

**CLINICAL PROTEOMICS AND THE HUPO  
PLASMA PROTEOME PROJECT PILOT  
PHASE: TECHNOLOGY PLATFORMS,  
REFERENCE SPECIMENS, DATABASE,  
LANDMARK FINDINGS**

**AACC Annual Meeting EduTrak  
Los Angeles, 27 July, 2004**

**Gilbert S. Omenn  
University of Michigan, Ann Arbor, MI**

1

## SUMMARY OF LAB EFFORTS (cont'd)

Numbers reporting different kinds of technology platforms:

- 6 – 2D gels
- 17 – Liquid chromatography
- 2 – Peptide digest first
- 20 – MALDI or LC/MS, MS/MS
- 9 – SELDI
- 4 – Microarrays, labeled proteins

11

## Analytical Protocols by Lab

Lab #	Depletion	Separation	MS method
1	albumin+ig	2d chromat	maldi-ms/ms, esi-ms/ms
2	none	1d scx+1d rpc	esi-ms/ms
4	none	2d gel	maldi-ms
7	none	2d gel	maldi-ms
8	none, albumin+ig	2d gel, 2d chromat, 1d sax+1d rpc	maldi-ms, esi-ms/ms
11	other	1d rpc	esi-ms/ms
12	albumin+ig	2d chromat	esi-ms/ms
17	albumin+ig	1d rpc	esi-ms/ms
21	agilent	3d chromat	maldi-ms, esi-ms/ms
22	none, agilent	2d gel, 2d chromat, 3d chromat, ief+1d rpc	maldi-ms, esi-ms/ms
23	none	1d scx+1d rpc, 2d chromat, 2d gel	maldi-ms, esi-ms/ms
24	none, albumin only	2d chromat	esi-ms/ms
26	none	not specified, ief+1d rpc	maldi-ms, esi-ms/ms

14

## Analytical Protocols by Lab continued

Lab #	Depletion	Separation	MS method
28	ig only	1d scx+1d rpc, 2d chromat	fticr-ms
29	agilent	2d chromat	esi-ms/ms
33	agilent	2d chromat, ief+1d rpc	esi-ms/ms
34	agilent	not specified	esi-ms/ms
37	none, albumin+ig	2d gel	maldi-ms, esi-ms/ms
38	size exclusion	2d chromat	esi-ms/ms
40	albumin+ig	3d chromat	esi-ms/ms
41	size exclusion	2d chromat	esi-ms/ms
43	albumin+ig	not specified	maldi-ms/ms, esi-ms/ms
44	none	2d gel	maldi-ms
52	other	1d rpc	maldi-ms, esi-ms/ms

15

## Comparison of Identifications Made by Different Laboratories

### Challenge:

Different laboratories used different search databases and different strategies for choosing protein accession numbers when mass spec peaks search returned a group of equally good hits.

### Solutions:

Laboratories submitted sequences of identified peptides and now raw spectra. These sequences and spectra will be subjected to intensive cross-lab, cross-platform, and cross-specimen analyses.

## Five Standard Levels for Protein Identification

### Member of a gene family

- A peptide matching only in this gene family
- Often useful biologically information

### Gene product (transcription unit)

- Multiple peptide matches
- Complete genome sequence available
- Peptide matches that are diagnostic between paralogs

### Post translational modification

- Modifications defined by number, type and location
- Varying levels of precision (e.g. residue vs. peptide level assignment)

### Transcriptional/splice variant

- Multiple peptide matches including matches defining the N and C termini
- Peptide matches that are diagnostic between paralogs
- Peptide matches or molecular weight data diagnostic for splice isoform

### Complete covalent structure

- Covering set of peptide matches
- Covering set of MS/MS data on all peptides

35

### **TENTATIVE CONCLUSIONS**

**Approximately 10,000 different proteins found in IPI database reported: a massive database requiring fastidious quality assurance**

**Approximately 2000 distinct proteins detected by 2 or more labs, enhanced when high abundance proteins depleted: several thresholds pending**

**Some single-hit proteins have high probability/confidence estimates**

**Annotations being developed for 2500 proteins, including selected interaction maps**

**Lots of promising findings and inferences about biological insights, especially subproteomes (low MW, glycoproteins, tissue of origin) and microbial proteins**

**Possible identification of genes from proteins and other annotation for the Human Genome**

43