

**THE HUPO PLASMA PROTEOME  
PROJECT PILOT PHASE:  
MONTREAL WORKING MEETING:  
AIMS & OUTCOMES/NEXT STEPS**

**HUPO World Congress Satellite  
11-12 October, 2003**

**Gilbert S. Omenn, M.D., Ph.D.  
University of Michigan**

# **AIMS & ACTION AT WORKING MEETING**

- 1. Report on overall progress on the PPP Pilot Phase (slides following)**
- 2. Update and evaluate the Data Submission Process**  
**Please visit intra-PPP website for finalized David States presentation if you are a registered lab; through 10/7, 10 labs submitted results with >120 distinct proteins identified**
- 3. Hear and discuss early laboratory results with PPP reference specimens (10 presentations: listed on slides # )**
- 4. Consider design requirements for cross-lab, cross-platform, cross-specimen analyses (8 topical groups: # )**
- 5. Discuss funds, criteria, and review process for internal PPP funding of special projects (both lab and data mining) (#**
- 6. Plan next activities: subgroup protocols; population cohort studies; June 1-4, 2004, “Jamboree workshop”; Beijing HUPO Congress satellite 23-24 October, 2004; extensive publication plans**

# **TIMETABLE FOR CORE ANALYSES**

**Initial submission: 9/15/03**

**Complete submission to be included in PPP  
analyses: 12/31/03**

**Feedback from Bioinformatics Units, re-  
checking, possible further experiments,  
before 5/15/04**

**Special projects: organize now, complete  
by 5/15/04**

**Jamboree Workshop: 1-4 June '04**

# PPP LABS MAKING PRESENTATIONS

1. **Speicher/Wistar: Sigma, Agilent depletion options, Zoom, narrow 2D gels or 1D gels to nLC-MS/MS: extensive identification**
2. **Paik/Yonsei-Korea: depletion to 1D, tryptic digestion, LC-MS/MS, Sequest, or 2D + lightning-fast MALDI: extensive ID, little overlap between 2 methods**
3. **Qian/Beijing: Agilent depletion, digestion, SCX with stepped gradient, RP, nESI-MS/MS: extensive IDs**
4. **Pounds/PNNL: LC-MS/MS and FTICR-MS after depletion of IgG, triplicate assays, multiple HUPO reference specimens: very extensive IDs**

5. Chan & Rai/Hopkins: multiple ref specimens, depletions, SELDI-TOF, many peaks
  6. Admon & Ziv/Haifa: dye and protein depletion, 2D LC, MS DEXA XP, PepMiner; many IDs
  7. Simpson & Moritz/Ludwig-Melbourne: Octopus FFE, RP or Agilent depletion, LC-MS/MS, Chomper Data Interrogation Tool; early IDs
  8. Leung/Singapore: SELDI surfaces, IgY depletion, IDs pending
  9. Tsugita/Nobatani: protease inhibitor cocktail, 2DE, MALDI-TOF; + Montage or BioRad depletion kits; early IDs across b1 specimens
  10. Peltier/Prolexys: Chromolith RP-18e column; early IDs
  11. Chromy/LLNL Calif: Amersham, ABI, Agilent depletion options; 2D DIGE & SELDI; large # spots
- Forsberg/Amersham-Harvard-Partners, Locke/Gradipore, Veenstra/NCI, others presented at Congress itself

# PPP PRELIMINARY ANALYSIS

**Bioinformatics Core cross-checks all protein IDs and unique IDs reported from each lab; complexities from multiple formats (XML, Excel, email) and varied databases, even versions of IPI; also multiple reference specimens; outstanding range of technologies, in combinations**

**8232 perfect IPI matches for unique peptides; >2000 distinct protein IDs with serum so far (other specimens not as extensively analyzed yet), only about 10% confirmed in a 2<sup>nd</sup> lab at this point.**

**Many cross-analyses illustrated.**

# **DATA ANALYSIS IMPRESSIONS**

**Relational database reasonably represents the submitted data, good QC, self-consistent sets of submissions**

**Total of >3300 distinct proteins from 1943 genes provisionally identified in at least one lab, 10% confirmed in 1+ more labs**

**Consider lab to lab differences at this juncture, some due to families of genes and proteins**

# HUPO PPP INTERNAL GRANTS

**AIM:** To maximize the scientific value of the Pilot Phase by:

a) **Conducting additional protocol-based experimental lab studies, beyond the initial characterization of Reference Specimens with each lab's established and emerging technological platforms, with maximal identification of proteins**

b) **Proposing and conducting analyses of the PPP database across specimens, platforms, and laboratories, with support from EBI and UM bioinformatics.**

# **TOPICS FROM TECHNOLOGY COMMITTEE: with Bioinformatics support; to generate SOPs**

- 1. Specimen comparisons/reliable dataset (Pounds)**
- 2. Stability protocols (Rai; Simpson; Vitzthum; Siest; BD, Wu)**
- 3. Database and search engine QA: software and algorithm comparisons (Fenyó, Kapp, Beer, Pandey, Eng)**
- 4. Software and repository development (Taylor, Hermjakob, States, Eddes, Kamijo)**
- 5. DB performance (States, Eddes, Henning)**
- 6. Comparison of results within sets of technologies (Speicher, Simpson, Paik, Beer)**
- 7. Comparison of results across types of technologies (Simpson, Wang, Tsugita)**
- 8. SELDI users group (Rai, Stemmer, Adam, Leung)**
- 9. Cost-effectiveness analyses of types and uses of technologies (Pisano, Apffel)**

# **Examples of Types of Special Studies (Omenn from general discussion)**

- 1. Specimen Handling Protocol: need clear aims and endpoints, and multiple platforms**
- 2. Depletion, pre-fractionation: dynamic range obtained with various technologies**
- 3. Protease inhibition**
- 4. Protein identification criteria/confidence; algorithms; DBs**
- 5. Protein abundance estimation**
- 6. Insight into biological significance of proteins detected**

# Process for Grants

**Eligible:** all labs, groups of labs, committees

**Topics:** preceding 2 slides illustrative

**Funds:** expect \$250K total; propose \$20k to 40K,  
expect average about \$25K

**Timeline:** Notice here and by email this week

Submit by 12/1 ( $\leq 4$  pp, incl. budget/just'n

Peer Review (committee members) and decision by  
Exec Comm by 1/5/04

**Work period:** Complete by 5/15/04 for Jamboree 6/1-  
4/04, publication submission by 7/15/04,  
presentations in Beijing 10/23-24/04.

# **FUTURE POPULATION COHORT STUDIES**

**Gerard Siest and Sophie Visvikus have drafted a brief questionnaire for investigators interested in linking their population cohorts and/or specimen banks with the HUPO Plasma Proteome Project and PPP laboratories.**

**This draft will be shared with other members of the PPP Population Cohorts & Specimen Banks Committee for review and approval.**

**Your suggestions of such studies and contact information are welcome at any time.**

# **OVERALL SCIENTIFIC GOALS OF PPP**

**Comprehensive analysis of plasma and serum protein constituents in people**

**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**

**Determination of the extent of variation across populations and within populations**

# **WHY 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**

# **CHALLENGES/ISSUES FOR PILOT PHASE**

## **Initial Planning Meeting, Bethesda, April, 2002**

- 1. Sensitivity of various techniques for dynamic range of proteins and of peptides**
- 2. Technical aspects of specimen collection, handling, storage: need to standardize**
- 3. Methods of depleting or fractionating high abundance proteins**
- 4. Comparisons of serum and plasma**
- 5. Enumeration and categorization of proteins: post-translational modifications, tissue of origin**

## **CHALLENGES/ISSUES (cont'd)**

- 6. Separation of digested proteins (peptides) or intact proteins**
- 7. Liquid phase multi-dimensional methods vs gel-based methods**
- 8. Parameters for high-throughput link to MS**
- 9. MALDI and SELDI comparisons**
- 10. Various labeling methods for comparison of paired samples**

# **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, Bethesda: Workshop with 35 labs to establish protocols**

**Oct 11-12, 2003, Montreal: Initial data to be discussed**

# AIMS FOR PILOT PHASE OF PPP

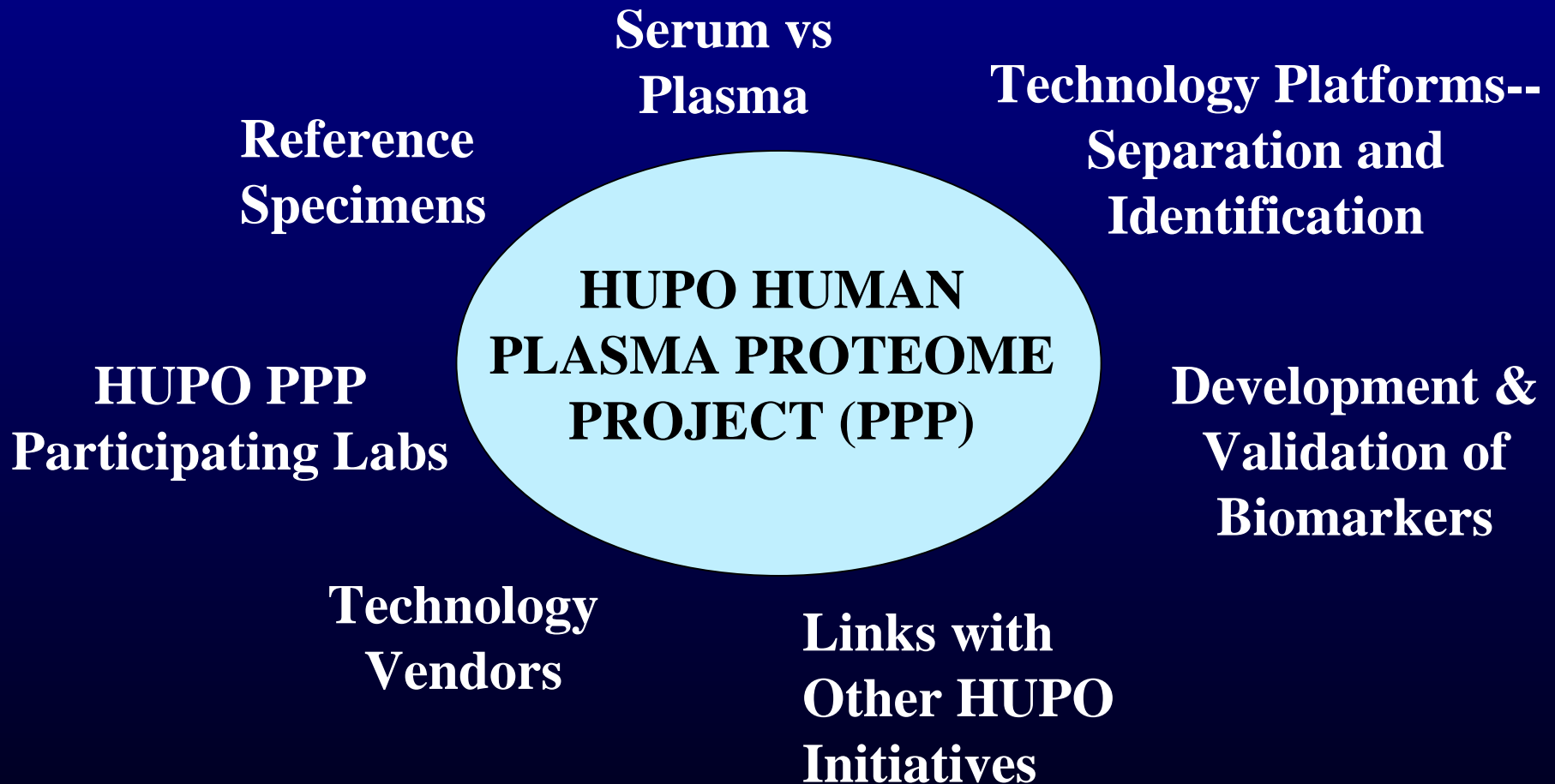
1. Compare a broad range of technology platforms for the characterization of proteins in human plasma and serum. Assess resolution, sensitivity, time, cost, volumes of sample required, practicality with reference specimens.
2. Clarify influence of various technical variables in specimen collection, handling, and storage, especially anti-coagulation and plasma vs. serum.
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 studies of plasma or serum biomarkers in health and disease across major ethnic groups.

# **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)**



## SERUM AND PLASMA REFERENCE SPECIMENS

1. BD Biosciences: 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.
2. 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.
3. Chinese Academy of Medical Sciences: Sets of three plasma + serum, according to BD protocol.

# **SUMMARY OF PPP PARTICIPATING LABS**

**47 Total Participating Labs (28 US, 19-International):**

**17 – US Academic**

**6 – US Federal**

**5 – US Corporate**

**Other Countries: 7 – Europe, 1– Israel,  
9 – Asia, 2 – Australia**

**Number requesting/running reference specimens:**

**41 – UK NIBSC**

**43 – BD b1**

**18 – BD b2**

**18 – CAMS**

## **SUMMARY OF LAB EFFORTS (cont'd)**

**27 – Number volunteering for multi-parameter specimen handling protocol**

**30 – Number testing depletion or pre-fractionation**

**21 – Number requesting vendor input**

**Numbers running different kinds of technology platforms:**

**31 – 2D gels**

**29 – Liquid chromatography**

**25 – Peptide digest first**

**32 – MALDI or LC/MS, MS/MS**

**15 – SELDI**

**18 – Labeled proteins**

**9 – Other**

# **HOW MANY PROTEINS CAN BE DETECTED IN PLASMA/SERUM?**

**Anderson/Anderson review: 289 (on 2D gels)  
MCP 2002; 1:845-68.**

**Adkins et al (PNL): 490 by RP microcapillary  
LC + ion trap MS following Ig depletion,  
digestion, and SCX chromatography MCP  
2002; 1:947-55.**

**Pieper et al: 325 (3700 spots on 2DGE, 1800 spots  
identified by MS) Proteomics 2003; 3:1345-64.**

# **MOST ABUNDANT PLASMA PROTEINS (mg/ml)**

From Chemical Rubber Handbook of Biochemistry 1970, pp C-36-39,  
and current data from GenWay, Inc, at (Sept 2002 PPP Workshop)

albumin 35-45

haptoglobin 0.3-2

IgG, IgA, IgM 12-18

alpha-1 acid

fibrinogen 2-6

glycoprotein 1

alpha-1 antitrypsin 2-5

hemopexin 1

alpha-2 macroglobulin 2-4

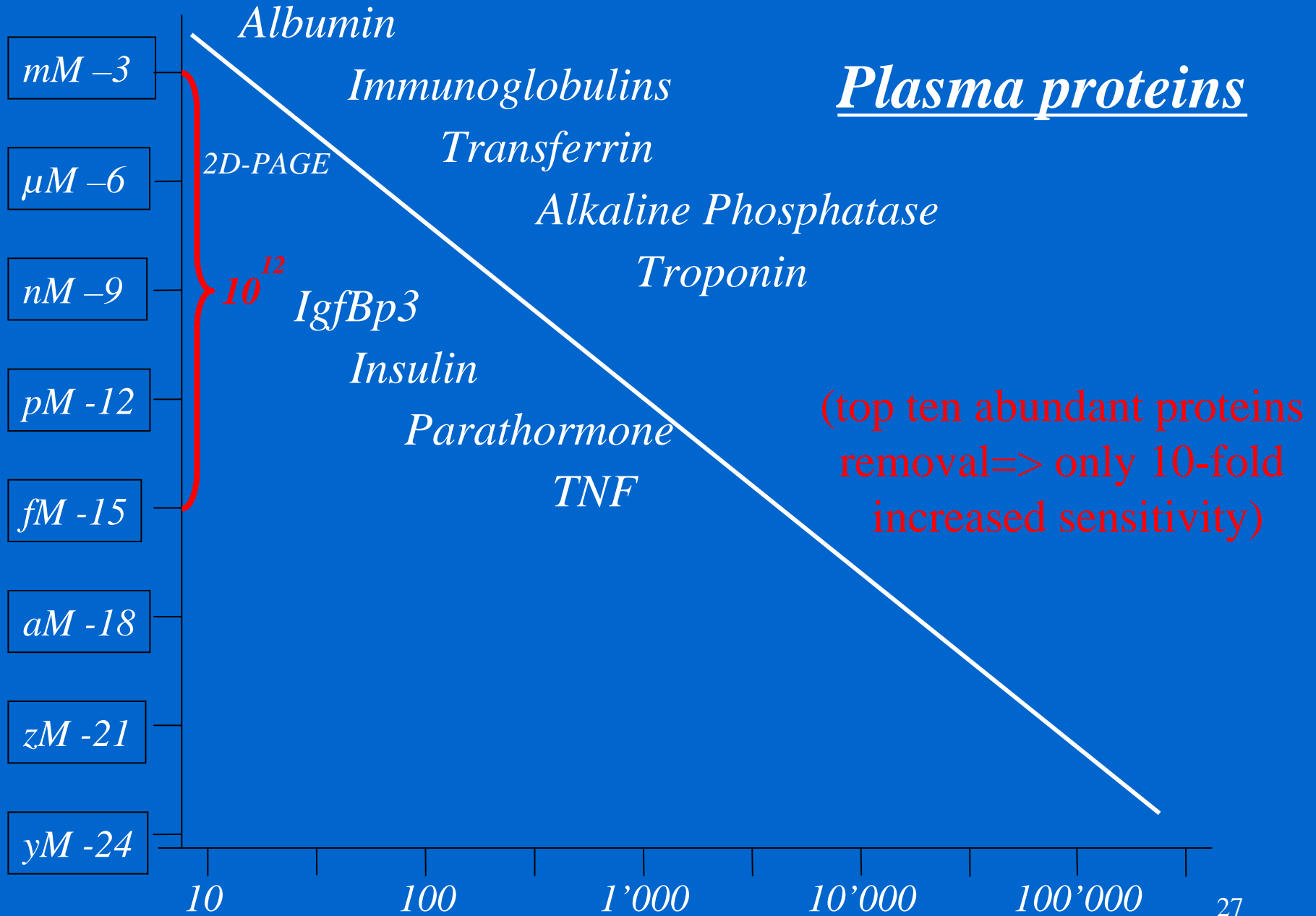
pre-albumin 0.3-0.4

transferrin 2-3

ceruloplasmin 0.3

alpha-2 + beta-lipoproteins (LDL) 4-7

alpha-lipoproteins (HDL) 0.6-1.5



# **CALIBRATION OF DYNAMIC RANGE**

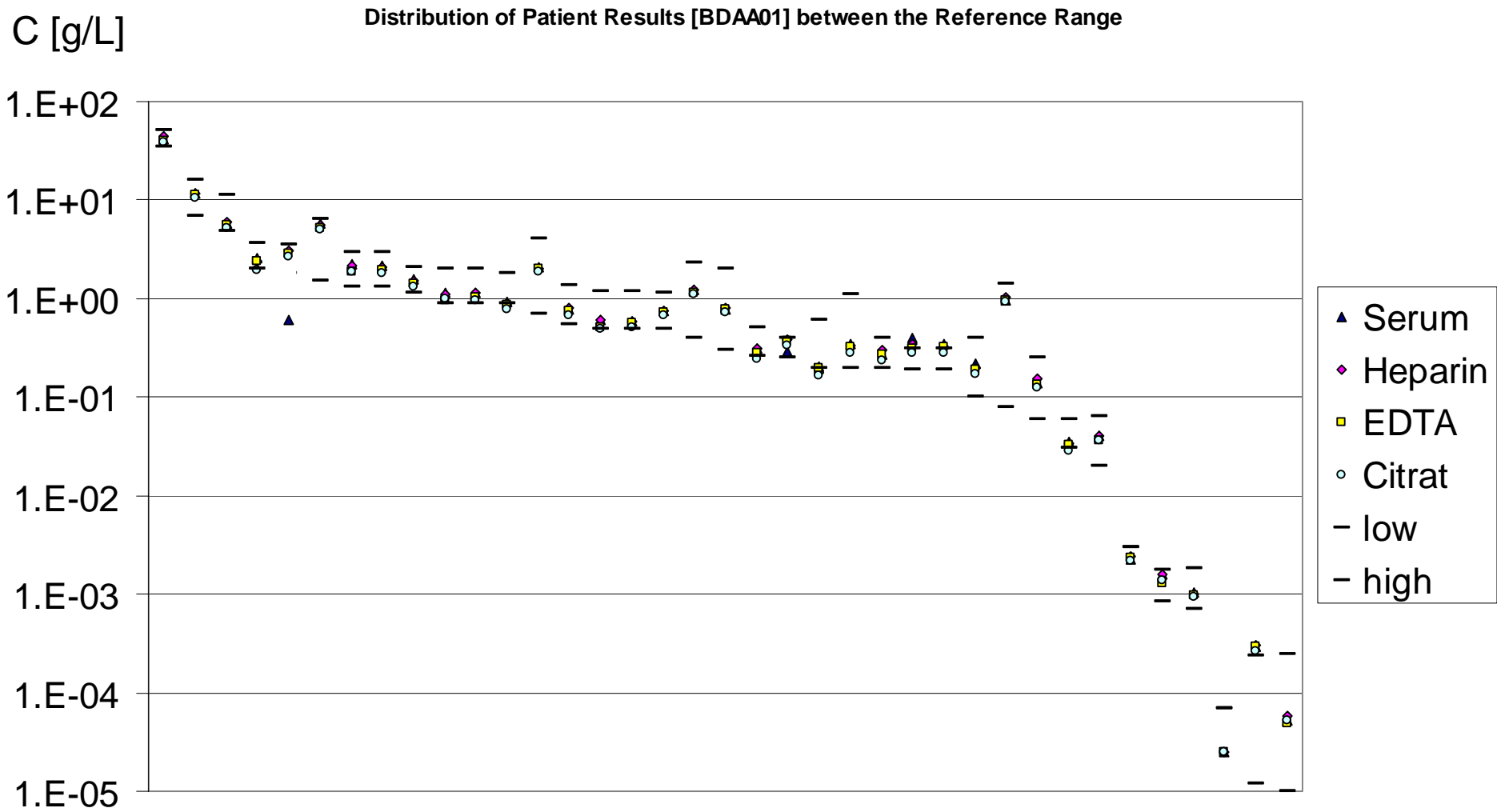
## **in HUPO PPP Reference Specimens**

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

**Additional 8 analytes from ng/ml to pg/ml  
(Chan/Hopkins)**

**Additional 20+ analytes, broad range  
(Hefta, Bristol Myers Squibb)**

# CODED SELECTED PROTEIN CONCENTRATIONS OF HUPO REFERENCE SPECIMEN



# PPP PILOT PHASE ACTIVITIES & MILESTONES

1. Prepared HUPO PPP reference specimens; BD and UK specimens distributed to participating labs in May 2003, Chinese specimens in Sept 2003.
2. All participating laboratories will analyze reference specimen set(s), as well as other specimens chosen by the Lab, using each lab's preferred technology platform(s), with guidance from PPP technical committees sent 16 May 2003, updated at July workshop and in 14 Aug memo.

## PILOT PHASE (cont'd)

### 3. Subsets of labs will:

- a) Test technical parameters of sample handling, e.g. duration at various temps before frozen, temp & duration of storage, freeze/thaw cycles
- b) Compare with vs without anti-proteases
- c) Compare with vs without depletion of albumin and/or other highly abundant proteins
- d) Run the three plasma preps (EDTA, heparin, citrate) & serum from other ethnic groups

## 4. Database development

- a) Develop methods for identification, enumeration, and comparison of proteins, accommodating a broad array of technology platforms for separation and then identification, including mass spectrometry and antibody arrays
- b) Assure specimen tracking and methods comparisons
- c) Link with databases of HUPO/PSI, Swissprot, EBI; agreed to use International Protein Index (IPI) and XML/PEDRo or Excel formats for data
- d) Build inventory of plasma/serum proteins from HUPO PPP studies

## Data Development (cont'd)

e) Consider system for identifying and classifying related proteins (in cooperation with EBI); an example:

1.0 Primary amino acid backbone sequence of gene-coded protein (including splice variants and SNPs)

1.1 Modification of backbone: cleavage  
(activation, inactivation)

1.2 Modification of side-chains

1.2.1 Phosphorylation (sites) and dephosphorylation

1.2.2 Glycosylation (sites)

1.2.3 Other classes of post-translation modifications

1.3 Aggregation (dimers, etc)

1.4 Interactions: protein-protein,-nucleic acid,-others

## PILOT PHASE (cont'd)

### 5. Administration and Dissemination

- a) Develop administrative hub (responsibility assigned by HUPO to U. Michigan)
- b) Clarify intellectual property matters: committed to public domain, with ample opportunity for proprietary development of biomarkers and targets/agents
- c) Link with related HUPO initiatives: liver, protein standards, brain, antibodies
- d) Raise funds for budget: private & public

# **CORPORATE SPONSORS**

**Johnson & Johnson**

**Abbott Labs**

**Pfizer**

**Novartis**

**Invitrogen**

**Amersham**

**Procter & Gamble**

**BD Clinical Preanalytical Solutions**

**Ciphergen**

**Bristol Myers Squibb**

**Agilent**

**Dade Behring**

# **HUPO PPP SUPPORT FROM NIH**

## **Trans-NIH Consortium**

**Natl Cancer Inst: Div Cancer Prevention;  
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**

## Administration & Dissemination (cont'd)

- e) Held PPP Workshop 16-17 July in Bethesda to drive pilot studies; set agenda for this working meeting
- f) Plan for the June 2004 “Jamboree”: emphasis on cross-laboratory interpretive analyses; also Beijing satellite meeting (Oct 2004)
- g) Publish comparisons of samples, of technologies, from different populations around the world, from participating laboratories: plan for Special Issue of Proteomics (end of 2004); invitations from J Proteome Research/ACS and from BioMed Central

# **POTENTIAL POPULATION COHORTS AND SPECIMEN RESOURCES FOR PPP**

**Stanislas Cohort: Nancy, France (Siest):**

**4600 in 1006 families: cardiovascular/metabolic disease emphasis; 700 families returning at 10 yrs 2003-04; with and without clot activator for serum; EDTA for plasma; 280 ul straws, 10 each**

**Women's Health Initiative (NHLBI/NIH, Fred Hutch, Seattle, CC): 67,000 women; CHD, breast/endometrial CA/osteoporosis emphases**

**CARET: Seattle, WA (Goodman et al):  
18,314 current & former smokers**

**PLCO (Prostate/Lung/Colon/Ovary) Study, US: NCI/Detroit site**

**General population: Leeds, UK (Grant)**

**Alcoholism Cohort: Colorado, US (Tabakoff)**

**Children and families, The Children's Hospital, Boston, US  
(Castillo)**

**Early Detection Research Network, US (NCI)**

# **PUBLICATION POSSIBILITIES**

**Proteomics (Mike Dunn): special issue offered and committed/end of 2004**

**J. Proteome Research (Bill Hancock): interest in sets of articles, offer to use American Chemical Society electronic archives**

**BioMed Central (Matt Day): electronic**

**Other**

