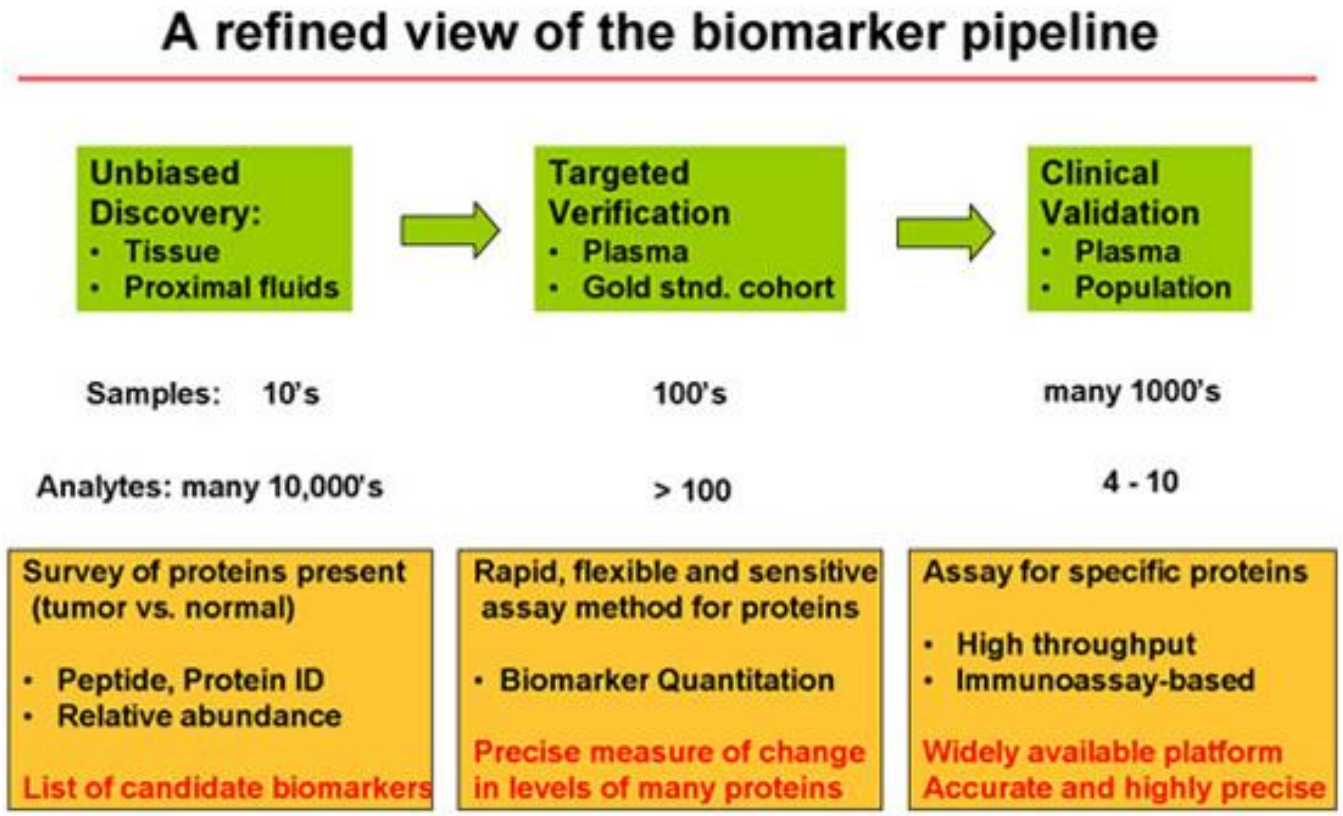
The proteomic Biomarker Pipeline

Discovery proteomics

- No prior knowledge
- Optimize # protein IDs
- More time per sample
- Sacrifice number of samples analysed
- High resolution instruments

Targeted proteomics

- Quantitation of discrete subsets of peptides
- Limits number of “features”
- Optimize for high sensitivity and low variability
- Thousands of samples
- Triple quad instruments



Broadinstitute.org, adapted from N. Rifai et al (2006) *Nature Biotech*. 24: 971-983.

Proteome methods

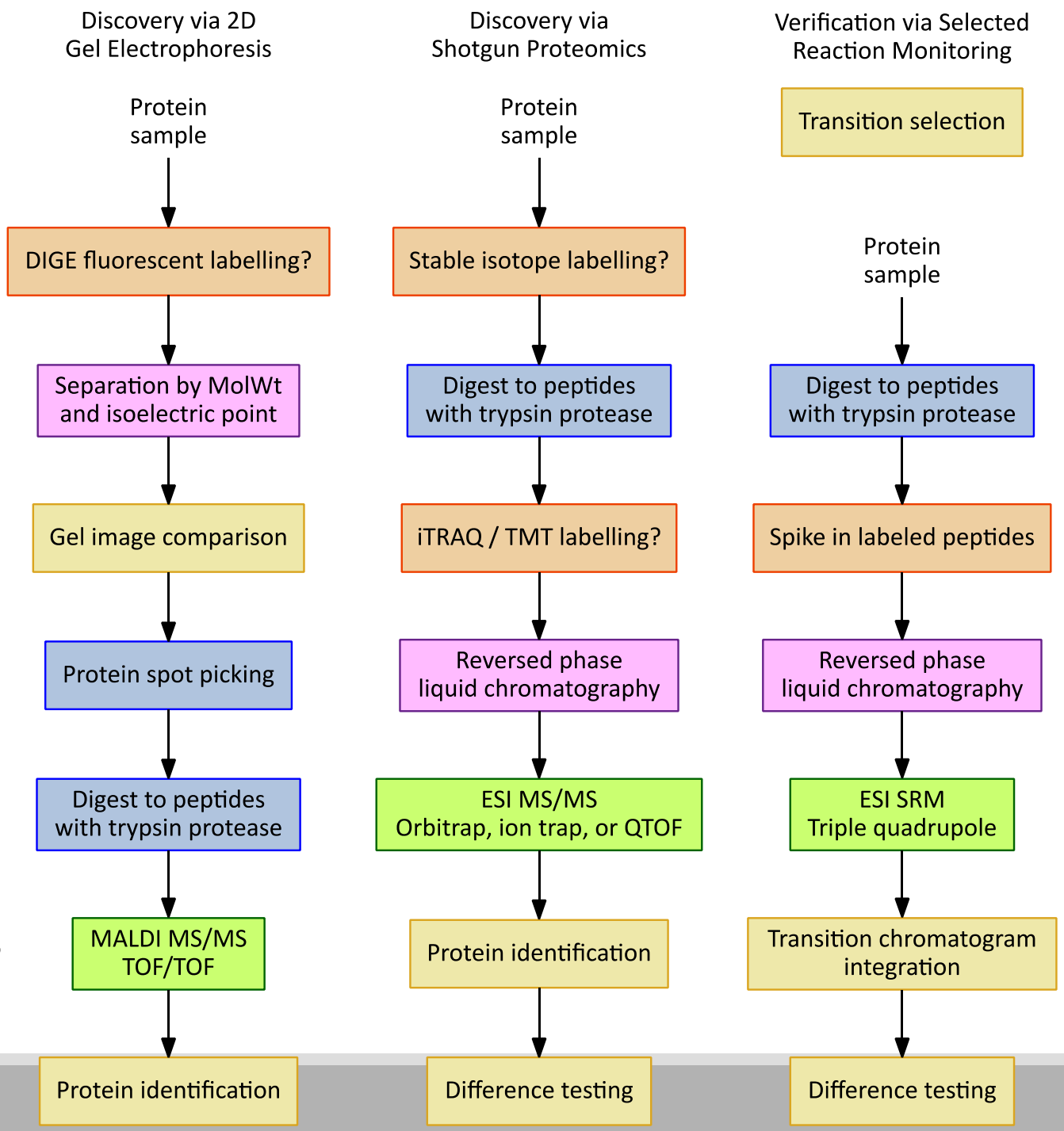
Comparative studies seek changing proteins, e.g. cases versus controls

Complex samples call for fractionation; biofluids may require depletion.

2D Gels identify only the protein spots that change intensity between cohorts.

Shotgun techniques identify proteins from MS/MS, then seek the proteins that have much more evidence in one cohort.

Targeted methods quantify only the peptides of proteins on a pre-specified list.



2D-GE Techniques

Separate proteins via IEF strip based on pH-dependent charge (isoelectric point).

Run the proteins into gel, with small proteins moving quickly and large proteins moving most slowly.

Gel imaging, alignment, and comparison finds differential spots.

Differential spots are excised, in-gel digested, and subjected to MS/MS for identification.

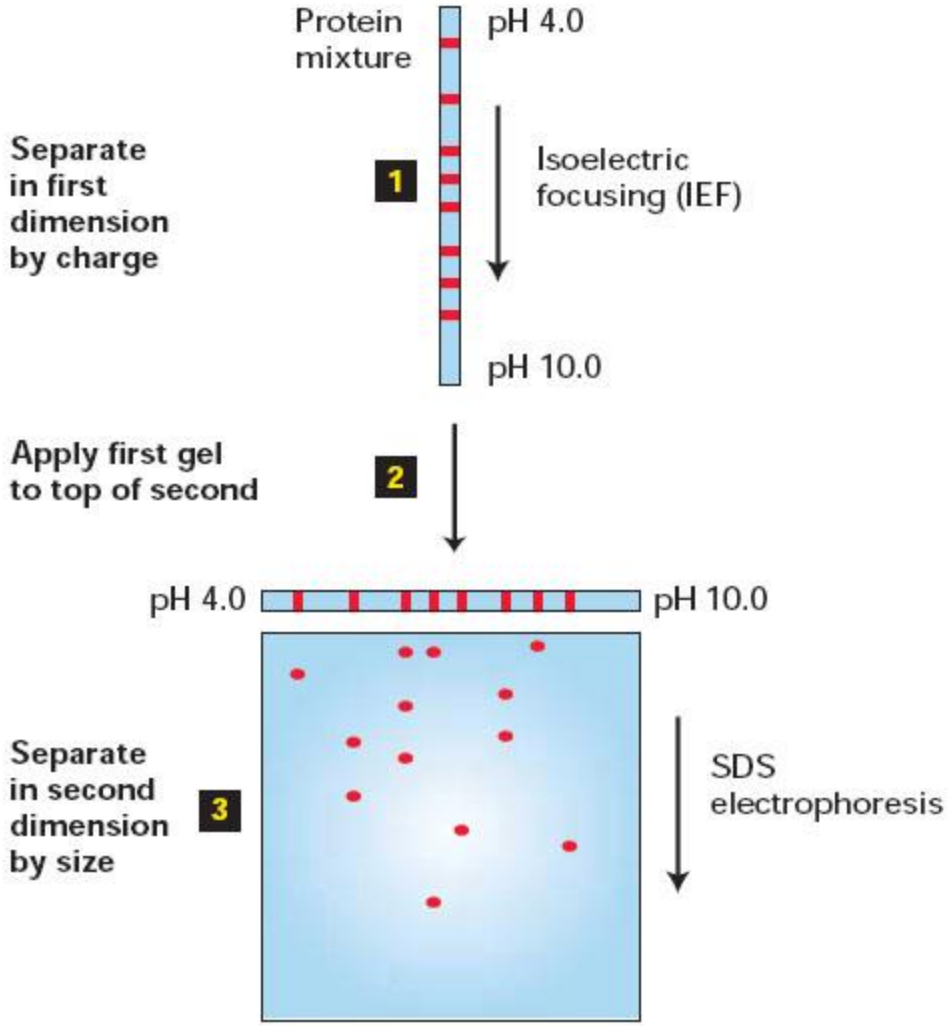
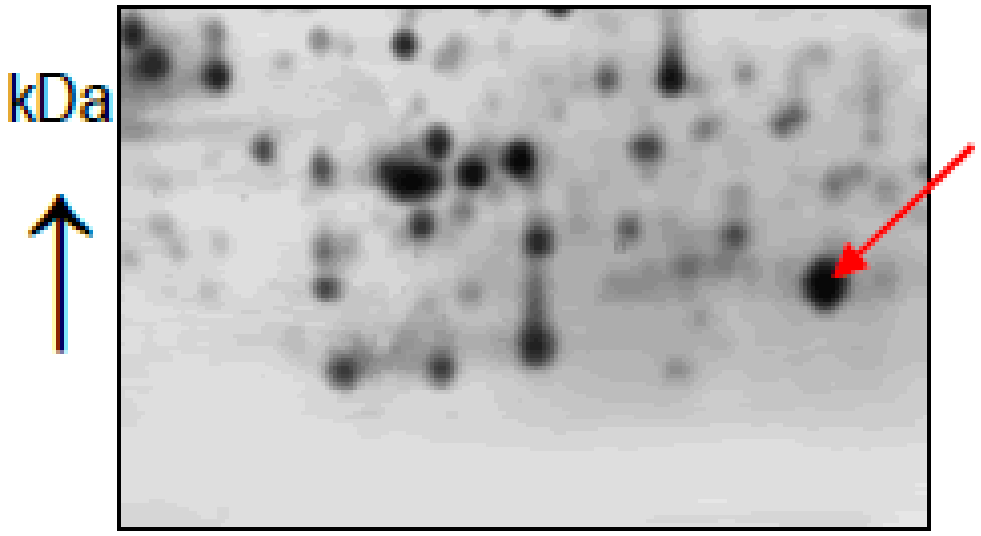
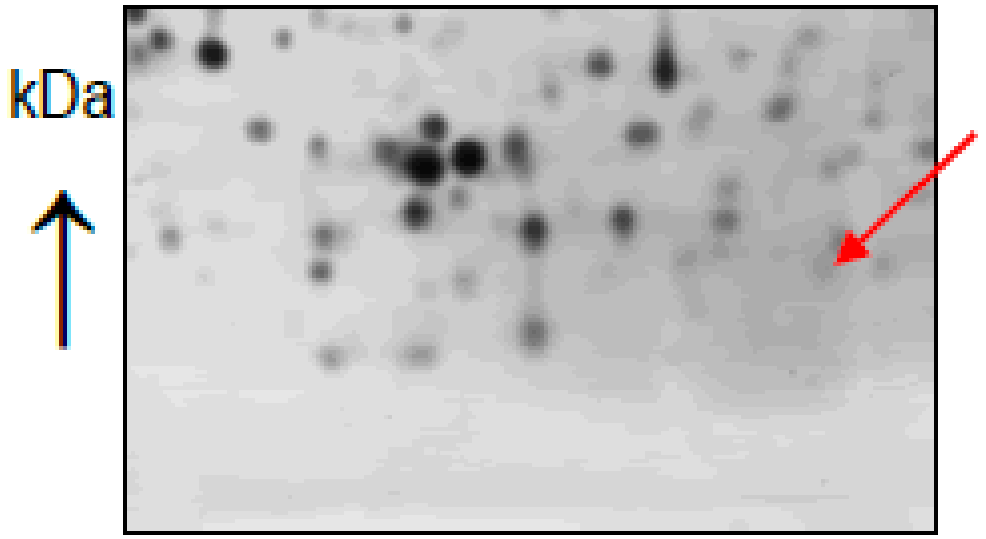


Image adapted from www.kendricklabs.com

2D-GE Techniques



Diseased



Normal

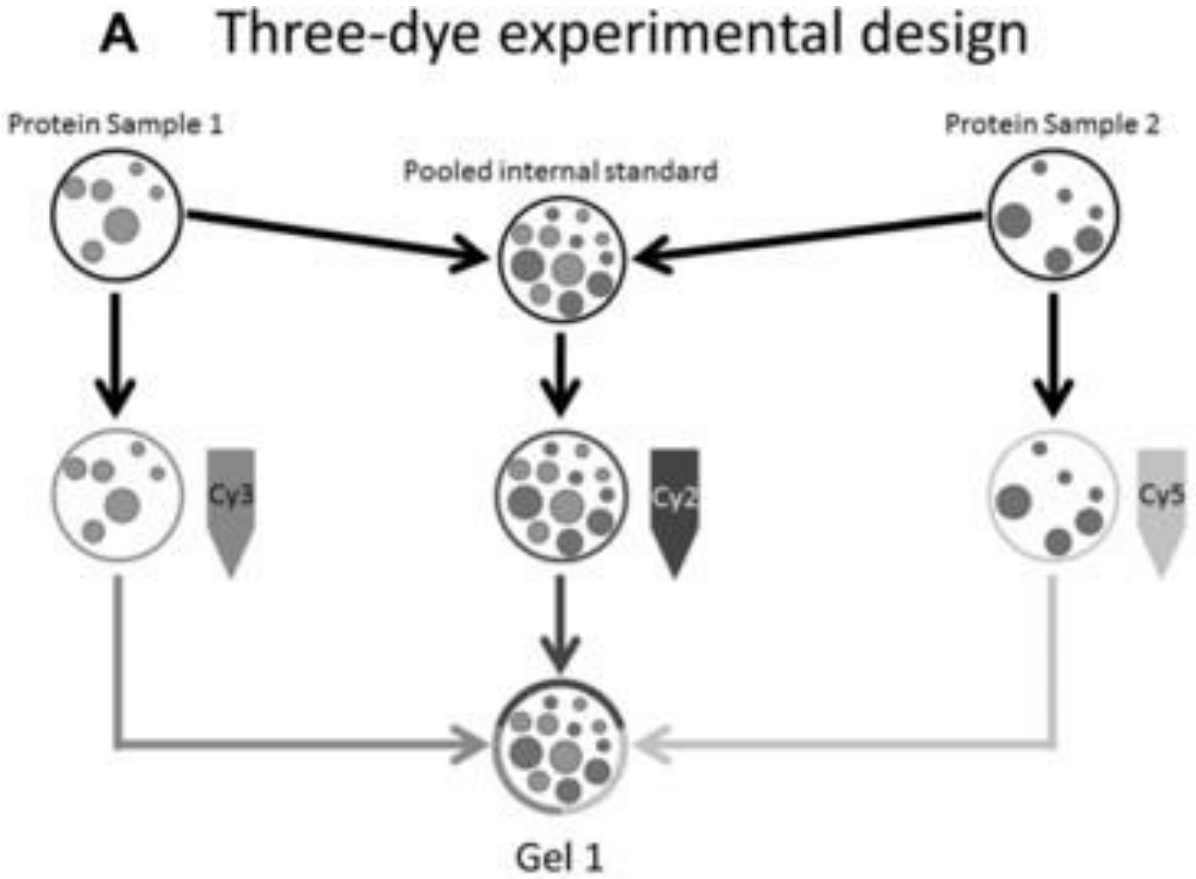
Image adapted from www.kendricklabs.com

2D-GE Techniques

Differential In-Gel Electrophoresis (DIGE) separates proteins of two samples in one gel.

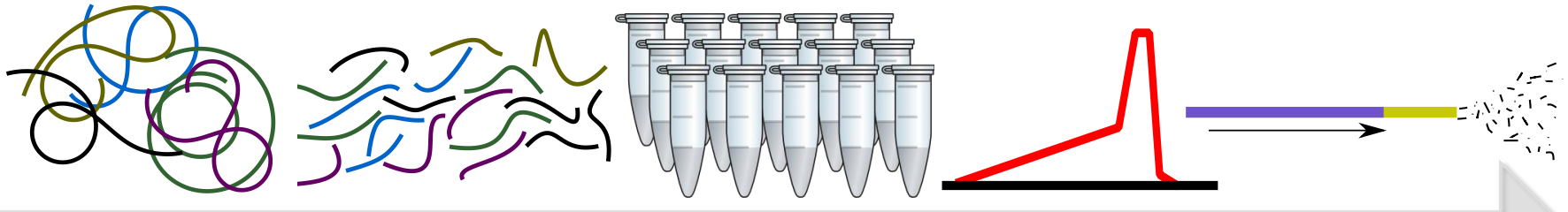
In a three-dye system, each of the two samples uses a different dye; a pooled internal standard uses a third dye.

Dyes have been chosen for minimum disruption of isoelectric point or mobility.

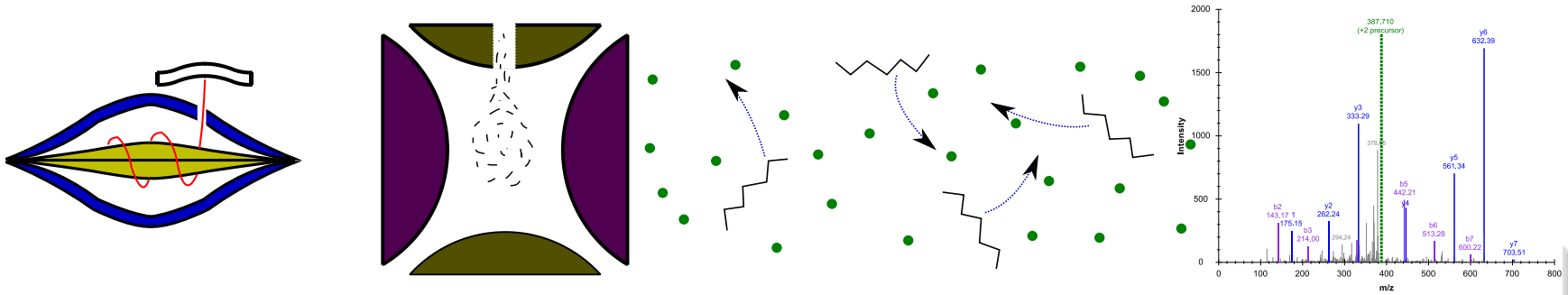


G. Arentz et al. *Proteomics Clin. Appl*. (2015) 9: 277-288.

Discovery Proteomics



Protein Mixture → Peptide Mixture → Peptide Fractionation → Liquid Chromatography → Electrospray Ionization



High-Resolution Mass Spectrometry → Isolate Ions of Peptide → Collide Ions to Dissociate → Collect Fragments in Tandem MS

Which proteins are present? How have they been modified?
Which proteins differ most between cohorts?

M. Bantscheff et al. *Anal. Bioanal. Chem.* (2007) 389: 1017

High-throughput Shotgun Proteomics

Mass spectrometry advances

- Sensitivity
- Resolution
- Mass accuracy

Test minute amounts of sample

High throughput: ID 1000s of proteins + quantitation

Labeled vs label-free

Relative vs absolute quantification

Labeled quantitative proteomics

Relative quantification: compare protein abundances

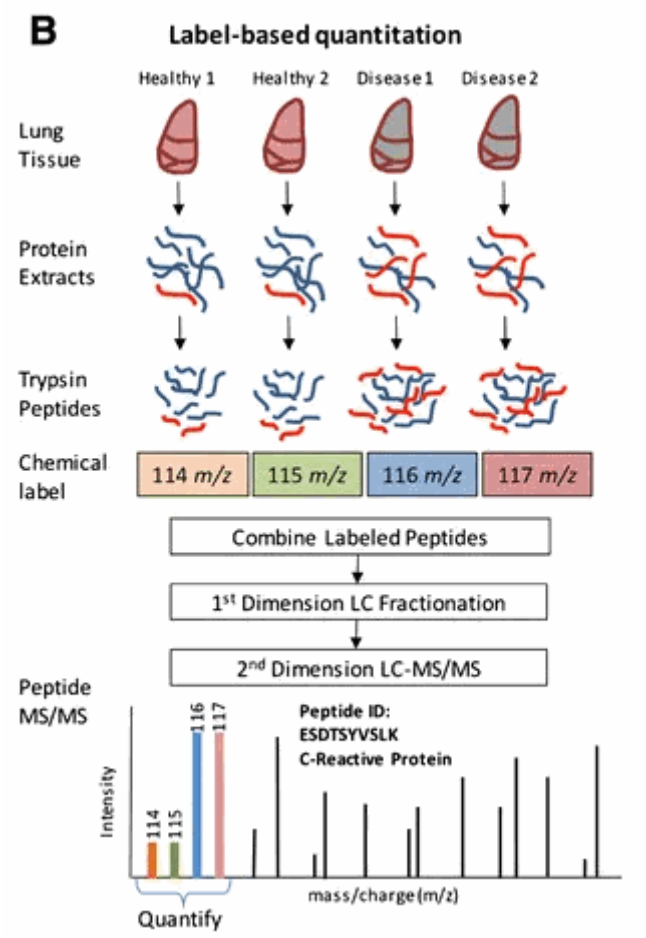
Many labelling options:

- Isobaric tags for relative and absolute quantification (iTRAQ / TMT)
- Stable isotope labeling with amino acids in cell culture (SILAC)

Pool samples:

- Multiplex
- Reduced variation between MS runs
- Reduced cost of MS runs

Costly reagents + sample processing



M. Bhargava et al. *Clin. Transl. Medicine* (2014) 3: 34

Label-free quantitative proteomics

"Straightforward" + inexpensive

Minimal sample manipulation

Comparison of "unlimited" samples

No multiplexing

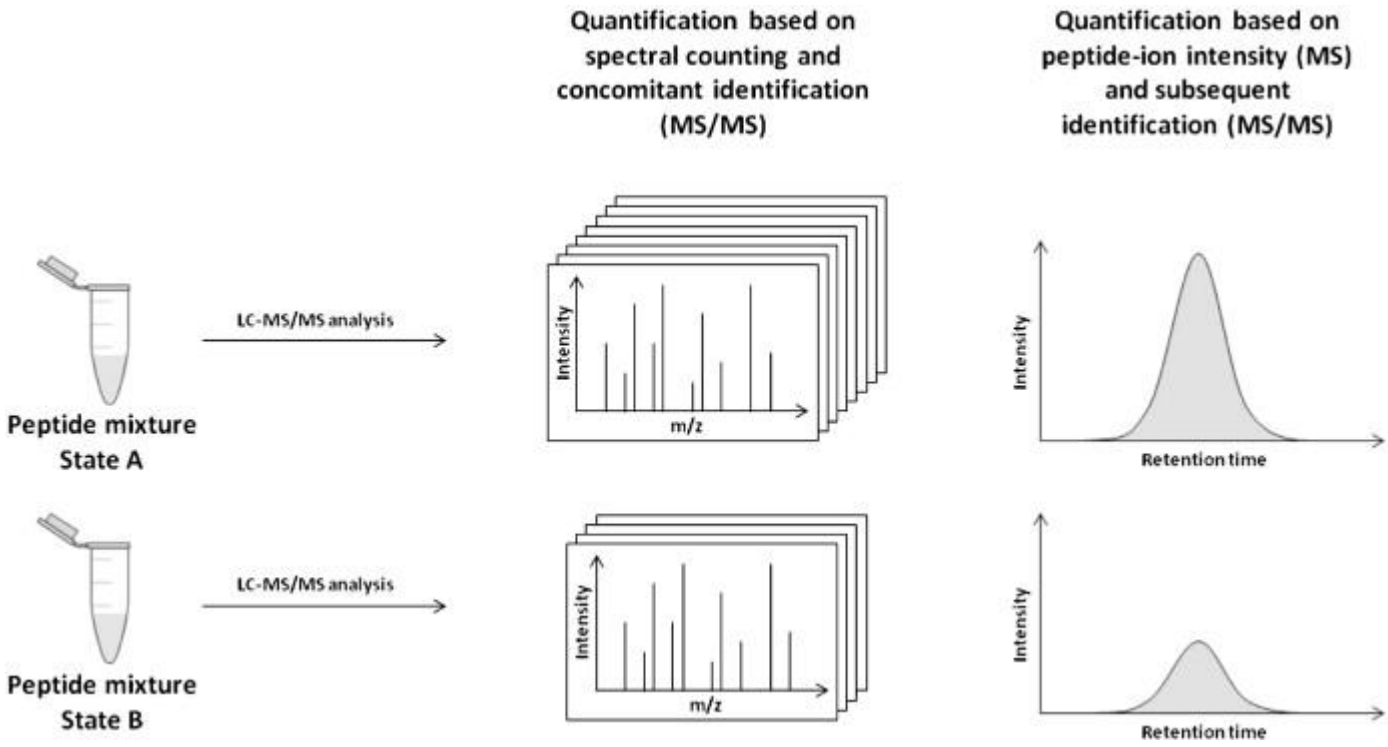
Spectral counting

Ion intensity (peak area)

"Instrument drift": increased measurement error

Normalize data

Less able to detect small fold-changes



DA Megger et al. *Biochimica et Biophysica Acta* (2013) 8: 1581-1590.

Which strategy should I choose?

No technology is ideal in all cases.

What is my sample type?

How complex are my samples?

How many samples do I want to compare?

What are the dynamic range limits for this technology?

What instrument is available?

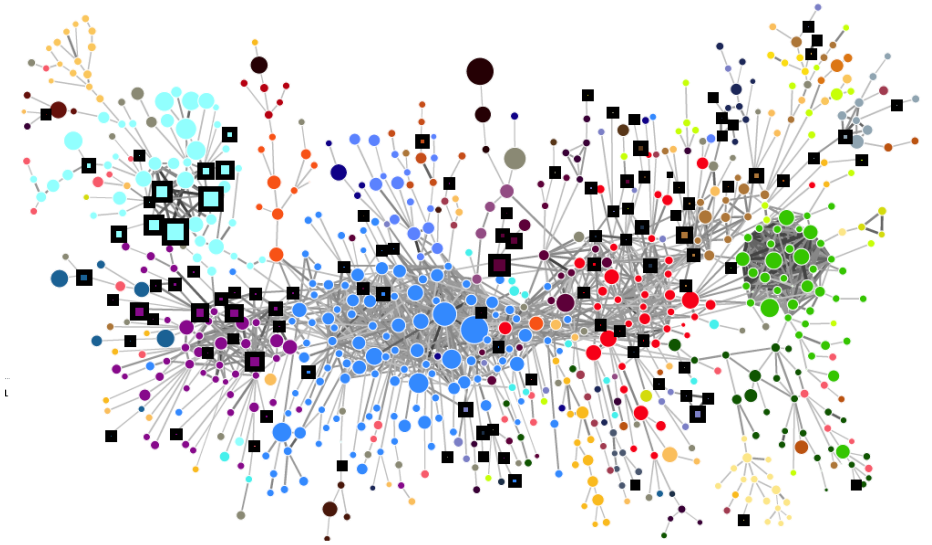
How much money is available?

Why can’t I see anything?

Large number of proteins + high dynamic range

(Interesting) Biomarkers are usually low abundance

- Highly abundant proteins =
  - 98% proteome in serum (albumin and IgG)
  - 40% proteome in heart tissue (myosin and actin)



Need to reduce complexity

- Depletion
- Enrichment
- Fractionation

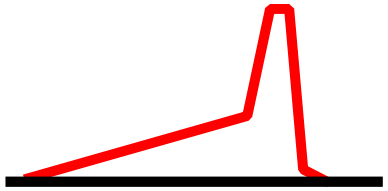
Consistent and careful sample preparation is important!

Consider “off-target” effects

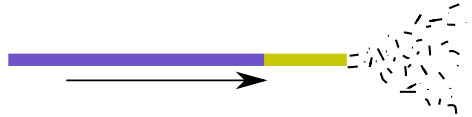
Image source: www.economplex.org/

Targeted Proteomics

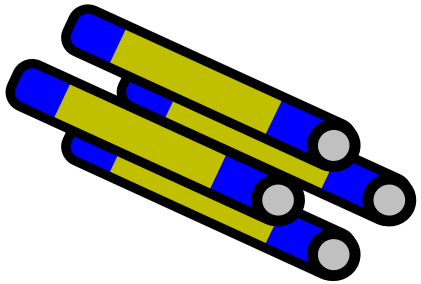
Peptide Mixture



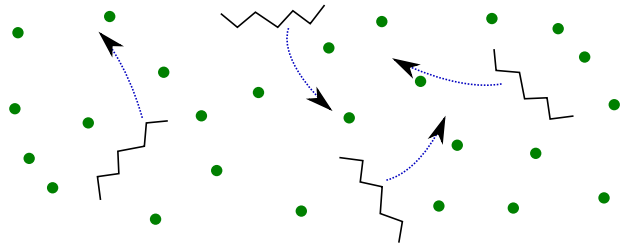
Liquid Chromatography



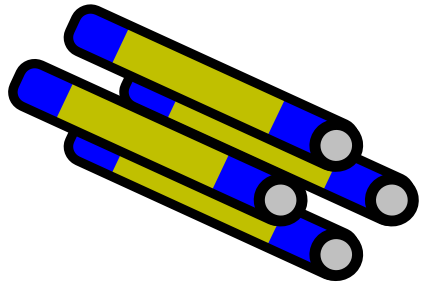
Electrospray Ionization



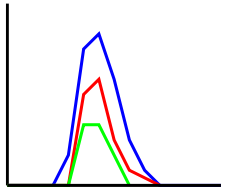
Screen out all but Target Mass



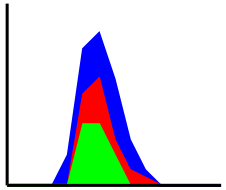
Collide Ions to Dissociate



Screen out all but Fragment Mass



Find Peaks from Related Traces



Integrate Peak Areas

Compare Areas to Reference Areas

Validation: targeted assays

Selected Reaction Monitoring (SRM) collects chromatograms.

Each chromatogram corresponds to a ***transition***: a unique combination of peptide m/z and fragment ion m/z.

Most triple-quads can measure 50 proteins for each LC-MS/MS:

- Each protein is measured in three different peptides.
- Each peptide is measured at three different fragment ions.
- 50 proteins = 150 peptides = 450 transitions

“Scheduling” allows the instrument to measure the chromatogram only near the best retention time for each peptide.

Take home message

1. Proteomic biomarkers are close to the “scene of the action.”
2. Prefractionation and depletion is often required to reduce sample complexity
3. Gel-based methods have gradually given way to gel-free techniques.
4. While differentiation is possible in “shotgun” experiments by both label-free and labeled approaches, targeted quantitation yields less variability.
5. A biomarker search begins with “shotgun” experiments (tens of people) and then transitions to targeted SRM experiments (hundreds of people).