

Proteomics and Mass Spectrometry 2014

The team

Stephen Barnes, PhD Matt Renfrow, PhD Helen Kim, PhD
4-7117, MCLM 452 6-4681, MCLM 570 4-3880, MCLM 460A
sbarnes@uab.edu Renfrow@uab.edu helenkim@uab.edu

Jeevan Prasain, PhD
6-2612, MCLM 456
jprasain@uab.edu

Peter E. Prevelige, PhD
5-5327, BBRB 416
prevelig@uab.edu

S Barnes BMG 744 1/06/14

Course plan

- Meet Mondays/Wednesdays/Fridays in MCLM 401 from 9-10:30 am (Jan 6-Mar 21)
- Graduate Students taking this course are required to attend each session (unless there is advance communication with the instructor)
- Evaluations will be made from exams and in-class presentations
- Where possible, class notes will be available on the UAB proteomics website (go to <http://www.uab.edu/proteomics/massspec/> - click on **Class**)

S Barnes BMG 744 1/06/14

Recommended general texts

- ***“The Expanding Role of Mass Spectrometry in Biotechnology”*** by Gary Siuzdak (a 2003 edition of the 1996 first edition)
- ***“Mass spectrometry data analysis in proteomics”***, (ed., Mathiesson, R) in *Methods in Molecular Biology*, vol 367.
- ***“Protein Mass Spectrometry, Volume 52”*** (Comprehensive Analytical Chemistry) (Julian Whitelegge (Editor))
- See also <http://en.wikibooks.org/wiki/Proteomics>

S Barnes BMG 744 1/06/14

Suggested readings

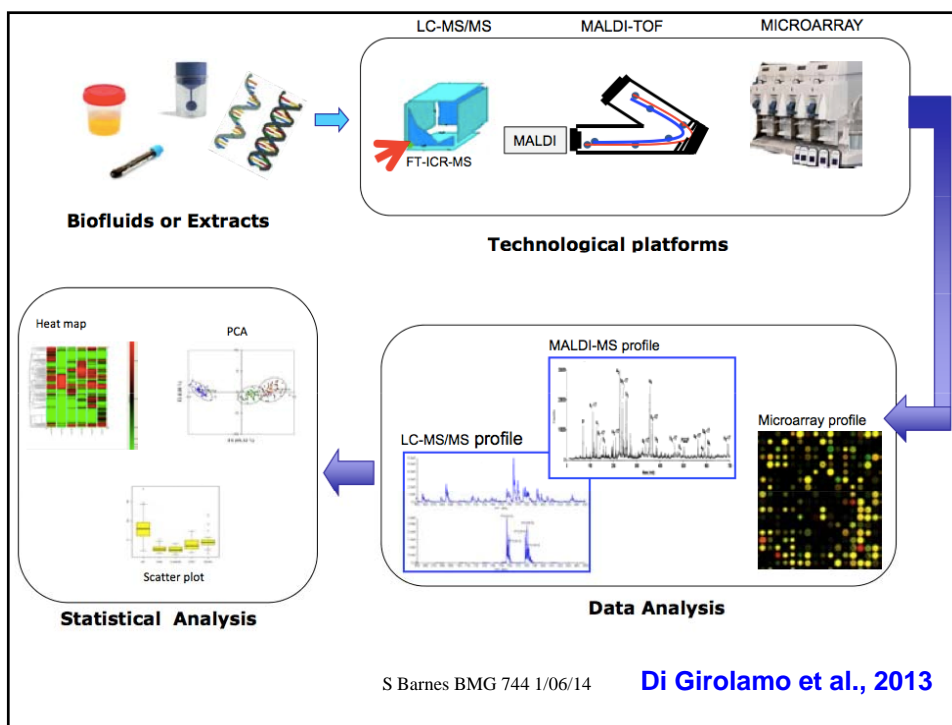
- Selvaraju S, Rassi ZE. *Liquid-phase-based separation systems for depletion, prefractionation and enrichment of proteins in biological fluids and matrices for in-depth proteomics analysis--an update covering the period 2008-2011*. *Electrophoresis*. 2012 Jan;33(1):74-88.
- Gillett et al. *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. *Mol Cell Proteomics*. 2012 Jun;11(6):O111.016717.
- Drissi et al. *Proteomics methods for subcellular proteome analysis*. *FEBS J*. 2013 Nov;280(22):5626-34.
- Liu et al. *Mass spectrometric protein maps for biomarker discovery and clinical research*. *Expert Rev Mol Diagn*. 2013 Nov;13(8):811-25.

S Barnes BMG 744 1/06/14

A useful overall review for this course

- Di Girolamo F, Lante I, Muraca M, Putignani L. The Role of Mass Spectrometry in the "Omics" Era. *Curr Org Chem.* 2013 Dec;17(23):2891-2905.

S Barnes BMG 744 1/06/14



BMG/PHR 744 - section 1

- Jan 6, Mon Barnes Analyzing biomolecules. The impact of -omics on biomedical research
- Jan 8, Wed H. Kim Simplifying the proteome - techniques of protein purification
- Jan 10, Fri H. Kim Protein separation by electrophoresis and other 2D-methods
- Jan 13, Mon M. Renfrow The Mass Spectrum: What does it show you? MS¹, MS², MSⁿ
- Jan 15, Wed M. Renfrow Mass Spectrometry: Detecting and moving ions in the gas phase. Instrumentation, Mass Analyzers, Ionization
- Jan 17, Fri M. Renfrow Biological Mass Spectrometry: Ionization, Calculating Mass, Charge
- Jan 20, Mon *Martin Luther King Holiday*
- Jan 22, Wed M. Renfrow Peptide/protein identification by MS, Search algorithms, false positives.
- Jan 24, Fri S. Barnes Qualitative burrowing of the proteome – identifying PTMs
- Jan 27, Mon S. Barnes Metabolomics – GC-MS, LC-MS and NMR
- Jan 29, Wed C. Crasto Web tools and the proteome/metabolome; Expaty, KEGG, NCBI, others
- Jan 31, Fri S. Barnes MRMPath; MRMutation; MRMass Space
- Feb 3, Mon S. Barnes Quantitative burrowing of the proteome – Labeling, label-free and absolute quantification
- Feb 5, Wed M. Renfrow *Take home Exam* (Possible lecture catch up and questions)

S Barnes BMG 744 1/06/14

BMG/PHR 744 - section 2

- Feb 7, Fri J. Prasain Lipidomics and other small molecules by LC-MS
- Feb 10, Mon J. Prasain Quantitative analysis/method validation in metabolomics
- Feb 12, Wed S. Barnes Enzymology, metabolism and mass spectrometry
- Feb 14, Fri E. Shonsey Application of MS in Forensics
- Feb 17, Mon M. Renfrow Analysis of protein-protein interactions by affinity purification and mass spectrometry
- Feb 19, Wed M. Renfrow Applications of FT-ICR-MS
- Feb 21, Fri J. Novak/ Mass spectrometry in glycomics research –
Renfrow Application to IgA nephropathy
- Feb 24, Mon S. Barnes/ Application of MS to tissue imaging
J. Kabarowski
- Feb 26, Wed [Student presentations](#)
- Feb 28, Fri [Student presentation](#)

S Barnes BMG 744 1/06/14

BMG/PHR 744 - section 3

- Mar 3, Mon H. Kim Use of proteomics and MS methods in the study of the brain proteome and neurodegenerative diseases
- Mar 5, Wed H. Kim/ Putting it all together – by-passing
S. Barnes pyruvate kinase
- Mar 7, Fri P. Prevelige Mass Spectrometry as a Tool for Studying Protein Structure
- Mar 10, Mon P. Prevelige Study of macromolecular structures – protein complexes
- Mar 12, Wed S. Barnes Isotopes in mass spectrometry
- Mar 14, Fri S. Barnes Applying mass spectrometry to Free Radical Biology
- Mar 21, Fri *Final report due*

S Barnes BMG 744 1/06/14

Course learning objectives

- Introduction to the concepts and practice of systems biology where it involves mass spectrometry
- Sample ionization and mass spectrometers
- Mass spectrometry and its principal methods
 - protein and peptide ID; peptide and metabolite ion fragmentation; stable isotope labeling; quantification

S Barnes BMG 744 1/06/14

Course learning objectives

- Importance of prefractionation in proteomics - 2DE, LC, capillary electrophoresis and arrays
- Applying mass spectrometry to protein modifications, function, structure and biological location, and to other biological molecules
- Informatics, statistics and quality control in mass spectrometry

S Barnes BMG 744 1/06/14

History of proteomics

- Essentially preceded genomics
- “Human protein index” conceived in the 1970s by Norman and Leigh Anderson
- The term “proteomics” coined by Marc Wilkins in 1994
- Human proteomics initiative (HPI) began in 2000 in Switzerland - <http://www.hupo.org>
- Human Proteome Organization (HUPO) had meetings in 2002 in Versailles; 2003 in Montreal; 2004 in Beijing; 2005 in Munich; 2006 in Long Beach; 2007 in Seoul; 2008 in Amsterdam; 2009 in Toronto; 2010 in Sydney; 2011 in Geneva; 2012 in Boston; 2013 in Yokohoma; 2014 to be in Madrid

S Barnes BMG 744 1/06/14

What -Omics is, what it isn't

-Omics is not just a 1D- or 2D-gel or a LC-MS run

-Omics is the identities, quantities, structures, and biochemical and cellular functions of all proteins and metabolites in an organism, organ or organelle, and how these vary in space, time and physiological state.

S Barnes BMG 744 1/06/14

Collapse of the single target paradigm - the need for systems biology

Old paradigm

Diseases are due to single genes - by knocking out the gene, or designing specific inhibitors to its protein, disease can be cured

But the gene
KO mouse
didn't notice
the loss of the
gene



S Barnes BMG 744 1/06/14

New paradigm

We have to understand gene, protein and metabolite networks - proteins don't act alone - effective systems have built in redundancy

Research styles

- **Classical NIH R01**
 - A specific target and meaningful substrates
 - Emphasis on mechanism
 - Hypothesis-driven
 - **Linearizes locally multi-dimensional space**
- **Example**
 - Using an X-ray crystal structure of a protein to determine if a specific compound can fit into a binding pocket - from this “*a disease can be cured*” - this approach ignores whether the compound can get to the necessary biological site, whether it remains chemically intact, and where else it goes

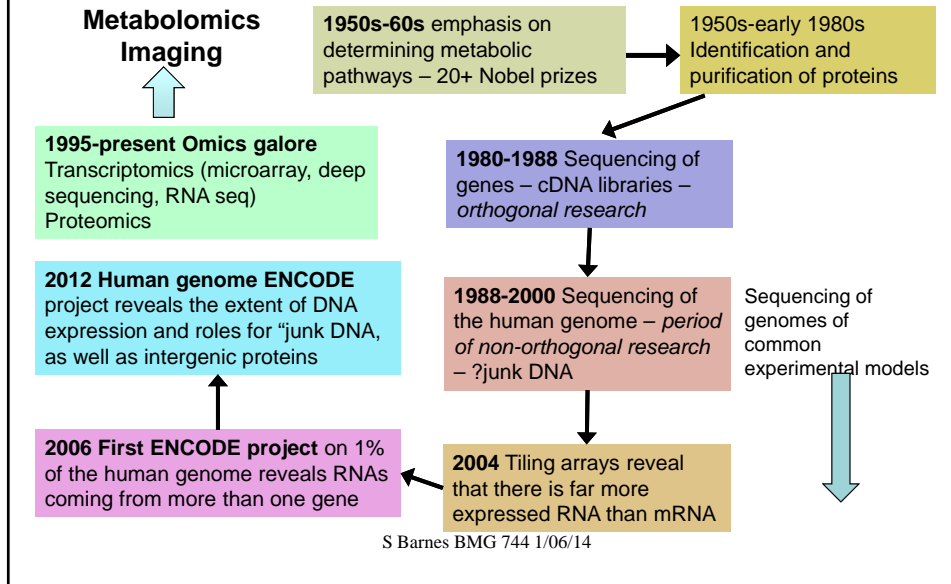
S Barnes BMG 744 1/06/14

From substrates to targets to systems - a changing paradigm

- **Classical approach** - one substrate/one target
- **Mid 1980s** - use of a pure reagent to isolate DNAs from cDNA libraries (multiple targets)
- **Early 1990s** - use of a reagent library (multiple ligands) to perfect interaction with a specific target
- **2000+** - effects of specific reagents on cell systems using DNA microarrays (500+ genes change, not just one)
- **2008** - integration of transcriptomics, proteomics, peptidomics, metabolomics (everything changes, just like in ecology)
- **2010** – NextGen and RNASeq analyses introduced (the canonical sequence is a myth- discovery of new transcriptome products)

S Barnes BMG 744 1/06/14

Direction of NIH Research 1950-2014



Exploring information space - the *Systems Biology* approach

- **Systems biology means measuring everything about a system at the same time**
- **For a long time, it was deemed as too complex for useful or purposeful investigation**
- **But are the tools available today?**

S Barnes BMG 744 1/06/14

Systems Biology

“To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism.”

“Properties of systems, such as robustness, emerge as central issues, and understanding these properties may have an impact on the future of medicine.”

“However, many breakthroughs in experimental devices, advanced software, and analytical methods are required before the achievements of systems biology can live up to their much-touted potential.”

Kitano, 2002

S Barnes BMG 744 1/06/14

The Biological Data of the Future

- Destructive
- Qualitative
- Uni-dimensional

- Low temporal resolution
- Low data density
- Variable standards
- Non cumulative

Current nature of data

- Non-destructive
- Quantitative
- Multi-dimensional and spatially resolved
- High Temporal resolution
- High data density
- Stricter standards
- Cumulative

Elias Zerhouni, FASEB 2004

S Barnes BMG 744 1/06/14

Techniques in Systems Biology

- DNA microarrays to describe and *quantify* the transcriptosome
 - Being replaced by NextGen sequencing and RNASeq
- Large scale and small scale proteomics
- Protein arrays
- Protein structure
- Metabolomics
- Imaging methods
- Integrated computational models

S Barnes BMG 744 1/06/14

Papers on systems biology

Deo RC, MacRae CA. *The zebrafish: scalable in vivo modeling for systems biology*. [WIREs Systems Biol 2011;3:335-346](#).

Gardy JL et al. *Enabling a systems biology approach to immunology: focus on innate immunity*. [Trends in Immunol 2011;30:249-262](#).

Kriete A et al. *Computational systems biology of aging*. [WIREs Systems Biol 2011;3:414-428](#).

Jorgenson JM, Haddow, PC. *Visualization in simulation tools: requirements and a tool specification to support the teaching of dynamic biological processes*. [J Bioinform Comp Biol 2011;9:579-595](#).

Gerdtsen ZP. *Modeling Metabolic Networks for Mammalian Cell Systems: General Considerations, Modeling Strategies, and Available Tools*. [Adv Biochem Engin/Biotechnol DOI: 10.1007/10_2011_120](#).

He JC et al. *Systems biology of kidney diseases*. [Kidney Int. 2012;81:22-39](#).

Meng Q, et al. *Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases*. [Curr Cardiovasc Risk Rep. 2013 Feb;7\(1\):73-83](#).

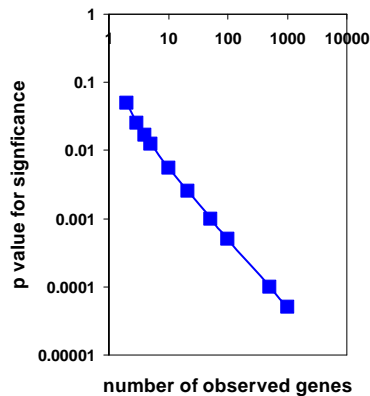
Altm e S, et al. *Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium*. [Hum Reprod Update 2014;20:12-28](#).

S Barnes BMG 744 1/06/14

High dimensionality of microarray, metabolomics and proteomics data means you must understand statistics

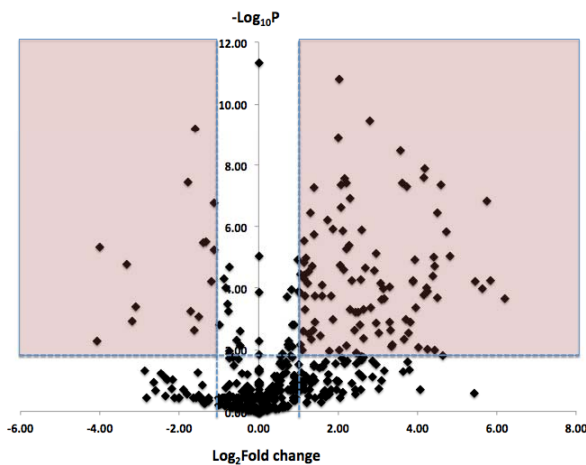
While reproducible data can be obtained, the large numbers of parameters (individual genes, metabolites or proteins) require large changes in expression before a change can be regarded as significant

Use of the Bonferroni correction:
A (too) conservative correction



S Barnes BMG 744 1/06/14

Limitation of univariate statistics

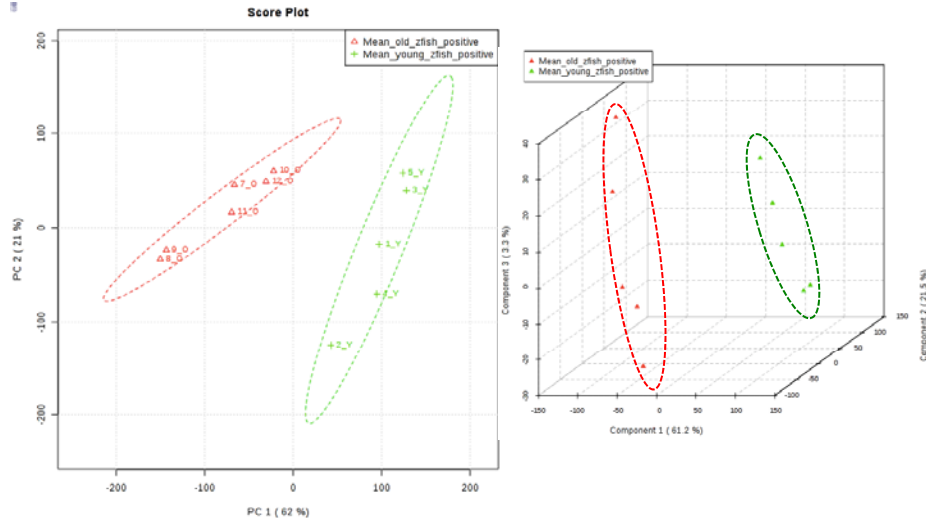


A Volcano Plot

Univariate statistics are not adequate to recover the important information from -Omics data

S Barnes BMG 744 1/06/14

Multivariate statistics and -Omics



S Barnes BMG 744 1/06/14

Vulnerability of a system

- **To really understand biological systems, you have to appreciate their dynamic state**
 - Read about control theory
 - Realize that systems are subject to rhythms
 - Subject them to fourier transform analysis to detect their resonance (requires far more data than we can currently collect)
- **A small signal at the right frequency can disrupt the system**
 - Analogies “the small boy in the bath” and “the screech of chalk on a chalk board”

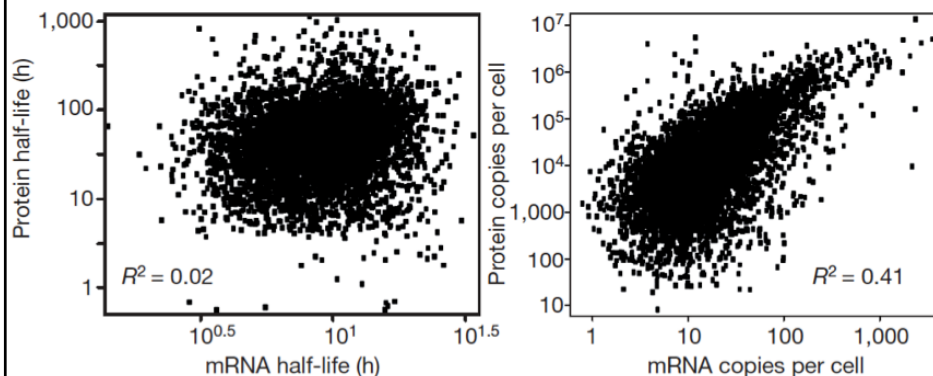
S Barnes BMG 744 1/06/14

Why study the proteome when we can study DNA and RNA?

- NextGen and RNASeq analysis allows one to examine the mRNA levels of thousands and thousands of genes
- However, the correlation between gene expression and protein levels is often poor, although that may be an issue of the timing of sampling
- Is this a new finding? No, before the age of molecular biology, it was well known

S Barnes BMG 744 1/06/14

Weak relationship between gene and protein expression



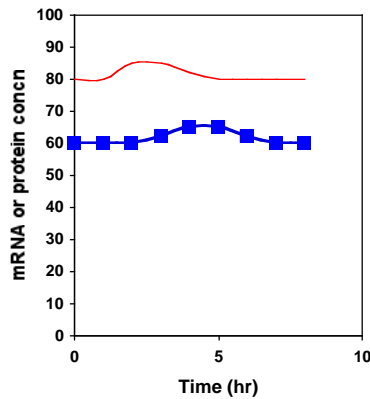
S Barnes BMG 744 1/06/14

Schwanhäusser et al.
Nature 473: 337 (2011)

Housekeeping genes and proteins are probably related

This is the relationship between mRNA (red) and protein (blue) levels expression of a house-keeping gene/protein, i.e., one that has to be expressed at all times

- Even with the small perturbation, the amounts of mRNA and protein are well correlated to each other

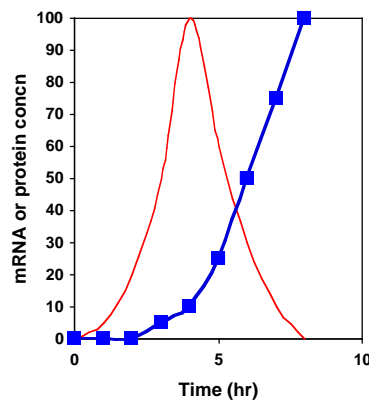


S Barnes BMG 744 1/06/14 (Barnes & Allison, 2004)

Sampling time affects interpretation of correlation between mRNA and protein expression for important proteins

Determining the relationship between mRNA (red) and protein (blue) levels depends totally on when you measure them - for the figure opposite, the ratio at 2.5 hr is 10:1, whereas at 7.5 hr it's 1:100

- better to measure the ratio over time and integrate the area under the curve



S Barnes BMG 744 1/06/14 (Barnes & Allison, 2004)

Issues in analyzing proteins with proteomics methods

- The fewer proteins in the proteome you analyze, the better the chances of detecting the ones that “matter.”
- Genomics data can complement proteomics data.
- Understanding the biological properties of the proteins of interest can enhance proteomics analysis.
- Intrinsic properties of proteins form the basis of invaluable prefractionation prior to proteomics analysis.
- Quality control is an issue that becomes increasingly important with large datasets and measurement of small changes

S Barnes BMG 744 1/06/14

Predicting the proteome

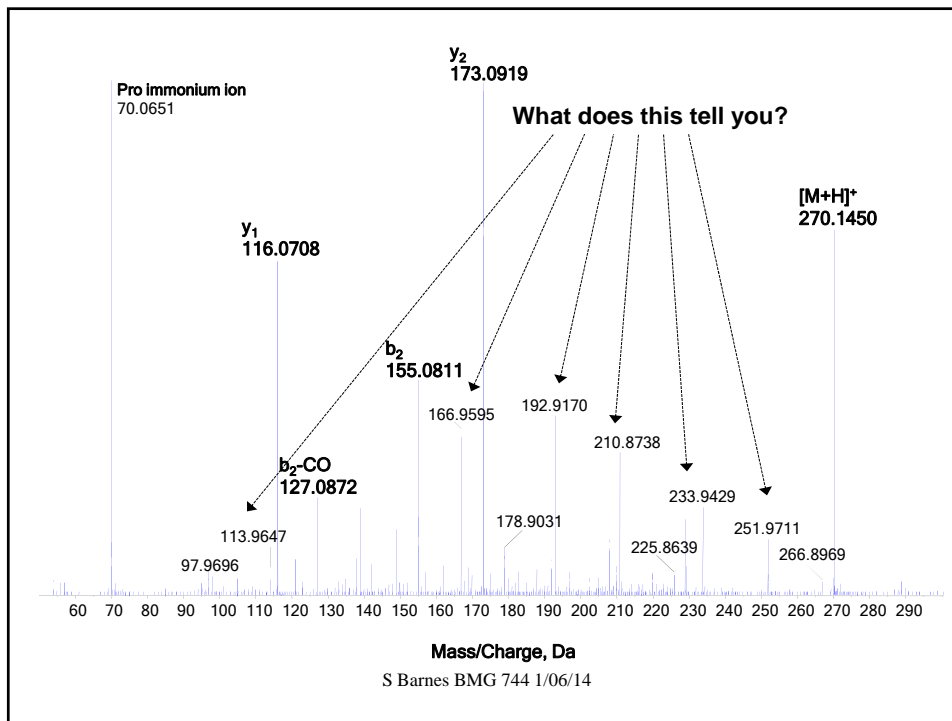
- *Bioinformatics* is the basis of high throughput proteome analysis using mass spectrometry. Protein sequences can be computationally predicted from the genome sequence
- However, *bioinformatics* is not able to predict with accuracy the sites or chemistry of posttranslational modifications - these need to be defined chemically (using mass spectrometry)
- Proteins in individuals will have different sequences – there are 161 known natural mutations of the LDL receptor and 1361 mutations of human p53

S Barnes BMG 744 1/06/14

Predicting the proteome

- *Predicting the proteome has elements of a circular argument*
 - protein sequences were initially determined chemically and were correlated with the early gene sequences. It then became easier to sequence a protein from its mRNA (captured from a cDNA library). This could be checked (to a degree) by comparison to peptide sequences. Now we have the human genome.
- *So, is it valid to predict the genes (and hence the proteome) from the sequence of the genome?*
 - We're doing this in current research. But as we'll see, the mass spectrometer is the ultimate test of this hypothesis -
 - why? because of its mass accuracy

S Barnes BMG 744 1/06/14



Protein structure

- Determined by folding - folding rules not yet defined - cannot predict structure *de novo*
- X-ray crystallography has been used to produce elegant structural information
- NMR and H-D exchange combined with mass spec enable the in-solution structure to be determined (see Peter Prevelige's lectures on March 7/9)

S Barnes BMG 744 1/06/14

Protein informatics

- The predicted sequences of the proteins encoded by genes in sequenced genomes are available in many publicly available databases (subject to the limitations mentioned earlier)
- The mass of the protein is less useful (for bottom up, **but not top-down analyses**) than the masses of its fragment ions - as we'll see later, the masses of tryptic peptides can be used to identify a protein in a matter of seconds

S Barnes BMG 744 1/06/14

So, what do we do with all these data?

- **Management of the data generated by DNA microarrays, NextGen Sequencing, RNASeq and proteomics/protein arrays**
 - High dimensional analysis
- **Beyond the capabilities of individual investigators**
- **Urgent need for visualization tools**
- **The importance of new statistical methods for analysis of high dimensional systems**

S Barnes BMG 744 1/06/14

PROTIG and Videocast

- **There is an NIH-based proteomics interest group (PROTIG)**
 - <http://proteome.nih.gov>
- **Proteomics and mass spec talks are available for viewing (using Real Player)**
 - Log on at <http://videocast.nih.gov>
 - Podcasts are also available

S Barnes BMG 744 1/06/14