CANCER PROTEOMICS: CONSTRUCTING A KNOWLEDGE BASE FOR THE HUMAN PLASMA PROTEOME

AACR Conference on Advances in Proteomics in Cancer Research

Key Biscayne, FL, 8 October,2004

Gilbert S. Omenn

University of Michigan, Ann Arbor, MI

POTENTIAL CONTRIBUTIONS OF PROTEOMICS IN CANCER RESEARCH

- Biomarkers for
  - Early diagnosis
  - Heterogeneity of etiology
  - Prognosis
  - Monitoring treatment response
- Targets for prevention and for treatment

ADVANCES NEEDED

- Extensive identification of the plasma and serum proteome in health and disease
- Discovery of candidate biomarkers—
  individual proteins, panels, patterns
- Validation studies with sensitivity, specificity, and positive predictive value
- Much higher throughput of fractionation and global analysis of proteins
- Application to epidemiologic studies and treatment and prevention trials

CLINICAL PROTEOMICS: WHY STUDY PLASMA AND SERUM?

- Accessible sources of specimens
- Extensive specimen archives (banks)
- Dynamic reflection of organ functions in health and disease
- Feasibly linked to clinical data
- Base for human systems biology program

OVERALL SCIENTIFIC GOALS OF HUPO PPP

1. Comprehensive analysis of plasma and serum protein constituents in people

2. Identification of biological sources of variation within individuals over time, with validation of biomarkers

   Physiological: age, sex/menstrual cycle, exercise
   Pathological: selected diseases/special cohorts
   Pharmacological: common medications

3. Determination of the extent of variation across populations and within populations

STEPS PLANNING & LAUNCHING THE PPP

April 2002, Bethesda—organizing HUPO

Sept 2002, Ann Arbor: organizing PPP

November 2002: Versailles: mobilizing laboratories and technical committees

Jan 2003, Bethesda: FDA mtg on standardization

May 2003: Reference specimens distributed worldwide

July 2003: Workshop with 35 labs to establish protocols

Steps in PPP Pilot Phase (cont'd)

Oct 11-12, 2003, Montreal: Initial data and plans for special projects

Dec-Jan 2004: Proposals/grants for Special Projects

April 22, 2004: PPP Roadmap Conference in Bethesda for NIH and corporate sponsors

June 1-4, 2004: JAMBOREE WORKSHOP

Oct 23-24, 2004: Beijing Congress

AIMS FOR PILOT PHASE

1. Compare a broad range of technology platforms for the characterization of proteins in human plasma and serum. Assess resolution and sensitivity, plus time, cost, and volumes of sample required, using reference specimens.

2. Clarify influence of various technical variables in specimen collection, handling, and storage, especially plasma vs. serum and choice of anticoagulant.

3. Determine whether the most abundant plasma proteins should be depleted, and whether anti-protease cocktails are necessary or desirable.

AIMS FOR PILOT PHASE (cont'd)

4. Develop a database structure and repository for HUPO PPP results

5. Lay groundwork--through evaluation of technology platforms and specimen handling and through established international collaborations--for large-scale studies of plasma or serum biomarkers in health and disease across major ethnic groups: a Plasma Proteome Roadmap.

These aims drive our working groups and publication plans.

Scheme Showing Aims and Linkages of the HUPO Plasma Proteome Project

HUPO HUMAN PLASMA PROTEOME PROJECT (PPP)

Serum vs Plasma

Technology Platforms-- Separation and Identification

Reference Specimens

Development & Validation of Biomarkers

HUPO PPP Participating Labs

Technology Vendors

Liver and Brain Proteome Projects

Omenn GS. The Human Proteome Organization Plasma Proteome Project Pilot Phase: Reference Specimens, Technology Platform Comparisons, and Standardized Data Submissions and Analyses. Proteomics 2004;4 (in press)

SERUM AND PLASMA REFERENCE SPECIMENS

1. BD: specially prepared male/female pooled samples, divided into EDTA-, Heparin-, and Citrate-anti-coagulated Plasma and Serum (250 ul x4 of each). No clot activator. No protease inhibitors. Three separate ethnic pools prepared. Shipped frozen.

2. Chinese Academy of Medical Sciences: Sets of three plasmas + serum, similar to BD protocol.

3. National Institute for Biological Standards & Control, UK: citrate-anti-coagulated, freeze-dried plasma, from 25 donors, prepared for Intl Soc Thrombosis & Hemostasis, 1 ml aliquots/ampoules.

UPDATED SUMMARY OF PPP LABS

31 Total Participating Labs (18 US, 13-International):

9 – US Academic

3 – US Federal

6– US Corporate

4 – Europe, 1– Israel, 6 – Asia, 2 – Australia

Number that analyzed various reference specimens:

9 – UK NIBSC

26 – BD b1, Caucasian-American

9 – BD b2/b3, African- and Asian-American

5 – CAMS

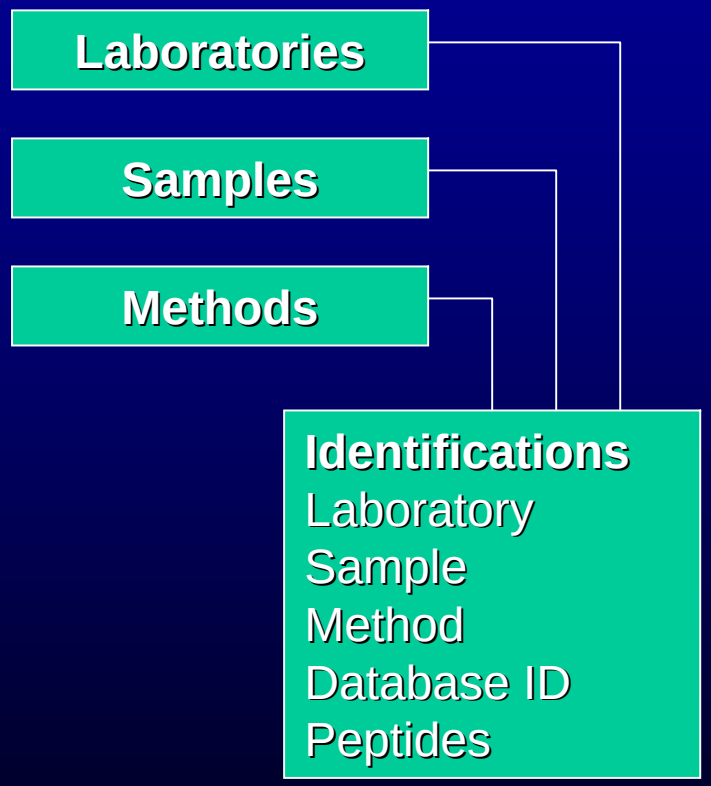
Database Design and Implementation

RDBMS

- Stable, proven technology
- Data validation

Commercial package

- Microsoft SQL Server
- Stable and supported
- Full RDBMS functionality
  - Transactions
  - Referential integrity checks
- Effective development tools
  - GUI
  - Cross-tab extension



Bioinformatics Issues in Data Gathering

Multiple formats permitted for submissions

- XML
  - Pedro data definitions
  - Only 2 labs used XML
- Web based submission of Excel templates
  - Multiple versions as the project evolved
- Email submission of Excel templates
  - Initial suggestion
  - Technical concerns, but did not really complicate matters

Identifications based on a variety of databases and variety of protocols

Some revised submissions; Core had to clarify whether new data or revised data.

SUMMARY OF LAB PLATFORMS (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

Analytical Protocols by Lab (MS; MS/MS)

Lab #	Depletion	Protein/Peptides Separation	MS method
1	albumin+Ig	none; rp/scx/rp	esi-ms/ms-deca-xp
2	none	CHO; scx/rp	esi-ms/ms-qTof
4	none	2d gel	maldi-ms
7	none	2d gel	maldi-ms
8	none, albumin+Ig	2d chr; sax/rp	maldi-ms
11	none	CHO; rp	esi-ms/ms-deca-xp
12	albumin+Ig	none; rp/scx/rp	esi-ms/ms-deca
17	albumin+Ig	1d sds; rp	esi-ms/ms
21	Top-6	rotofor-ief, rp,1d-sds (IPAS);rp	maldi-ms, esi-ms/ms
22	none;Top-6	2d gel, ief+1d rpc, or 3d chr	maldi-ms, esi-ms/ms
23	None	2d gel, 2d chr; scx/ rp	maldi-ms
24	none, alb only	rp; rp	esi-ms/ms
26	none	rotorof ief+ 1d sds; rp	esi-ms/ms

Analytical Protocols by Lab

Lab #	Depletion	Protein/peptides Separation	MS method
28	Ig only	none; scx/rp	esi-fticr-ms
29	Top-6	none; scx/rp	esi-ms/ms
33	Top-6	FFE + rp; rp	esi-ms/ms
34	Top-6	Zoom-IEF,1d-sds	esi-ms/ms, lcq vs ltq
37	none, albumin+Ig	2d gel	maldi-ms, esi-ms/ms
40	albumin+Ig	Rp; scx/rp	esi-ms/ms
41	size exclusion	Gradiflow chr; scx/rp	esi-ms/ms
43	albumin+Ig	none; rp	maldi-ms/ms, esi-ms/ms
44	None	2d gel	maldi-ms
46	Top-6	none; scx/rp	esi-ms/ms
55	None	sax	esi-ms/ms-ltq

SPECIFICATIONS FOR DATA SUBMISSION

Each lab was instructed to provide

a) a detailed experimental protocol

b) peptide sequences, rated as “high” or lower (“low”) confidence, based on MS/MS criteria

c) protein IDs from IPI 2.21 (July 2003) and search engine used to align peptide sequences with proteins in human database

Later, we requested m/z peak lists and raw spectra (by CD or DVD); search parameters

Quality Assurance and Validation

Can we confirm the protein ID in the human protein databases?

- In nearly all cases, yes.
- Remainder may be data entry typographic errors
- Had to convert some datasets to IPI 2.21

Confirm that the database-identified protein contains precisely the peptides used to identify it

Concordance between labs and specimens

Are the assignment & confidence criteria equivalent?

Did identical specimens yield the same proteins?

How good is agreement across all specimens?

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, later, raw spectra. These sequences and spectra are being subjected to intensive cross-lab, cross-platform, and cross-specimen analyses.

Five Levels for Protein Identification

Member of a gene family

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

Gene product (transcription unit)

- 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, incl 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

CRITERIA FOR "HITS"

1. Peptide sequence with threshold score or threshold probability (Sequest, Mascot, other)
2. Protein ID in two labs

   or ID in one lab with compelling features
3. Save one-peptide hits in database separate from "confirmed" hits
4. Respect special features of certain methods (low MW, glycoproteins)
5. Report results with a series of cutpoints

CRITERIA FOR HIGH CONFIDENCE IDENTIFICATION OF PEPTIDES, ILLUSTRATED WITH SEQUEST

Xcorr: singly-charged ion, >=1.9

doubly-charged ion, >=2.2

triply-charged ion, >=3.75

Delta Cn >= 0.1; Rsp <= 4

Fully tryptic

Threshold for low confidence:

Xcorr >= 1.5, 2.0, 2.5

Protein Integration Workflow: Rationale

Following mass spectrometry of tryptic digests of proteins, peptide m/z peak lists and peptide sequences were matched to protein databases with search engines.

Often, searches returned a cluster of proteins, each containing the same matching peptides. Or peak list matches returned 2 or more equally good protein hits (different accession numbers).

Participating laboratories used different versions of several protein databases to perform the searches. Labs applied different strategies for choosing representative protein accession numbers from a cluster of equivalent hits. Result: multiple accession numbers reported for identifications based on the same peptide sequences.

Integration Algorithm (cont'd)

Further steps were thus needed to:

- o Integrate results from such disparate sources
- o Evaluate concordance between results from different laboratories
- o Reduce ambiguity and redundancy of the identifications
- o Select accession numbers of the most likely present protein.

In order to more uniformly select a representative protein accession number, we designed a workflow that uses sequences of identified peptides, rather than submitted protein accession numbers.

(1) Assembly of peptide sequence lists: **Protein identifications submitted by the labs were accompanied by lists of sequences of matching peptides. All lists were pooled to build a set of distinct peptide sequence lists. Each sequence preserves all references to its origin - e.g. if a particular list was identified in more than one specimen by one lab, it has more than one reference.**

(2) Search of the peptide lists: **Each of the peptide sequences next was searched against the database selected as the standard database of the project - IPI version 2.21 (July 2003). We required 100% identity between sequences.**

(3) Selection of representative protein from each cluster of equivalent protein hits: **Often, more than one entry in the reference protein database matches all of the components of a peptide sequence and peak list. We call this set of matching entries a "cluster of equivalent protein hits" for that peptide sequence. Clusters for different lists may overlap. When they do, we choose one protein entry from the intersection of several clusters to represent all of the proteins in each of the overlapping clusters, as follows:**

Protein Integration Workflow (cont'd)

Each protein entry in the reference database receives three scores:

(a) The number of distinct laboratories **reporting peptide sequence lists whose clusters include this protein**

(b) The number of distinct experiments **(number of laboratories × specimens × protocols) reporting these peptide sequences**

(c) The number of identifications **(number of laboratories × specimens × protocols × clusters), for clusters including this protein.**

The cluster member with the largest value of score (a) is chosen as its representative protein entry. Scores (b) and (c), followed by criteria (d - g), serve to break numeric ties.

(d) Well described protein - product of a well described gene in Ensembl: proteins and genes with not-empty description line, no words like 'fragment', 'similar to', 'putative' in their description.

(e) Well described protein - product of any gene.

(f) Well described protein not assigned to any gene.

(g) Protein described as a fragment, by similarity to another protein, or not described at all.

Protein Integration Workflow (cont'd)

RESULT: The same protein is generally chosen as the representative entry from several overlapping clusters of equivalent protein IDs.

This simplifies the process of comparing the protein identifications reported from different laboratories and different experiments.

Justification: The representative protein ID supported by the largest number of independent experiments is the protein most likely to be actually present in the specimen. Score (a) counts each laboratory only once, no matter from how many specimens or with how many different peptide sequence lists the laboratory identified this protein.

Protein Integration Workflow - cont.

Effectiveness of the integration: 12,388 submitted accession numbers were replaced by 9,506 uniformly chosen IPI entries

category	number of IPI entries matching a peptide sequence list		
	one	**>1**	one + >1
submitted peptide sequence lists	**12,303**	**5,795**	**18,098**
submitted protein accession numbers	**7,000**	**5,388**	**12,388**
matching entries in IPI database	**6,601**	**9,668**	**15,710**
matching entries in IPI database after the integration	**6,601**	**3,273**	**9,506**
average number of peptide sequence lists per IPI entry after the integration	**1.9**	**1.8**	**1.9**
reduction from level of submitted accession numbers to IPI entries	6%	39%	23%

Numbers of Proteins Identified (LC-MS/MS or FTICR-MS)

After integration algorithm restricted IDs to those in IPI 2.21, eliminated duplications, and chose representative protein for clusters:

(a) all protein IDs (high and lower conf)

9506 = 1 or more peptide matches

[5946 in serum; 6378 in plasma specimens]

3020 = 2 or more peptide matches

1325 = 2+ peptides, same specimen/same lab

505 = 2+ different peptides in 2+ diff labs

[474 in serum; 428 in plasma]

Numbers of Proteins Identified (LC-MS/MS or FTICR-MS)

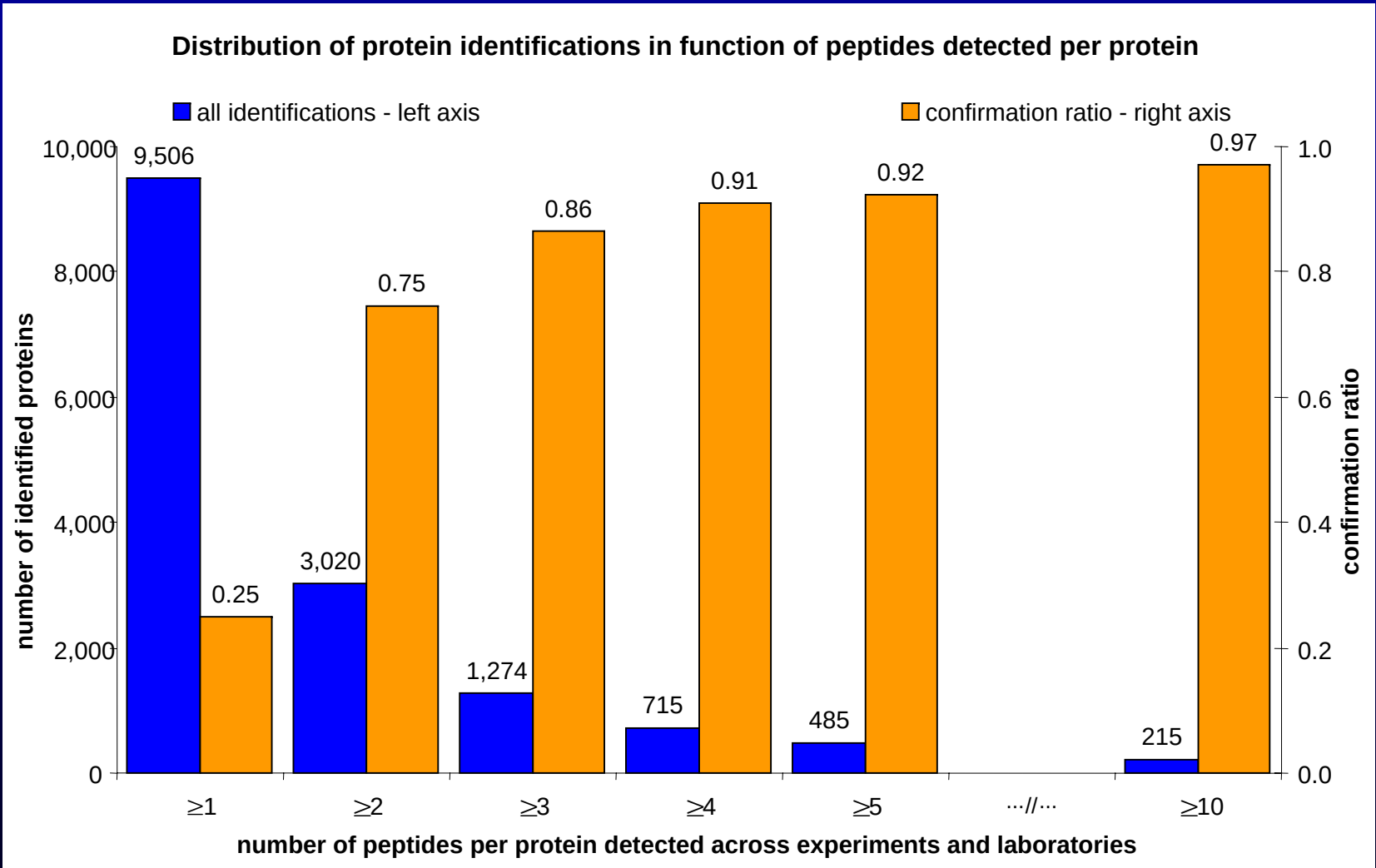
(b) all protein IDs (high conf peptides, only)

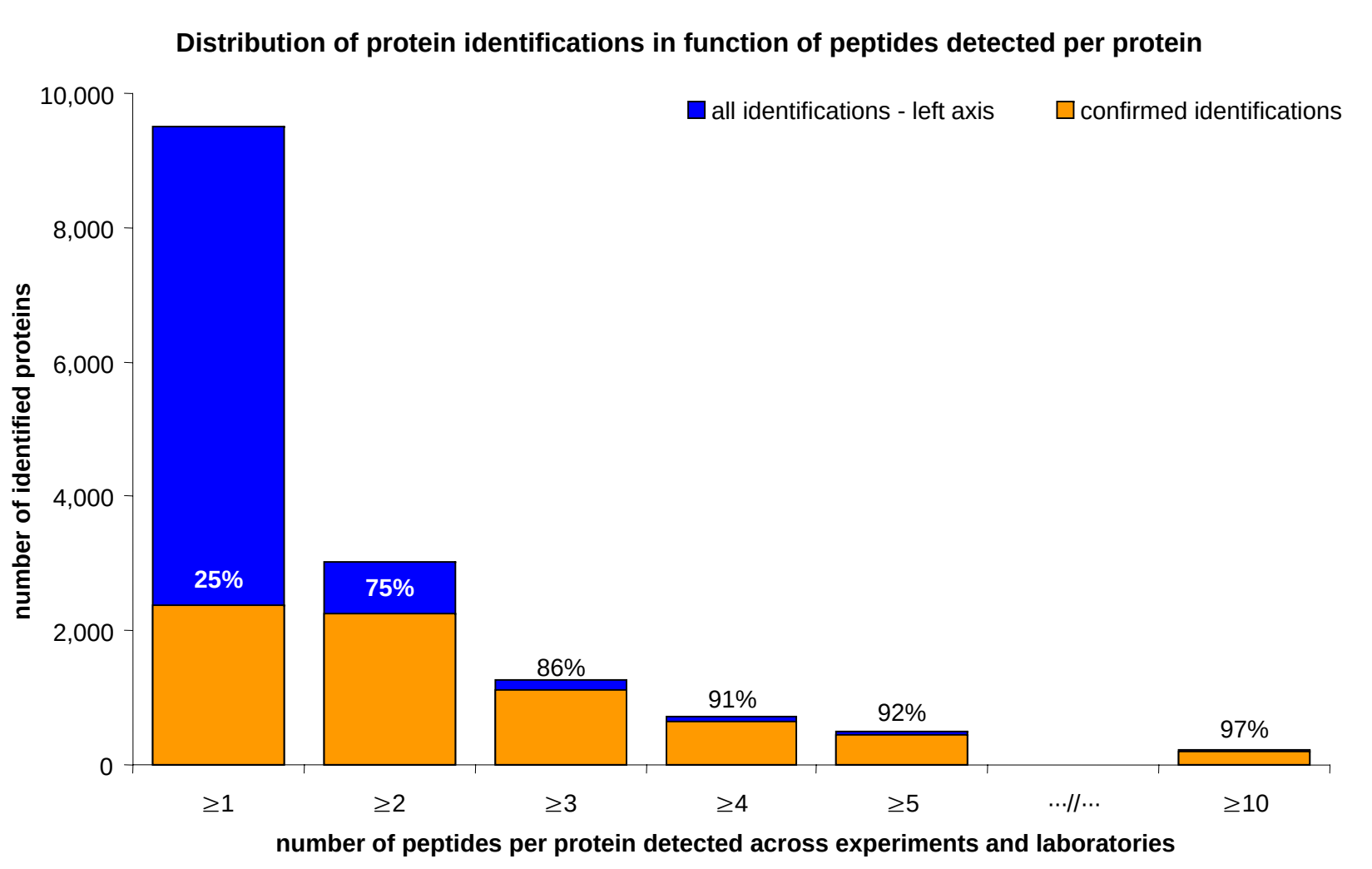
2852 = 1 or more peptide matches

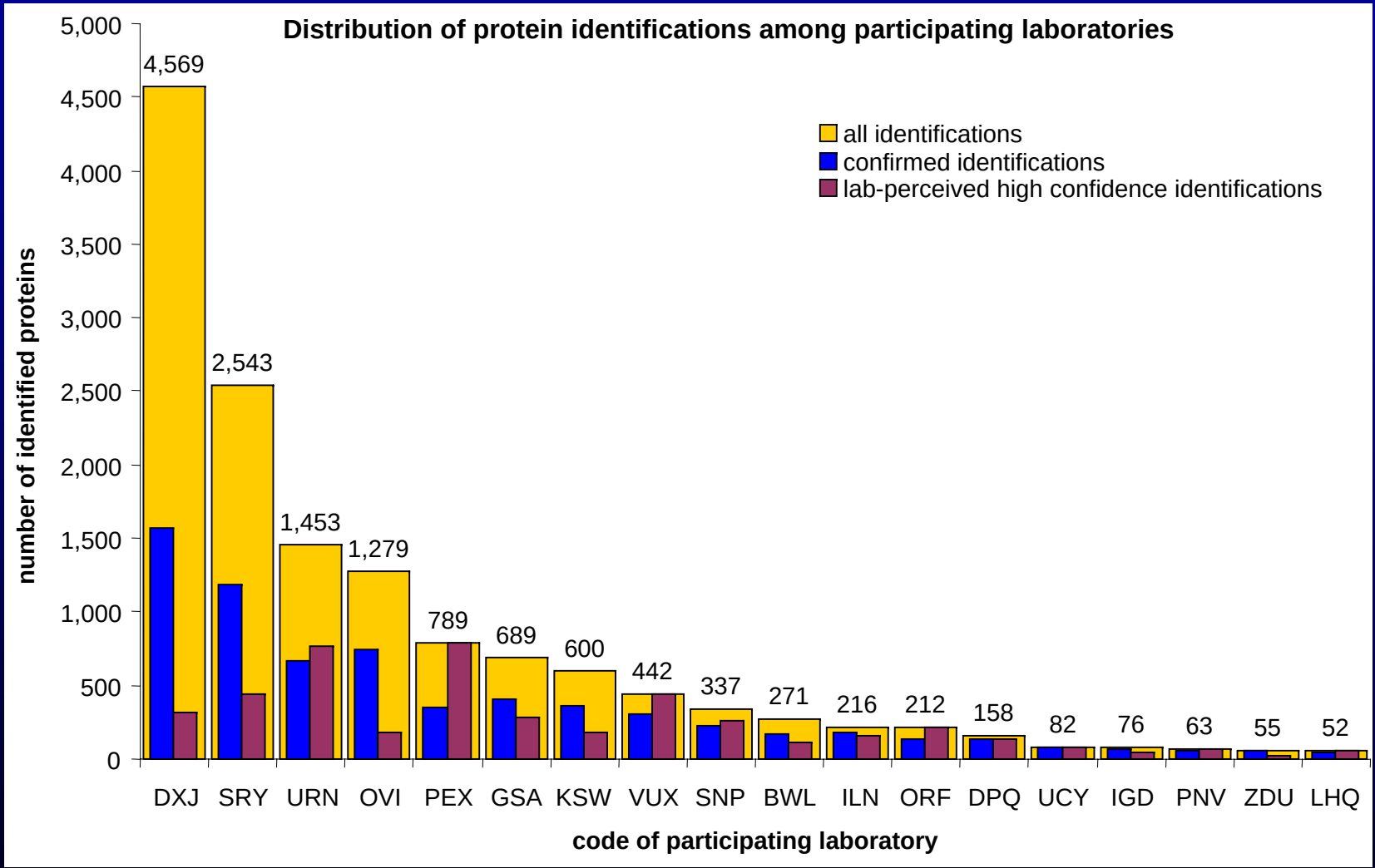
794 = 2+ peptides, same specimen/same lab

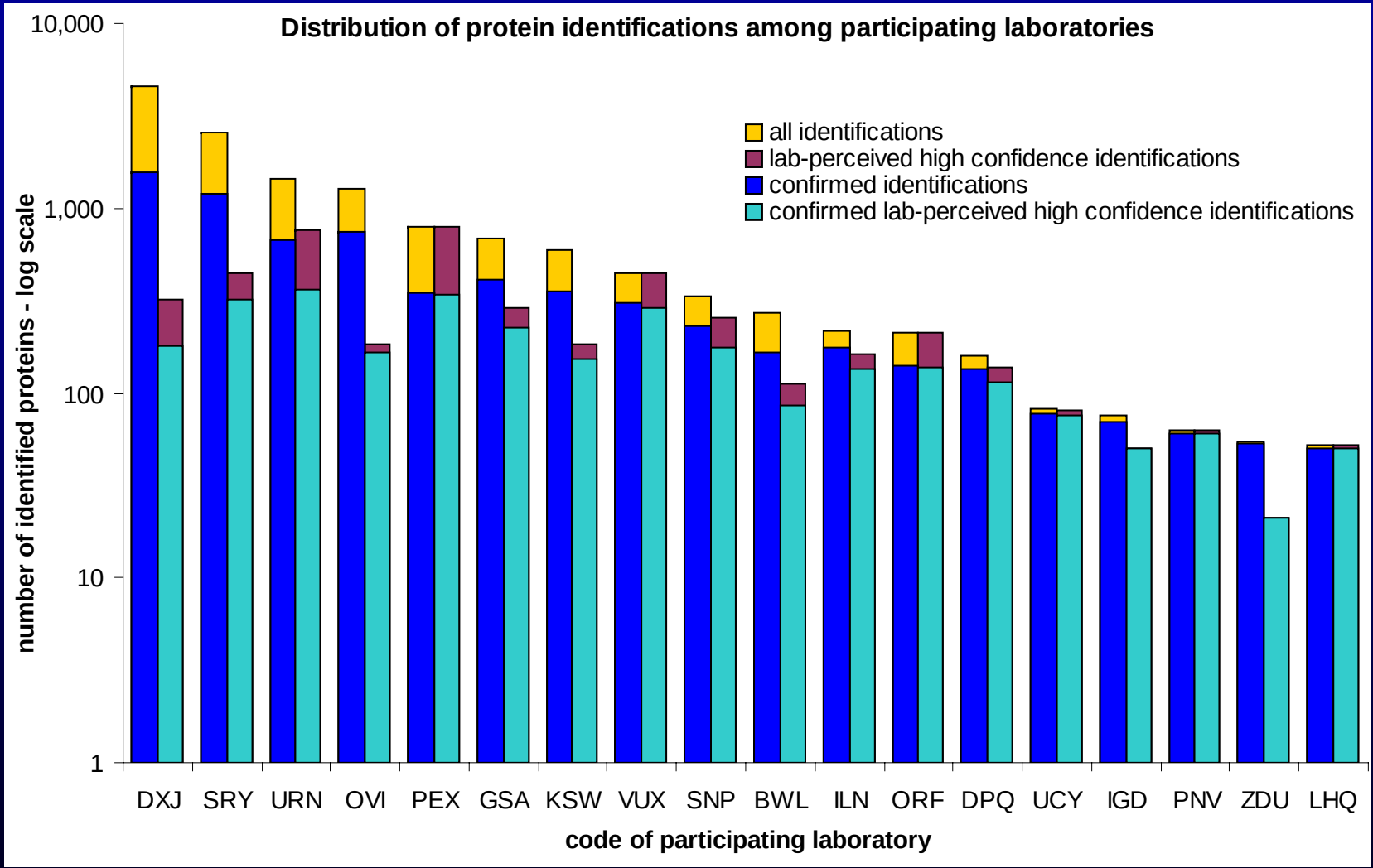
259 = 2+ different peptides in 2+ diff labs

(405 if confirming peptide does not have to meet stringent high-conf criteria)

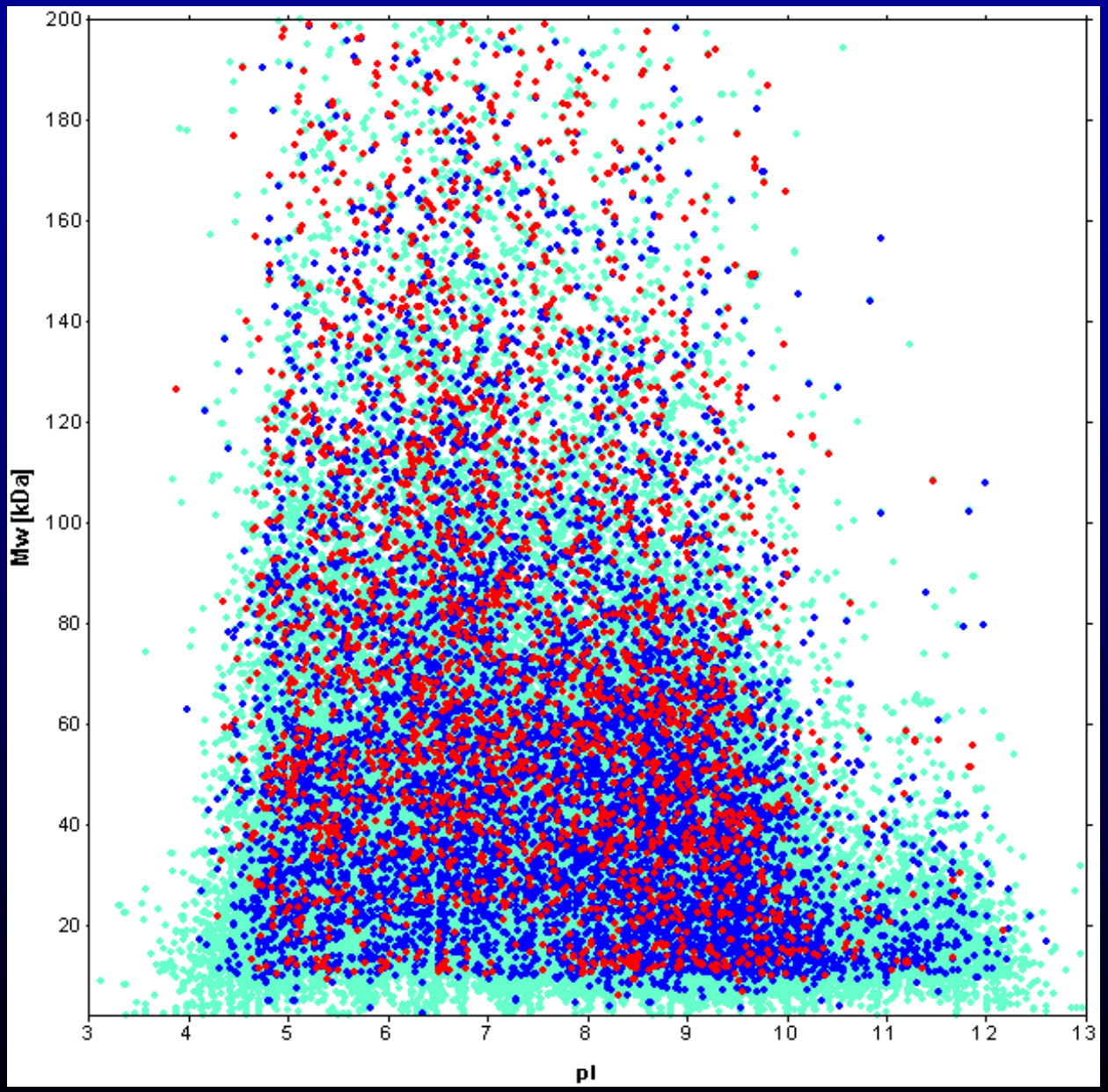

Distribution of protein identifications in function of peptides detected per protein
all identifications - left axis
confirmation ratio - right axis
number of identified proteins
confirmation ratio
10,000
8,000
6,000
4,000
2,000
0
1.0
0.8
0.6
0.4
0.2
0.0
9,506
0.25
3,020
0.75
1,274
0.86
715
0.91
485
0.92
215
0.97
≥1
≥2
≥3
≥4
≥5
···//···
≥10
number of peptides per protein detected across experiments and laboratories


Distribution of protein identifications in function of peptides detected per protein
all identifications - left axis
confirmed identifications
number of identified proteins
10,000
8,000
6,000
4,000
2,000
0
25%
75%
86%
91%
92%
97%
≥1
≥2
≥3
≥4
≥5
…//…
≥10
number of peptides per protein detected across experiments and laboratories


Distribution of protein identifications among participating laboratories
number of identified proteins
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
all identifications
confirmed identifications
lab-perceived high confidence identifications
4,569
2,543
1,453
1,279
789
689
600
442
337
271
216
212
158
82
76
63
55
52
DXJ
SRY
URN
OVI
PEX
GSA
KSW
VUX
SNP
BWL
ILN
ORF
DPQ
UCY
IGD
PNV
ZDU
LHQ
code of participating laboratory


Distribution of protein identifications among participating laboratories
all identifications
lab-perceived high confidence identifications
confirmed identifications
confirmed lab-perceived high confidence identifications
number of identified proteins - log scale
10,000
1,000
100
10
1
DXJ
SRY
URN
OVI
PEX
GSA
KSW
VUX
SNP
BWL
ILN
ORF
DPQ
UCY
IGD
PNV
ZDU
LHQ
code of participating laboratory

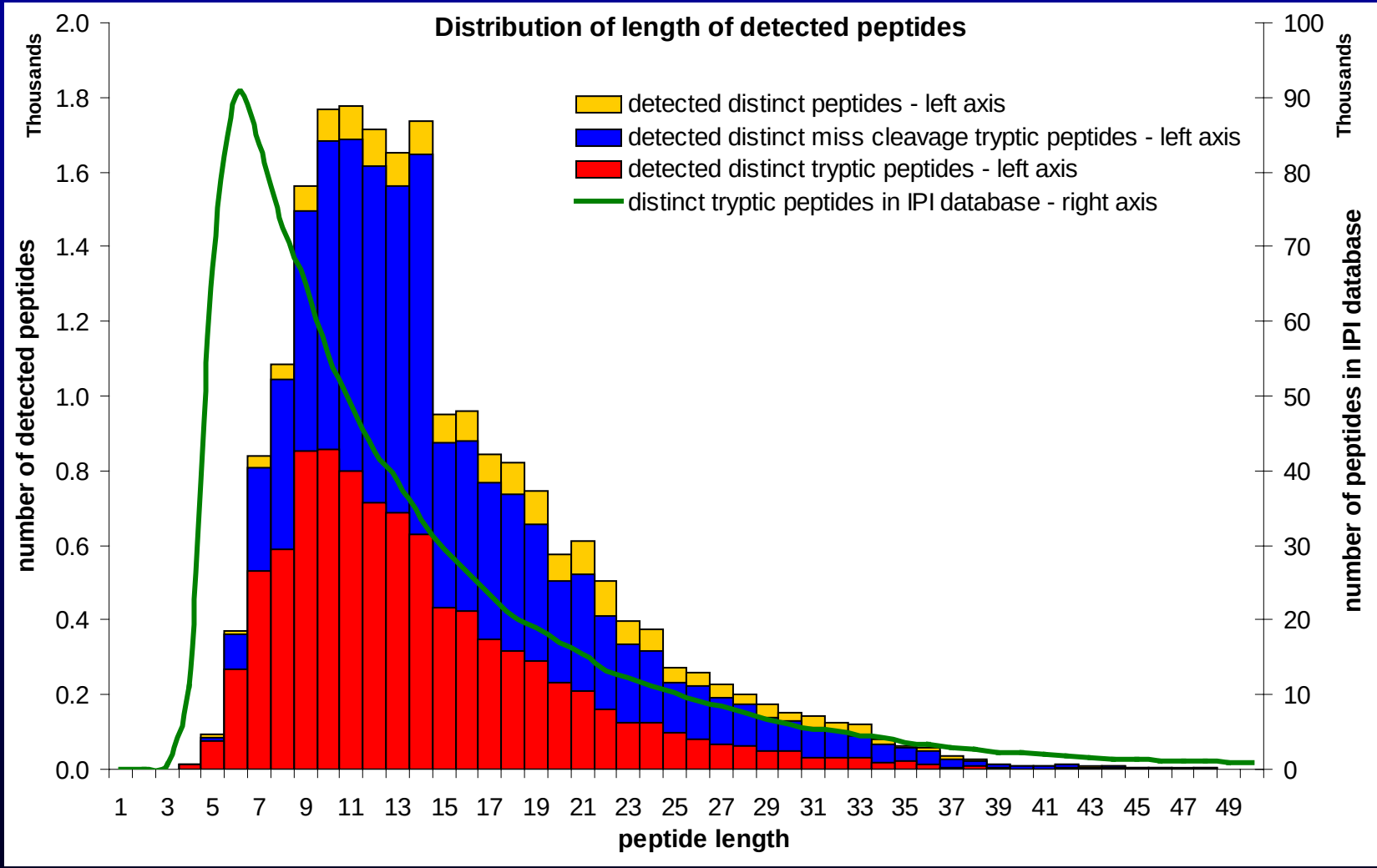
Virtual 2D gel



Proteins detected with at least 2 peptides

All Detected Proteins

All proteins in IPI 2.21


Distribution of length of detected peptides
detected distinct peptides - left axis
detected distinct miss cleavage tryptic peptides - left axis
detected distinct tryptic peptides - left axis
distinct tryptic peptides in IPI database - right axis
number of detected peptides
Thousands
number of peptides in IPI database
Thousands
peptide length
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
100
90
80
70
60
50
40
30
20
10
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

Summary

Database reasonably represents the submitted data

- RDBMS built, tested and used extensively within the group
- Quality assurance checks show few discrepancies or problems
- Overall self consistent set of submissions

Composition of human plasma/serum proteome

- 9506 distinct proteins identified
- 3020 protein identifications confirmed by more than one laboratory or more than one specimen
- With more stringent criteria for high-confidence peptides, 2872 proteins identified

Number of confirmed identifications adds significantly to the published literature

Origins of lab to lab differences being investigated further: depletion, fractionation, MS methods, intensity of effort

SPECIFIC OBSERVATIONS: GREATEST RESOLUTION AND SENSITIVITY

The most extensive high-confidence yield was from combined use of immunoaffinity ("top-6") depletion, 2 or 3-D high-resolution fractionation, and then ESI-MS/MS with ion-trap LTQ instrument. LTQ gave several fold more IDs than did LCQ in same hands (B1-serum vs B1-heparin).

SPECIFIC OBSERVATIONS: SPECIMENS

- The sets of four specimens from a given donor pool yielded rather similar numbers of proteins when analyzed identically. Further analyses of the non-overlapping lower abundance proteins pending.
- Quantitative immunoassays do show generally lower values for citrate-plasma, but this may be due to a dilution effect in the preparation, not interference with or loss of identifiable proteins.

SPECIFIC OBSERVATIONS: DEPLETION

Many investigators depleted albumin and/or immunoglobulins, and several obtained access early via HUPO PPP to the Agilent immunoaffinity column to remove albumin, IgG, IgA, haptoglobin, transferrin, and alpha-1 antitrypsin (top-6). There were much higher numbers of identifications after depletion. Inadvertent removal of other proteins may be more extensive (with LC-MS/MS) than was detected by gels.

SPECIFIC OBSERVATIONS: PROTEASES

There is a choice to be made whether or not to use anti-protease cocktails in primary specimen collection, or in a later step.

Advantages: reduce proteolytic degradation ex vivo; reduce complexity of peptides after tryptic digestion.

Disadvantages: adds peptides, as well as small molecules, to the mix to be analyzed; may modify the proteins (ABERF shown to do so).

INDEPENDENT ANALYSES FROM SPECTRA

Multiple (but not all) datasets submitted by the various participating laboratories are being independently analyzed from raw spectra, using (a) PeptideProphet/ ProteinProphet (Eng, USA); (b) Digger and other search engines (Kapp, Australia); and (c) PepMiner (Beer, Israel). These results will be compared with the primary analyses presented here from the submitting labs plus the University of Michigan Bioinformatics Core.

SPECIFIC OBSERVATIONS: INFLUENCE OF ABUNDANCE

In general, we presume a relationship between concentration of a protein and the number of matching peptide sequences identified (modified, of course, by number of tryptic sites, number of isoforms, and many other factors).

Using quantitative immunoassays and microarrays (generally unknown epitopes), we have found very high rates of detection of the more abundant proteins, less in the mid-range, and occasional detection of very low abundance proteins.

CALIBRATION OF DYNAMIC RANGE
in HUPO PPP Reference Specimens

Quantitative Assays of 33 Selected Proteins from 40 mg/ml to about 40 ng/ml (Vitzthum/Ackermann, Dade Behring)

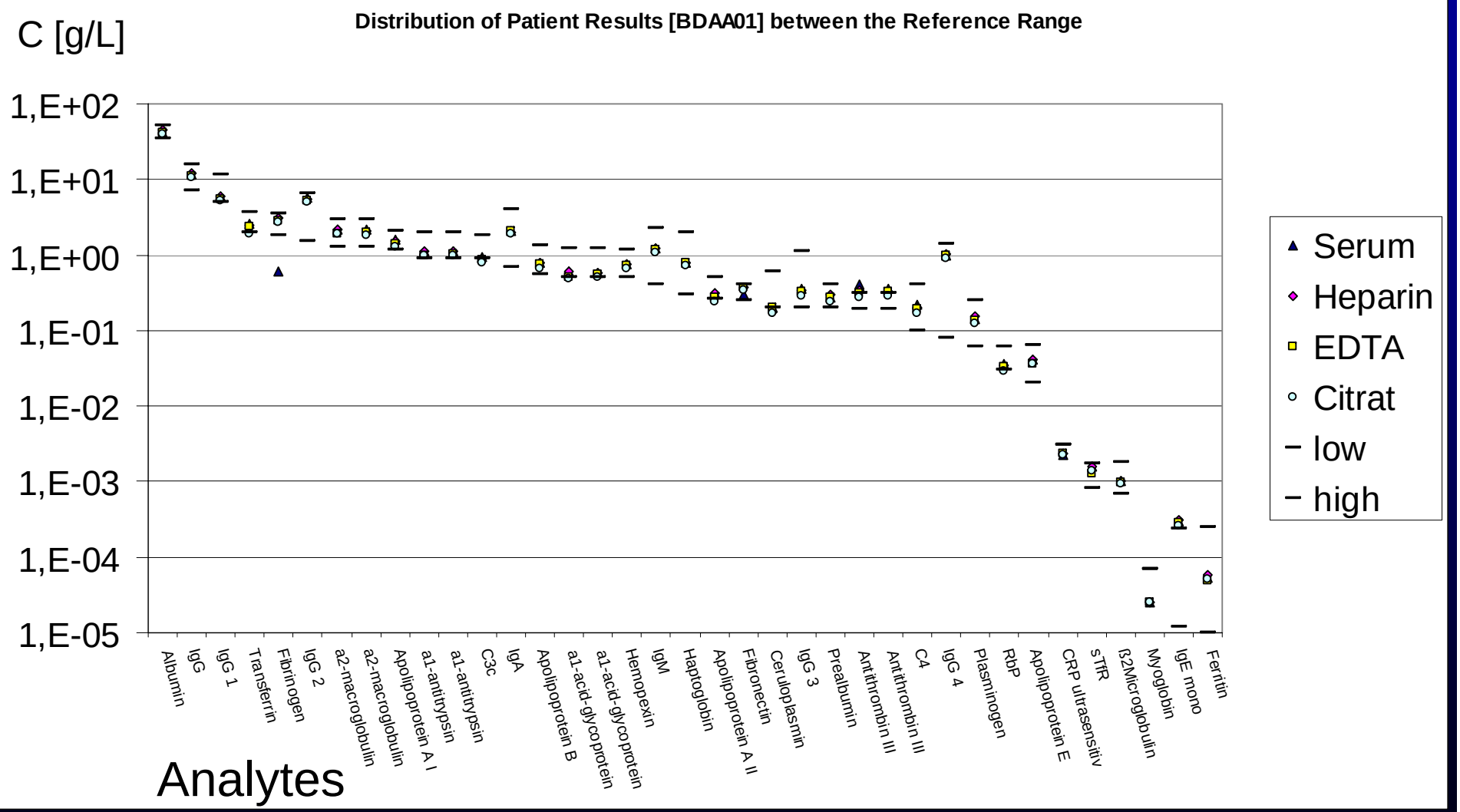
8 analytes from ng to pg/ml Chan/Hopkins)

20+ analytes, broad range (Hefta, BMS)

141 analytes, Molecular Staging (Tschernev)

142 analytes, Novartis Genomics Institute (Geierstanger)

Protein Concentrations – HUPO PPP Specimen BDAA01



SPECIFIC OBSERVATIONS: ORIGIN OF THE PROTEINS

- The vast majority of circulating proteins readily detected arise in the liver. That is certainly true for most of the traditional "plasma proteins".
- Other sites of origin can be deduced from matching to Novartis Atlas or published gene expression datasets by organ.
- Platelets, white blood cells, and red blood cells may be sources of proteins detected in plasma or serum. Further work is needed on this point.

SANBI Gene Expression Data

665 IPI proteins from our 3020 proteins matched with the proteins found in blood; 392 of them were matched with various blood cell types.

They are categorized as follows:

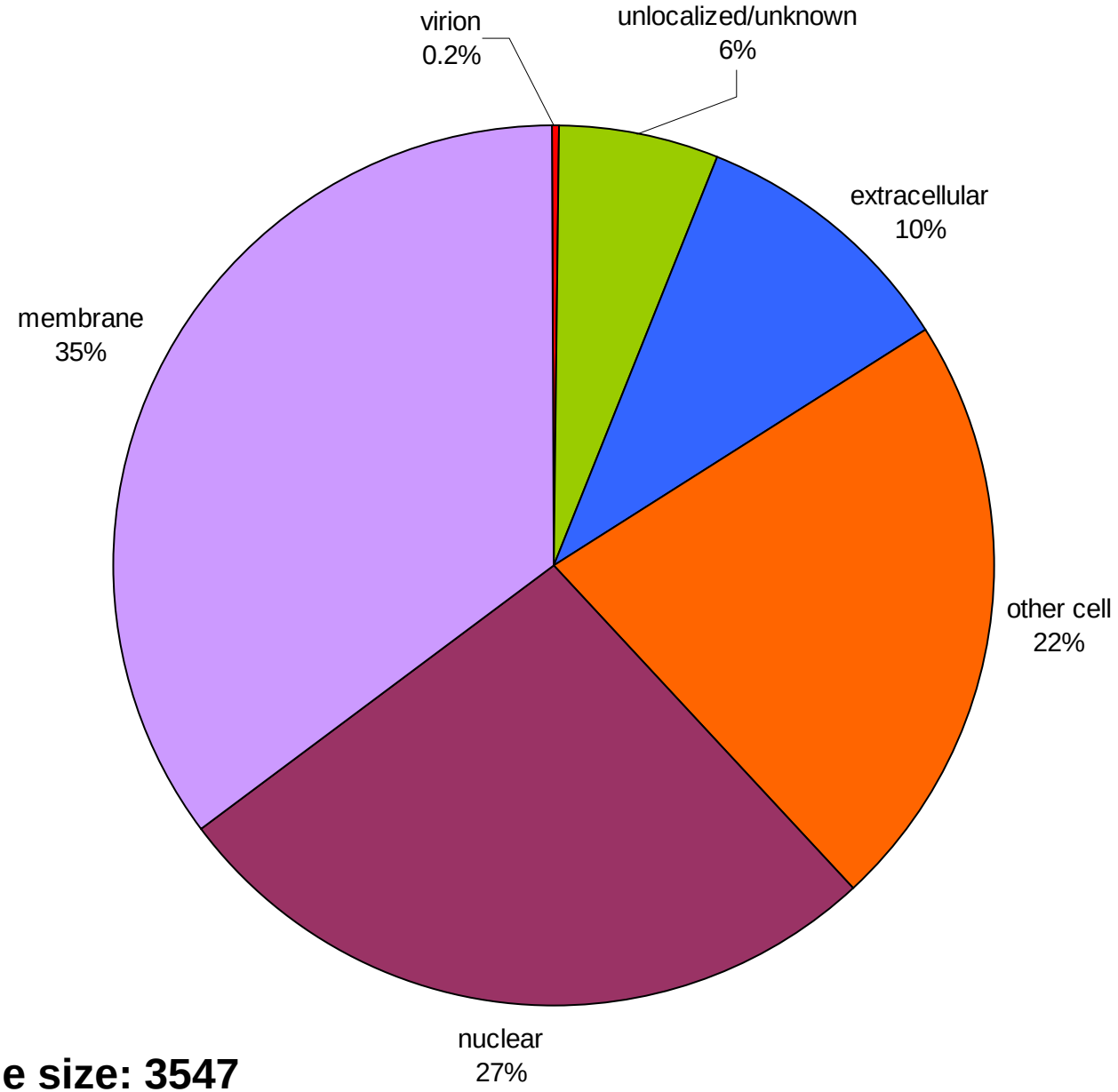
leukocyte	49
Leukocyte/lymphocyte	115
leukocyte -> lymphocyte -> natural killer cell	83
leukocyte -> lymphocyte -> T-lymphocyte	12
leukocyte -> macrophage	1
monocyte	1
myeloid cell	113
proerythroblast	114
promonocyte	1
promyeloblast	1
stem cell	6

No data for platelets or megakaryocytes

BIOLOGICAL INSIGHTS

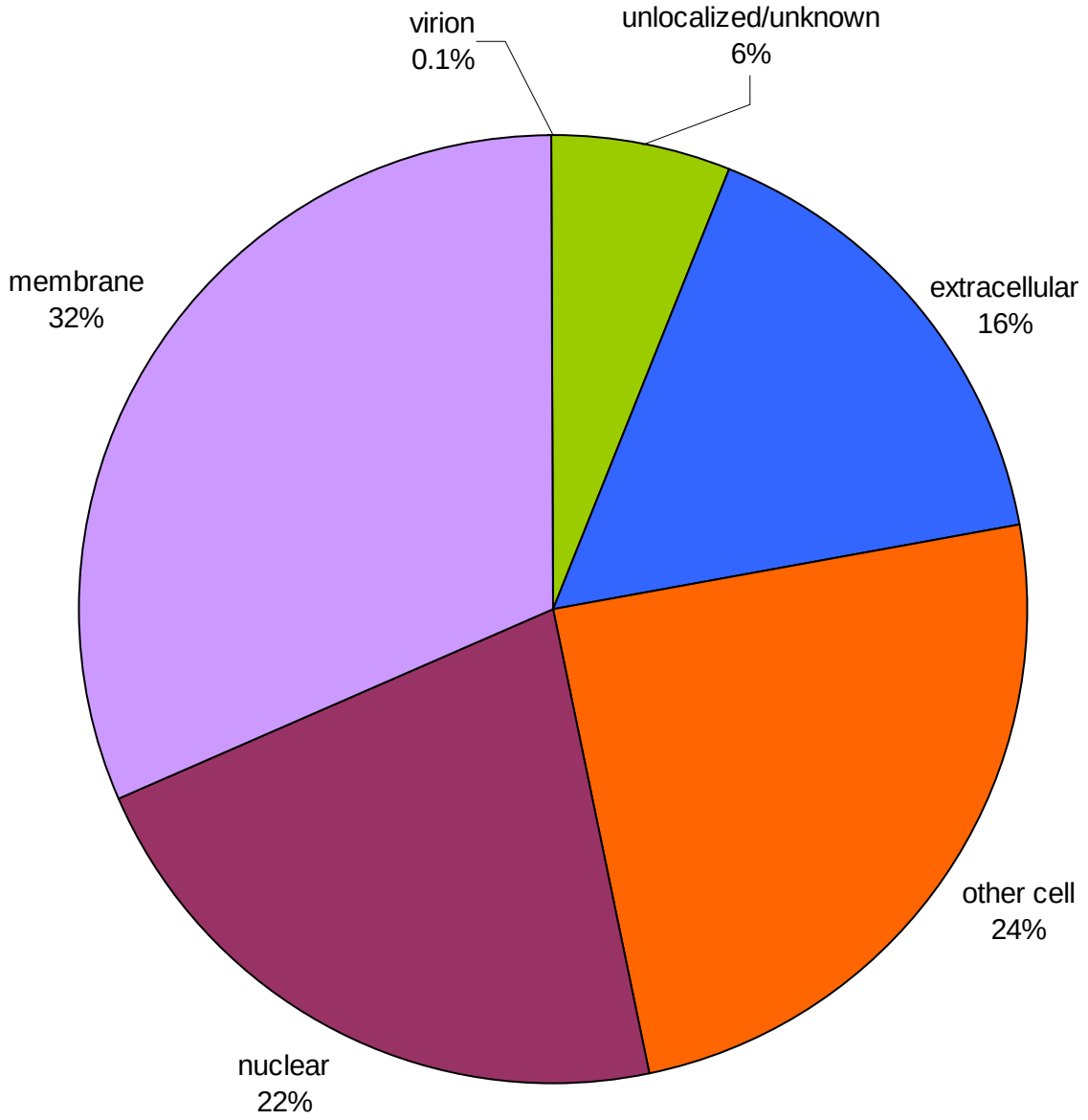
The proteins identified can be annotated by many methods. We are pursuing searches against multiple databases, including Gene Ontology, Novartis Atlas, OMIM, incomplete or unidentified sequences in the human genome, microbial genomes, protein domains, etc. Some examples follow.

GO Cellular Component Classification
All Identified Proteins



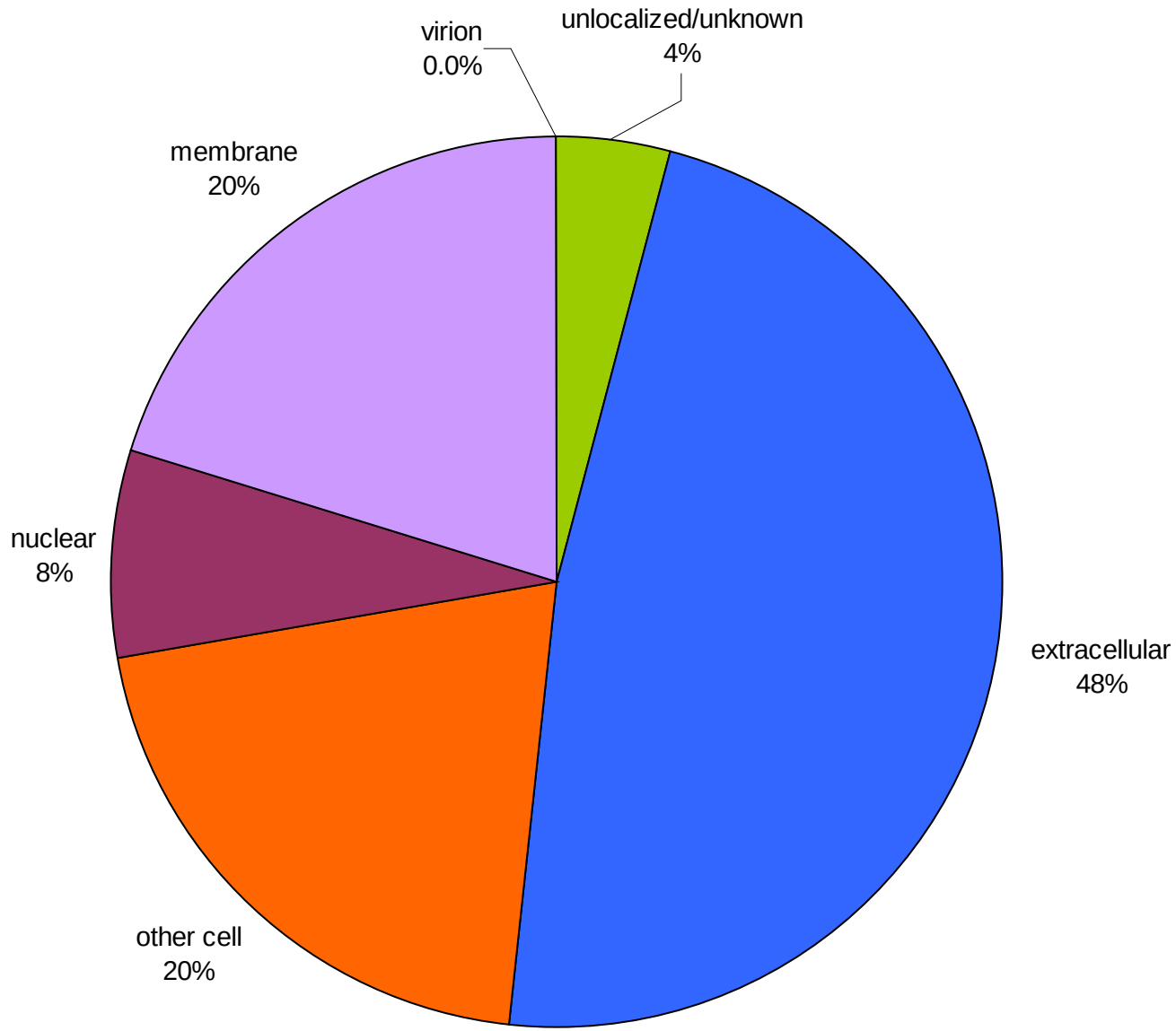
Pie size: 3547

GO Cellular Component Classification
Proteins Confirmed by at least 2 labs



Pie size: 1094

GO Cellular Component Classification
Proteins Confirmed by at least 4 labs



Pie size: 172

PROTEINS FROM INHERITED CANCER DISORDERS
Linking IPI IDs and Mendelian Inheritance in Man (OMIM)

IPI	Cancer Types	Protein	Labs	No of Peptides
IPI00012391.1	Colorectal	APC	2	2
IPI00017303.1	Colorectal , NHPCC; Ovarian	DNA mismatch repair protein Msh2	2	2
IPI00020732.2	Medullary or papillary thyroid	Tyrosine kinase ret receptor precursor	2	3
IPI00025087.1	Colorectal	Cellular tumor antigen p53	1	3
IPI00031036.1	Colon	Chloride anion exchanger	2	4
IPI00164713.1	Breast, Endometrial, Gastric, Ov	Epithelial-cadherin precursor	2	4
IPI00181932.1	Prostate	Zinc phosphodiesterase	2	5
IPI00185027.1	Pancreatic	Arg-Glu dipeptide (RE) repeats	2	2
IPI00218982.1	Breast , Ovarian	BRCA1	3	5
IPI00257731.1	Prostate	N33 protein.	2	2
IPI00289819.1	Hepatocellular	Cation-ind mannose-6-P receptor precursor	2	3
IPI00293471.1	Breast 2, Pancreatic	BRCA2	4	8
IPI00294982.1	Breast	Estrogen receptor	2	2
IPI00329643.1	Endometrial	DNA mismatch repair protein Msh3	3	3

OUTCOMES: HUPO HUMAN PLASMA PROTEOME-2004

1. **STANDARD OPERATING PROCEDURES from workgroups**
   - Specimens, depletion and fractionation, technology platforms, database structure, curation, access, quality assurance
   - SOPs will be directly useful for follow-on HUPO PPP population cohort studies, NIH & other grant proposals

2. **PUBLICATIONS**
   - Landmark PPP publication
   - Overview and database development papers
   - Papers from each workgroup above
   - Papers from individual labs with sufficient data and variables

3. **PRESENTATIONS**
   - Beijing 3rd HUPO World Congress on Proteomics, 23-24, 24-27 October 2004

PPP WORKGROUPS & PAPERS

A. SPECIMENS

1. Stability
2. Protease inhibition A1 + A2 (combined)
3. Immunoassays A3 + C3 (input to C5)

B. BIOINFORMATICS

1. Search engine performance B1 + C1a
2. Database performance B2
3. Annotation B3-4-5 combined Ms
4. Biological insights
5. Subproteomes

C. PROTEIN ID RESULTS

1. MS/MS: a) Kapp b) Eng c) Beer
2. Direct MS/SELDI
3. Antibody Arrays/ Immunoassays
4. Depletion & fractionation
5. Comparison of protein IDs across specimens, platforms, labs

One or more papers from each

D. COST-EFFECTIVENESS ANALYSIS D

TOTAL about 12 manuscripts from workgroups

Bioinformatics Acknowledgements

University of Michigan

David States

Marcin Adamski

Thomas Blackwell

Yin Xu

Rajasree Menon

EBI - England

Henning Hermjakob

Chris Taylor

Ludwig Institute - Australia

Richard Simpson

James Eddes

Eugene Kapp

Institute for Systems Biology

Jimmy Eng

Technion/IBM, Ilan Beer

PPP TECHNICAL COMMITTEE STRUCTURE

- Reference Specimens and Specimen Handling Issues (Dan Chan, chair)
- Technology Platforms & Protocols (Richard Simpson)
- Database Development and Links with EBI (HUPO/PSI) (Henning Hermjakob)
- Population Cohorts/Specimen Banks (Gerard Siest)
- Education & Training Committee (Peipei Ping)
- Executive Committee (including Partnerships) (Omenn)

HUPO PPP SUPPORT FROM NIH

Trans-NIH Consortium

Natl Cancer Inst: Div CancerPrevention;
Div Cancer Treatment

Natl Institute on Aging

Natl Inst on Alcoholism & Alcohol Abuse

Natl Inst on Diabetes, Digestive, & Kidney Diseases

Natl Inst for Environmental Health Science

Natl Inst for Neurologic Diseases & Stroke

CORPORATE SPONSORS

Johnson & Johnson	Abbott Labs
Pfizer	Novartis
Invitrogen	Amersham
Procter & Gamble	
BD Biosciences	BioVision
Ciphergen	Molecular Staging
Bristol Myers Squibb	Sigma-Aldrich
Agilent	Dade Behring

HUPO HUMAN PLASMA PROTEOME PROJECT (PPP)

Serum vs Plasma

Technology Platforms-- Separation and Identification

Reference Specimens

HUPO PPP Participating Labs

Development & Validation of Biomarkers

Technology Vendors

Links with Other HUPO Initiatives