UAB University of Alabama at Birmingham
School of Medicine

"Designing the Proteomics Experiment; From Protein Mixture to Systems Biology and Validation"

(An Overview)

James Mobley, Ph.D.
Assistant Professor, Surgery & Pharmacology
Director, UAB Urologic and Clinical Proteomics Facility
Associate Director, UAB Mass Spectrometry Shared Facility

Topics…….

- A quick overview of Genomics Vs. Proteomics
- A General Look at the Field of Proteomics; MS vs. Other approaches.
- Overview of Instrumentation, generating MS for PMF vs. Tandem MS and why.
- Separations when needed, Top-Down overview Vs. Bottom up and MuDPIT.
- MS Data Analysis using Matching Algorithms (a focus!).
- Statistical Assessment of the data…..n=?
- Systems Biology….Taking the "So What" Factor out of large studies.
- Validation??

Genomics Vs. Proteomics?

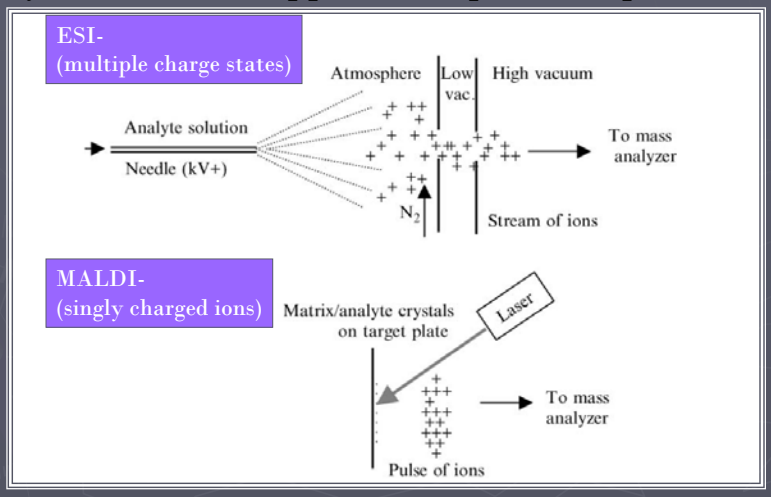
- Genomic studies will always remain important!
- However, we're finding that gene transcription does not necessarily equate to gene translation.
- Therefore, many genes are transcribed that do not lead to translation, or to functional proteins!
- Finally, post-translational modification (PTM) events will dictate the 1) cellular location, 2) biological process, or 3) molecular function.

What Tools Encompass Proteomics?

- Relative or Absolute Quantification of Known Proteins…….
  - Mass Spectrometry (Heavy Labeled Stds, i.e. AQUA)
  - Immuno-Directed (examples)
    - 1) Protein Arrays, 2) Western Blot, 3) ELISA, 4) Bioplex/ Luminex
- Characterization of Unknown Proteins/ Mapping Post Translational Modifications (PTM's)……..
  - Sequencing of Unknowns, Mass Spectrometry alone!

Soft Ionization Techniques

(Question;
why are these two approaches special for proteomics?)



Baldwin M.A., Mass spectrometers for the analysis of biomolecules, Methods in Enzymology, Vol 402

Molecular Characterization

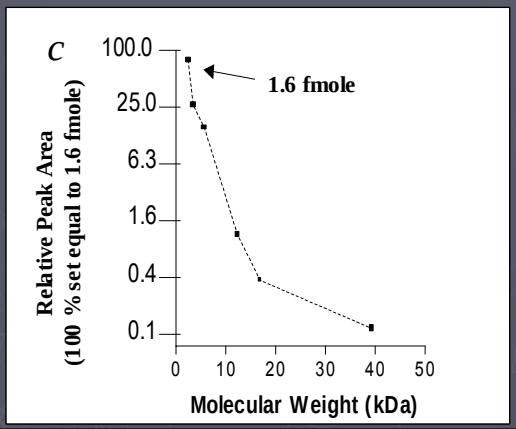
- So….There are just three simple points to remember in the business of molecular characterization.
  - 1) Generation of a parent & daughter ion mass spectra.
    - [MS^1 only]; for digested proteins the peptide mass fingerprinting [PMF] (protein must be nearly pure)
    - [MS^2]; sequence data (high purity is not necessarily required)
  - 2) Analysis.
    - Matching algorithms based on *in-silico* digestion and predicted size of peptides from proteins for example (MS^1) & fragmentation (MS^2).
    - *Denovo* Sequencing (MS^2), based on the known size of fragments, for example daughter ions from amino acids.
  - 3) Validation.
    - Immuno-affinity techniques (Western analysis, Immuno-histochemistry, ELISA, Luminex, etc).
    - Mass Spectrometry (standard heavy isotope tags)

Proteins to Peptides….

- Even today, we are highly limited by decreased detection, resolving power, and poor fragmentation of "whole" proteins!
- Therefore, we "digest" proteins to peptides prior to MS analysis.
- Many chemical and enzymatic techniques have been published; however, trypsin remains the most commonly utilized enzyme for use in proteomics!
- This enzyme cleaves at arginine and lysine, yielding peptides that are easily detected and fragmented in the most common mass analyzers today.
- Keeping in mind that utilizing multiple digestion procedures carried out on the same sample can be very complementing!

http://donatello.ucsf.edu/
A lot of information here……
Take a look at Protein Prospector…….

Sensitivity is Inversely Proportional to Mass
(MALDI-ToF Example)



Student Lab Example!

So….digestion is often necessary, but does increase complexity!

Common Enzymes:

Trypsin*	K-X, R-X
Chymotrypsin*	X-L, -F, -Y, -W
Lys-C*	K-X
Arg-C*	R-X
Asp-N	X-D
Glu-C*	E-X

Common Chemicals:

Cyanogen Bromide	X-M
HCL	X-X

* a tailing proline inhibits digestion

Concept of PMF [MS^1]

Question; What's the downfall of this approach, upswing?

- MS^1 approaches – spectra containing peptide parent molecules only!
- This type of unambiguous protein ID is referred to as "peptide mass fingerprinting" (PMF) introduced ~1990.
- In this case, no sequence information is generated, but it is very sensitive when very little sample is available!
- The downfall is that the sample must be very pure! Highly complementing for 2D PAGE work.
- However, high mass accuracy is a must as well!
- Overall, these days……unless absolutely necessary…… PMF should not be used!
- There are simply too many matches possible with this technique even with access to high resolution instruments.

Protein Fragmentation [MS^2]

- Both ESI and MALDI based tandem instruments are common in most core settings, each with a combination of mass analyzers.
- Each source and combination of mass analyzers have their selective advantages worthy of a second talk!
- Fragmentation is generally similar, primarily with the generation of either…….
  - b and y ions; collision induced decay (CID) & infrared multiphoton dissociation (IRMPD)
  - z and c ions; electron transfer dissociation (ETD) & electron capture dissociation (ECD)

Many Instruments for Many Applications!

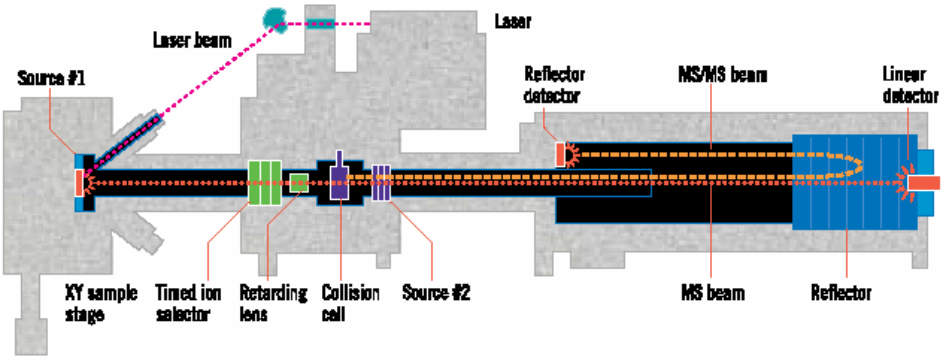
Quad, ToF, Ion Trap, and…or FT?

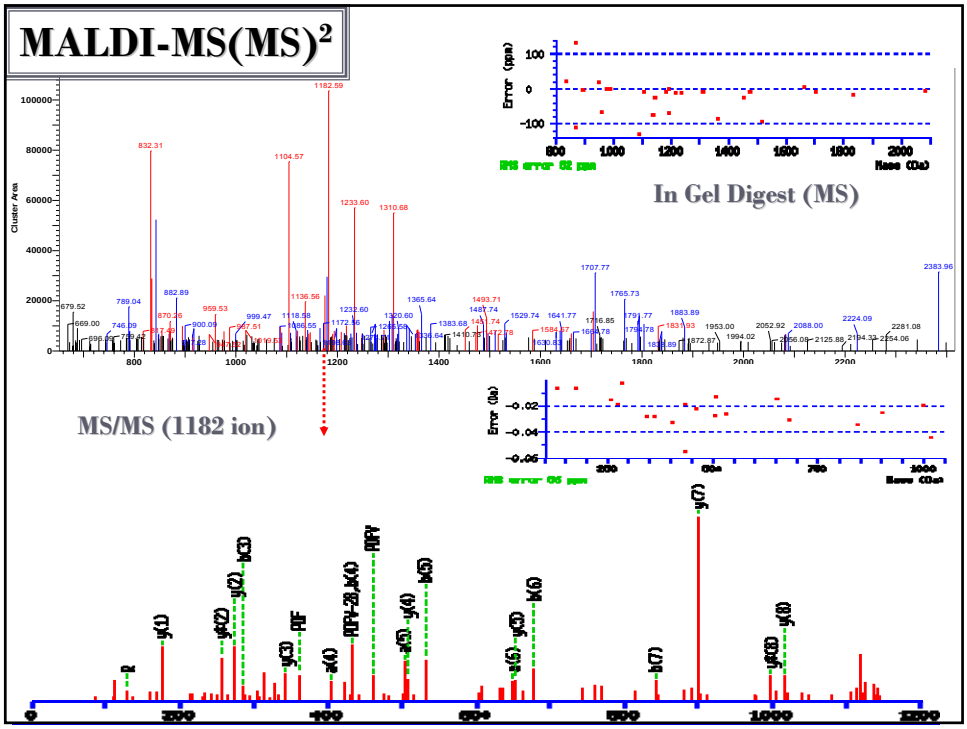
	IT-LIT	Q-Q-ToF	ToF-ToF	FT-ICR	Q-Q-Q	QQ-LIT
Mass accuracy	Low	Good	Good	Excellent	Medium	Medium
Resolving power	Low	Good	High	Very high	Low	Low
Sensitivity (LOD)	Good		High	Medium	High	High
Dynamic range	Low	Medium	Medium	Medium	High	High
ESI	✓	✓		✓	✓	✓
MALDI	(✓)	(✓)	✓			
MS/MS capabilities	✓	✓	✓	✓	✓	✓
Additional capabilities	Seq. MS/MS			Precursor, Neutral loss, MRM		
Identification	++	++	++	+++	+	+
Quantification	+	+++	++	++	+++	+++
Throughput	+++	++	+++	++	++	++
Detection of modifications	+	+	+	+		+++

Domon & Aebersold, Science, V312, 14 April, 2006

Principals of the MALDI-ToF/ToF 4700

(for peptide sequencing)





ESI-Linear Ion Trap

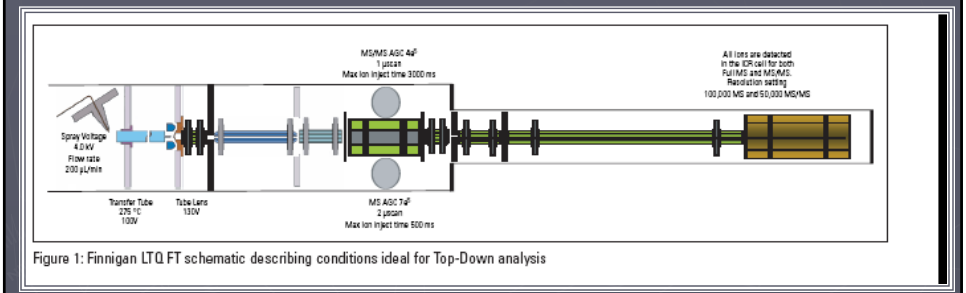
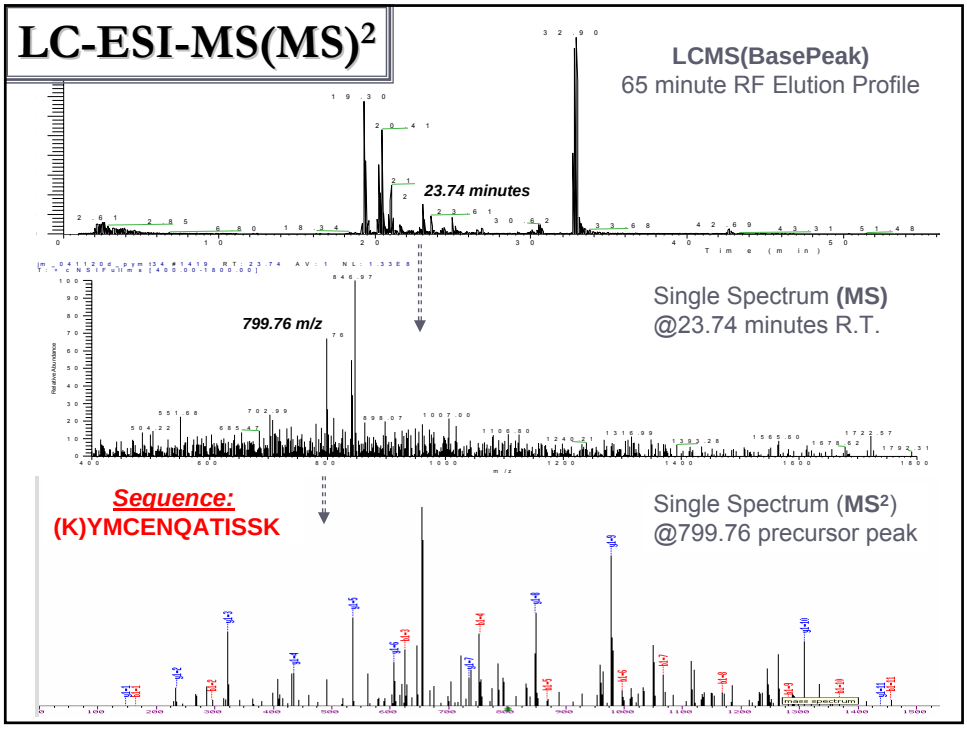


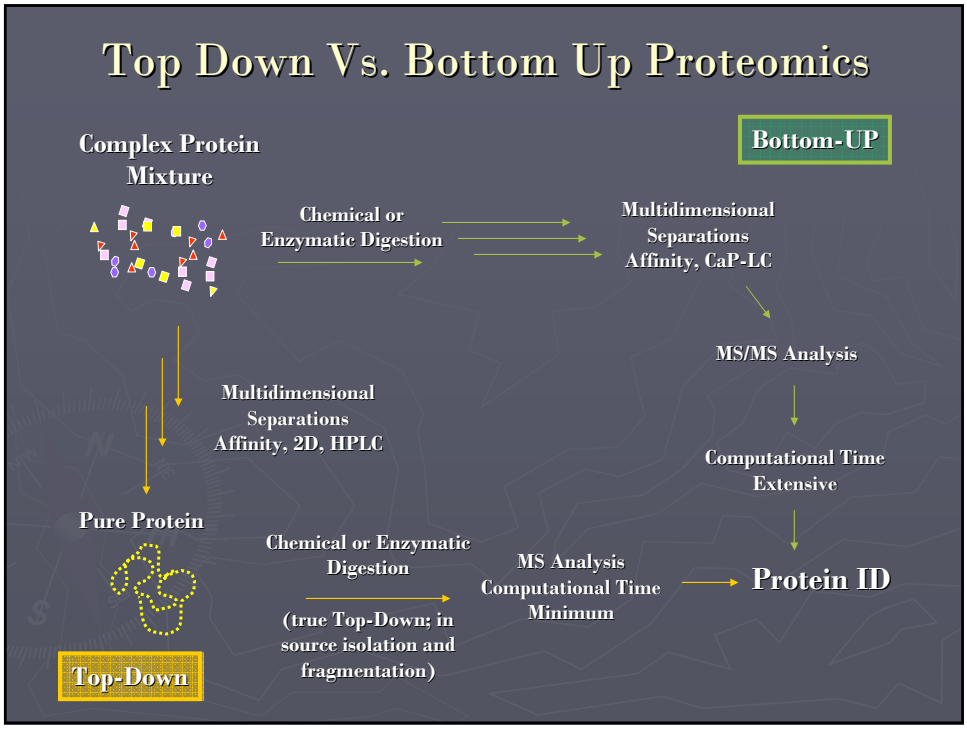
Figure 1: Finnigan LTQ FT schematic describing conditions ideal for Top-Down analysis

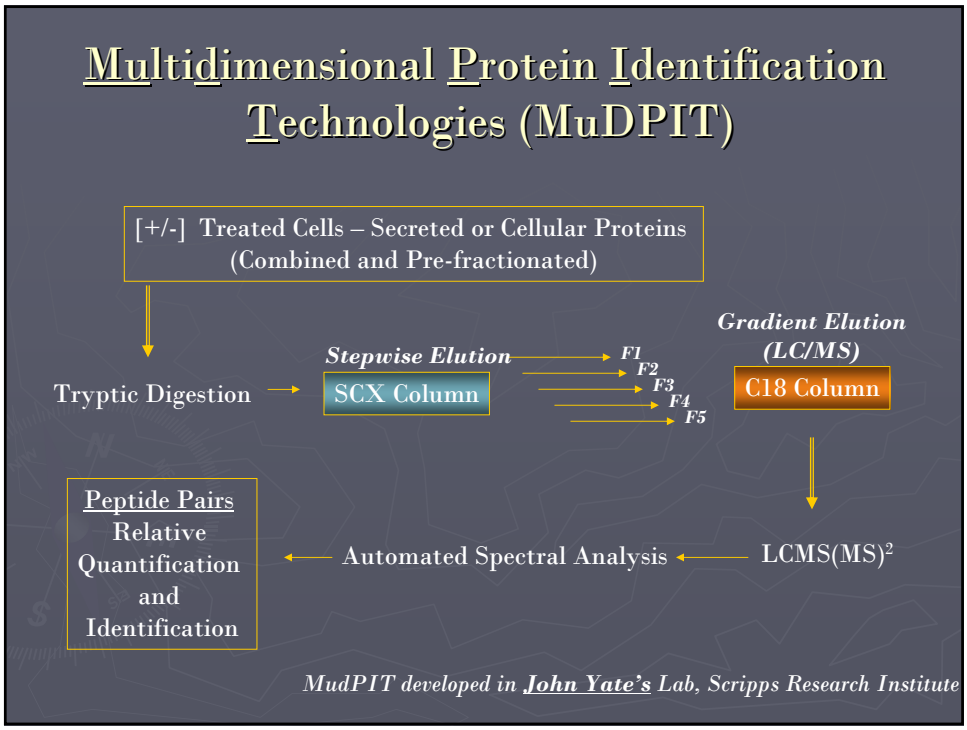
FT-ICR or Orbitrap



Question.

- In-General, Why is one Instrument better than the other, for what experiments??
- What would you do for a……..?
- In-Gel 1D experiment (complex mix)
- In-Gel 1D experiment (after an IP experiment)
- In-Gel 2D experiment (complex mix)
- In-Solution Digestion of any kind other than a pure protein.
- In-Solution, whole Proteins Vs. Large Native Peptides
- Quantitative Proteomics Experiments (many to choose from)?
- What about Price for each, is that important?


Top Down Vs. Bottom Up Proteomics
Complex Protein Mixture
Bottom-UP
Chemical or Enzymatic Digestion
Multidimensional Separations Affinity, CaP-LC
MS/MS Analysis
Computational Time Extensive
Multidimensional Separations Affinity, 2D, HPLC
Pure Protein
Chemical or Enzymatic Digestion
(true Top-Down; in source isolation and fragmentation)
MS Analysis Computational Time Minimum
Protein ID
Top-Down

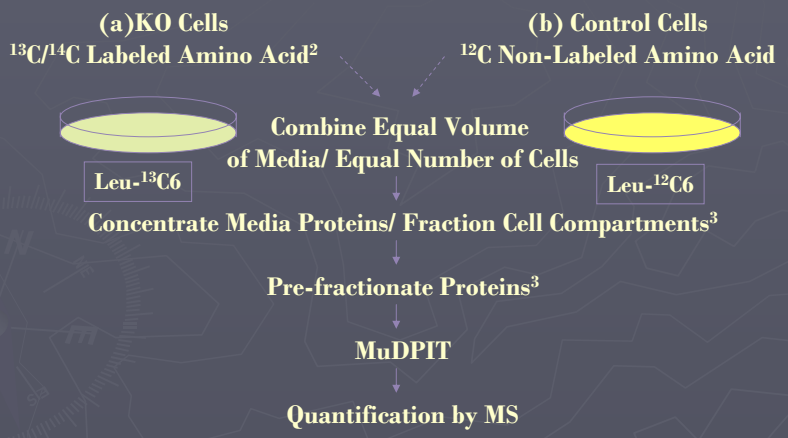

Multidimensional Protein Identification Technologies (MuDPIT)
[+/-] Treated Cells – Secreted or Cellular Proteins (Combined and Pre-fractionated)
Gradient Elution (LC/MS)
Stepwise Elution
F1
F2
F3
F4
F5
Tryptic Digestion
SCX Column
C18 Column
Peptide Pairs Relative Quantification and Identification
Automated Spectral Analysis
LCMS(MS)²
MudPIT developed in John Yate's Lab, Scripps Research Institute

Quantitative Proteomics

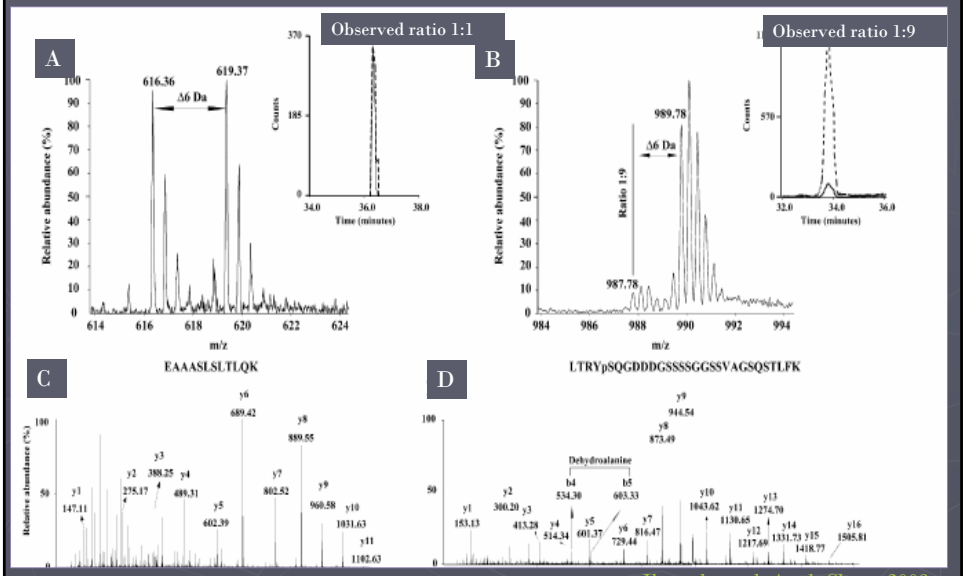
Using a Bottom UP or Shotgun Approach!

- Stable Isotope Tags:
  - ICAT (isotope coded affinity tag)
  - SILAC (stable isotope labeled AA in cell culture)
  - iTRAQ (Isotope tags for relative and absolute and quantification)
  - ^{18}O Digests (labels trypsin digested peptides at lysine and arginine)
  - Many Other Chemical Tags

Stable Isotope Labeling with Amino Acids (SILAC)[1]

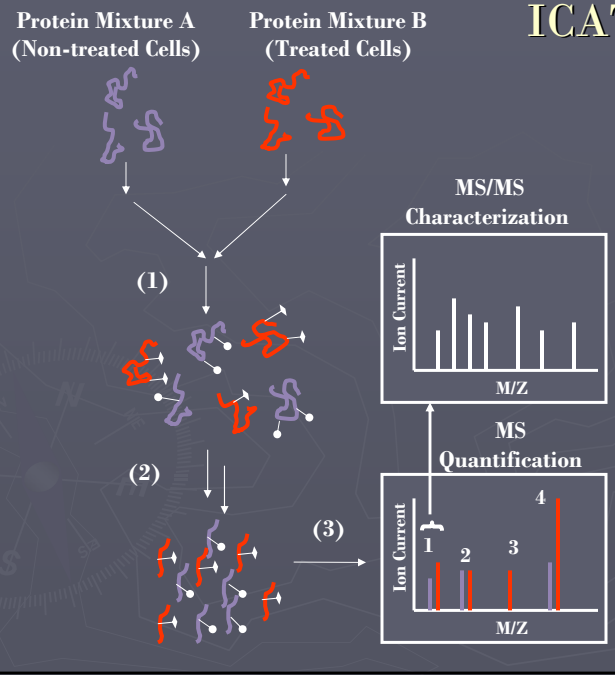


Protein Quantitation and Identification Using LC-MS/MS



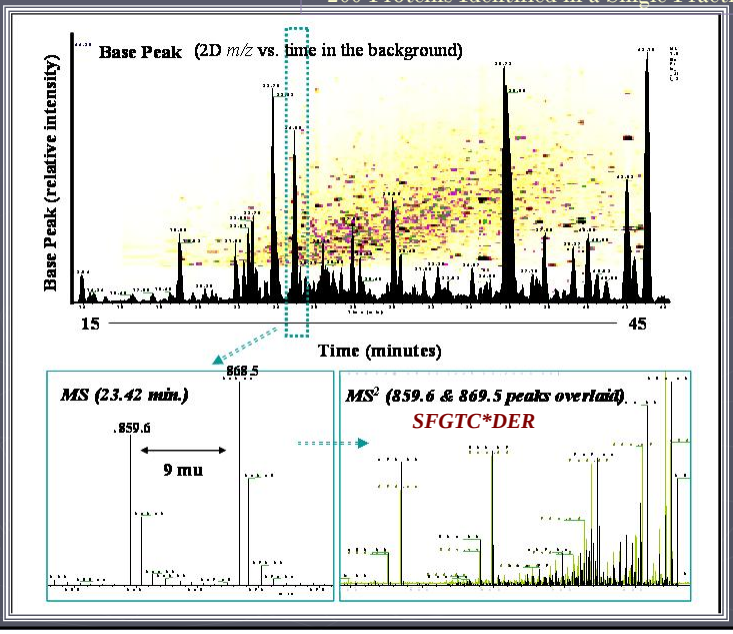
Ibarrola et al, Anal. Chem. 2003

ICAT Approach



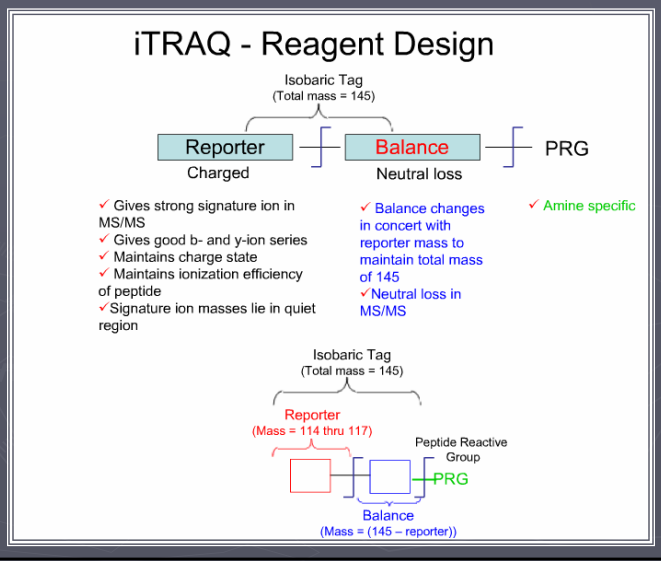
Example of Spectra Taken From an ICAT Run

>200 Proteins Identified in a Single Fraction

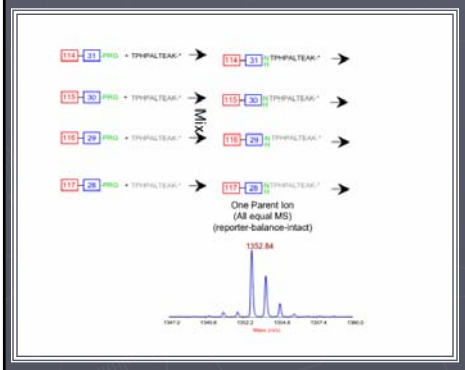


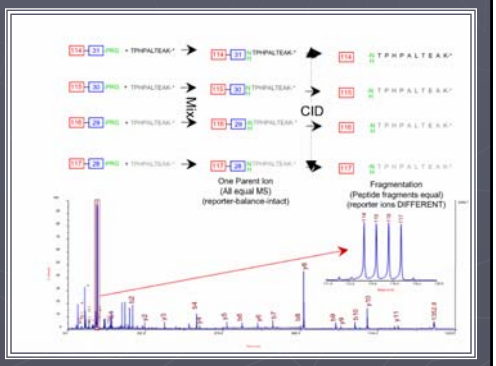
iTRAQ

(can label up to 8 samples in a single run!)



Reporter Tags for iTRAQ





Any Other Examples?

- Question……..What are we missing?
- Non-Tagged Proteomics (Top-Down) or Peptidomics (shotgun, Bottom-up)
- This can be carried out in a Mass Spectrometer or via 2D PAGE for example.

Data Analysis

- A standard 1D LC-ESI run may have as many as 4,000-6,000 MS files!!
- A MuDPIT run may contain 25,000-60,000 files!!
- While LC-MALDI runs generate far fewer data files, they still contain too-much data to analyze by hand!
- Therefore, automated data analysis is required!!

Data Analysis

- Common Matching Algorithms;
  - *Sequest, MASCOT, XTandem*
- Automated Denovo Sequence Tools;
  - *Peaks, Rapid Denovo, DenovoX, Mascot Distiller, PepNovo, others…..*
- Statistical Software
  - *Scaffold, Protein Profit, Finnigan, others…*
- Standardizing the Field!
  - *Trans Proteomic Pipeline (Sashimi Project; mzXML based universal software package)*

Matching Algorithms

- All matching algorithms (i.e. SEQUEST, MASCOT, X!TANDEM) are based on generating a score based on "closeness of fit" between the peptide fragments measured in the mass spectrometer and the *in-silico* digestion/ fragmentation of known genes or proteins in a database.
  - The two most commonly used databases include: NCBI-NR and Uniprot

Getting at the Good Stuff; Overview of Scaffold and Such…..

- Protein Profit…..or Scaffold?
- Great Overview at……..
- http://www.proteomesoftware.com/
- http://www.proteomesoftware.com/Proteome_software_pro_interpreting.html

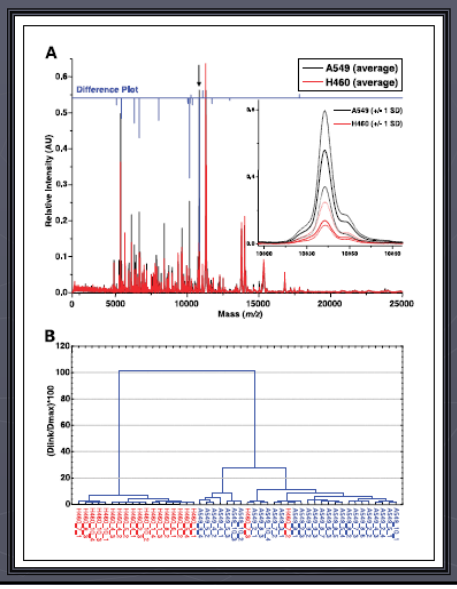
Overview of Stats Approaches (n=?)

- MS Based Analysis can Include;
  - Quantitative Tags
  - AQUA (knowns)
  - Non-Tagged (MS or 2D PAGE)
- Either way, Filter MS Data with High Confidence Cut off's.
- Combine the common hits from each arm of the experiment (i.e. how common is of interest, what do we do with the zero's?).
- Carry out basic statistics for each hit, i.e. non-parametric tests to filter out less significant hits.

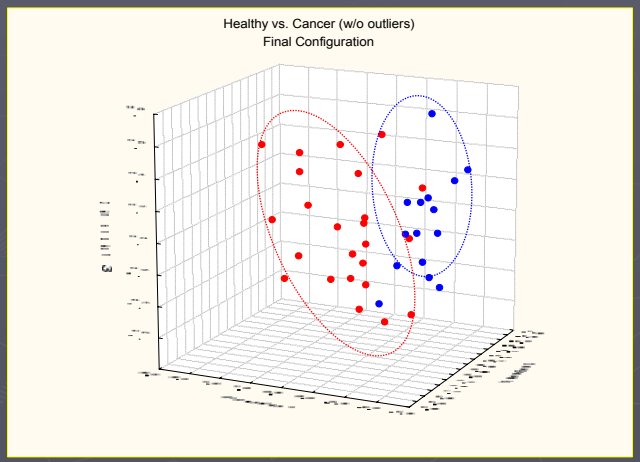
Multivariate and Classification Techniques

- Visual Assessment of differentially expressed proteins, post filtered as discussed with application of Clustering Approaches.
  - Principal Component or Multidimensional Scaling
  - Hierarchical Clustering Analysis
- Classification of larger data sets can be carried out using KNN/ Boot Strapping/ LOOCV……

Example (HCA)



Example (MDS)

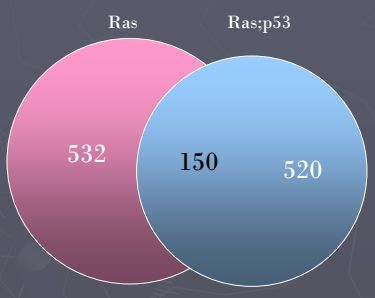


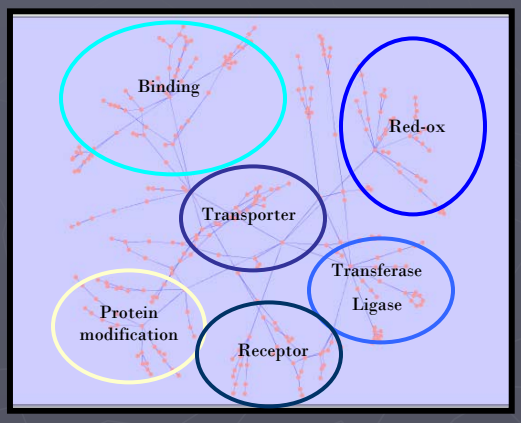
Overview of Systems Biology Approaches

- The "So What" Factor……..how do we side step this after a large experiment?
- Great Overview at……..
- http://www.cytoscape.org/
- http://cytoscape.org/cgi-bin/moin.cgi/Presentations

Example:

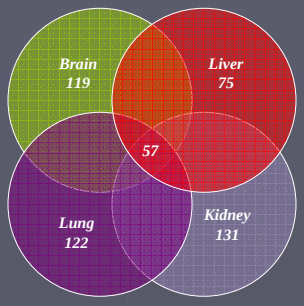
Protein bands were excised from the 1D gel for Ras and Ras:p53-63, digested with trypsin and run with LTQ-XL/CID mode. The results from LTQ-XL were run through SEQUEST to identify proteins. The proteins unique to Ras:p53-63 was run through *Cytoscape to see how those proteins relate to one another based on their molecular functions.

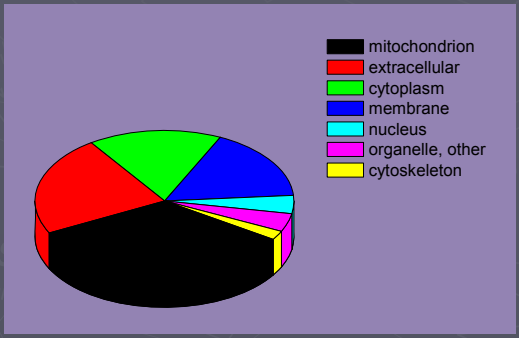




*Cytoscape is an open source bioinformatics software platform for ***visualizing*** molecular interaction networks and ***integrating*** these interactions with gene expression profiles and other state data.

Total Unique Proteins	= 277
Brain (B)	119
Liver (Li)	75
Lung (Lu)	122
Kidney (K)	131
Common (>2)	57
*Common (all 4)	33





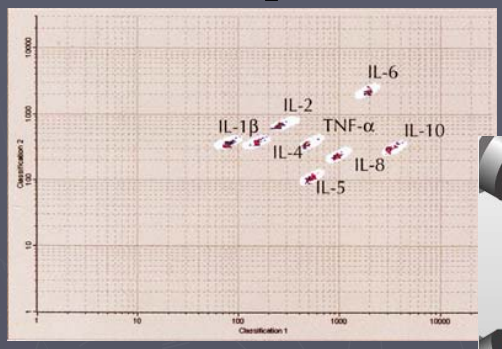
Example of Common Proteins

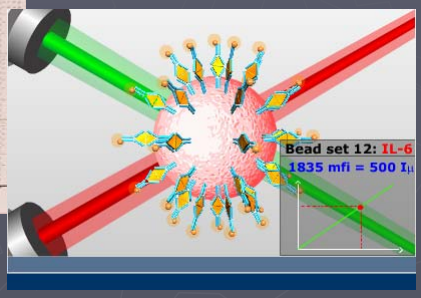
*Acyl-CoA-binding protein
Aldose reductase
*Alpha-enolase
***Apolipoprotein D precursor**
*ATP synthase-coupling factor
Bisphosphoglycerate mutase
*Copper transport protein ATOX1
*Cytochrome c oxidase
*D-dopachrome decarboxylase
Dynein light chain roadblock-type 1
*Glutathione S-transferase P 1
Isopentenyl-diphosphate isomerase 1
L-xylulose reductase
*Peptidyl-prolyl cis-trans isomerase A
*Phosphatidylethanolamine-binding protein 1
Polyadenylate-binding protein 1
Protein S100-A6
*Ptms protein
*Selenium-binding protein 1
*Superoxide dismutase [Cu-Zn]
Thioredoxin
*Thymosin beta-4
*TSC22 domain family protein 1
***Albumin**
***Apolipoprotein A-II**
Calmodulin 1
*Caronic anydrases (2)
***Hb-alpha**
***Hb-beta**
*Keratin Subtypes (10)
*MIF
Mitochondrial Membrane Tim8 A
Mitochondrial Membrane Tim8 B
*Niemann Pick type C2
*Prosaposins (3)
Ubiquitin C, full insert sequence
*Unknowns or Not Annotated (8)

Validation…..

- **Relative or Absolute Quantification of Known Proteins…….**
  - Mass Spectrometry (Mass Tags/ AQUA)
  - Immuno-Directed (examples)
    - 1) Protein Arrays
    - 2) Western Blot
    - 3) ELISA
    - 4) Bioplex/ Luminex

Ex: HTP Validation of Novel Markers with Multiplex Bead Assays (Luminex)





Multiplex Bead Assay for cytokines

The highlighted area represent populations of fluorescent beads, distinctively labeled, and carrying capture antibodies for sandwich assay of different cytokines.
All detection antibodies carry the same fluorophore, which is read in a third channel to quantify sample cytokine concentration

Bosch I. et. al. work in progress!

Summary

- Whether or not you do the MS work yourself……..
  - Know the specifics……!
  - Know the limitations…..!
- Sample Prep is always important…..but which instrument you have access to is also important!
- Similarly important….how is your data analyzed??
  - Denovo?
  - Matching Algorithm??
  - Mix of techniques??
  - What are the cut off points and does it make sense?
- Keep in mind that validation must be part of your workflow!!
- So, make sure you have confidence in your choices before going forward!

Useful Links!

- i-mass.com
- spectroscopynow.com
- expasy.ch/tools
- cprmap.com
- psidev.sourceforge.net
- prospector.ucsf.edu
- jeolusa.com/ms/docs/ionize.html
- asms.org (become a member!)
- hupo.org
- matrixscience.com
- proteomecenter.org/software.php

ionsource.com
bruker.com
thermo.com
appliedbiosystems.com
shimadzu.com
luminexcorp.com

Questions??