The Plasma Proteome:
Enabling a Revolution in Diagnostics

Leigh Anderson Ph.D.
Founder & CEO, Plasma Proteome Institute

PPI

Introduction to the Plasma Proteome

- Why the plasma proteome represents such an enduring technical challenge
- What it takes to develop a comprehensive list of candidate markers in plasma
- The next step: what we need to carry out systematic validation of markers to support a revolution in diagnostics

PPI

Plasma is the largest and deepest version of the human proteome

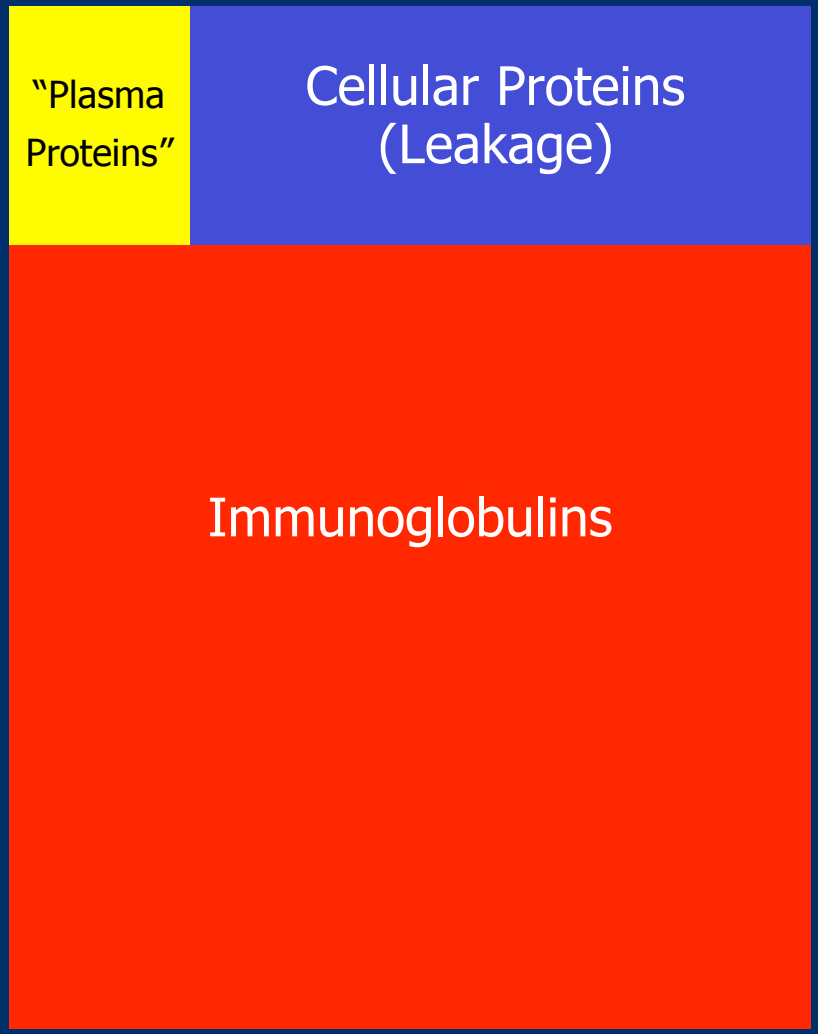
- Largest = Most proteins
- Deepest = Widest dynamic range

PPI

Major Components of the Plasma Proteome

- ~40,000 forms of proteins secreted to function in plasma, most glycoproteins
  - Assume 500 gene products x 2 splice variants x 20 glycoforms x 2 clip forms
- ~500,000 forms of tissue proteins
  - Essentially all tissue proteins x splice and PTM variants
- ~10,000,000 clonal forms of immunoglobulin

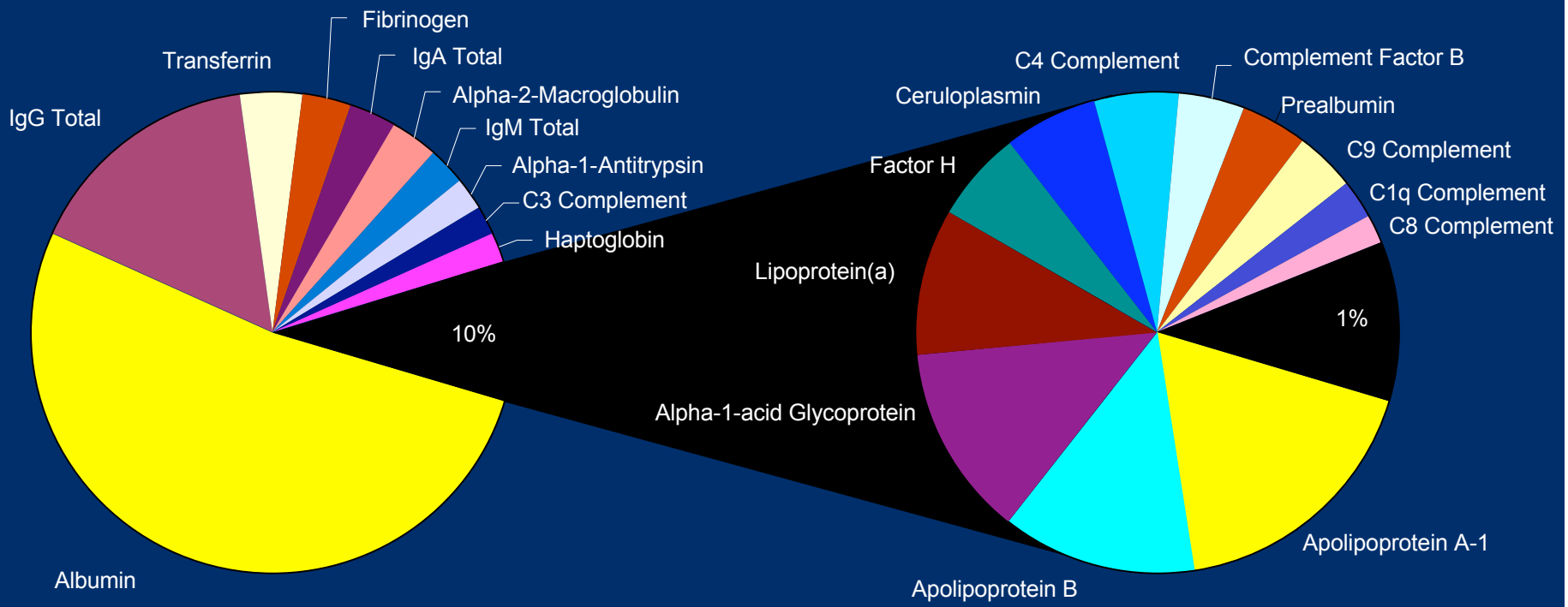
Total: the largest version of the human proteome



PPI

Major Plasma Proteins

99% of plasma protein mass

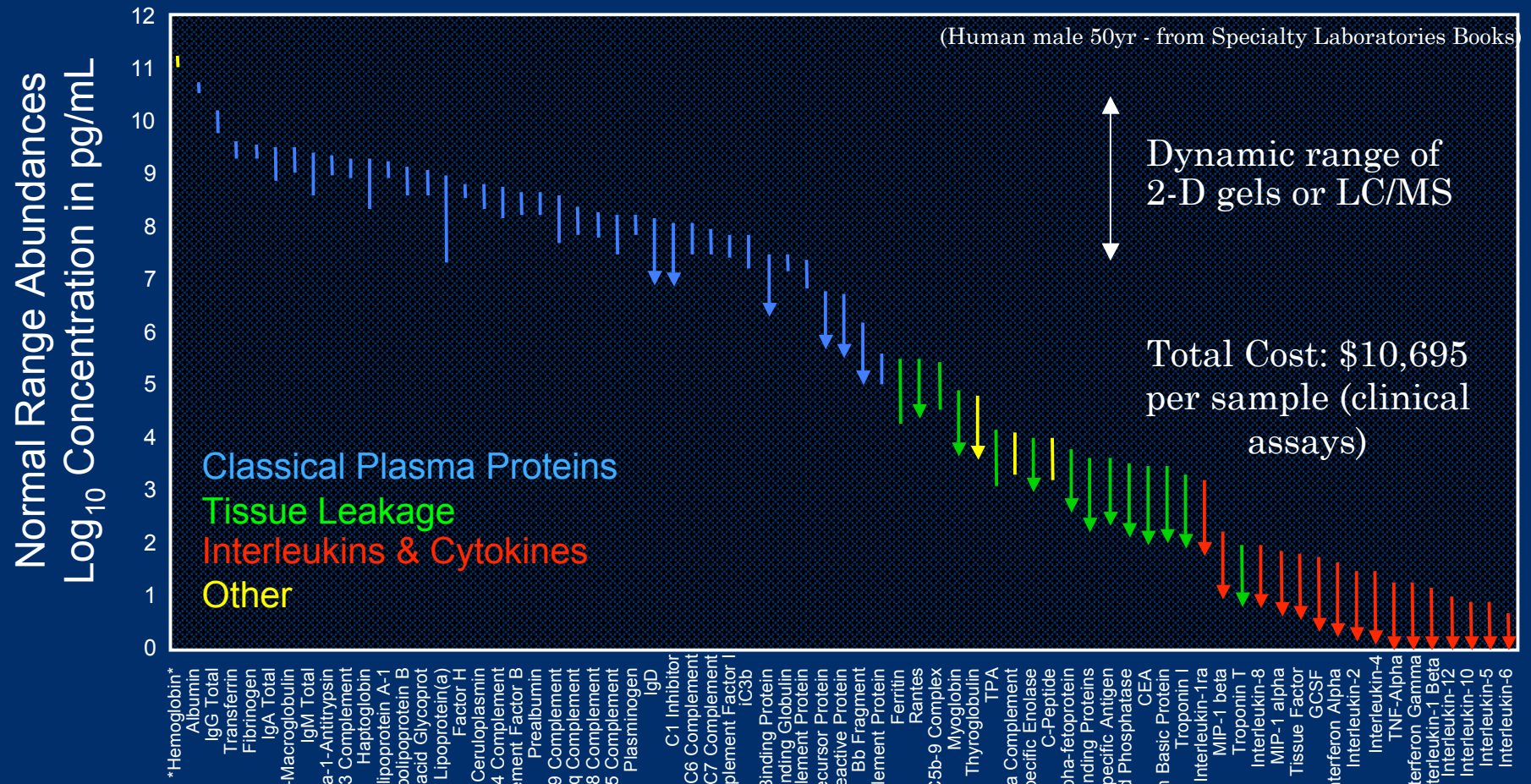


0 - 90%

90 - 99%

PPI

Proteins Measured Clinically in Plasma Span > 10 Orders of Magnitude in Abundance



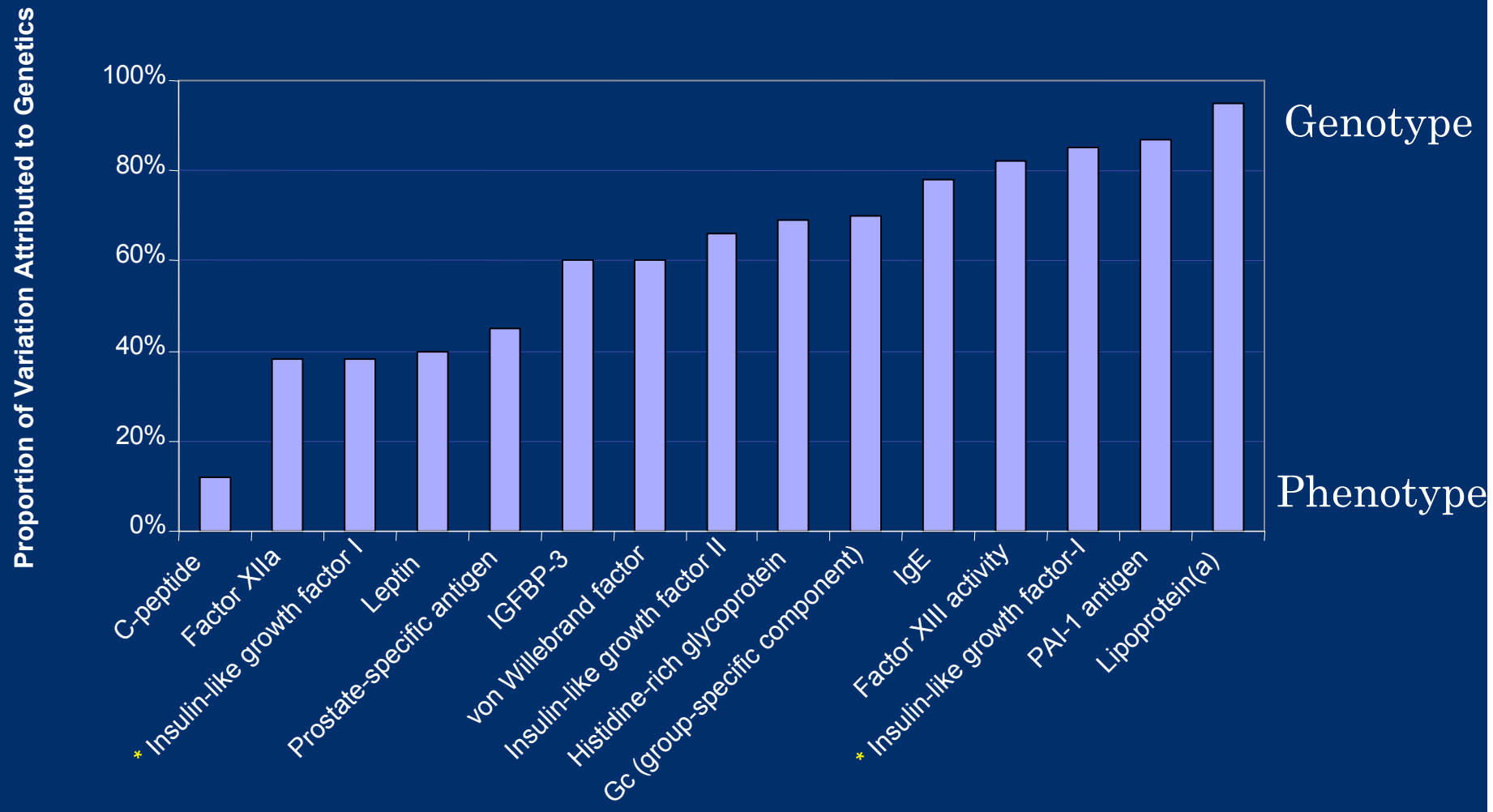
From:The human plasma proteome: history, character, and diagnostic prospects. Anderson, N. L. Anderson, N. G., Mol Cell Proteomics (2002) 1:845-67.

PPI

Nucleic Acids Exist Free in Plasma

- Genomic DNA is present in plasma (~1ug/ml)
  - Released by apoptosis & necrosis, e.g., by tumor cells
  - Mutations (e.g., in P53) can be detected
- mRNA can be detected in plasma
  - Fetal mRNA can be found free in maternal circulation
- Current and likely future utility confined to qualitative tests (genotyping)
  - mRNA concentrations are very poor indicators of the amount of protein products
- Proteins remain the focus of the search for disease and response markers

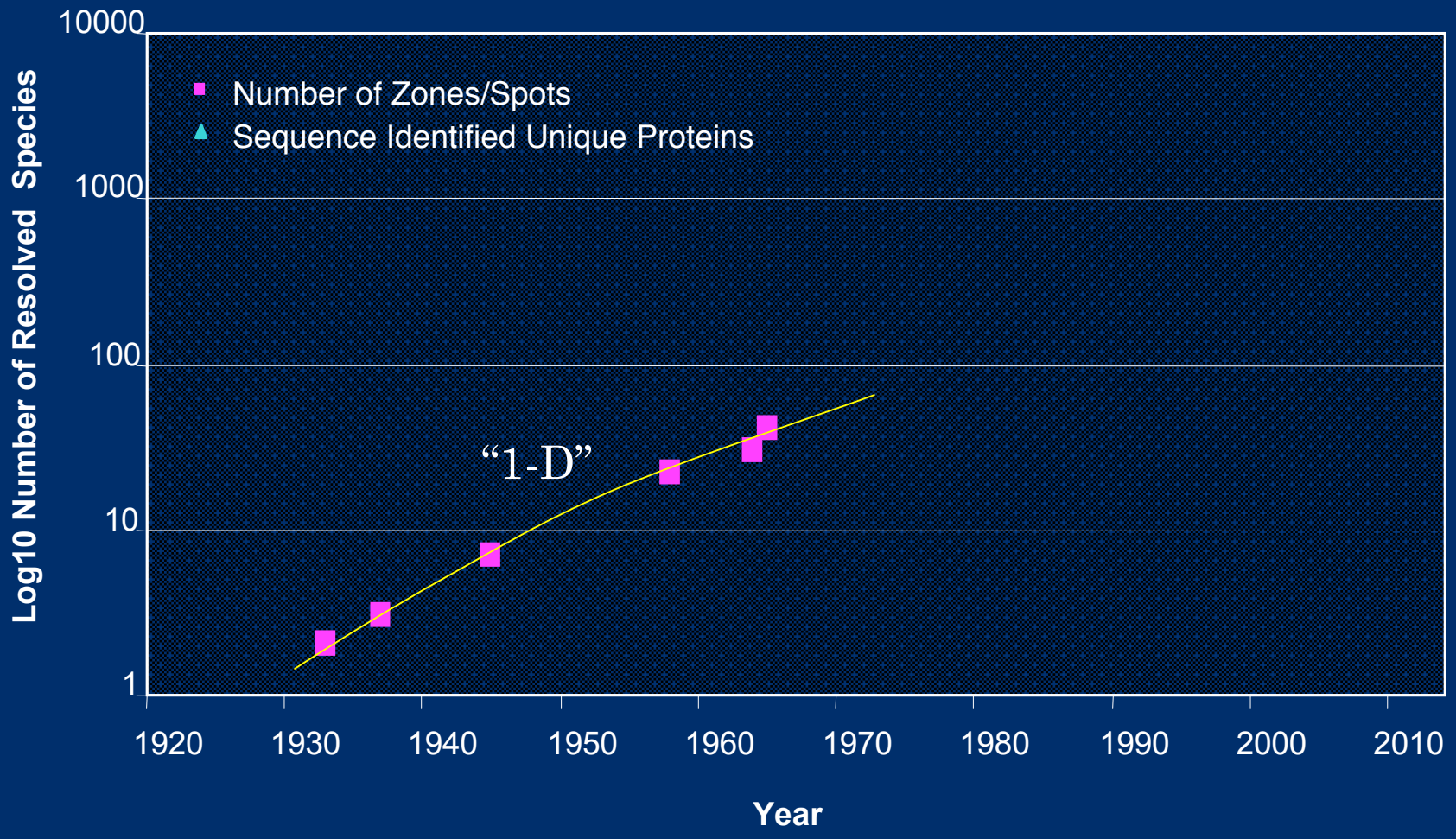
Genetic Component of Variation in Abundance of 15 Proteins in Plasma



(Published values from many sources
*= two discordant studies)

Plasma Proteomics Began in Sweden

Svedberg, Tiselius, Laurell, et al



From:The human plasma proteome: history, character, and diagnostic prospects. Anderson, N. L. Anderson, N. G., Mol Cell Proteomics (2002) 1:845-67.

PPI

The Second Phase of Plasma Exploration Began with 2-D Gels (c. 1976)

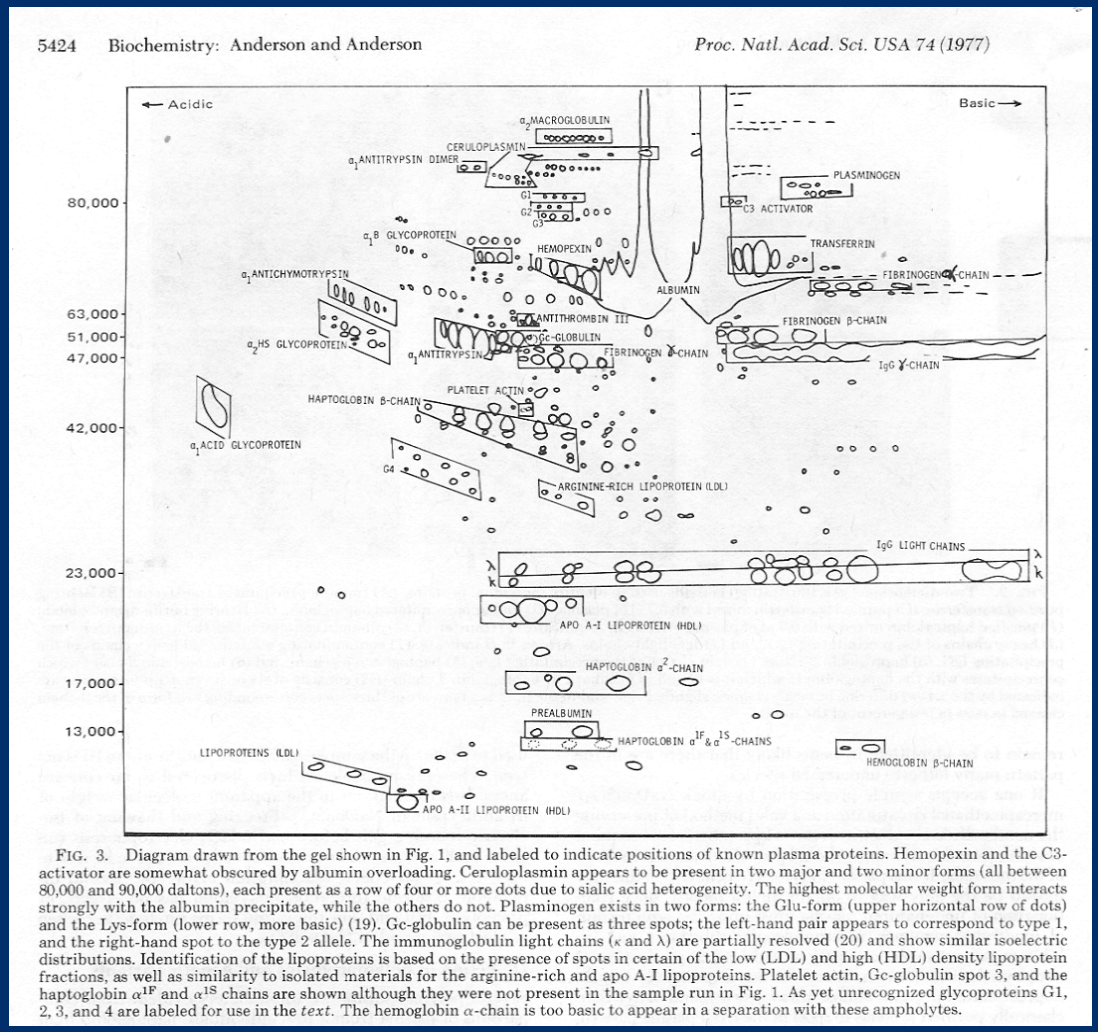


FIG. 3. Diagram drawn from the gel shown in Fig. 1, and labeled to indicate positions of known plasma proteins. Hemopexin and the C3-activator are somewhat obscured by albumin overloading. Ceruloplasmin appears to be present in two major and two minor forms (all between 80,000 and 90,000 daltons), each present as a row of four or more dots due to sialic acid heterogeneity. The highest molecular weight form interacts strongly with the albumin precipitate, while the others do not. Plasminogen exists in two forms: the Glu-form (upper horizontal row of dots) and the Lys-form (lower row, more basic) (19). Gc-globulin can be present as three spots; the left-hand pair appears to correspond to type 1, and the right-hand spot to the type 2 allele. The immunoglobulin light chains (κ and λ) are partially resolved (20) and show similar isoelectric distributions. Identification of the lipoproteins is based on the presence of spots in certain of the low (LDL) and high (HDL) density lipoprotein fractions, as well as similarity to isolated materials for the arginine-rich and apo A-I lipoproteins. Platelet actin, Gc-globulin spot 3, and the haptoglobin α^{1F} and α^{1S} chains are shown although they were not present in the sample run in Fig. 1. As yet unrecognized glycoproteins G1, 2, 3, and 4 are labeled for use in the *text*. The hemoglobin α-chain is too basic to appear in a separation with these ampholytes.

Anderson,L., Anderson,N. G. High resolution two-dimensional electrophoresis of human plasma proteins. (1977) PNAS 74, 5421-5

2-D Electrophoresis

300+ resolved spots

40 identified proteins

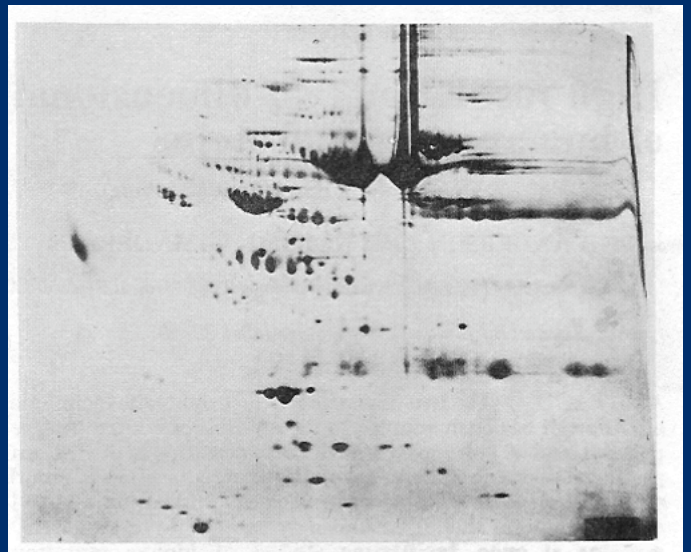
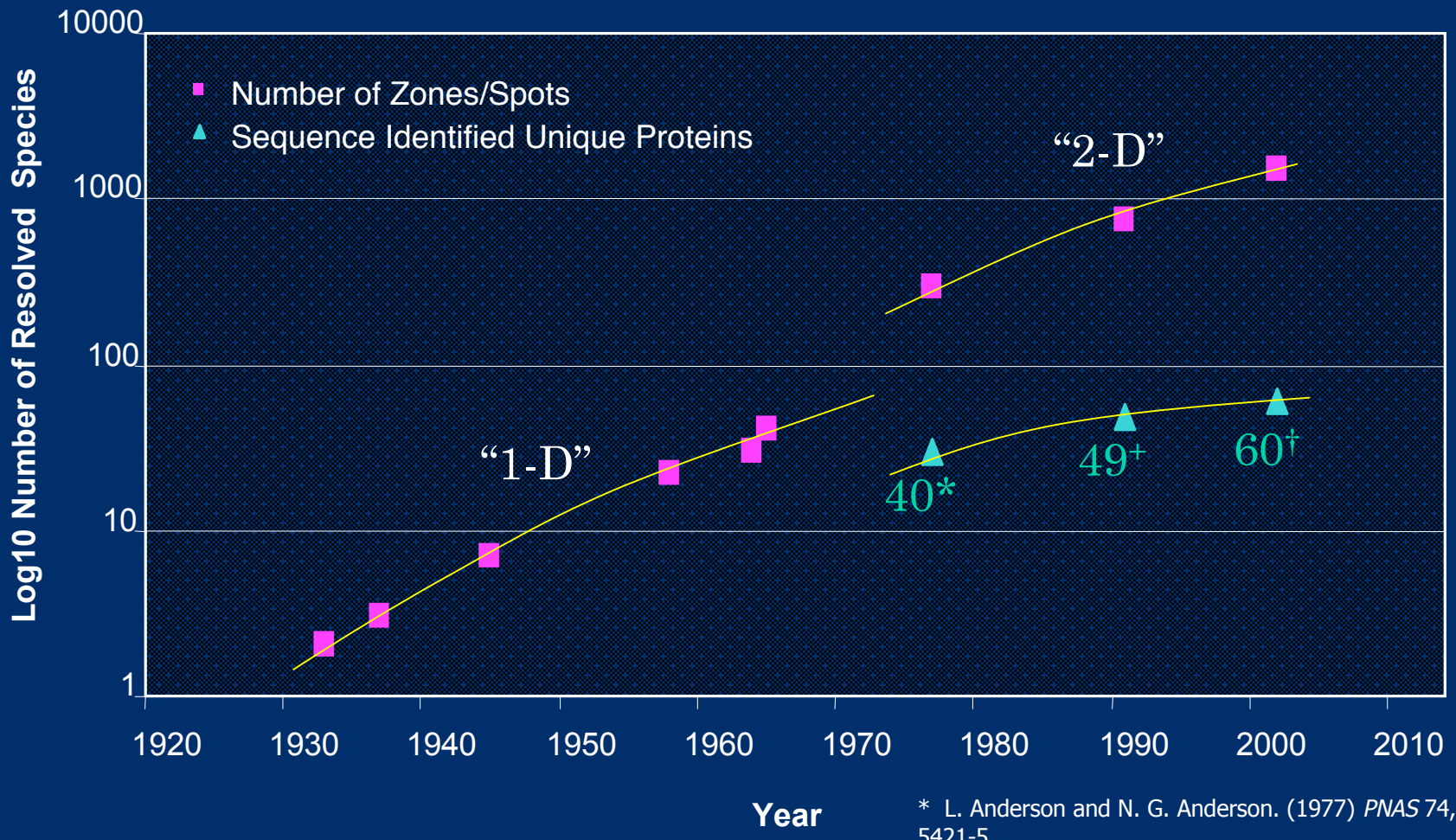


FIG. 1. Two-dimensional gel of human plasma proteins. The sample was 10 μl of fresh heparinized plasma denatured in Na-DodSO$_4$/mercaptoethanol.

PPI

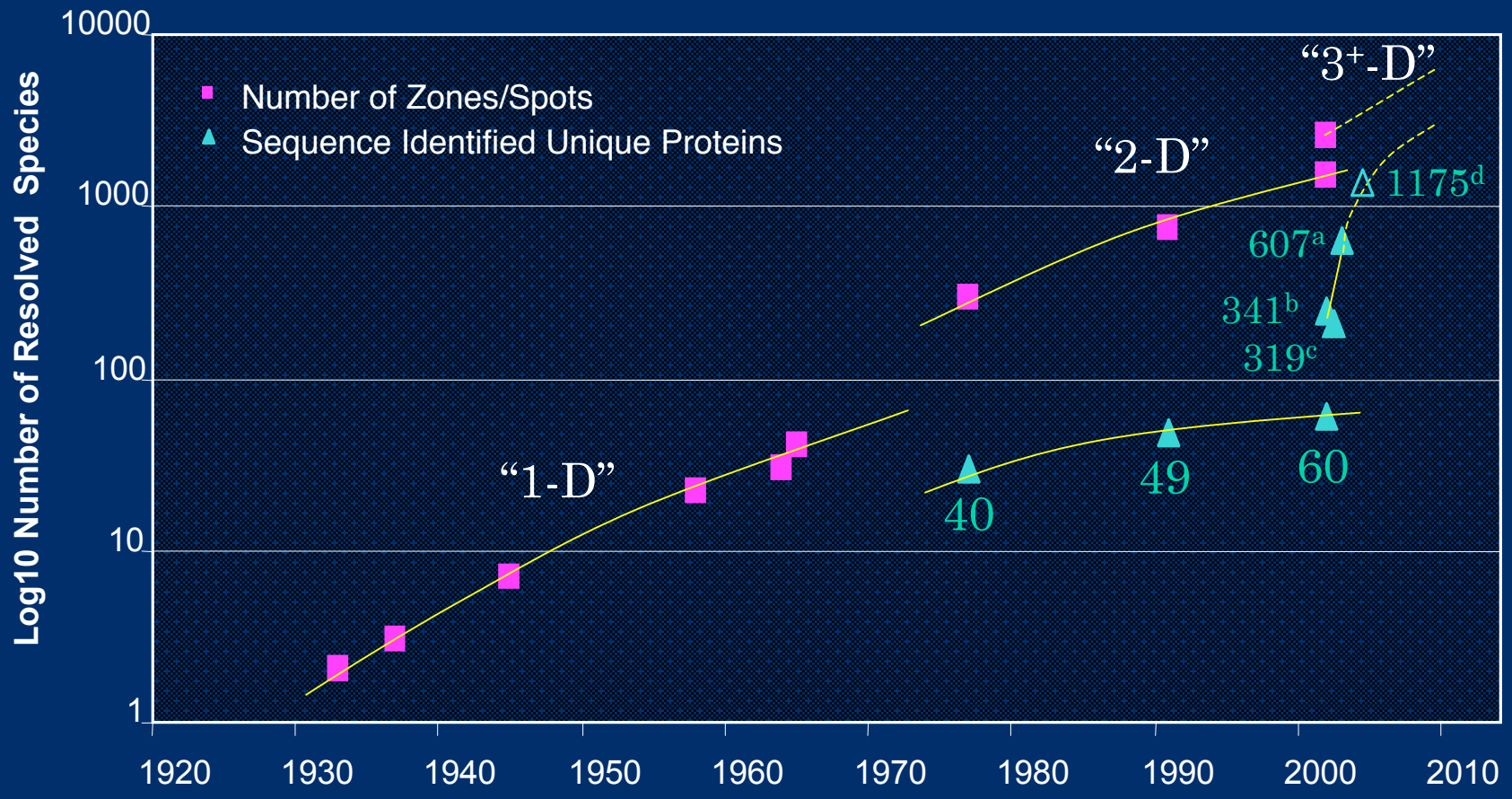
2-D Electrophoresis Continued the Growth in Number of "Spots" but Not Many New Proteins



* L. Anderson and N. G. Anderson. (1977) *PNAS* 74, 5421-5.
\+ G. J. Hughes, et al (1992) *Electrophoresis* 13, 707-14.
† Swiss 2DPage website

From:The human plasma proteome: history, character, and diagnostic prospects. Anderson, N. L. Anderson, N. G., Mol Cell Proteomics (2002) 1:845-67.

>2 Dimensions of Separation Radically Increases the Number of Detected Proteins

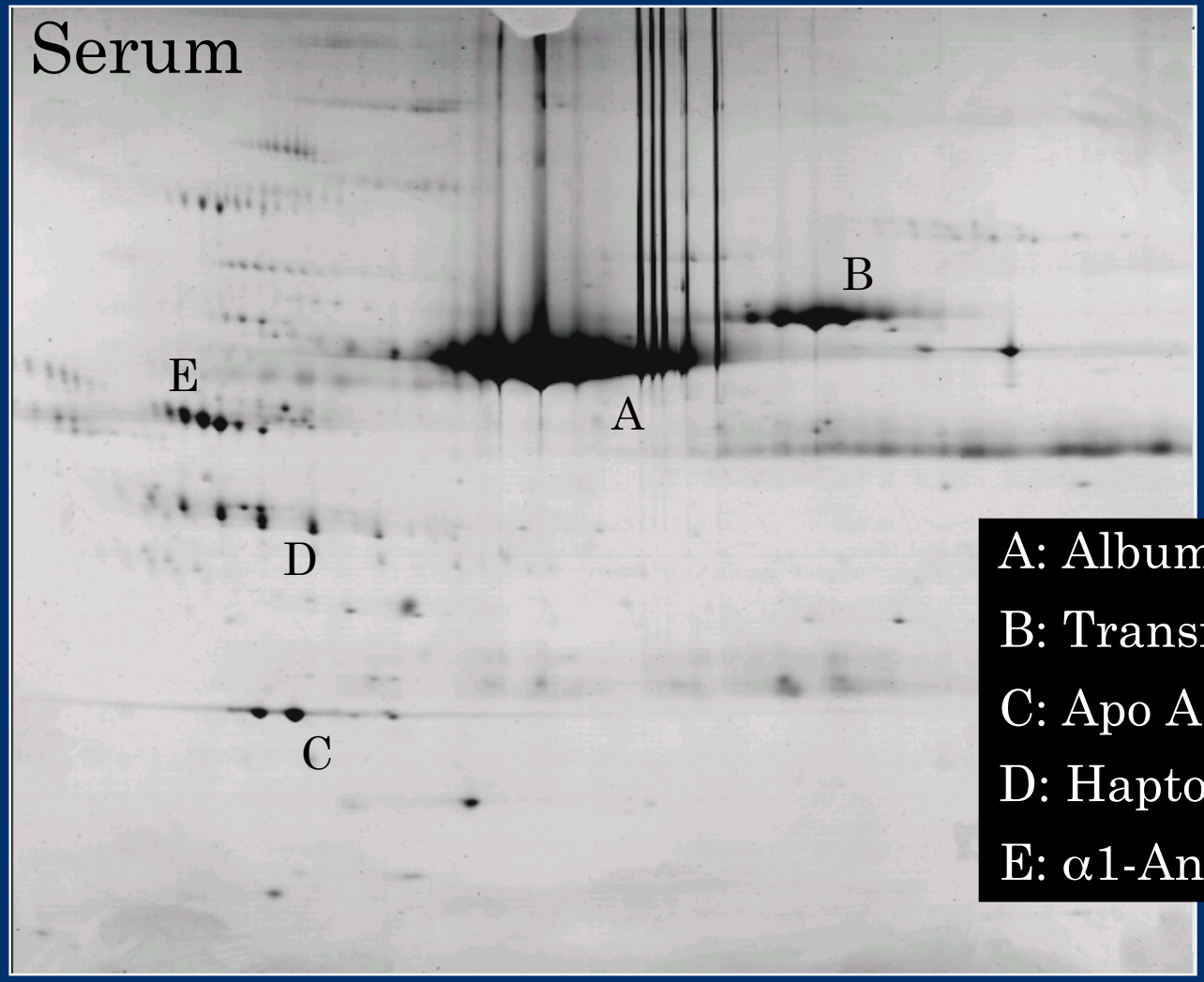


a. J. N. Adkins, et al, *Mol Cell Proteomics* **1,** 947-55.
b. R. S. Tirumalai, et al, *Mol Cell Proteomics* **2,** 1096-103.
c. R. Pieper, et al, *Proteomics* **3,** 422-32.
d. H_Plasma_NR-v2

From:The human plasma proteome: history, character, and diagnostic prospects. Anderson, N. L. Anderson, N. G., Mol Cell Proteomics (2002) 1:845-67.

PPI

High Abundance Proteins in Plasma or Serum Limit Detection of Minor Components



A: Albumin

B: Transferrin

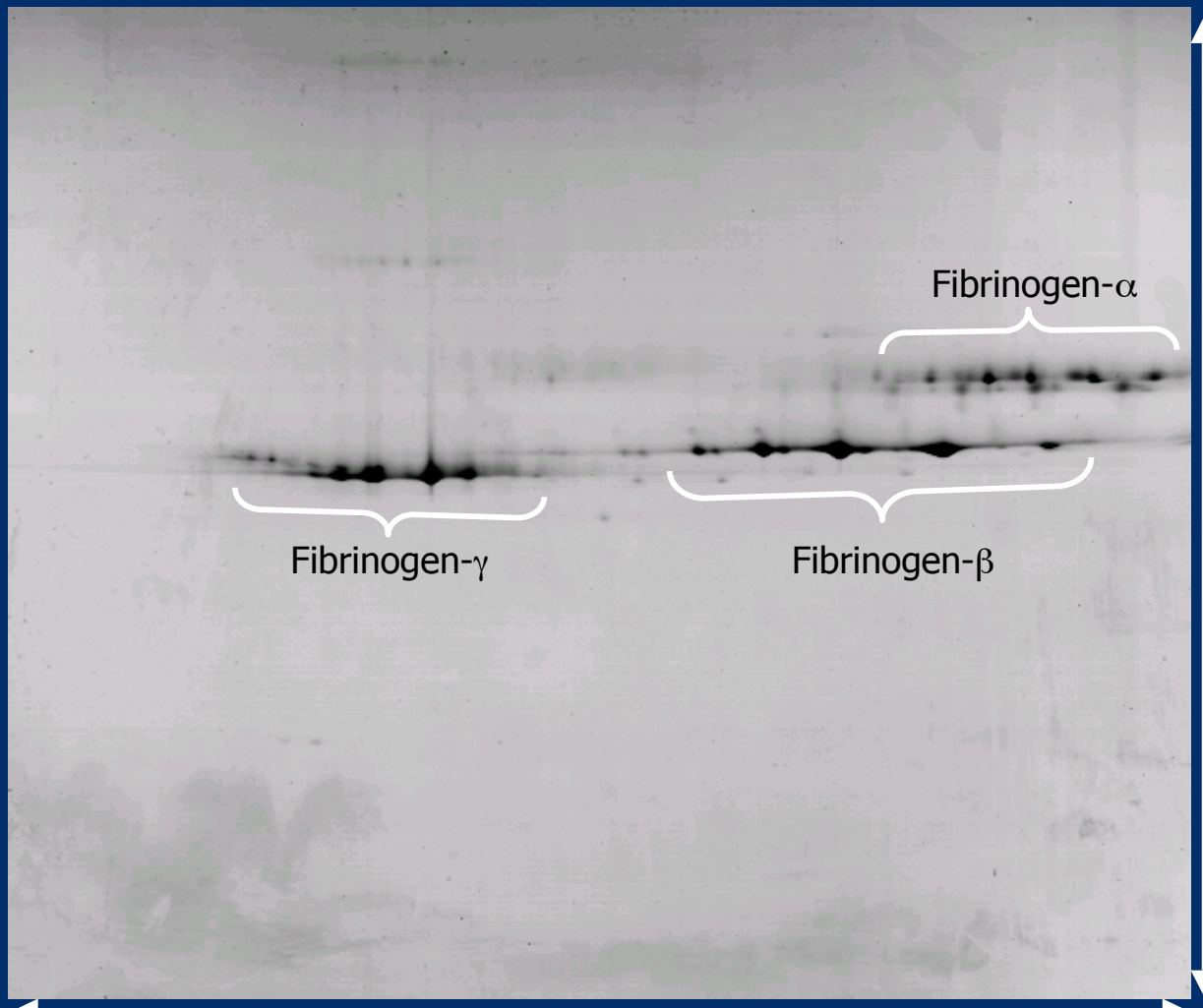
C: Apo A-I lipoprotein

D: Haptoglobin β-chain

E: α1-Antitrypsin

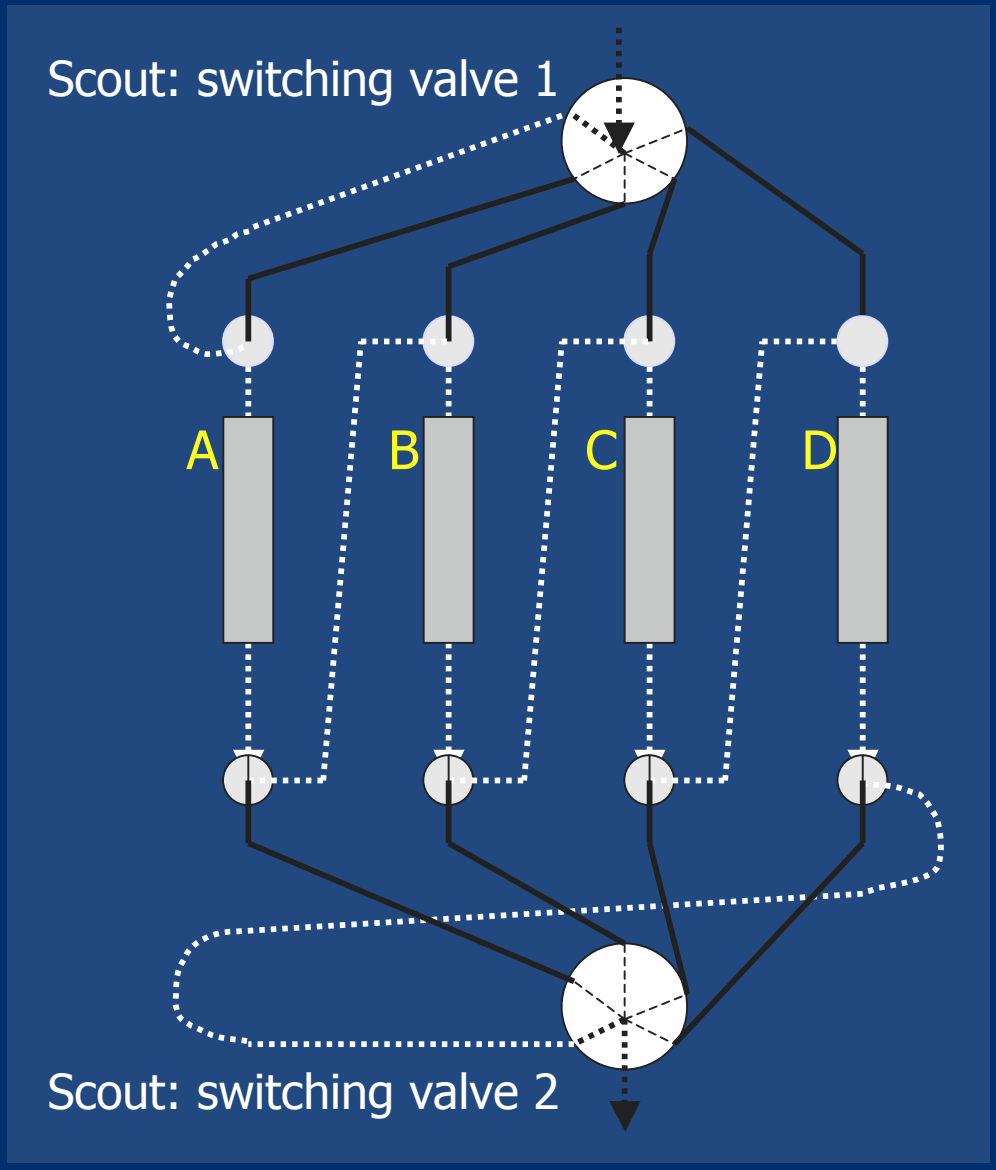
Polyclonal Antibodies Specifically and Repeatably Remove Target Proteins from Serum/Plasma

Proteins Bound from Plasma by Affinity-Purified anti-Fibrinogen Ab



PPI

Multicolumn Implementation of Plasma Immunosubtraction



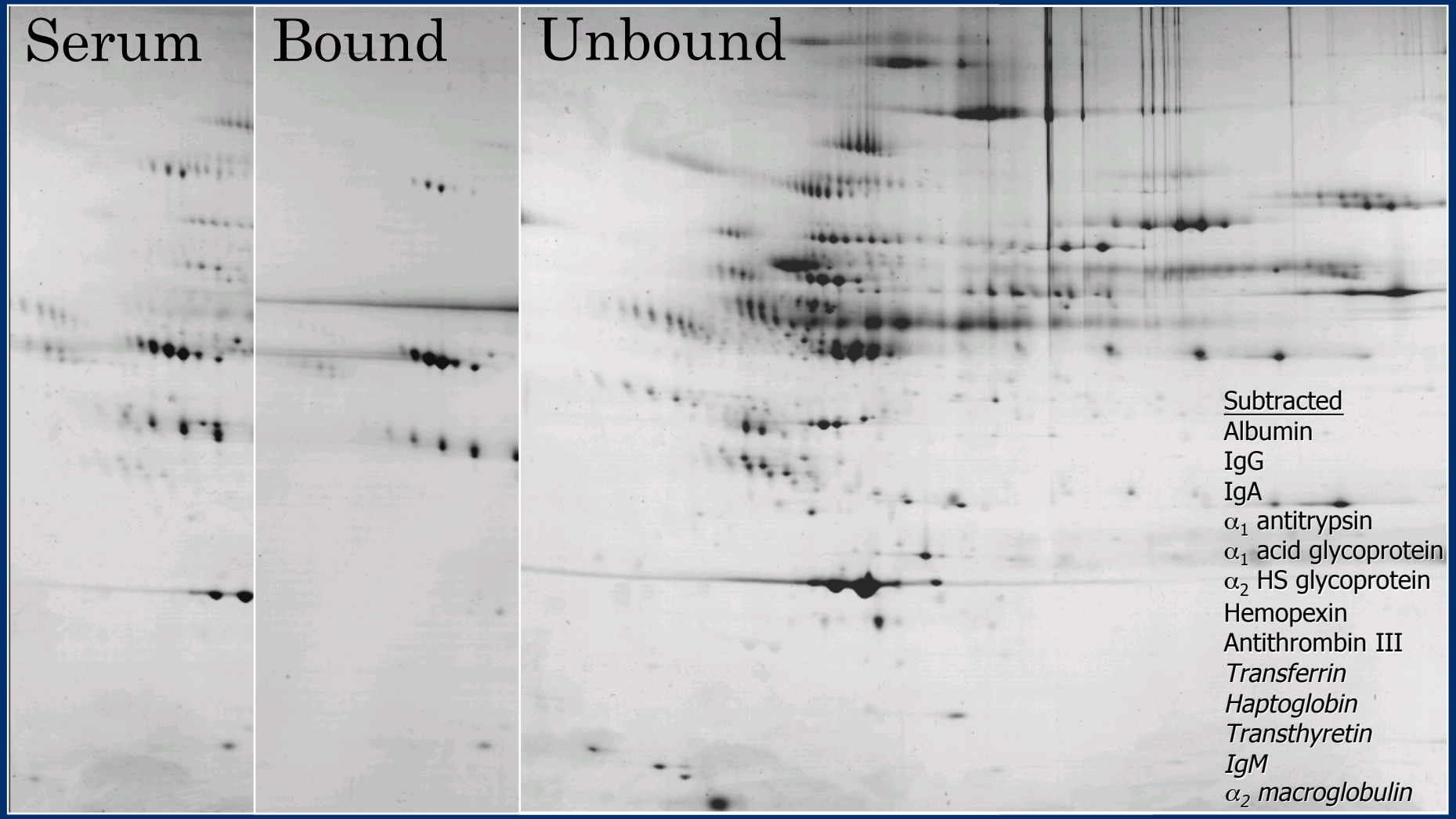
A: anti-albumin, anti-transferrin, anti-haptoglobin, anti-α-1-antitrypsin, anti-α-1-acid glycoprotein, anti-α-2-HS glycoprotein, anti-hemopexin, anti-transthyretin, anti-antithrombin-III.

B: 50% anti-IgA and 50% anti-IgM

C: anti-α-2-macroglobulin

D: anti-apolipoprotein A1

Antibody Affinity Subtraction of 10 High Abundance Proteins from Serum

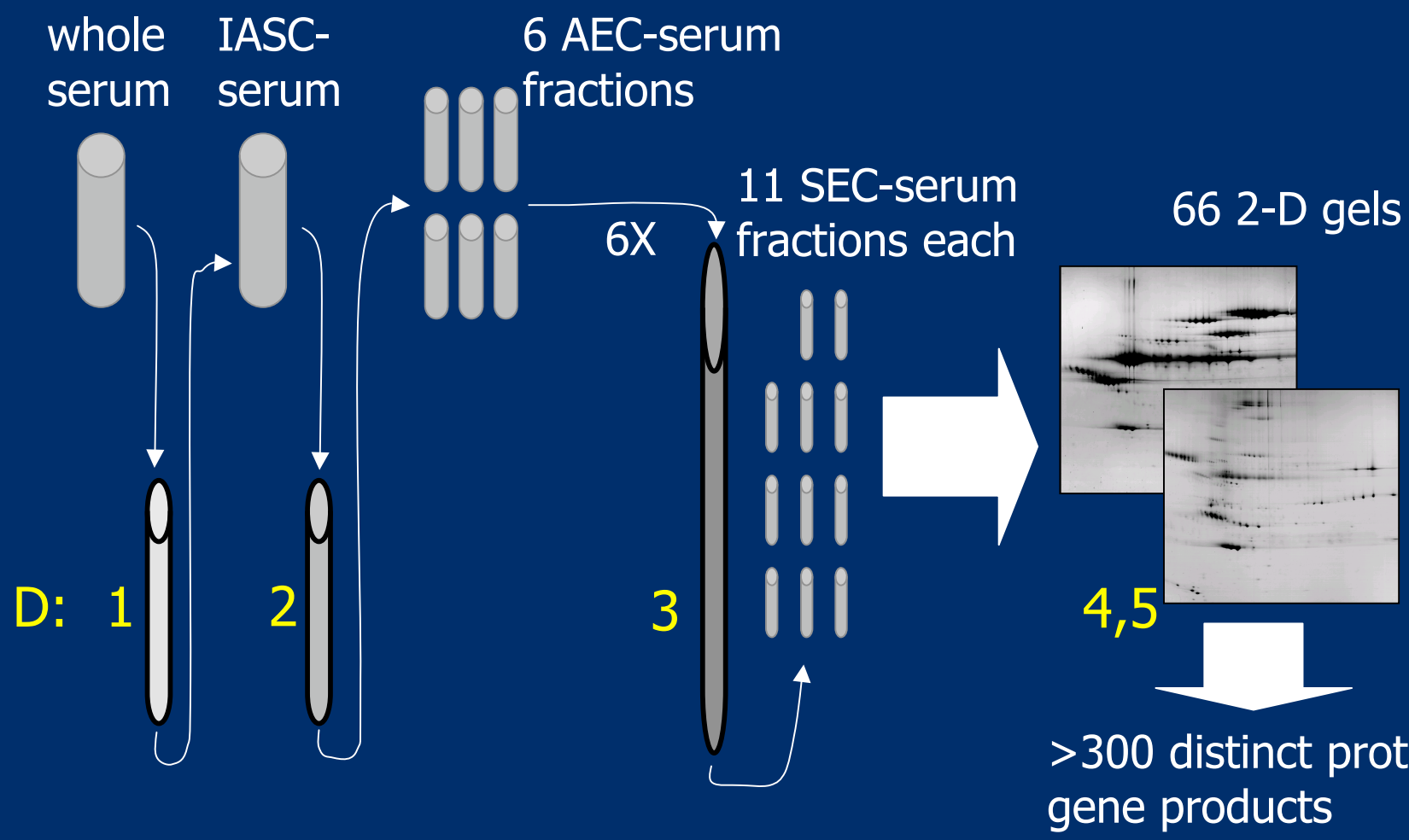


Multi-Component Immunoaffinity Subtraction Chromatography, An Innovative Step Towards A Comprehensive Survey Of The Human Plasma Proteome, Rembert Pieper, et al, Proteomics, (2003) *Proteomics* **3,** 422-32.

PPI

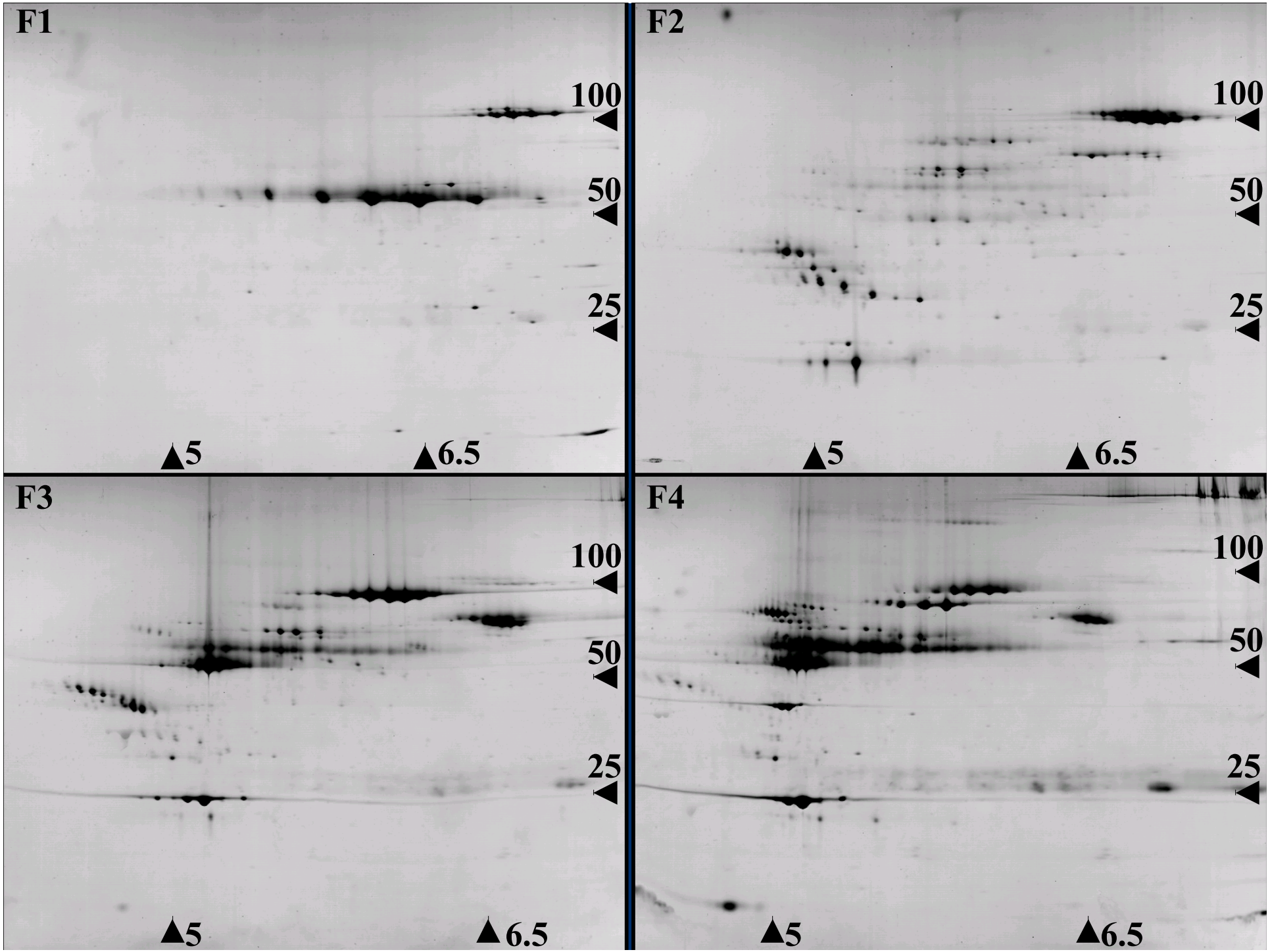
The Current Phase of Plasma Proteomics Employs Multi-Dimensional (>2D) Approaches

(e.g., 3-D Chromatography + 2-DE + LC/MS)



The Human Serum Proteome: Display Of Approximately 3,500 Chromatographically Separated Distinct Protein Spots On 2-DE Gels And Identification Of 307 Uniquely Annotated Proteins, Rembert **Pieper, et al, Proteomics 3(7): 1345-64. (2003)**

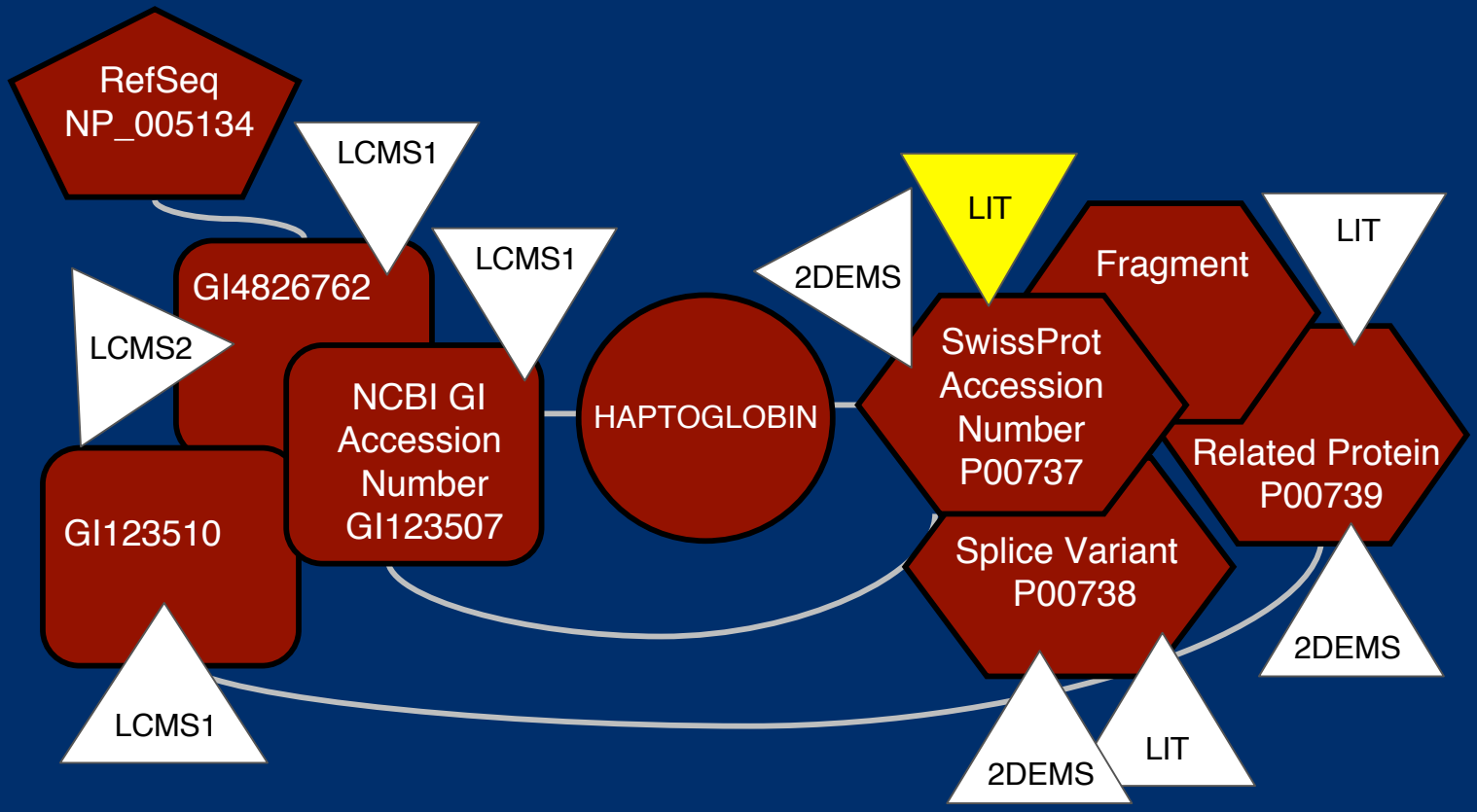
PPI


F1
100
50
25
▲5
▲6.5
F2
100
50
25
▲5
▲6.5
F3
100
50
25
▲5
▲6.5
F4
100
50
25
▲5
▲6.5

So How Many Proteins Are There in Plasma?

- Different methods yield different sets of proteins
  - Reasons include physicochemical biases of techniques, and statistics of peptide choice in MS/MS
- The most comprehensive approach is therefore to combine data from different approaches
  - We used four input datasets:
    - Base list of ~450 proteins reported in "non-proteomics" literature as measured/detected in plasma or serum
    - Three sets of 300-600 proteins each from proteomics surveys (2-D gels + MS/MS; LC/LC-MS/MS)
- Combined data can be made non-redundant using methods of genomics
  - Redundancy definition: >95% homology over >=15 amino acid subsequence
- Results from: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, Anderson et al, Molec. Cell Proteomics, in press

Achieving Non-Redundancy: MS Identifications of Haptoglobin in Plasma Proteome Datasets



From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

Sources of H_Plasma_NR_v2

	Lit	LCMS1	LCMS2	2DEMS	Total
Beginning Accessions	468	607	341	319	1735
Minus non-human	458	580	330	312	1680
Minus intra-source redundancy and non human accessions	433	475	318	283	1509
Unique to source in NR	284	334	221	141	980
Total combined NR list	-	-	-	-	1175

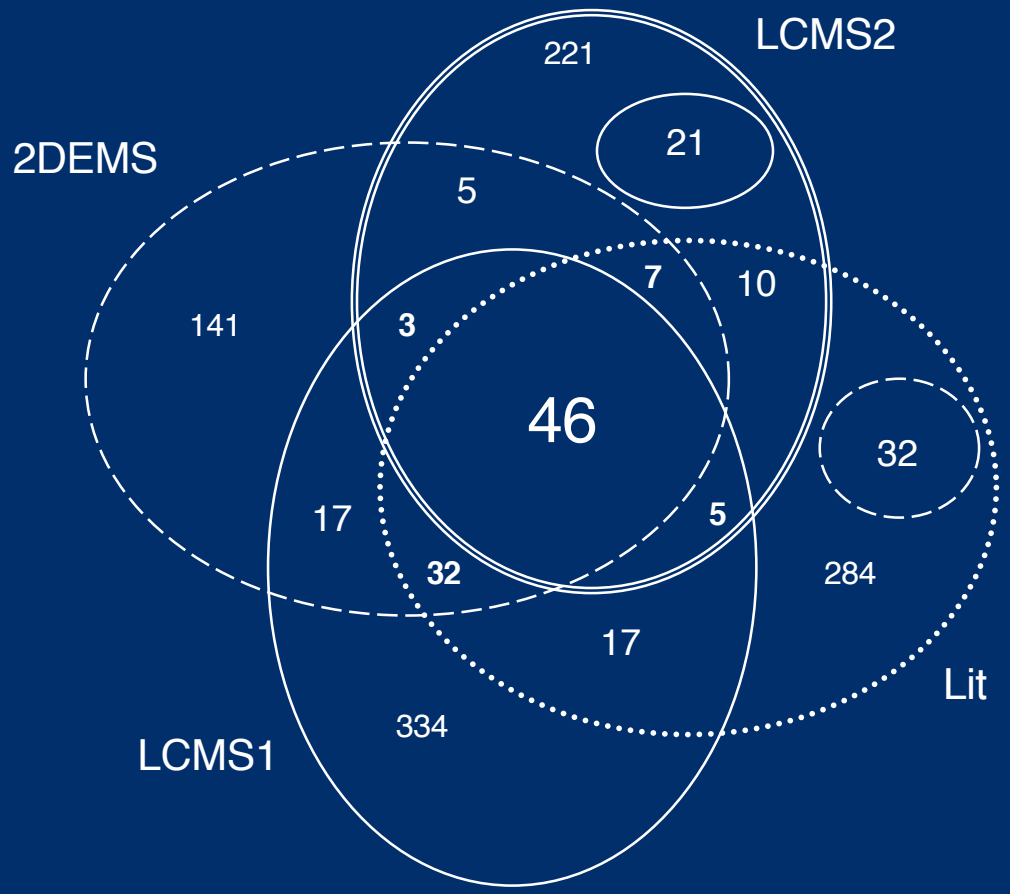
- Lit: N.L. Anderson and M. Polanski, result of literature search for proteins detected in plasma or serum.
- LCMS1: J. N. Adkins, S. M. Varnum, K. J. Auberry, R. J. Moore, N. H. Angell, R. D. Smith, D. L. Springer and J. G. Pounds. (2002) Toward a human blood serum proteome: Analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **1,** 947-55.
- LCMS2: R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads and T. D. Veenstra. (2003) Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics* **2,** 1096-103.
- 2DEMS: R. Pieper, Q. Su, C. L. Gatlin, S. T. Huang, N. L. Anderson and S. Steiner. (2003) Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **3,** 422-32.

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

PPI

Overlap of Four Plasma Proteome Datasets

(Number of NR proteins)



- 46 proteins in all four lists
- 195 proteins in 2 or more lists
- 1175 NR proteins total

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

Some Proteins of H_Plasma_195

- adiponectin (involved in the control of fat metabolism and insulin sensitivity),
- atrial natriuretic factor (a potent vasoactive substance synthesized in mammalian atria and thought to play a key role in cardiovascular homeostasis),
- various cathepsins (D, L, S),
- centromere protein F (involved in chromosome segregation during mitosis),
- creatine kinase M chain (an abundant muscle enzyme),
- glial fibrilary acid protein (GFAP: distinguishes astrocytes from other glial cells),
- psoriasin (S-100 family, highly up-regulated in psoriatic epidermis),
- interferon-induced viral-resistance protein MxA (confers resistance to influenza virus and vesicular stomatitis virus),
- melanoma-associated antigen p97 (a proposed cancer marker also expressed in multiple normal tissues),
- mismatch repair protein MSH2 (involved in post-replication mismatch repair, and whose defective forms are the cause of hereditary non-polyposis colorectal cancer type 1),

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

PPI

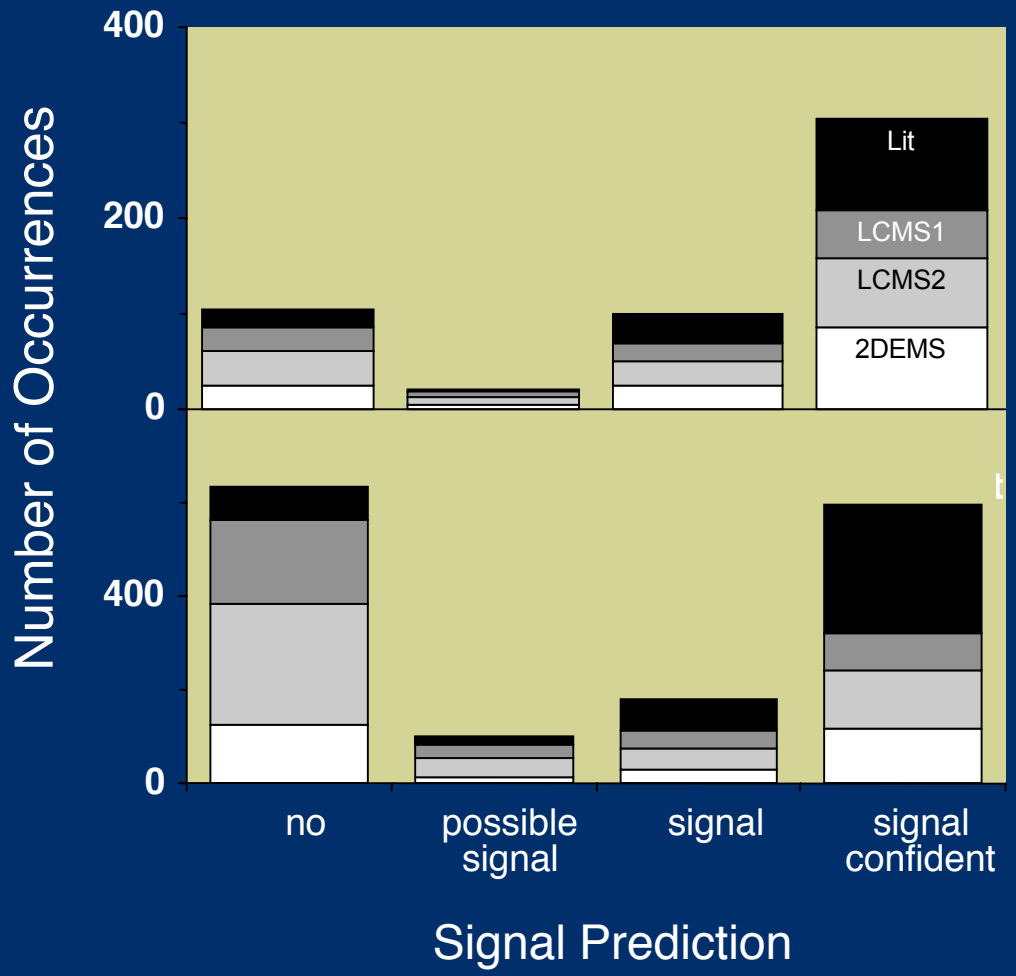
Some Proteins of H_Plasma_195

- oxygen regulated protein (which plays a pivotal role in cytoprotective cellular mechanisms triggered by oxygen deprivation),
- peroxisome proliferator-activated receptor (PPAR) binding protein (which plays a role in transcriptional coactivation),
- prostate-specific antigen (a protease involved in the liquefaction of the seminal coagulum, and one of the few successful cancer diagnostics),
- selenoprotein P (contains selenocyteines encoded by the opal codon, UGA),
- signal recognition particle receptor alpha subunit (an integral membrane protein ensuring, in conjunction with srp, the correct targeting of the nascent secretory proteins to the endoplasmic reticulum membrane system),
- squamous cell carcinoma antigen 1 (which may act as a protease inhibitor to modulate the host immune response against tumor cells),
- V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (the receptor for stem cell factor).

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

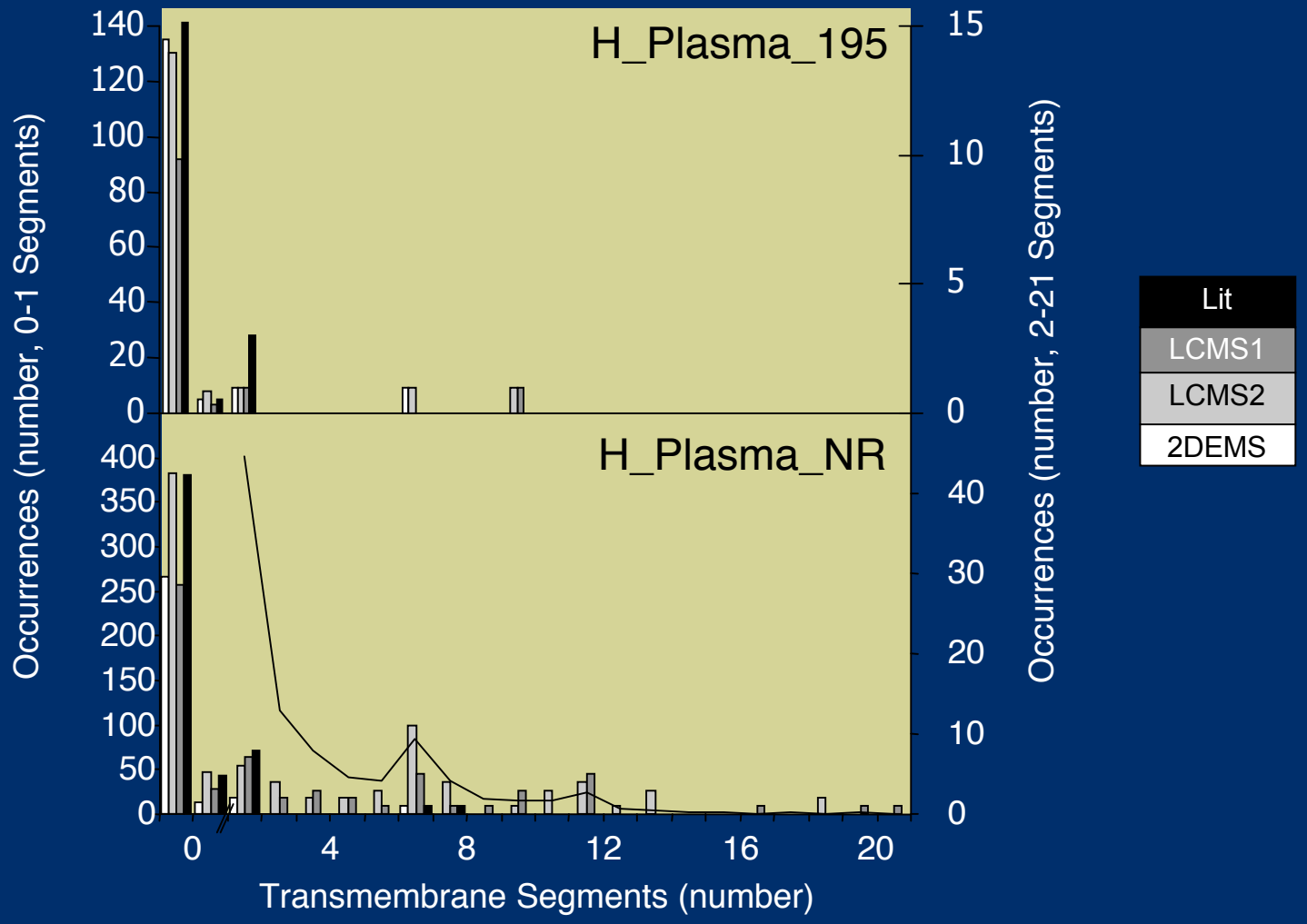
PPI

Signal Sequences in the Plasma Proteome



From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

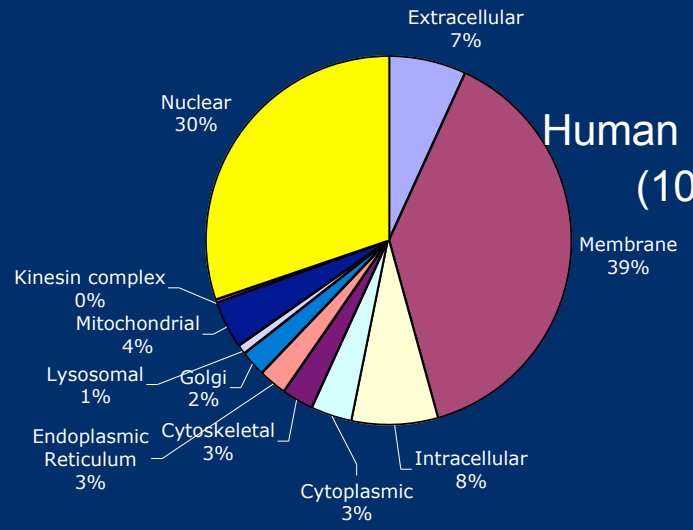
Predicted Transmembrane Segments



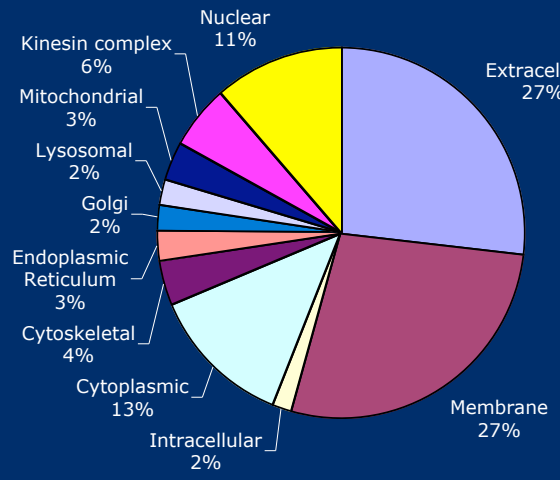
From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

PPI

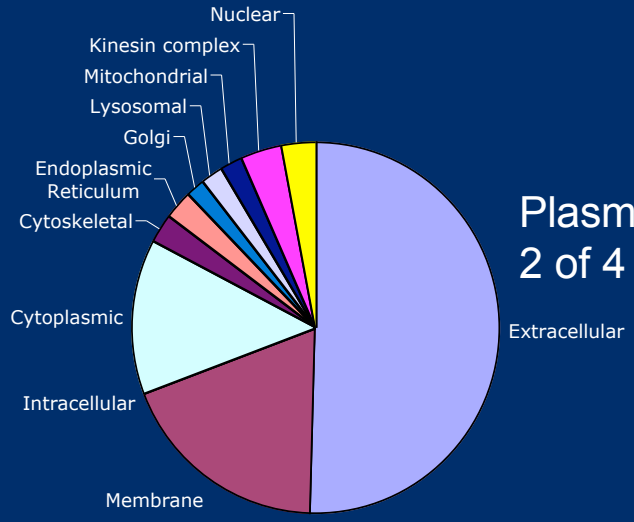
GO Component Annotations for Subsets of the Human Proteome



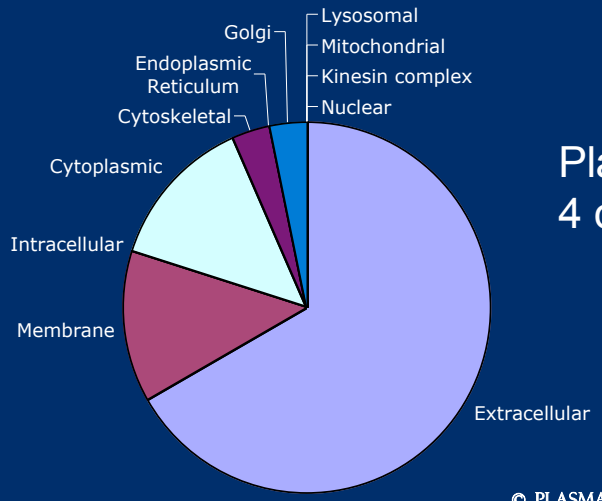
Human Proteome (10,911)



Plasma 1 of 4 (1,175)



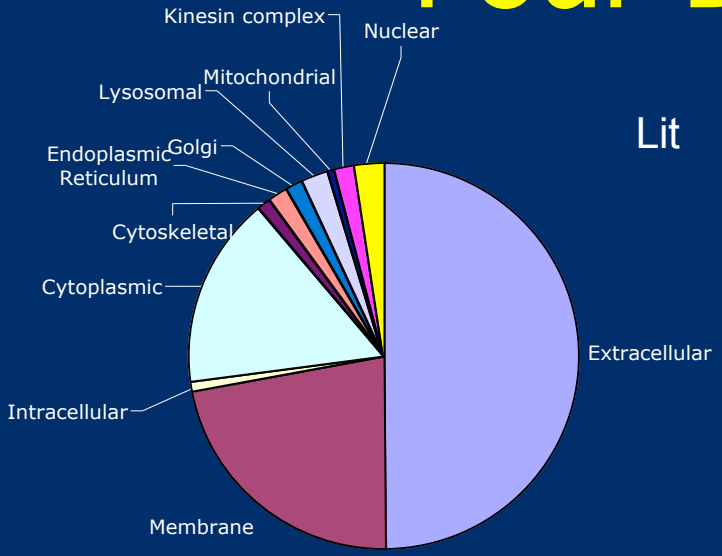
Plasma 2 of 4 (195)



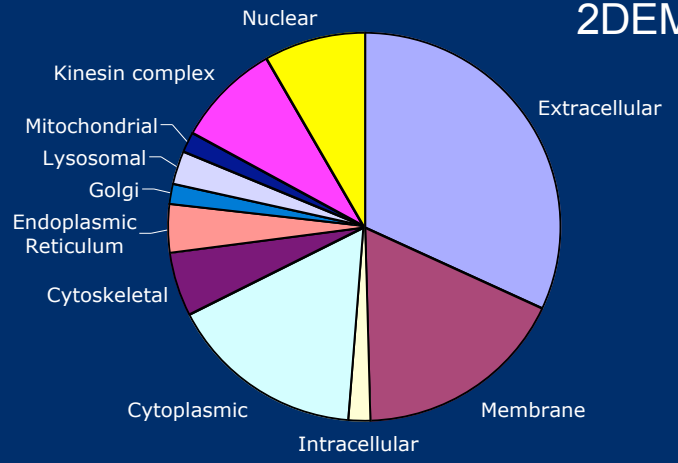
Plasma 4 of 4 (46)

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

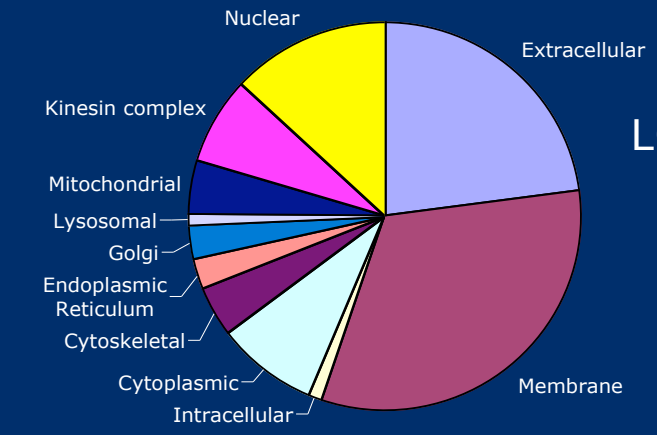
GO Component Annotations for Four Data Sources



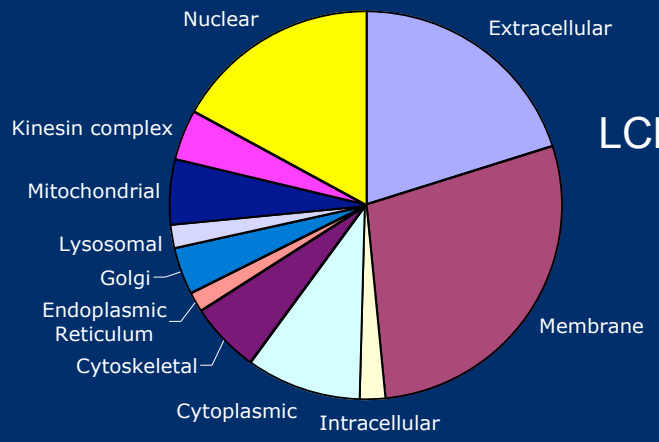
Lit



2DEMS



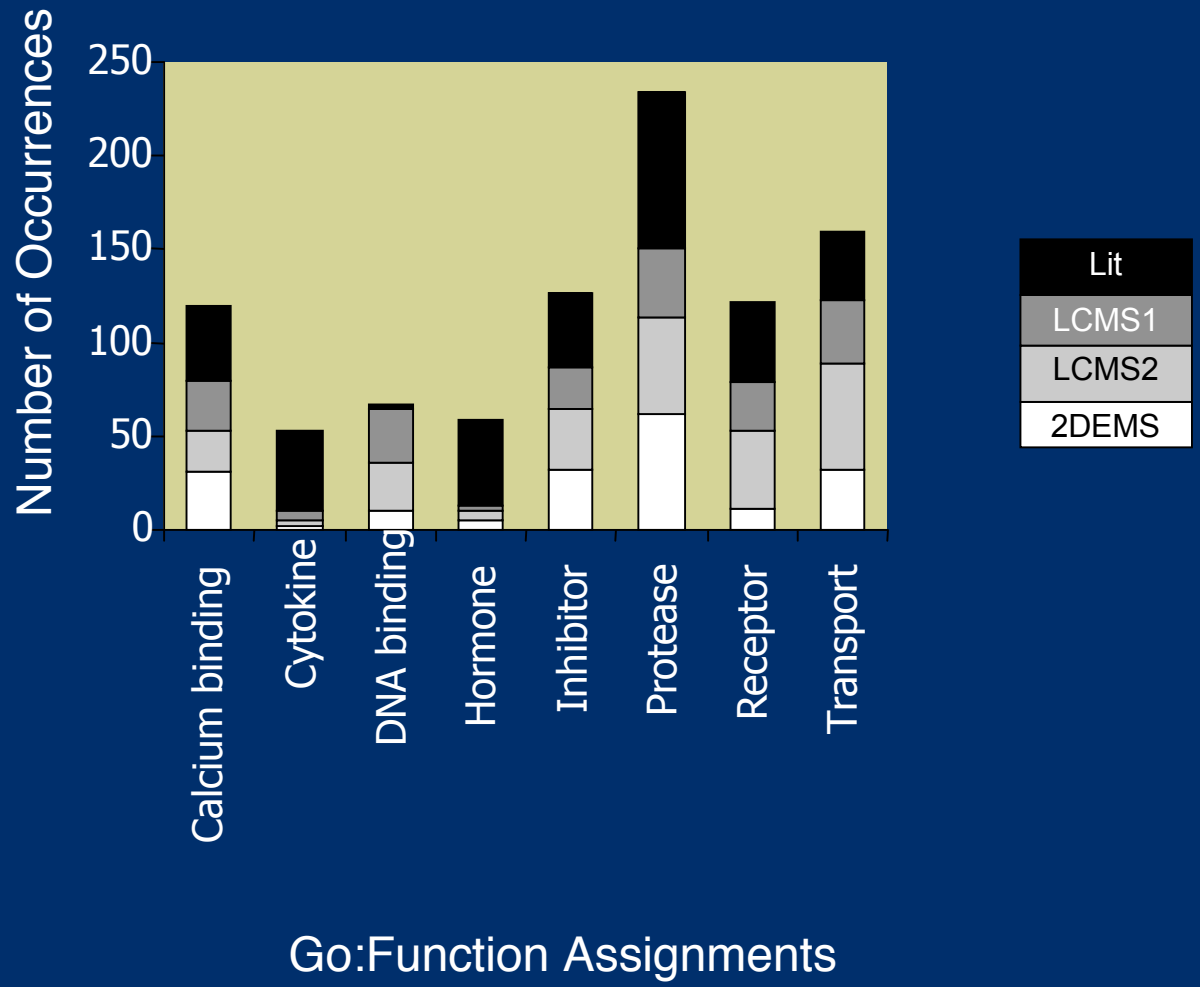
LCMS1



LCMS2

From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

GO Functional Categories of Proteins in H_Plasma_NR



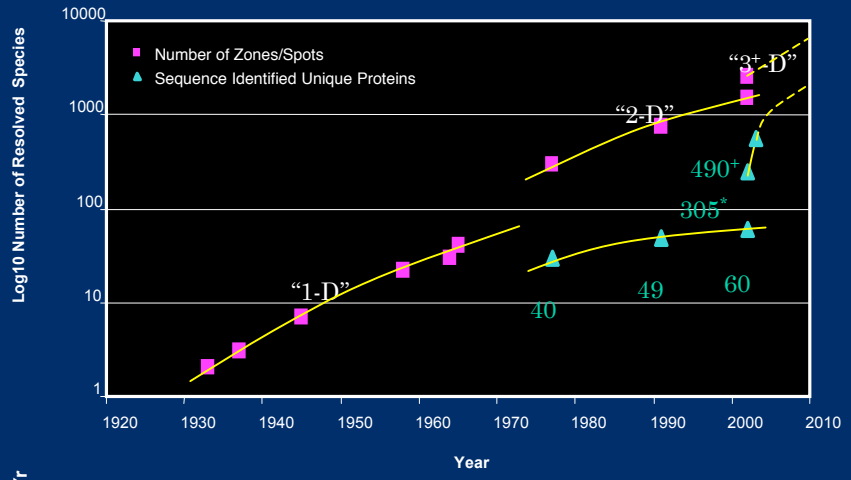
From: The Human Plasma Proteome: A Non-Redundant List Developed by Combination of Four Separate Sources, N. L. Anderson et al, Molec. Cell Proteomics, in press

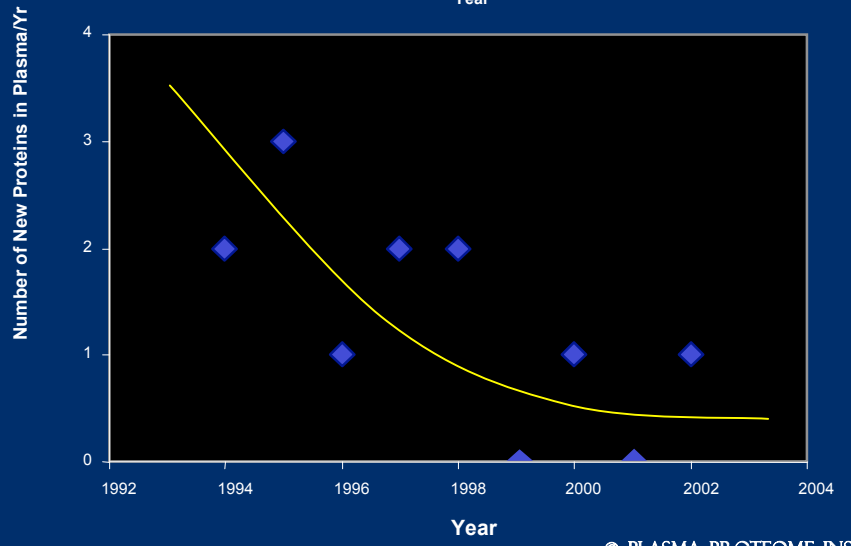
Reasons for Optimism re Discovery of Protein Disease Markers

- Discovery-type proteomics can now detect 500-1,000 distinct proteins in serum
  - Steady improvements in LC-LC/MS-MS
- Many of these proteins fall in classes likely to be informative re tissue status
  - Secreted proteins, extracellular domains of plasma membrane proteins, leaked intracellular proteins
- An open-source database of proteins of proteins observed in plasma is emerging
- Single protein markers are not required
  - Disease associated panels exist where no accurate single-protein disease marker is available
- A series of measurements over time detects change more accurately than comparison with simple reference interval

More New Proteins = Fewer New Diagnostics?

- The gap between what can be measured on a lab scale and what can be used effectively in clinical diagnostics is widening
- The number of tests for new proteins in plasma approved by US FDA over the last decade has declined, and now approaches zero
- Problems include
  - Technology mismatch
  - Regulatory costs
  - Demonstration of medical value
  - Medical economics



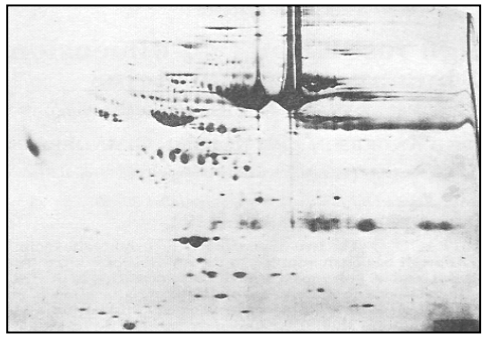


PPI

Challenges Facing Marker/Diagnostic Proteomics: Translation into Diagnostic Tests

- Lack of protein measurement platforms geared to validation (high-throughput, low-cost)
- Access to large, well-organized sample sets for validation
- Falling rate of new protein tests over last decade
- Low expectation of diagnostic profitability impairs commercial investment
- Potential IP traffic jam

A Major Technology Gap Exists Between Discovery and Routine Diagnostic Proteomics



?

Discovery		Routine IVD
50-700	# proteins	1-20
$1,000-$10M	$ per analysis	$2-100
25-50%	CV	3-5%
4-52 wks	time required	~15 min
2-50	# samples	100-10,000

PPI

"The appealing notion that research advances travel from bench to bedside is laudable, but conceptually flawed. Even though the U.S. Congress fully anticipates that funding to the National Institutes of Health (NIH) will result in advances in clinical medicine and that other forces, presumably non-governmental, will translate the latest in exciting science into health technologies, under the system of healthcare we have today, this advancement is not likely to happen."

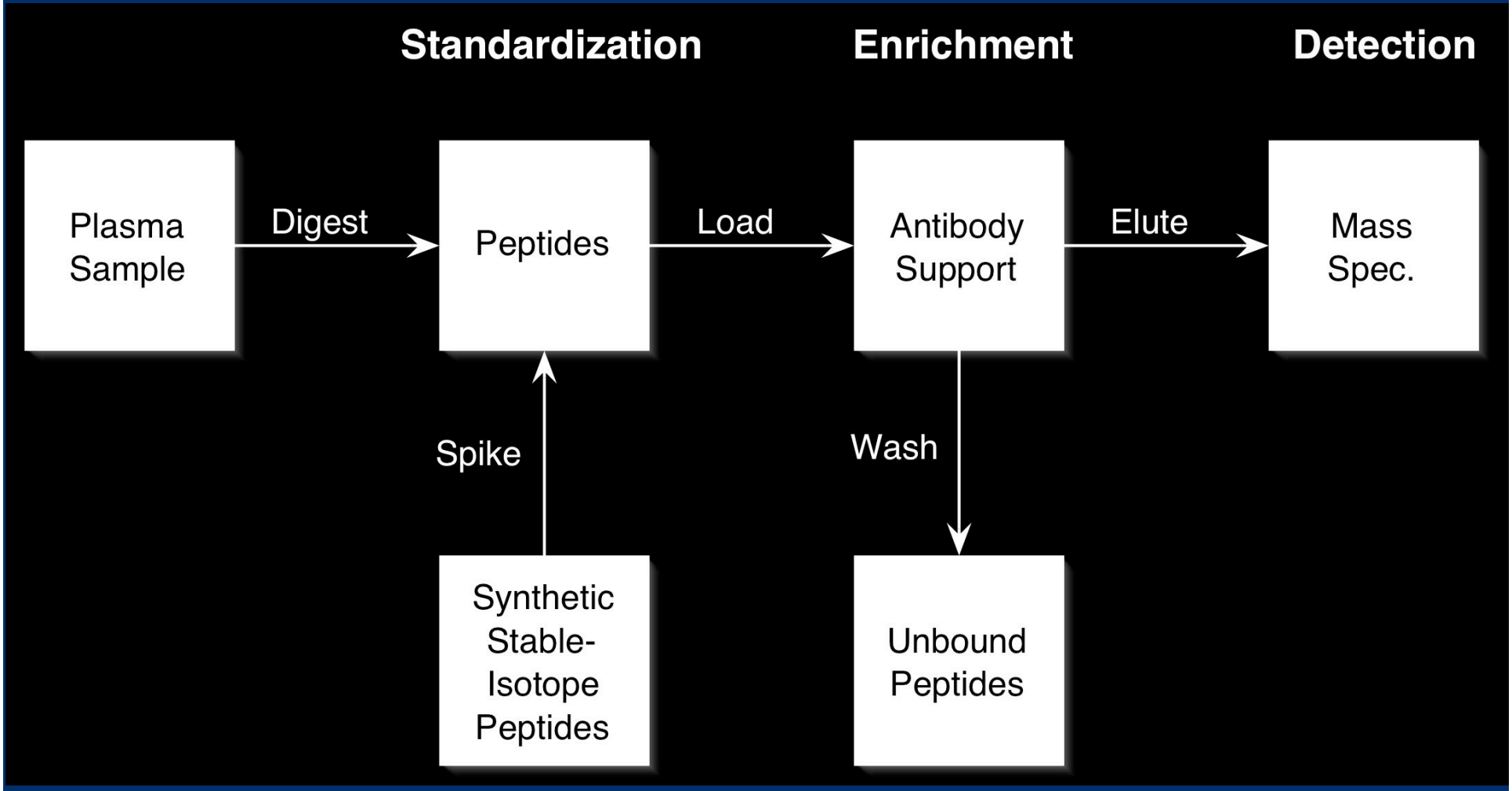
Floyd Bloom
President, AAAS
Science 300:1680-1685 (2003)

Towards a Flexible, High-Throughput Quantitative MS Platform

- Goals
  - Assay pre-selected proteins, i.e., identified candidate disease markers for validation studies
  - Combine specific enrichment with MS quantitation
  - Increase speed and throughput by decreasing reliance on gradient LC
  - Avoid method bias towards a class of proteins

SISCAPA*: A New Method Combining The Specificity of MS Detection with Sensitivity of Antibody Capture

(SISCAPA = Stable Isotope Standards with Capture by Anti-Peptide Antibody)



* patent pending

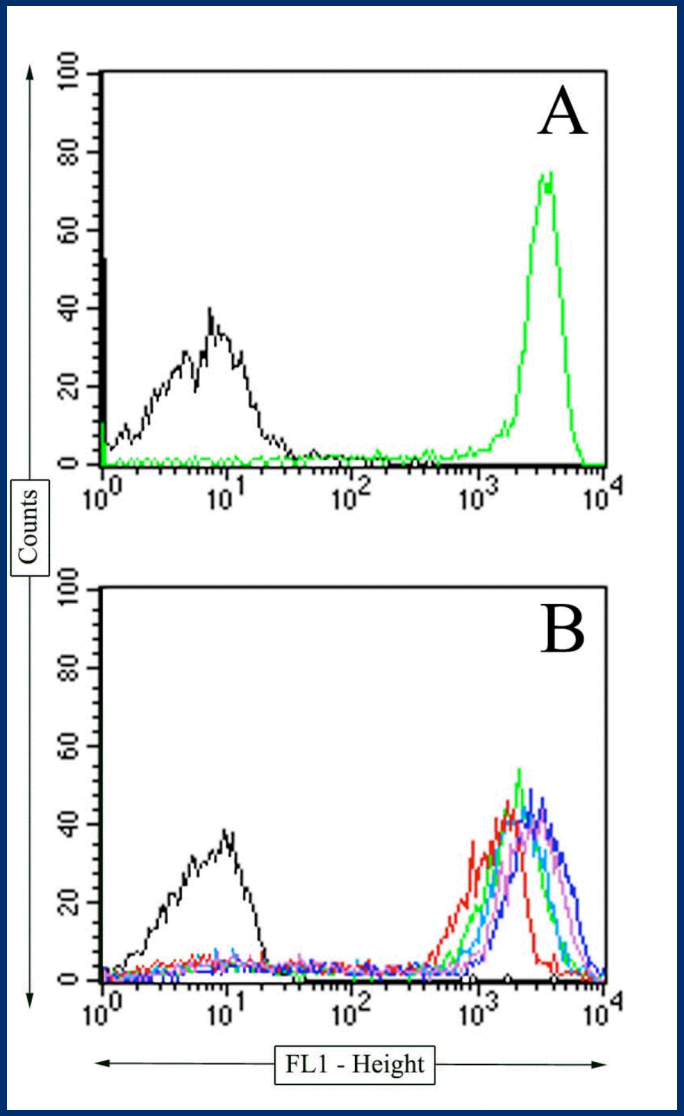
PPI

Selection of Peptides and Anti-Peptide Antibodies

- 10,203 peptides generated *in silico* from 237 known plasma sequences
- Monitor peptides selected based on physical/antigenicity parameters. >80% of proteins had 1 or more "good" peptides.
- Selected peptides synthesized, coupled to albumin carrier, used in 38-day rabbit immunization protocol. Ab's affinity purified on same peptide immobilized on agarose.

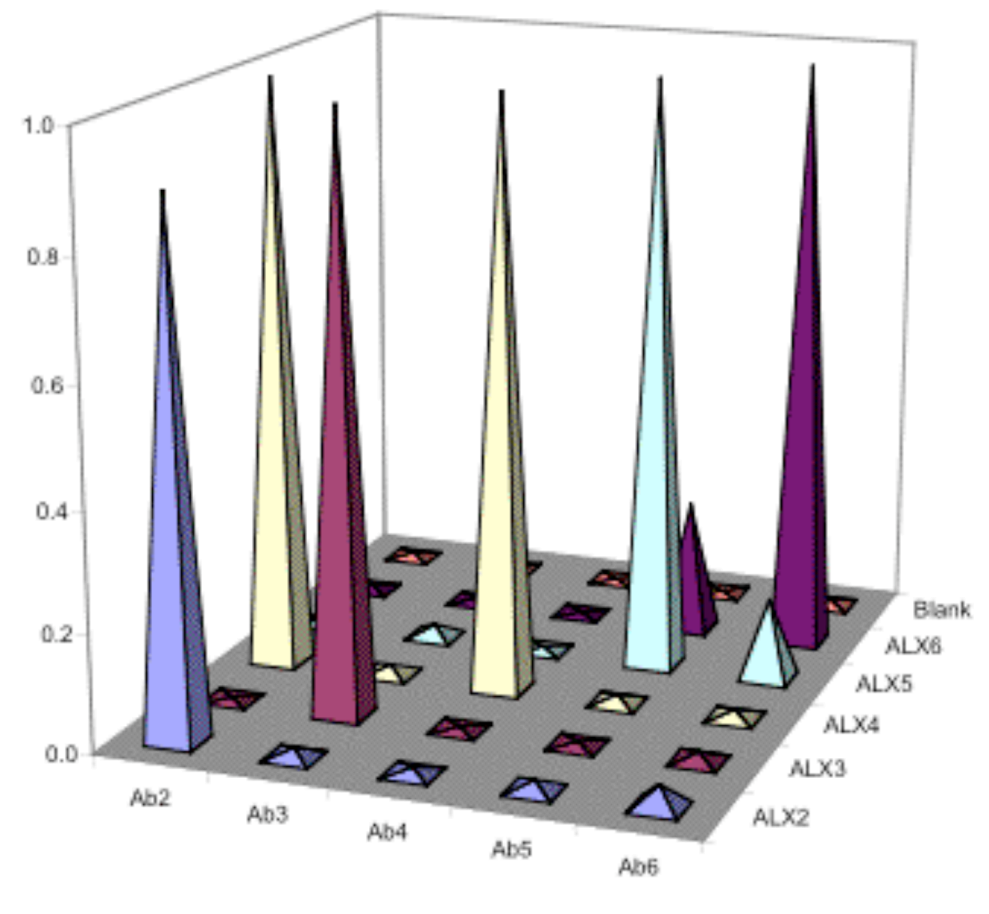
		Identification code	
Protein	Peptide	Immunogen	Rabbit Antibody
Interleukin-6 (IL-6)	EALAENNLNLPKGSGC	IMM2	Ab 2
Hemopexin (Hx)	NFPSPVDAAFRGSGC	IMM3	Ab 3
α_1-Antichymotrypsin (AAC)	EIGELYLPKGSGC	IMM4	Ab 4
Tumor necrosis factor alpha (TNFα)	DLSLISPLAQAVRGSGC	IMM5	Ab 5
Tumor necrosis factor alpha (TNFα)	CGSGDLSLISPLAQAVR	IMM6	Ab 6

Flow cytometric detection of rabbit anti-peptide antibodies coupled to POROS® –Protein G beads



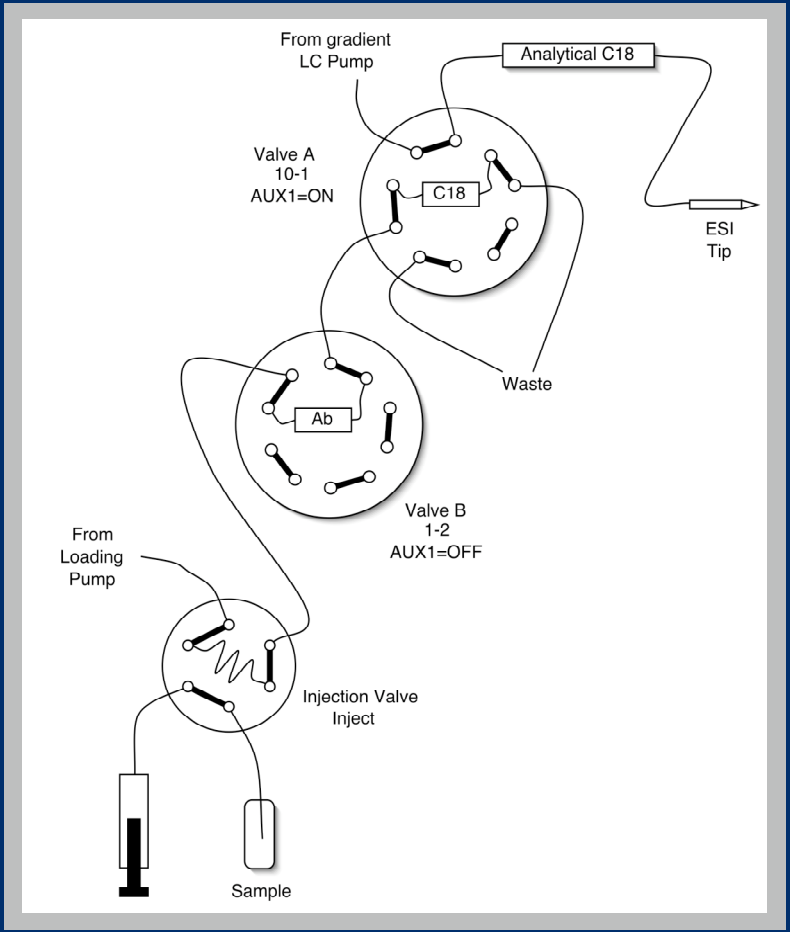
- A: Black profile; POROS® protein G beads incubated with biotinylated protein L and fluorescein-labeled streptavidin (negative control). Green profile; POROS® Streptavidin beads incubated with biotinylated Protein L and detected with fluorescein-conjugated streptavidin (positive control).
- B: Detection of covalently coupled rabbit anti-peptide antibodies on POROS® Protein G beads. Beads covalently coupled with 5 rabbit affinity-purified Abs incubated first with biotinylated Protein L, followed by detection with fluorescein-conjugated streptavidin.

Relative binding of Alexa Fluor®488-labeled peptides to five affinity purified anti-peptide antibodies



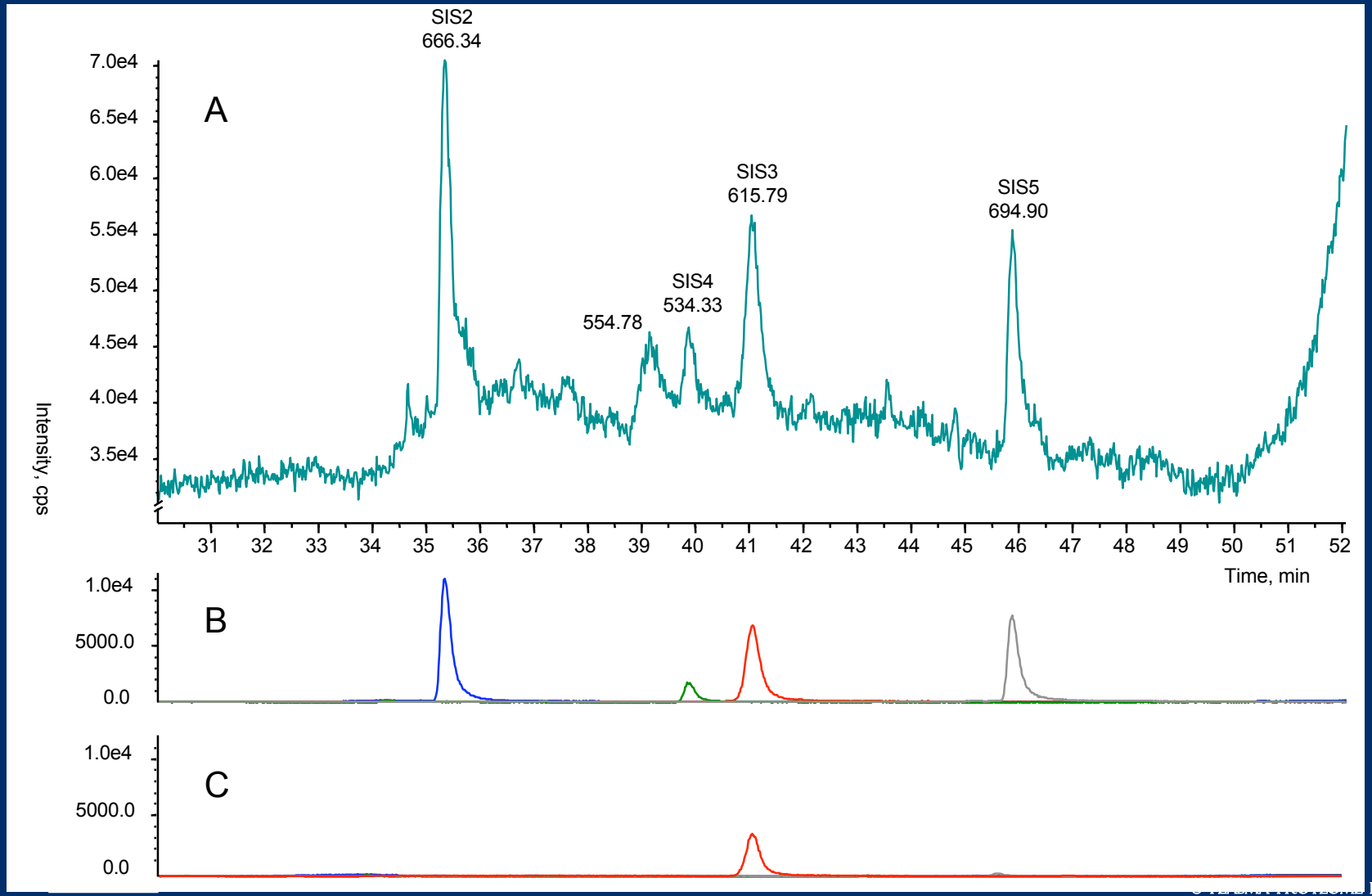
- The binding of four peptides, ALX 2-6, by POROS® affinity matrices containing either their homologous (specific) or heterologous (non-specific) antibodies was analysed. The values for each antibody are normalized to the maximum fluorescence intensity for that antibody. Each value is the median fluorescence intensity for 1200 flow cytometer events. Ab's 5 and 6 are against opposite ends of the same peptide.

SISCAPA: Initial LC-MS Experiments



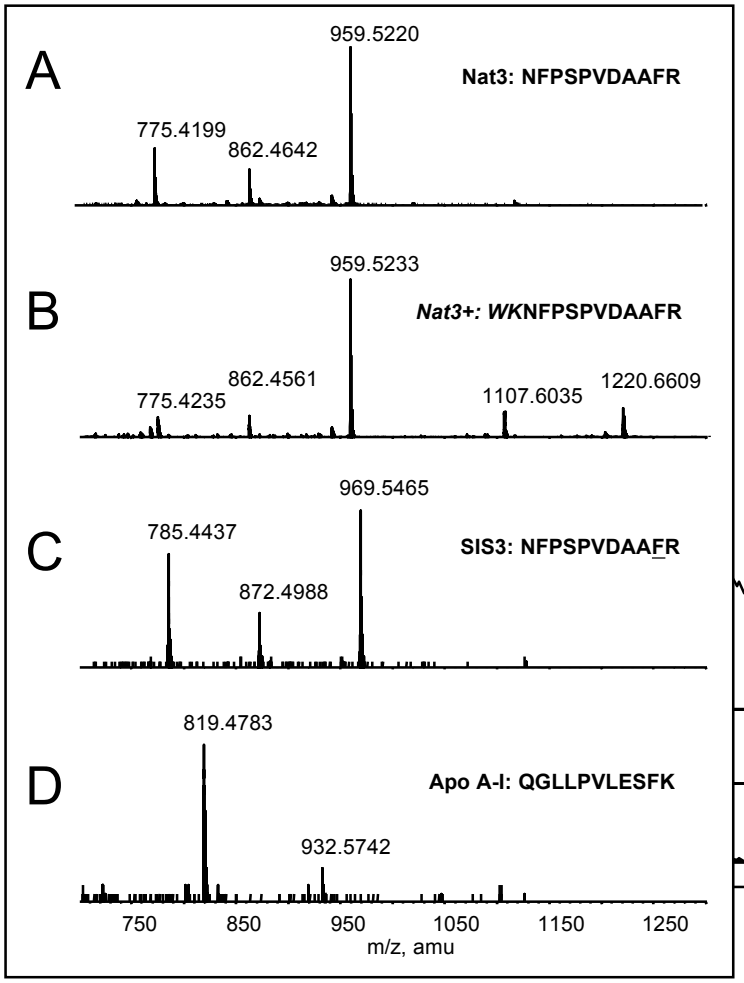
- 100nl Ab-POROS columns (100u ID x 1cm)
- Acid-eluted peptide (from Ab column) captured on C18 and eluted into MS by gradient LC
- N. L. Anderson et al, Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA), J. Proteome Research, in press

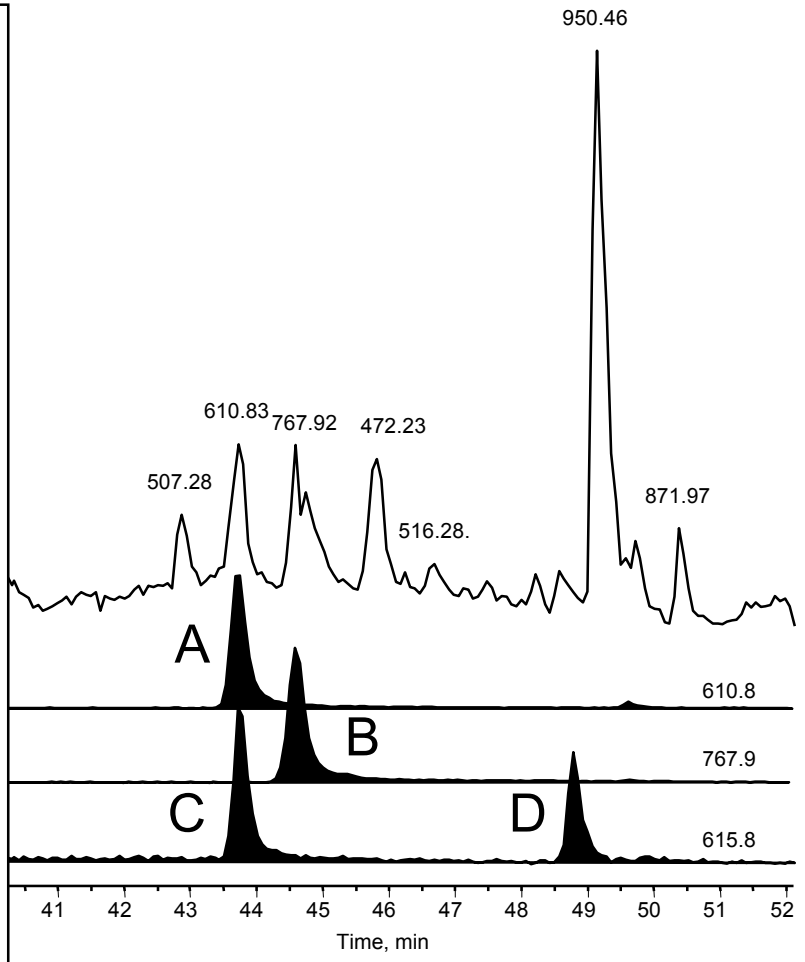
Ab Capture from SIS Peptide Mixture with Selected Ion Monitoring



PPI INSTITUTE

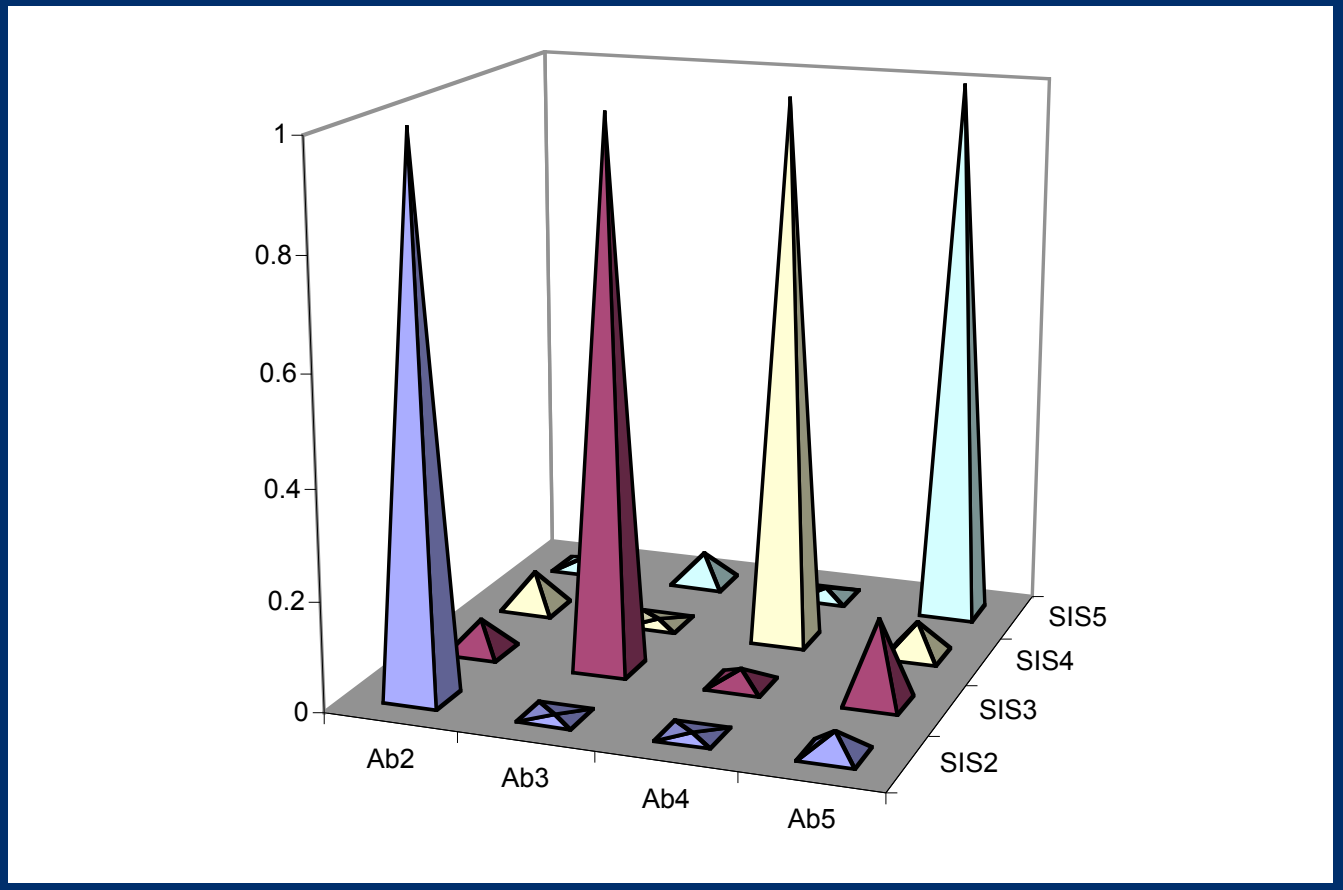
Enrichment and Quantitation of a Hemopexin Monitor Peptide by SISCAPA





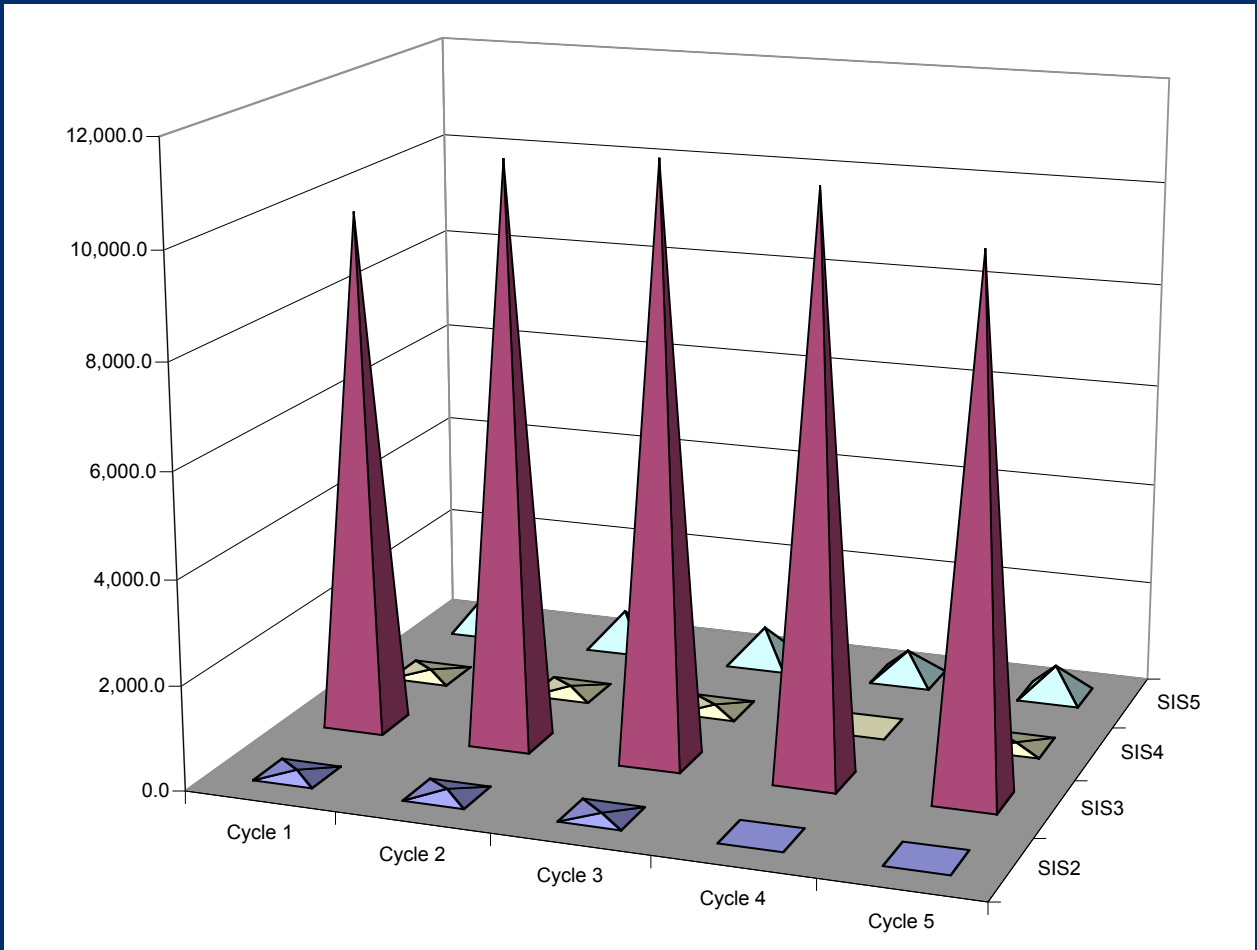
Relative Quantities of Four SIS Peptides Bound by Four Anti-Peptide Antibodies, Using Two-stage MS Selection (SRM)

Average Peptide Enrichment by Ab > 100-fold



The signals (vertical axis) for each antibody are normalized to the largest signal for that antibody

Rabbit Polyclonal Anti-Peptide Ab's Can Be Recycled After Acid Elution



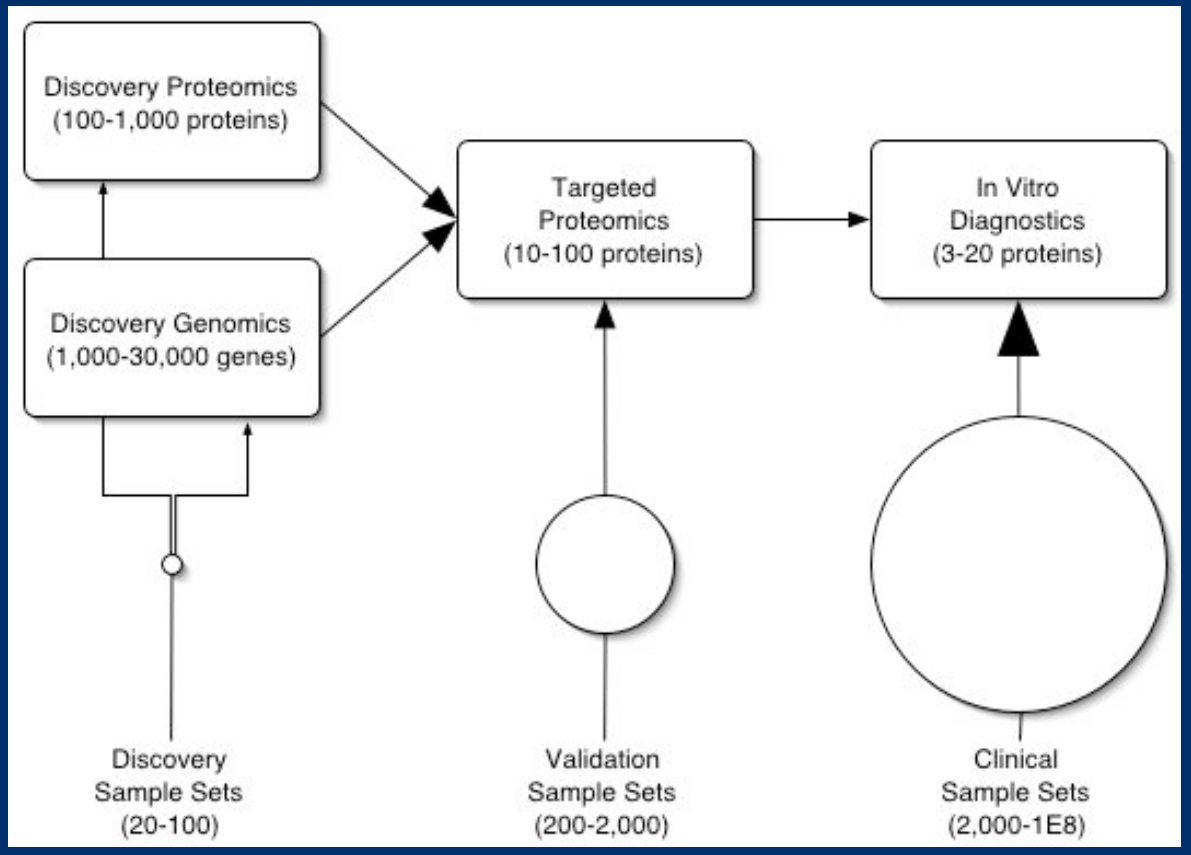
SRM integrated ion current measurements of four SIS peptides eluted on five successive cycles from Ab 3

.

PPI

Targeted Proteomics (e.g., SISCAPA) Enables Marker Validation

- Focus on measurement of identified proteins
- Accepts input from both proteomics and genomics
- Bridges gap between discovery and routine diagnostic use
- Hybrid methods, e.g., SISCAPA



PPI's Plasma Marker Validation Project

- Identify marker protein candidates from any source (literature, proteomics, genomics)
- Develop high-throughput MS-based validation assays
- Assay candidates in large well-characterized plasma/serum sample sets (disease vs control and population studies)
- Place validation data in public domain
- Advance protein measurement technology for plasma

PPI

Acknowledgements

- SISCAPA Experiments
  - Bob Olafson, Derryl Hardy, UVIC-Genome B.C. Proteomics Centre
  - Terry Pearson, Lee Haines, Department of Biochemistry and Microbiology, University of Victoria, B.C, Canada
  - John Rush, Cell Signaling
- Plasma Proteome Database
  - Malu Polanski (PPI)
  - Richard Fagan, Anna Lobley, Inpharmatica Ltd., London
  - Rembert Pieper, Tina Gatlin, present address: The Institute for Genomic Research
  - Radhakrishna S. Tirumalai, Timothy D. Veenstra, Mass Spectrometry Center, U. S. National Cancer Institute
  - Joshua N. Adkins, Joel G. Pounds, Biological Sciences Department, Pacific Northwest National Laboratory
- Graphics
  - Arkitek Studios, Seattle

- LSBC Proteomics Alumni

Ahmed, N
Anderson, N G
Aponte, A
Aponte, J
Braatz, J
Brooke, E
Das, T
Davis, T
Eidbo, E
Esquer, R
Field, E
Gatlin, C
Goodman, J
Hardin, R
Hofmann, J-P
Huang, S-T
Jett, G
Kho, K
Kiersarsky, K
Lennon, J
Love, C
Makusky, J
Matthews, J
McGrath, A
McCrea, C
Michael, S
Miller, S
Mondal, M
Montgomery, R
Myers, T
Norouzi, T
Parmar, P
Piasecki, M
Russo, P
Schatz, C
Seonarain, M
Sims, C
Steiner, S
Stewart, D
Su, Q
Sun, Q
Taylor, A
Taylor, J
Tran, M
Vizzi, B
Wallgren, L
Wang, F
Wannberg, S
Zhou, J

PPI