




## 약 력

### 1. 인적사항

	성 명	권 경 훈
	소속기관	한국기초과학지원연구원
	직 위	책임연구원
	전자메일	khoon@kbsi.re.kr

### 2. 학력/경력

연 도	학교 / 기관	전공 / 직위	학위 / 비고
1984	서울대학교	물리학	학사
1987	서울대학교 대학원	물리학	석사
1993	서울대학교 대학원	통계물리학	박사
1993	포항공과대학교 자연과학연구소	박사후 연구원	
1993 - 현재	한국기초과학지원연구원	책임연구원	
2005 - 현재	생물정보학회 Proteome Informatics 연구회	회장	

### 3. 주요연구실적(개조식, 간단하게)

- 프로테오믹스 실험 데이터의 분석 알고리즘 개발
- FT-ICR/MS 질량분석기 개발을 위한 분석 시스템 구축

### 4. 발표시 사용 기자재

\* LCD projector

# Data Analysis Methods for Quantitative Proteomics Research

한국기초과학지원연구원 질량분석기개발팀 권 경 훈

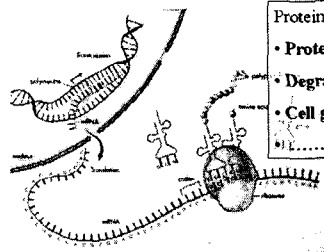
프로테오믹스는 생물체 안에 포함되어 있는 단백질을 통합적으로 연구하는 학문이다. 단백질을 동정(Protein identification)하고, 단백질의 상태를 분석(Protein characterization)하며, 단백질의 양적 변화를 관찰(Protein quantitation)한다. 유전자로부터 mRNA 로 복제되고 codon 의 규칙에 따라 합성되는 단백질이 세포 내에 얼마나 존재하는가라는 단백질의 양적인 변화는 세포 내의 환경에 따라 시시각각 변화할 수 있으며, 이러한 변화의 추적은 단백질의 기능을 밝히는 기초자료로서 중요성을 가진다. 특히 질병의 조기 진단을 위한 바이오마커를 발굴하기 위한 스크리닝 역할로서, 단백질의 발현 양상을 비교하는 프로테오믹스는 기대를 모으고 있다. 단백질에 대한 분석, 특히 질량분석기에 의해 초고속으로 대량의 단백질 데이터를 생산하는 프로테오믹스의 연구는 정량적인 단백질 발현양상 분석의 정확도를 높이기 위해 다양한 실험기법과 데이터 분석기법을 동원하고 있다.

이번 발표에서는 프로테오믹스에서 단백질의 양을 측정하기 위한 실험 방법들과 그에 따른 데이터 분석 방법들을 소개하고자 한다. 프로테오믹스 연구의 초창기부터 사용되어온 2차원 전기영동법에 의해 생성되는 2D-gel image 에서의 spot 분석법으로부터, 탄뎀 질량분석기를 사용하는 ICAT, iTRAQ 등의 labeling 방법에 의한 정량분석, 그리고 질량분석기의 정확도를 최대한으로 활용하는 label-free 방법에 대한 기본 개념을 살펴보고 데이터 분석 기술의 적용 방법을 알아본다.

## Data analysis methods for quantitative proteomics research

한국기초과학지원연구원  
질량분석기개발팀 권경훈

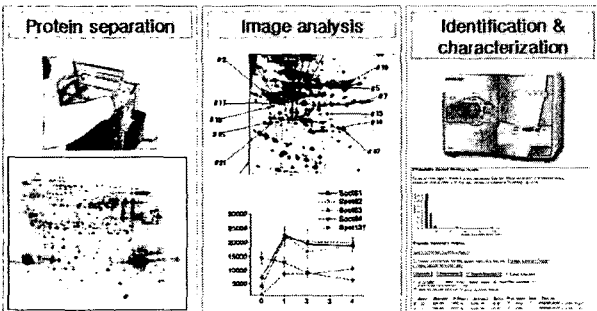
## Gene expression



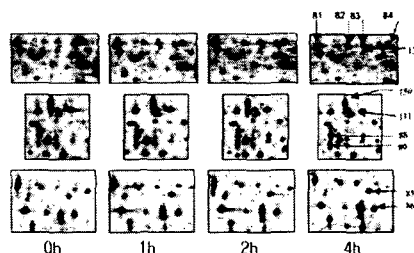
Protein amount depends on

- Protein synthesis rate
- Degradation rate
- Cell growth rate → protein dilution

## Proteomics / Quantitative proteomics

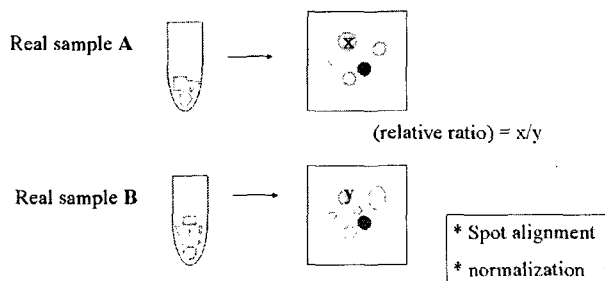


## 2D Electrophoresis Protein Expression Profile

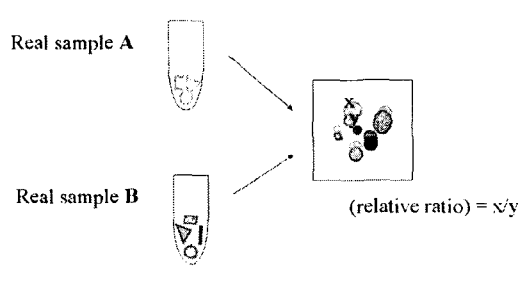


Nam, M.H., et al., Proteomics, 2003, 3, 2351-2367.

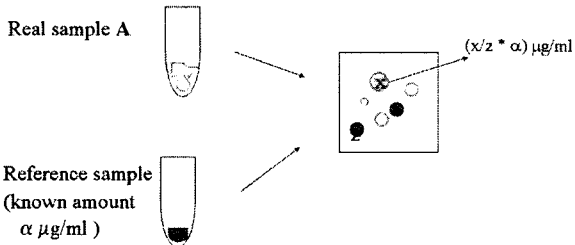
## Relative quantitation



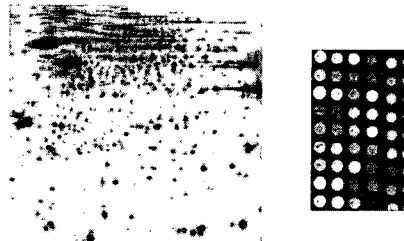
## Relative quantitation



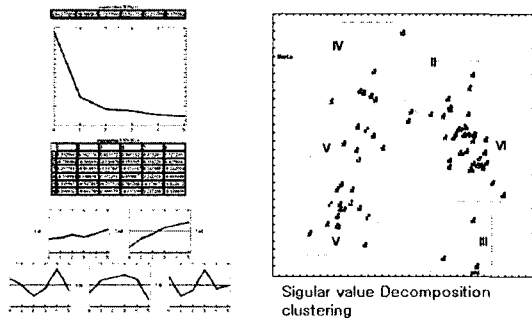
## Absolute quantitation



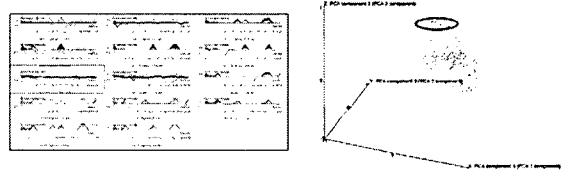
## 2D-gel spot vs. microarray



## Clustering 2D-gel spots

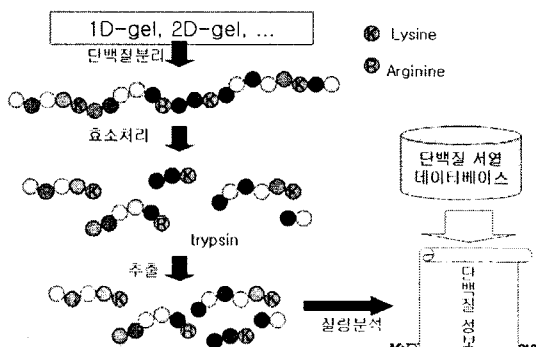


## Biomarker discovery by 2D-gel analysis

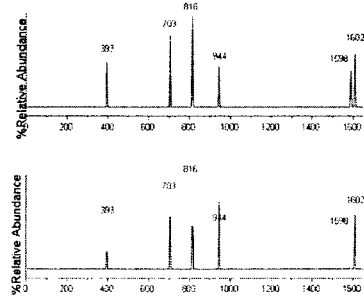


Ref) I.C. Gerling, S. Singh, N.I. Lenchik, D.R. Marshall, J. Wu, New Data-Analysis and Mining Approaches Identify Unique Proteome and Transcriptome Markers of Susceptibility to Autoimmune Diabetes, MCP (2005)

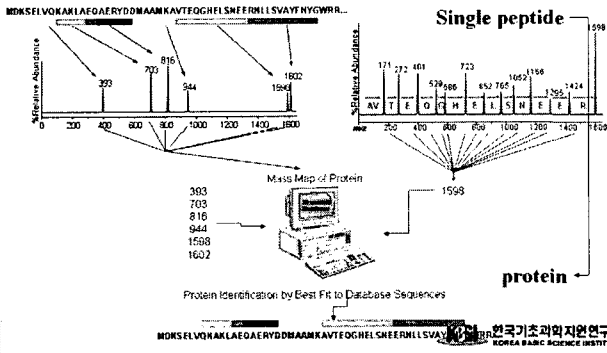
## Peptide Mass Fingerprinting



## Quantitation by PMF



### MS and MS/MS spectrum



### Protein Abundance Index (PAI)

$$PAI = \frac{\text{(the number of observed peptides per protein)}}{\text{(the number of observable peptides per protein)}}$$

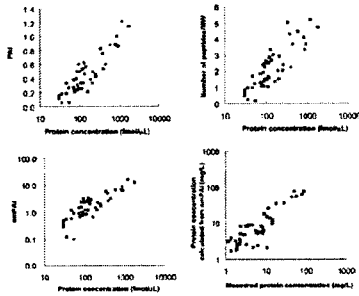
Exponentially modified PAI

$$emPAI = 10^{PAI} - 1$$

Protein content

$$(\text{mol } \%) = \frac{emPAI}{\sum (emPAI)} \times 100$$

### Evaluation of PAI

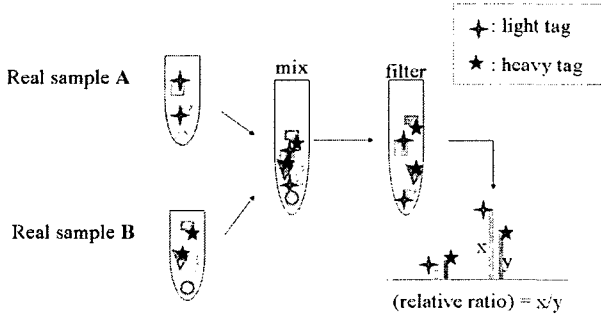


Y Ishihara et al., *Mol. Cell. Proteomics*, 2005, 4(9):1265-72.

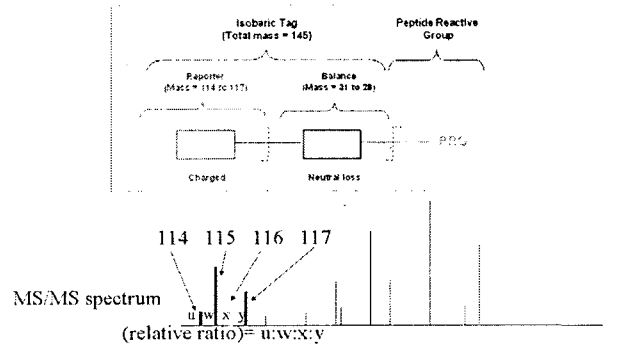
### Labeled Quantitation by mass spectrometry

- Isotope Coded Affinity Tag (ICAT)
- Stable isotope labeling with amino acids in cell culture (SILAC)
- iTRAQ

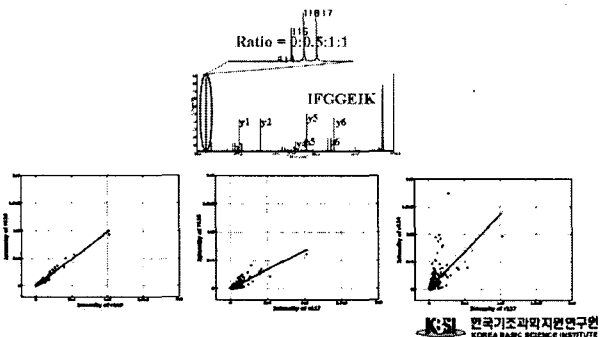
### Relative quantitation by tagging



### iTRAQ for 4 samples

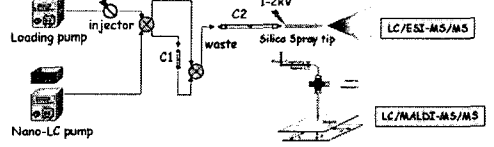


### iTRAQ data analysis

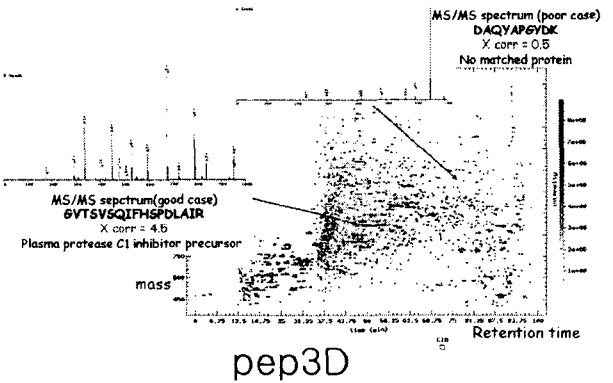


### ESI-MS analysis

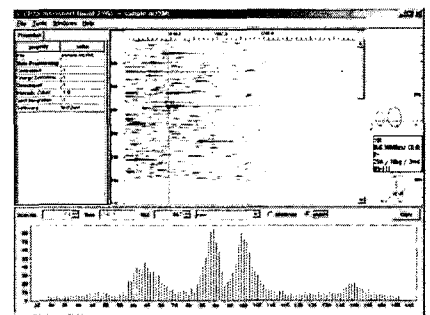
1. In gel or solution digested protein
2. Nanoflow ESI-LC/MS setting.



3. TIC, MS and MS/MS



### XIC (extracted ion chromatogram)

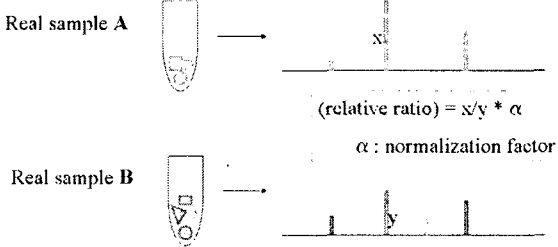


Snapshot of msInspect (from Fred Hutchinson Cancer Research Center)

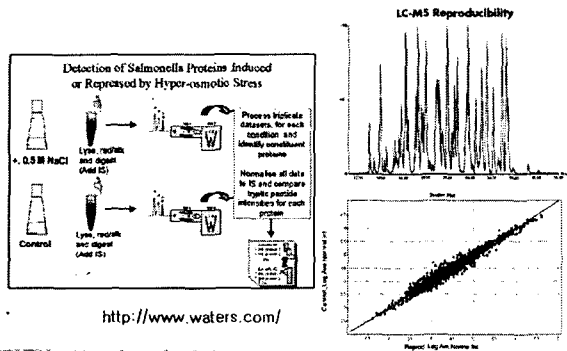
### LC/MS/MS quantitation tools

- ASAPRatio / XPRESS
- msQuant
- msInspect
- ...

### Relative quantitation from two experiments



## Label-free quantitation



<http://www.waters.com/>

## mRNA vs. Protein

### (References)

**Correlation between Protein and mRNA Abundance in Yeast**  
S P Gygi, Y Rochon, B R Franza, R Aebersold (1999) *Mol. Cell Biol.* (1999) 19: 1720

**Insights Into the Relation Between mRNA and Protein Expression Patterns**  
- A Mehra, K H Lee, V Hatzimanikatis (2003) *Biotech. and Bioeng.* 84: 823  
- P Lee, et al. (2003) *Biotech. and Bioeng.* 84: 834

**Comparing protein abundance and mRNA expression levels on a genomic scale.**  
D Greenbaum, C Colangelo, K Williams, M Gerstein (2003) *Genome Biol* 4: 117

## Discussion

- New high-throughput quantitative proteomics techniques continue to emerge.
- New quantification methods are accompanied with large-scale data analysis algorithms.
- The improvement of data reproducibility and accuracy at mass spectrometry enhances the reliability of quantitative analysis result.
- The integrated data analysis of the proteome, transcriptome and genome is expected to provide new information and knowledge in life science.

## Acknowledgement

- KBSI
  - 유종신 박사
  - 박영록 박사
  - 김진영 박사
  - 조 건 연구원
  - 박건욱 연구원
  - 이정화 연구원
- 고려대학교 유전체통계분석연구실
  - 이재원 교수
  - 이정복 박사