

Proteomics and Mass Spectrometry 2012

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

Jim Mobley, PhD Jeevan Prasain, PhD Peter E. Prevelige, PhD
6-6363, THT 521 6-2612, MCLM 456 5-5327, BBRB 416
mobleyja@uab.edu jprasain@uab.edu prevelig@uab.edu

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

Course plan

- Meet Mondays/Wednesdays/Fridays in MCLM 401 from 9-10:30 am (Jan 6-Mar 23)
- 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**)

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

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))

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

BMG/PHR 744 - section 1

Jan 6, Fri	Barnes/Kim	The world of biomolecules. The proteome, proteomics and other -omics and where to start
Jan 9, Mon	M. Renfrow	Mass spectrometry – gas phase transfer and instrumentation – include ETD
Jan 11, Wed	M. Renfrow	Methods for the identification of proteins: MALDI-TOF of proteins and peptide mass fingerprinting; LC analysis and peptide sequencing
Jan 13, Fri	M. Renfrow	Ion fragmentation in mass spectrometry; application to proteomics
Jan 16, Mon	S. Barnes/ H. Kim	Sample preparation for proteomics and mass spectrometry
Jan 18, Wed	H. Kim	Simplifying the proteome - techniques of protein purification
Jan 20, Fri	H. Kim	Protein separation by electrophoresis and other 2D-methods
Jan 23, Mon	S. Barnes	Qualitative burrowing of the proteome
Jan 25, Wed	S. Barnes	Quantitative burrowing of the proteome
Jan 27, Fri	M. Renfrow	MUDPIT and SEQUEST; false discovery rates in complex systems
Jan 30, Mon	C. Crasto	Web tools and the proteome; Expasy, KEGG, NCBI, others
Feb 1, Wed	S. Barnes	MRMPATH; MRMutation; MRMass Space
Feb 3, Fri	J. Prasain	Lipidomics and other small molecule LC-MS
Feb 6, Mon	<i>Exam</i>	

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

BMG/PHR 744 - section 2

Feb 8, Wed	J. Prasain	Metabolomics – LC-MS and GC-MS
Feb 10, Fri	J. Prasain	Quantitative analysis/method validation in metabolomics
Feb 13, Mon	S. Barnes	Enzymology and mass spectrometry
Feb 15, Wed		Student presentations
Feb 17, Fri	M. Renfrow	Analysis of protein-protein interactions by affinity purification and mass spectrometry
Feb 20, Mon	P. Prevelige	Mass Spectrometry as a Tool for Studying Protein Structure
Feb 22, Wed	P. Prevelige	Study of macromolecular structures – protein complexes
Feb 24, Fri	E. Shonsey	MS in Forensics

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

BMG/PHR 744 - section 3

Feb 27, Mon	TBA	Tissue and body fluid proteomics and mass spectrometry
Feb 29, Wed		Student presentations
Mar 2, Fri	D. Stella	Applications of MS to tissue imaging – the lens
Mar 5, Mon	M. Renfrow	Applications of FT-ICR-MS
Mar 7, Wed	J. Novak/	Mass spectrometry in glycomics research – M. Renfrow Application to IgA nephropathy
Mar 9, Fri	H. Kim	Use of proteomics and MS methods in the study of the brain proteome and neurodegenerative diseases
Mar 12, Mon	H. Kim/ S. Barnes	Putting it all together – by-passing pyruvate kinase
Mar 14, Wed	S. Barnes	Isotopes in mass spectrometry
Mar 16, Fri	S. Barnes	Applying mass spectrometry to Free Radical Biology
Mar 19, Mon		Final report due

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

Course learning objectives

- Introduction to the concepts and practice of systems biology
- 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/H Kim/M Renfrow BMG 744 1/06/12

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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Hopes and hazards of biomedical research

It boils down to whether having taken life apart into its distinct pieces, can we reassemble it in new ways? [*strong analogies to what have been the central quests of physics since the early 20th century*]

Can we create a form of life that might live in a very hostile extra-terrestrial environment and thereby save humanity?

Or will we (*as well as our enemies*) instead create life forms that can terrorize, or even eliminate us?
Will Einstein's and Oppenheimer's moral dilemmas surface in biomedical science?

S Barnes/H Kim/M Renfrow BMG 744 1/06/12

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, France; 2003 in Montreal, Canada; 2004 in Beijing, China; 2005 in Munich, Germany; 2006 in Long Beach, CA; 2007 in Seoul, Korea; 2008 in Amsterdam, The Netherlands; 2009 in Toronto, Canada; 2010 meeting in Sydney; 2011 meeting in Geneva, Switzerland; 2012 to be held in Boston

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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.

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

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)

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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.

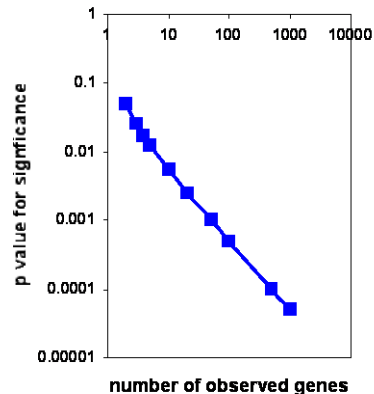
Jorgenson JM, Haddow, PC. *Visualization in simulation tools: requirements and a tool specification to support the teaching of dynamic biological processes*. J Bioinform Comp Biol 2011;9:579-595.

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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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



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

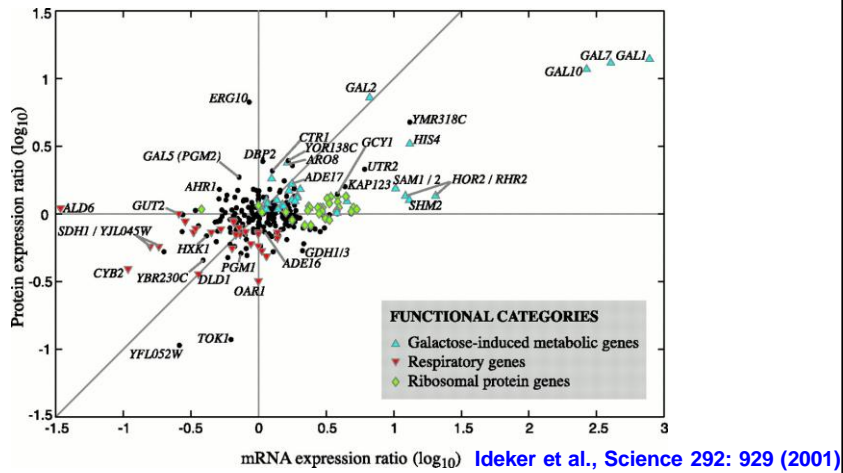
S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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

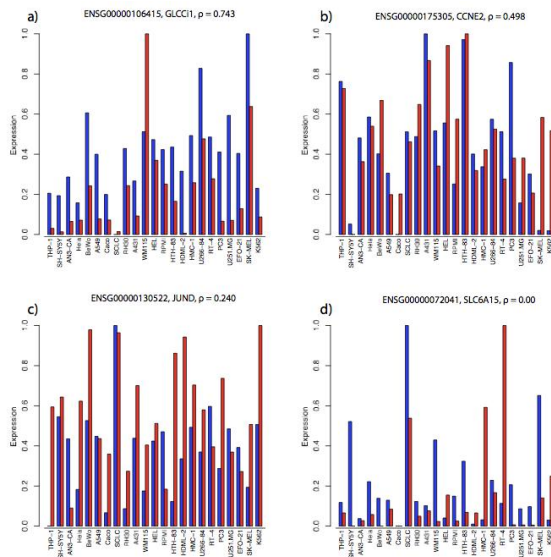
- **NextGen and RNASeq analysis allows one to examine the mRNA levels of thousands and thousands of genes**
- **However, the correlation between gene expression and protein levels is often poor, although that may be an issue of the timing of sampling**
- **Is this a new finding? No, before the age of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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Apparent structured relationship between gene expression and protein content



S Barnes/H Kim/M Renfrow BMG 744
1/06/12



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

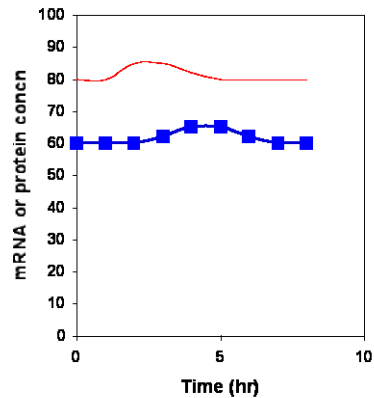
S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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



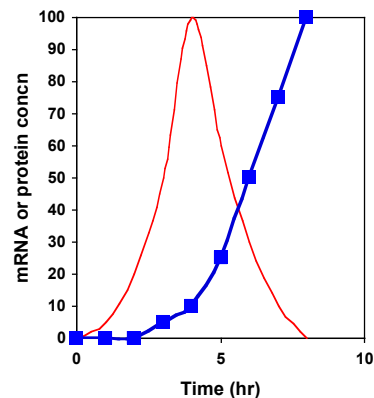
S Barnes/H Kim/M Renfrow BMG 744
1/06/12

(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/H Kim/M Renfrow BMG 744
1/06/12

(Barnes & Allison, 2004)

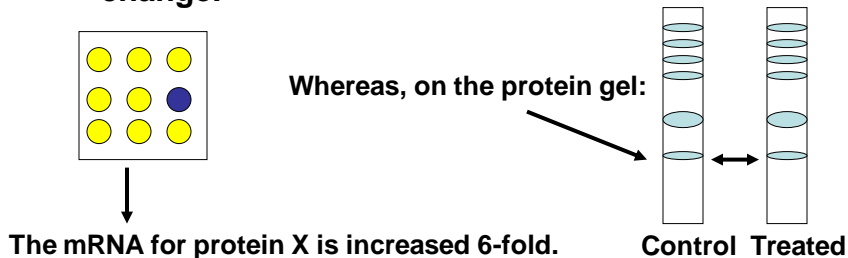
Defining disease using proteomic approaches

- Numerous examples of updating understanding of disease using -Omic approaches
 - Aging
 - Cancer
 - Cardiovascular disease
 - Neurodegeneration
 - Infectious disease (microbial proteomes)

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

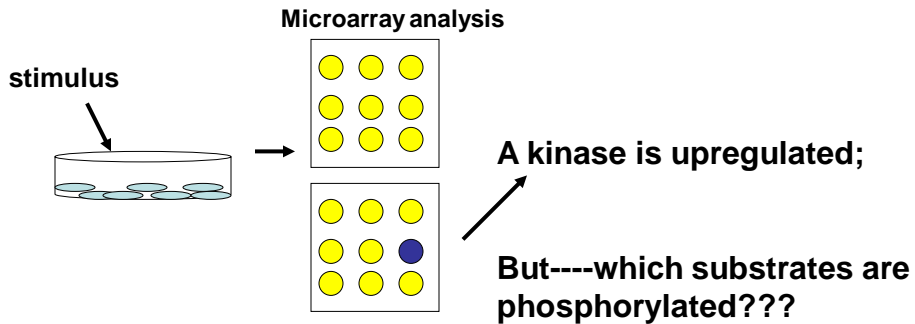
Upregulation of a gene for a crucial protein may have poor correlation at the protein level: example A.

- May be manifested as little or no change in amount, because the cell “damps” out the change.



S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Gene expression data may predict the nature of protein modifications, *but not which proteins are modified.*



S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Summary: Why do we need to do broad scale “proteomics”?

- When you don't have a clue;
- When you have only a very small clue;
i.e.. you've done a microarray experiment, and you have a list of 3,284 genes that are differentially regulated in your system;
- When you knock out a gene (and hence a protein) that you're convinced is essential for life and health, and the animal pees as usual.

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Rationales for proteomics approaches in today's research

- Identify a “marker” protein(s);
 - Cancer detection/Monitor response to chemotherapy
 - Identify one pathogen from others;
 - Distinguish a virulent strain of pathogen from nonvirulent.
- Characterize protein differences between disease and normal tissues--
 - For understanding the disease process;
 - To develop drug targets;
- In cancer, there may be novel proteins due to chromosome instability (ETV6-ABL and BCR-ABL), or inappropriate expression may occur (proteins from embryonic or fetal stages of development)

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Where there is pathology, but the genetic basis unknown, proteomics can have critical role in identifying proteins to target for therapeutic intervention

Two major disease examples:

- HIV: protease is targeted today;
 - are there other proteins, either viral or host, that could be targeted to better deal with the disease?
- Alzheimer's disease: 3 known mutations (APP, PS1, PS2) and risk factors (ApoE, estrogen loss);
 - 50% of AD patients do not have any of the known genetic abnormalities, yet all become demented, all have amyloid plaques and NFT in their brains.
 - Remember, every AD patient has AD 100%.

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Challenges in proteomics

- So many protein forms! How many 100-mers are possible using 20 different amino acids?
- Can protein forms be predicted from genomic sequence data?
 - Posttranslational modifications
 - Differential splicing
 - Unknown transcriptional mechanisms
- The dynamic range ($>10^9$)
- No equivalent to the PCR reaction
 - Avogadro's number

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Avogadro's number comes back to haunt us: a reality check for proteomics

- One gram-mole of anything is 6.02×10^{23} molecules
- For mass spec detection, we need at least 100 femtomole (10^{-13} moles) or 6.02×10^{10} molecules

<u>protein expression per cell</u>	<u>need # cells for 100 fmoles</u>
10	6.02×10^9
100	6.02×10^8
1,000	6.02×10^7
10,000	6.02×10^6
100,000	6.02×10^5

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Central issue in proteome complexity: dynamic range

Cellular proteomes involve a very wide *dynamic range*: proteins can differ in their amounts by nine (9) orders of magnitude;

- In blood, albumin is 0.5×10^{-3} M, whereas cytokines are pM (10^{-12} M)
- A 2D gel that is overloaded with respect to an abundant protein, may have *barely detectable* amounts of a low abundance protein.....And if you can't see it, you don't know it's changed.

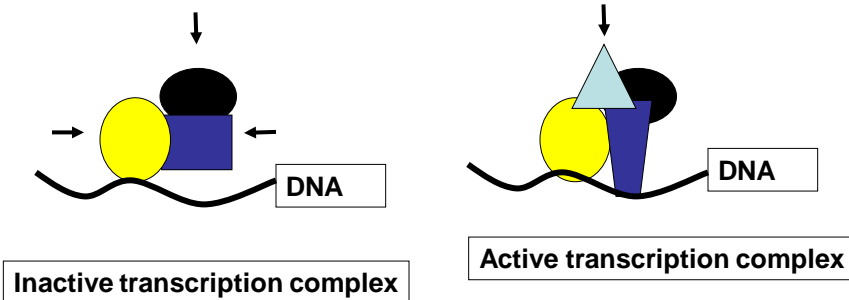
S Barnes/H Kim/M Renfrow BMG 744
1/06/12

The need to enrich for subproteomes, and/or isolate the lower abundant proteins

- See lecture by Helen Kim on January 18th in the course
- Biological properties:
 - Intracellular location; Protein-protein interactions; Posttranslational modifications
- Intrinsic properties:
 - Net charge; Size; Extent of tertiary structure; Hydrophobicity

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

A valuable role of the proteomics approach:
Discovery & analysis of protein-protein interactions



(The arrows indicate the proteins that could be antigens for immunoaffinity purification of the complex.)

See also the class by Matt Renfrow on February 17th

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Experimental design and quality control issues

- How do we carry out a proteomics experiment?
 - Randomize sample analysis
 - Process samples blinded to identities
 - Standardize procedures and vendors of disposable plastics used in experiment-- minimize variation where possible
 - Consult with statistician before experiment; ensure enough “power” for the experiment so that statistical analysis yields significant data.

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

Predicting the proteome

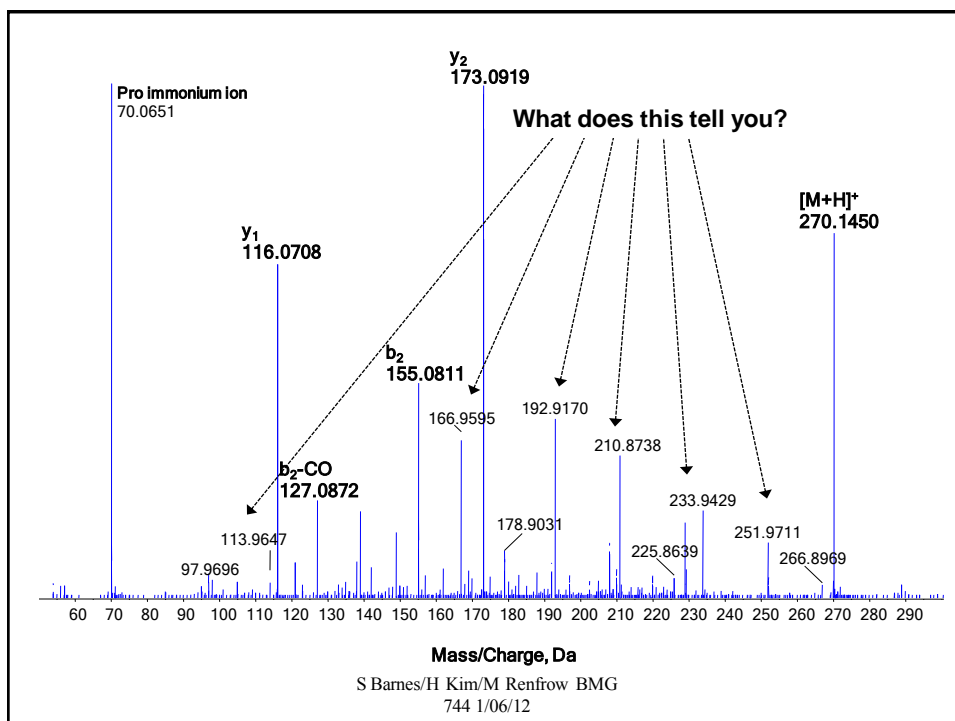
- **Bioinformatics** is the basis of high throughput proteome analysis using mass spectrometry. Protein sequences can be computationally predicted from the genome sequence
- However, **bioinformatics** is not able to predict with accuracy the sites or chemistry of posttranslational modifications - these 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/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12



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 February 20/22)**

S Barnes/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

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/H Kim/M Renfrow BMG 744
1/06/12

PROTIG and Videocast

- **There is an NIH-based proteomics interest group (PROTIG)**
 - <http://proteome.nih.gov/calendar.html>
- **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/H Kim/M Renfrow BMG 744
1/06/12