Proteomics and Mass Spectrometry 2013

The team

Stephen Barnes, PhD
4-7117, MCLM 452
sbarnes@uab.edu

Matt Renfrow, PhD
6-4681, MCLM 570
Renfrow@uab.edu

Helen Kim, PhD
4-3880, MCLM 460A
helenkim@uab.edu

Jim Mobley, PhD
6-6363, THT 521
mobleyja@uab.edu

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

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

Course plan

- **Meet Mondays/Wednesdays/Fridays in MCLM 401 from 9-10:30 am (Jan 7-Mar 22)**
- **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/index2.php - click on Class)**

Recommended general texts

- **Suggested text - "*Introduction to Proteomics*" by Daniel C. Liebler, 2002**
- **Also see "*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**

Suggested readings

- **Kenyon G, et al. *Defining the mandate of proteomics in the post-genomics era: workshop report*. Mol Cell Proteomics, 1: 763-80 (2002)**
- **Righetti P. et al. *Prefractionation techniques in proteome analysis: the mining tools of the third millennium*. Electrophoresis, 26: 297-319 (2005)**
- **Anderson NL. *The roles of multiple proteomic platforms in a pipeline for new diagnostics*. Mol Cell Proteomics, 4:1441-4 (2005)**
- **Venkatesan et al. *An empirical framework for binary interactome mapping*. Nat Methods, 6:83-90 (2009) PMID: 19060904**
- **Yan W et al. *Evolution of organelle-associated protein profiling*. J Proteomics, 72:4-11 (2009) PMID: 19110081**
- **Pan S, et al. *Mass Spectrometry Based Targeted Protein Quantification: Methods and Applications*. J Proteome Res, 8:787-97 (2009) PMID: 19105742**
- **Compton PD et al. *On the Scalability and Requirements of Whole Protein Mass Spectrometry.* Anal Chem, 83:6868–74 (2011) PMID:21744800**

BMG/PHR 744 - section 1

- Jan 7, Mon Barnes Analyzing biomolecules. The impact of –omics on biomedical research
- Jan 9, Wed H. Kim Simplifying the proteome - techniques of protein purification
- Jan 11, Fri H. Kim Protein separation by electrophoresis and other 2D-methods
- Jan 14, Mon M. Renfrow The Mass Spectrum: What does it show you? MS^1, MS^2, MS^n
- Jan 16, Wed M. Renfrow Mass Spectrometry: Detecting and moving ions in the gas phase. Instrumentation, Mass Analyzers, Ionization
- Jan 18, Fri M. Renfrow Biological Mass Spectrometry: Ionization, Calculating Mass, Charge
- Jan 21, Mon *Martin Luther King Holiday*
- Jan 23, Wed M. Renfrow Peptide/protein identification by MS, Search algorithms, false positives.
- Jan 25, Fri S. Barnes Qualitative burrowing of the proteome – identifying PTMs
- Jan 28, Mon S. Barnes Quantitative burrowing of the proteome – Labeling, label-free and absolute quantification
- Jan 30, Wed C. Crasto Web tools and the proteome/metabolome; Expasy, KEGG, NCBI, others
- Feb 1, Fri S. Barnes MRMPath; MRMutation; MRMass Space
- Feb 4, Mon J. Prasain Lipidomics and other small molecules by LC-MS
- Feb 6, Wed M. Renfrow *Exam* (Possible lecture catch up and questions)

BMG/PHR 744 - section 2

- Feb 8, Fri J. Prasain Metabolomics – LC-MS, GC-MS and NMR
- Feb 11, Mon J. Prasain Quantitative analysis/method validation in metabolomics
- Feb 13, Wed S. Barnes Enzymology, metabolism and mass spectrometry
- Feb 15, Fri Student presentations
- Feb 18, Mon M. Renfrow Analysis of protein-protein interactions by affinity purification and mass spectrometry
- Feb 20, Wed M. Renfrow Applications of FT-ICR-MS
- Feb 22, Fri J. Novak/Renfrow Mass spectrometry in glycomics research – Application to IgA nephropathy
- Feb 25, Mon E. Shonsey Application of MS in Forensics
- Feb 27, Wed S. Barnes Applications of MS to tissue imaging – the lens and the metabolome

BMG/PHR 744 - section 3

- **Mar 1, Fri** **Student presentations**
- **Mar 4, Mon** **P. Prevelige Mass Spectrometry as a Tool for Studying Protein Structure**
- **Mar 6, Wed** **P. Prevelige Study of macromolecular structures – protein complexes**
- **Mar 8, Fri** **H. Kim Use of proteomics and MS methods in the study of the brain proteome and neurodegenerative diseases**
- **Mar 11, Mon** **H. Kim/S. Barnes Putting it all together – by-passing pyruvate kinase**
- **Mar 13, Wed** **S. Barnes Isotopes in mass spectrometry**
- **Mar 15, Fri** **S. Barnes Applying mass spectrometry to Free Radical Biology**
- **Mar 22, Fri** ***Final report due***

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**

Course learning objectives

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

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 to be in Yokohoma

What proteomics is, what it isn't

"Proteomics is not just a mass spectrum of a spot on a gel"

George Kenyon,
2002 National Academy of Sciences Symposium

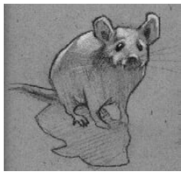
Proteomics is the identities, quantities, structures, and biochemical and cellular functions of all proteins in an organism, organ or organelle, and how these vary in space, time and physiological state.

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



New paradigm

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

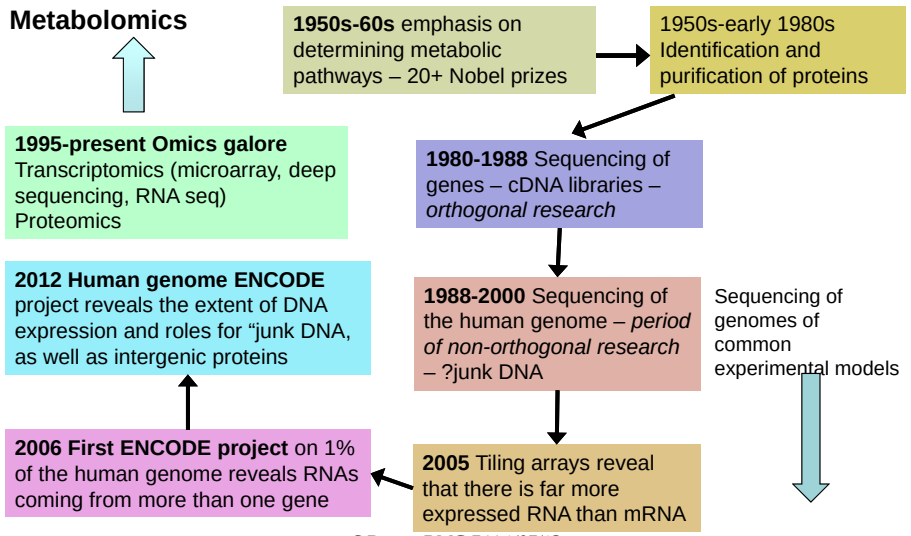
Research styles

- **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**

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 and new transcriptome products)**

Direction of NIH Research 1950-2012



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?**

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 1/07/13

The Biological Data of the Future

Current nature of data	Elias Zerhouni, FASEB 2004
• Destructive	• Non-destructive
• Qualitative	• Quantitative
• Uni-dimensional	• Multi-dimensional and spatially resolved
• Low temporal resolution	• High Temporal resolution
• Low data density	• High data density
• Variable standards	• Stricter standards
• Non cumulative	• Cumulative

S Barnes BMG 744 1/07/13

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**
- **Integrated computational models**

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.

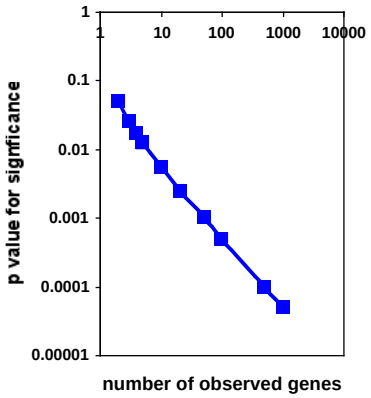
Shapira SD, Hacohen N. *Systems biology approaches to dissect mammalian innate immunity*. Current Opin Immunol 2011;23:71-77.

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.

Gerdtzen 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

High dimensionality of microarray or proteomics data means you must understand statistics

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



Use of the Bonferroni correction: A conservative correction

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"**

Hazards of interpreting transcriptomics (proteomic) data

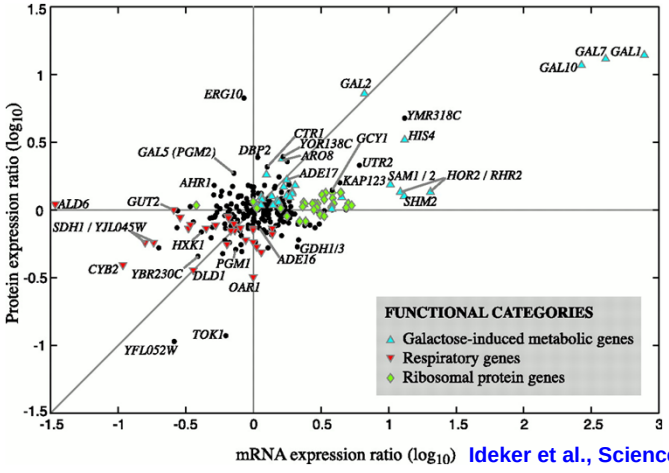
- "Expression patterns are the place where environmental variables and genetic variation come together. Environmental variables will affect gene expression levels."
- "Don't we need to be very careful to understand the environmental inputs that might have an impact on that expression? Perhaps an over-the-counter herbal supplement might cause an expression pattern that looks like that of a very aggressive tumor."

Abridged from Karen Kline, 2002

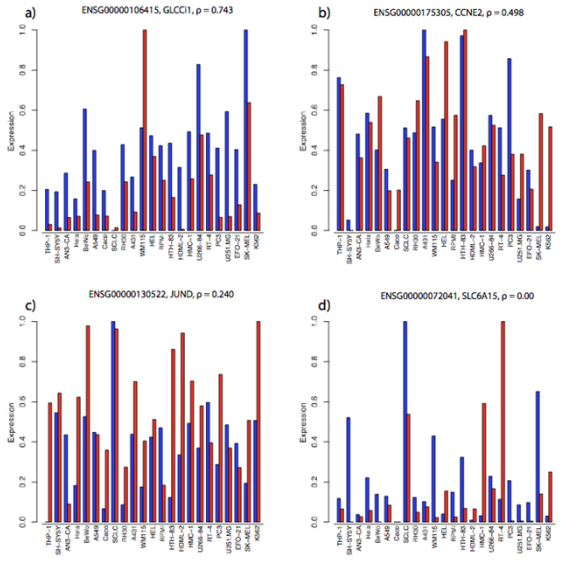
Why study the proteome when we can study DNA and RNA?

- 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

Apparent structured relationship between gene expression and protein content



Ideker et al., Science 292: 929 (2001)



Differing correlations between RNA expression of a gene and its associated protein amount in 23 different cell lines

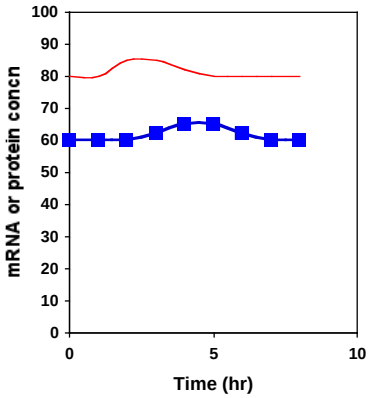
- **a) glucocorticoid-induced transcript 1 protein**
- **b) G1/S-specific cyclin-E2**
- **c) jun-D, a transcription factor**
- **d) neurotransmitter transporter NTT73**

From Gry et al. BMC Genomics 10:365 (2009)

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

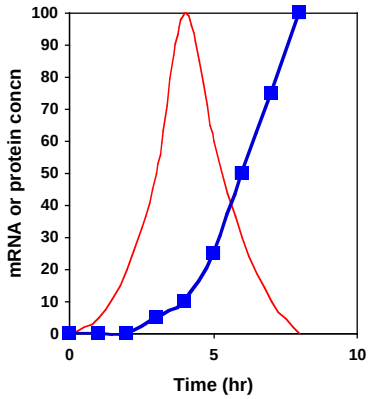


(Barnes & Allison, 2004)

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

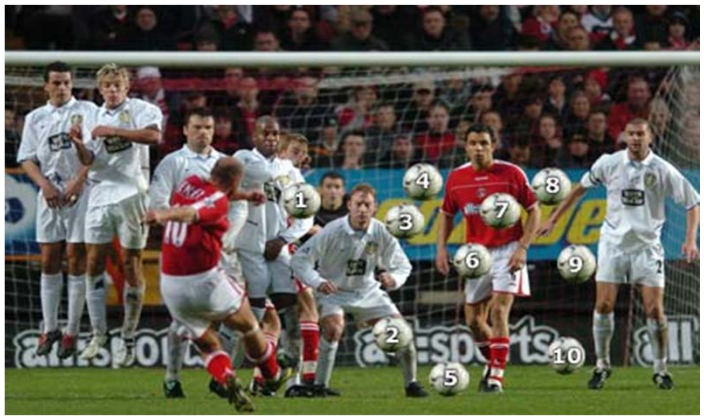
Determining the relationship between mRNA (red) and protein (blue) levels depends totally on when you measure them - for the figure opposite, the ratio at 2.5 hr is 10:1, whereas at 7.5 hr it's 1:100

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



(Barnes & Allison, 2004)

Figuring the right answer
- the disadvantage of a static set of data



Which ball in this picture is the real one? You have ten choices

Take home lessons in analyzing proteins with proteomics methods

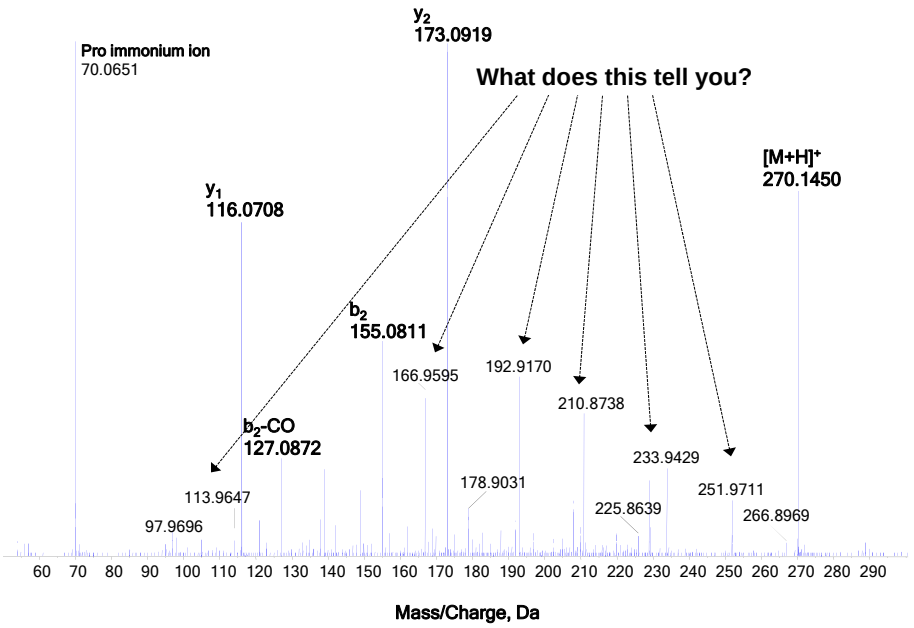
- **The fewer proteins in the proteome you analyze, the better the chances of detecting the ones that "matter."**
- **Genomics data can complement proteomics data.**
- **Understanding the biological properties of the proteins of interest can enhance proteomics analysis.**
- **Intrinsic properties of 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**

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**

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? <u>because of its mass accuracy</u>**



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 6/8)**

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

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

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 1/07/13