Proteomics Standards Research Group

sPRG2006: A Proteomics Standard

www.abrf.org/sprg

Proteomics Standards Research Group

Proteomics Standards Research Group (sPRG) Members

Phil Andrews - University of Michigan

David Arnott - Genentech, Inc.

Mary Ann Gawinowicz - Columbia University

Jeffrey A. Kowalak - National Institutes of Health (Chair)

William S. Lane - Harvard University (EB Liaison)

Kathryn S. Lilley - University of Cambridge

Larry Martin - East West University

Stephen E. Stein - NIST

Proteomics Standards Research Group

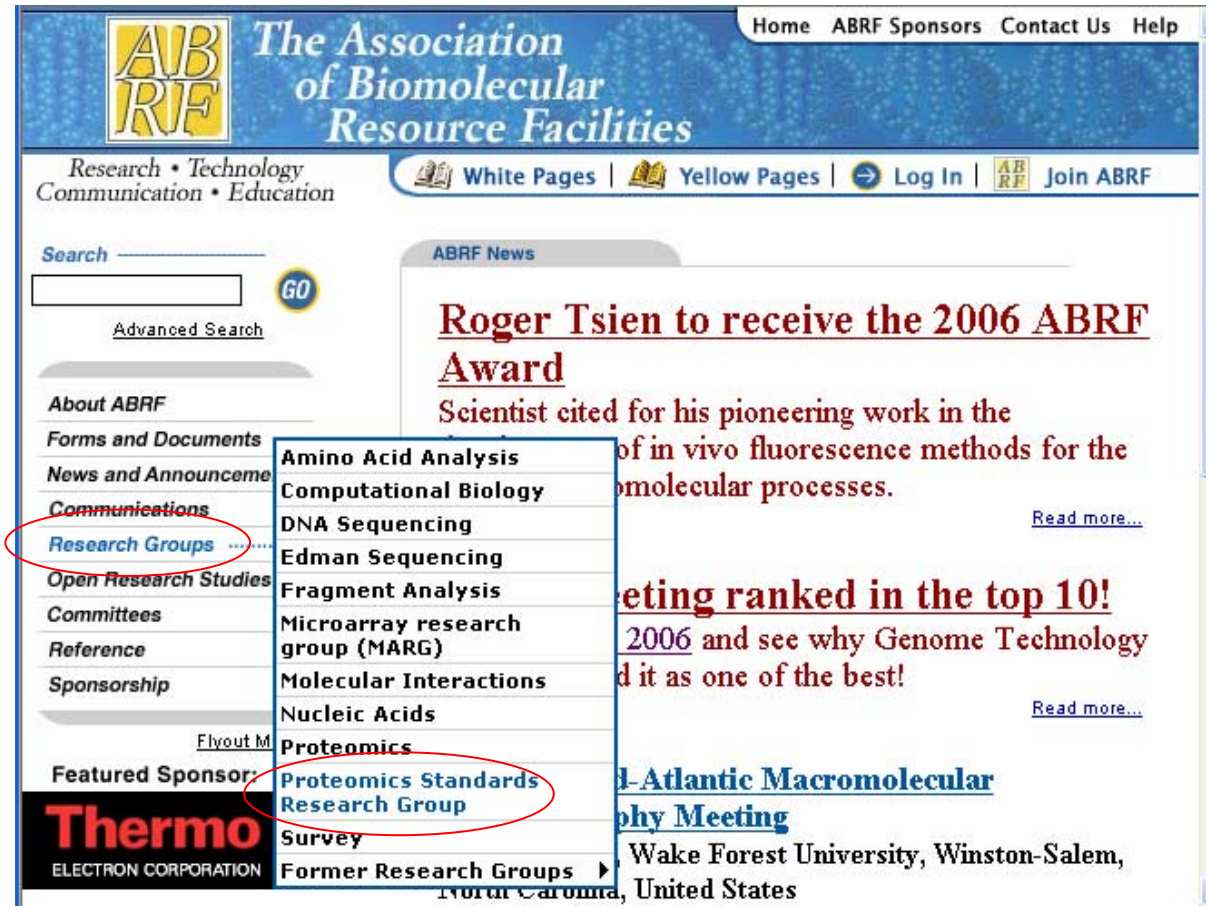
Mission Statement

- Promote and support the development and use of standards in Proteomics.
- Identify and implement technical standards that reflect the ABRF's commitment to accuracy, clarity, and consistency.
- Standards include, but are not limited to, reference materials, data sets, conditions and procedures that give researchers independent criteria to evaluate their abilities to produce predictable, consistent results.
- Strongly support efforts for standardization of the recording and reporting of proteomics experiments.

Proteomics Standards Research Group

sPRG Website

www.abrf.org/sprg



Proteomics Standards Research Group

Proteomics Standards in the Literature

Keller, A., Purvine, S., Nesvizhskii, A.I., Stolyar, S., Goodlett, D.R. and E. Kolker, E. (2002) “Experimental Protein Mixture for Validating Tandem Mass Spectral Analysis” *OMICS* **6:** 207-212

Authors described a data set of peptide tandem mass spectra generated from a control mixture of18 purified proteins that can be employed to evaluate strategies for tandem mass spectral analysis.

Purvine, S., Picone, A.F., and E. Kolker, E. (2004) “Standard Mixtures for Proteome Studies” *OMICS* **8:** 79-92

Authors introduced a mixture of 23 peptides and 12 protein digests to serve as possible standard for proteome analyses.

Materials

- Human proteins, purified from their biological source or recombinantly expressed.
- Subjected to multiple analytical methodologies, e.g., 1D PAGE, IEF, and RP-HPLC, in order to assess required levels of purity ($\geq$ 95%).
- Protein concentration was determined by amino acid analysis.
- Five picomole aliquots of each protein were combined and lyophilized in a 1 mL polypropylene tube.

ABRF

Proteomics Standards Research Group

Standard Validation

- Prior to distribution to participating labs, the prototype protein standard was distributed to the sPRG committee members and analyzed by 5 different member laboratories for validation.
- In three RG laboratories, the standard was analyzed using a shotgun approach by digestion with trypsin of the entire protein mixture followed by on-line 1D LC-MS/MS, on-line 2D LC-MS/MS, off-line capillary RP-HPLC, TOF/TOF
- In two RG laboratories, the protein mixture was partially separated by 1D SDS-PAGE, followed by in-gel tryptic digestion and on-line 1D LC-MS/MS.

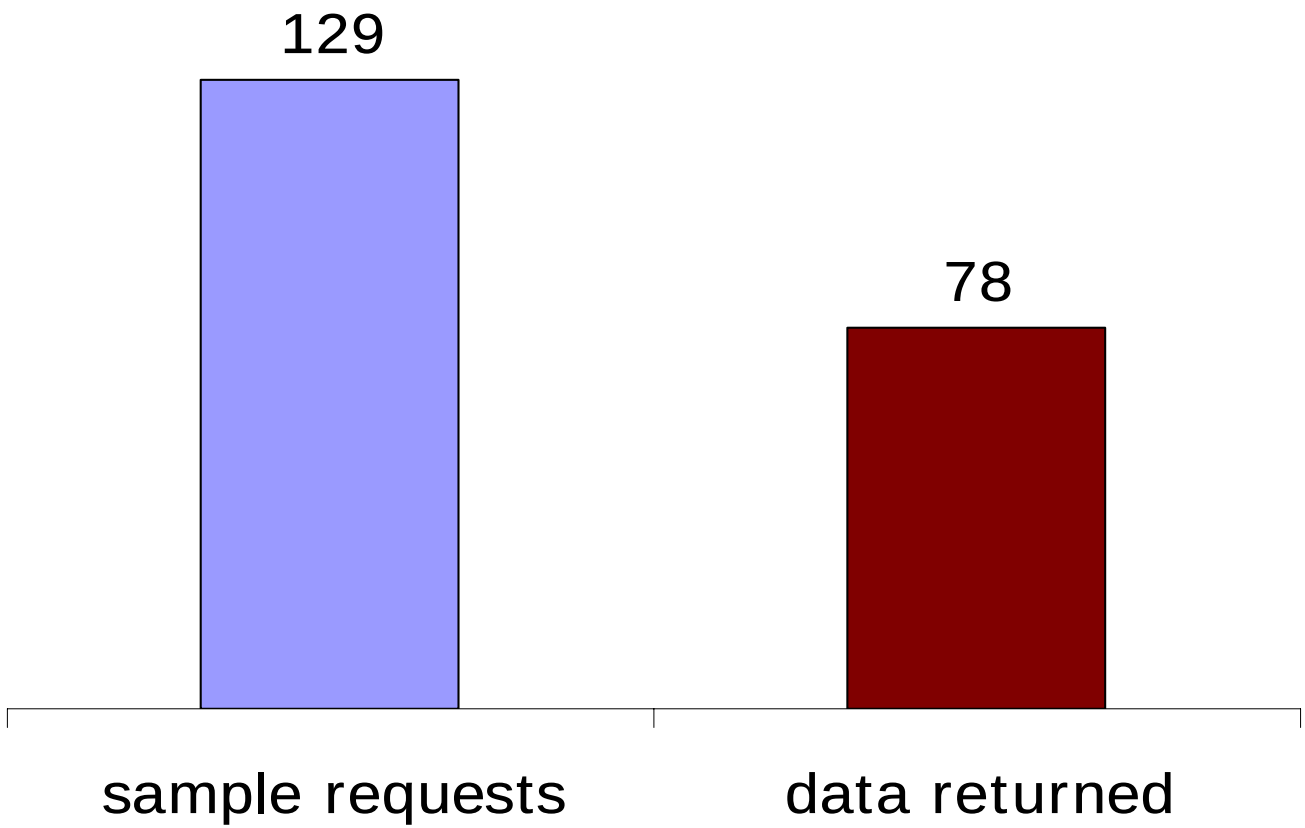
Proteomics Standards Research Group

Results Submission

- Results were submitted anonymously through an on-line survey and included experimental details, instrument(s) used, and a complete list of all proteins identified with accession numbers.
- A web site (www.proteomecommons.org/dev/abrf/) was provided for the voluntary submission of raw mass spectral data to be made available for future studies.

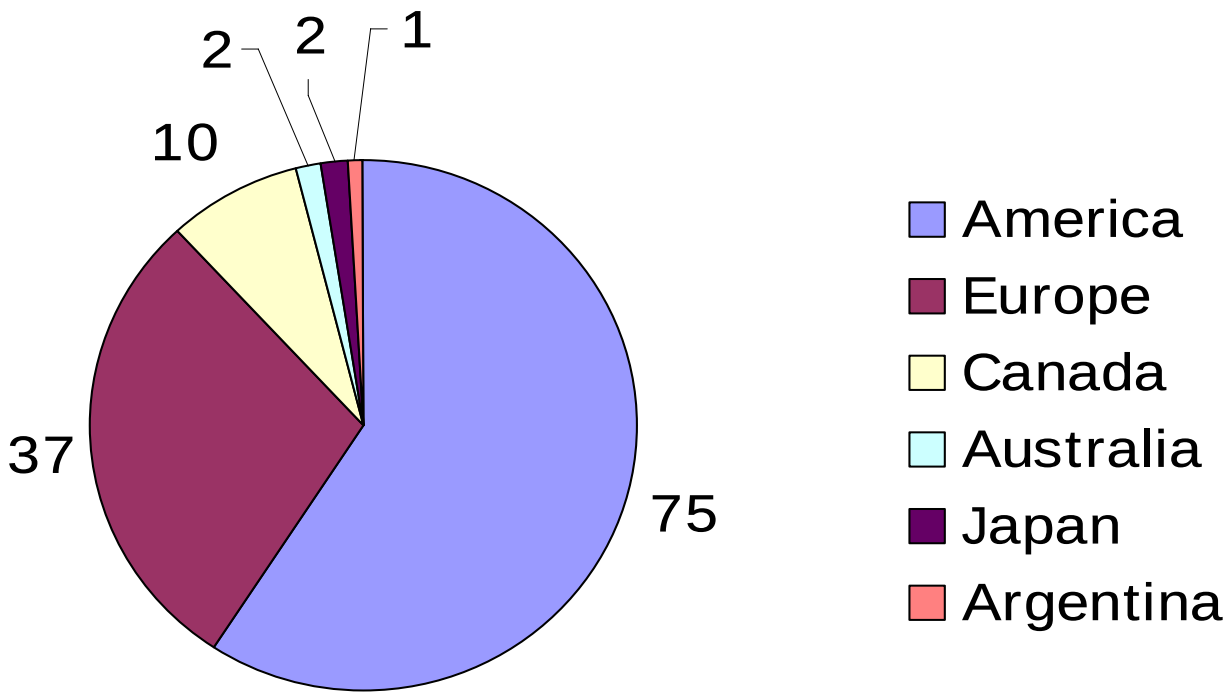
Proteomics Standards Research Group

60.5% Data Return



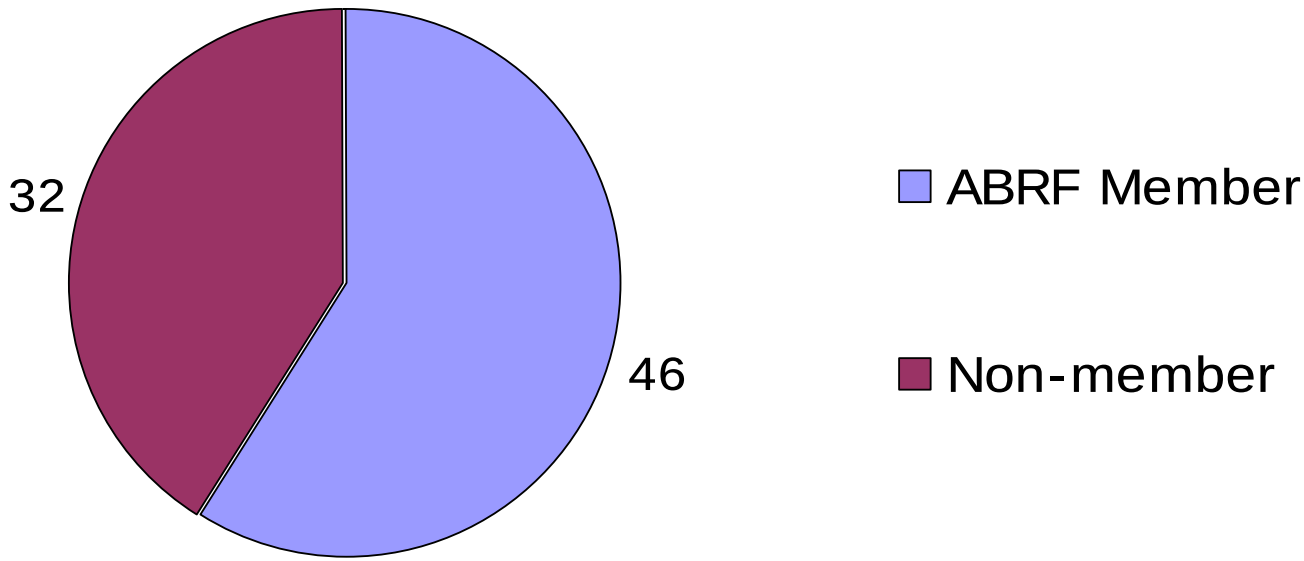
Proteomics Standards Research Group

sPRG2006 Requester Demographics



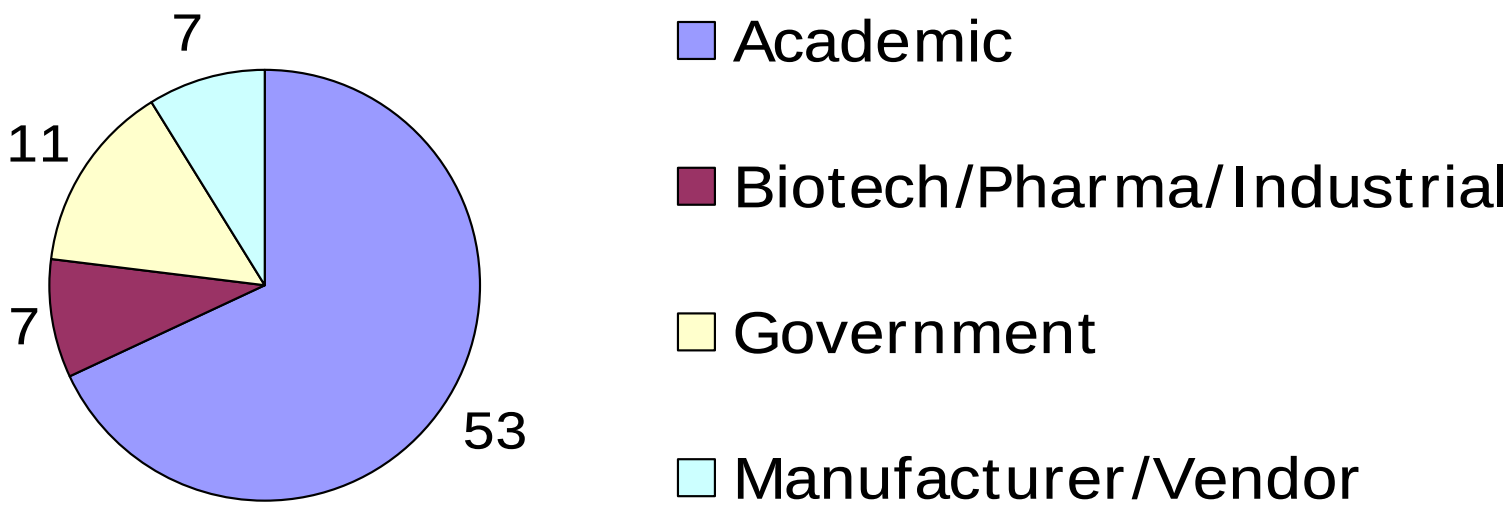
Proteomics Standards Research Group

sPRG2006 ABRF Membership



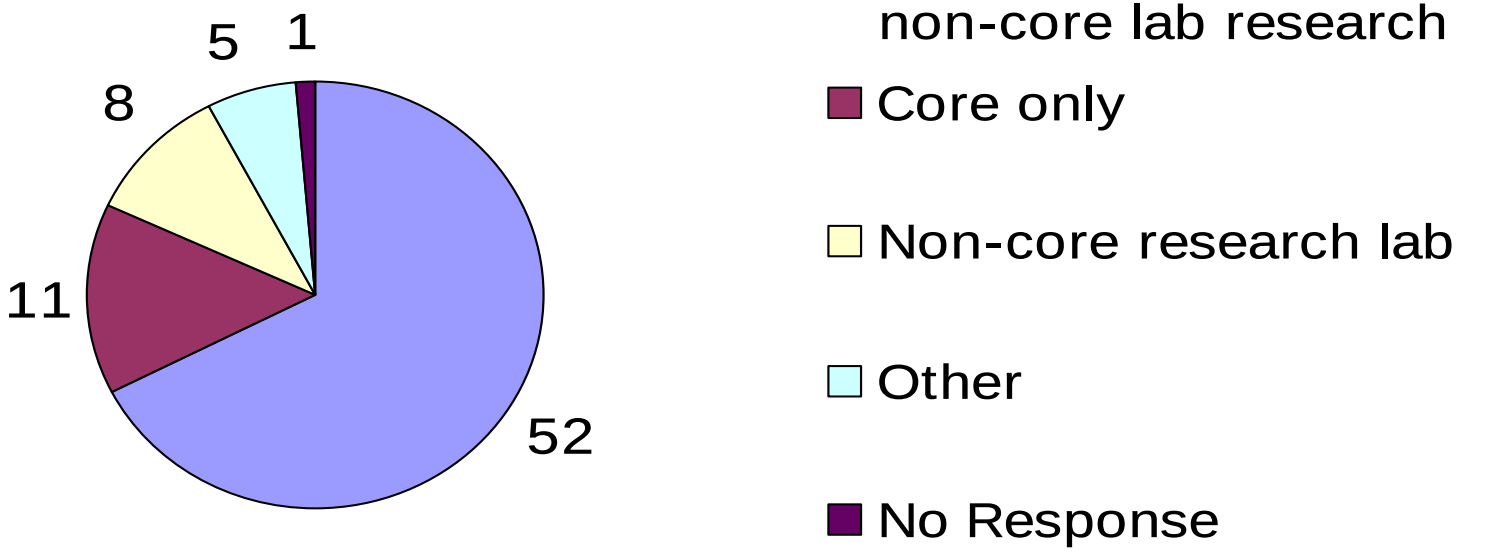
Proteomics Standards Research Group

sPRG2006 Laboratories



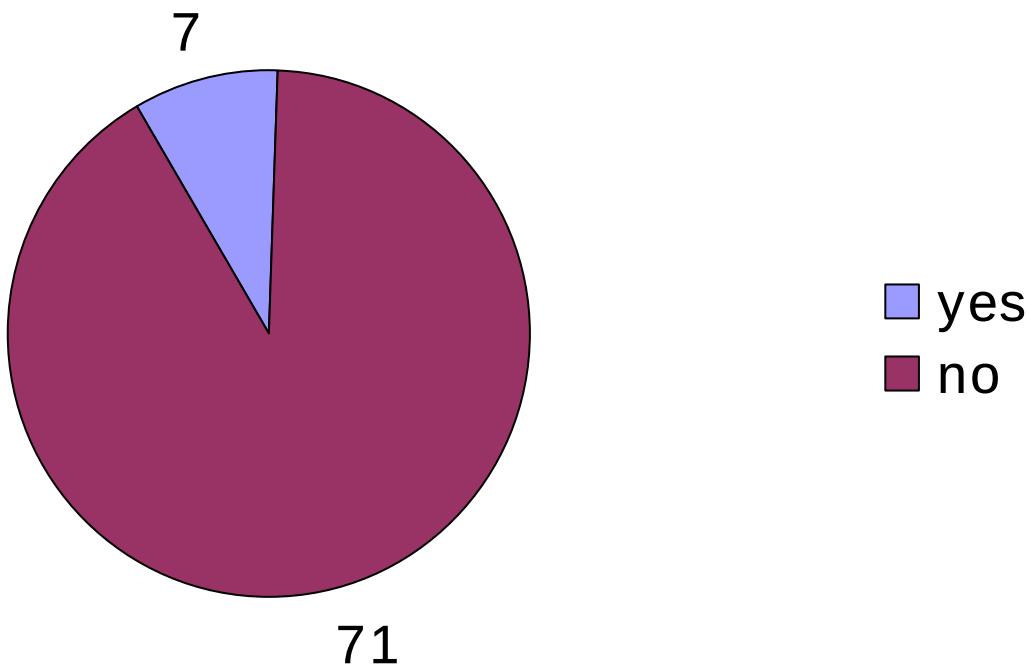
Proteomics Standards Research Group

sPRG2006 Facility Type



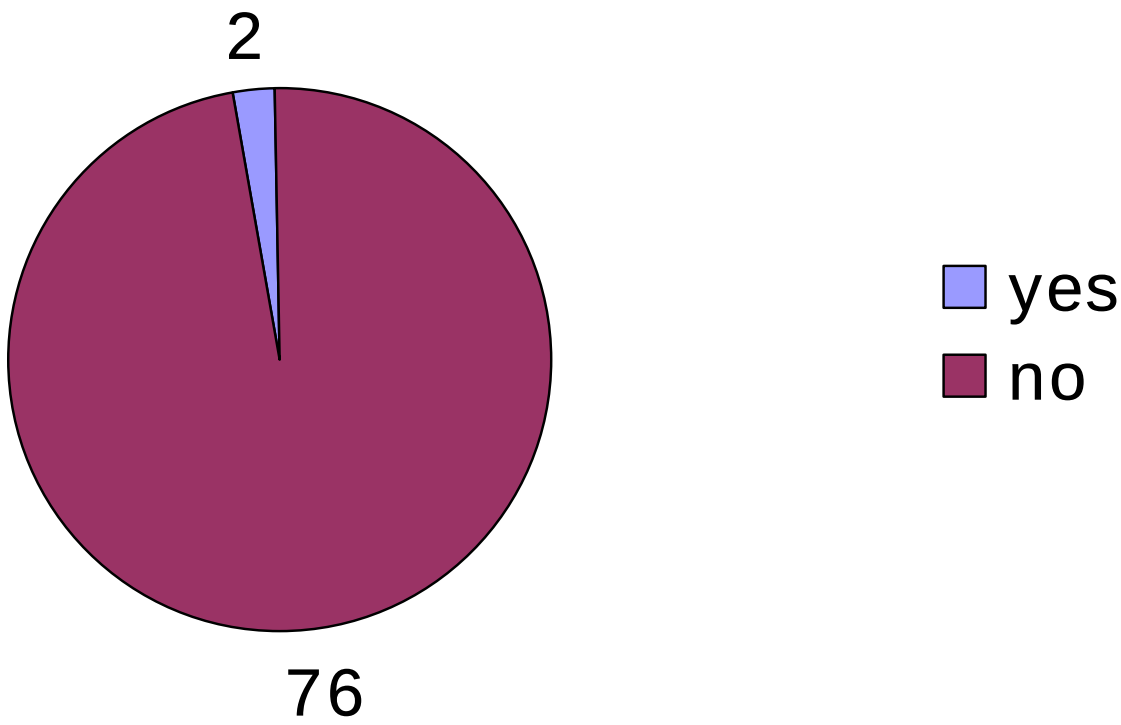
Proteomics Standards Research Group

Did you receive more than 1 sample?



Proteomics Standards Research Group

Did you have prior knowledge of another lab's results?



sPRG2006 Proteomics Standard

www.abrf.org/sprg

Proteomics Standards Research Group

Accsn	Entry	Protein
P01008	ANT3_HUMAN	Antithrombin-III precursor
P02768	ALBU_HUMAN	Serum albumin
P08758	ANXA5_HUMAN	Annexin A5
P55957	BID_HUMAN	BH3 interacting domain death agonist
P00167	CYB5_HUMAN	Cytochrome b5
P99999	CYC_HUMAN	Cytochrome c
P04040	CATA_HUMAN	Catalase
P62937	PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A
P02741	CRP_HUMAN	C-reactive protein
P00915	CAH1_HUMAN	Carbonic anhydrase 1
P08311	CATG_HUMAN	Cathepsin G
P01031	CO5_HUMAN	Complement C5
P00918	CAH2_HUMAN	Carbonic anhydrase 2
P07339	CATD_HUMAN	Cathepsin D
P06732	KCRM_HUMAN	Creatine kinase M-type
P15559	NQO1_HUMAN	Quinone reductase 1
P01133	EGF_HUMAN	Pro-epidermal growth factor
P05413	FABPH_HUMAN	Fatty acid-binding protein
P06396	GELS_HUMAN	Gelsolin
P08263	GSTA1_HUMAN	Glutathione S-transferase A1
P09211	GSTP1_HUMAN	Glutathione S-transferase P
P69905	HBA_HUMAN	Hemoglobin alpha chain
P68871	HBB_HUMAN	Hemoglobin beta chain
P10145	IL8_HUMAN	Interleukin-8
P01344	IGF2_HUMAN	Insulin-like growth factor II
P01343	IGF1A_HUMAN	Insulin-like growth factor IA
P12081	SYH_HUMAN	Histidyl-tRNA synthetase
P02788	TRFL_HUMAN	Lactotransferrin
P41159	LEP_HUMAN	Leptin
P61626	LYSC_HUMAN	Lysozyme C
P00709	LALBA_HUMAN	Alpha-lactalbumin
P61769	B2MG_HUMAN	Beta-2-microglobulin
P02144	MYG_HUMAN	Myoglobin
Q15843	NEDD8_HUMAN	Neddylin
P01127	PDGFB_HUMAN	Platelet-derived growth factor B-chn
Q06830	PRDX1_HUMAN	Peroxiredoxin 1
P16083	NQO2_HUMAN	NRH dehydrogenase [quinone] 2
P02753	RETBP_HUMAN	Retinol-binding protein
P01112	RASH_HUMAN	GTPase HRas
P63165	SUMO1_HUMAN	Small ubiquitin-related modifier 1
P00441	SODC_HUMAN	Superoxide dismutase [Cu-Zn]
P02787	TRFE_HUMAN	Serotransferrin
P01375	TNFA_HUMAN	Tumor necrosis factor
P10599	THIO_HUMAN	Thioredoxin
P10636	TAU_HUMAN	Microtubule-associated protein tau
P62988	UBIQ_HUMAN	Ubiquitin
P51965	UB2E1_HUMAN	Ubiquitin-conjugating enz E2 E1
P63279	UBE2I_HUMAN	Ubiquitin carrier protein 9
O00762	UBE2C_HUMAN	Ubiquitin-conjugating enzyme E2 C

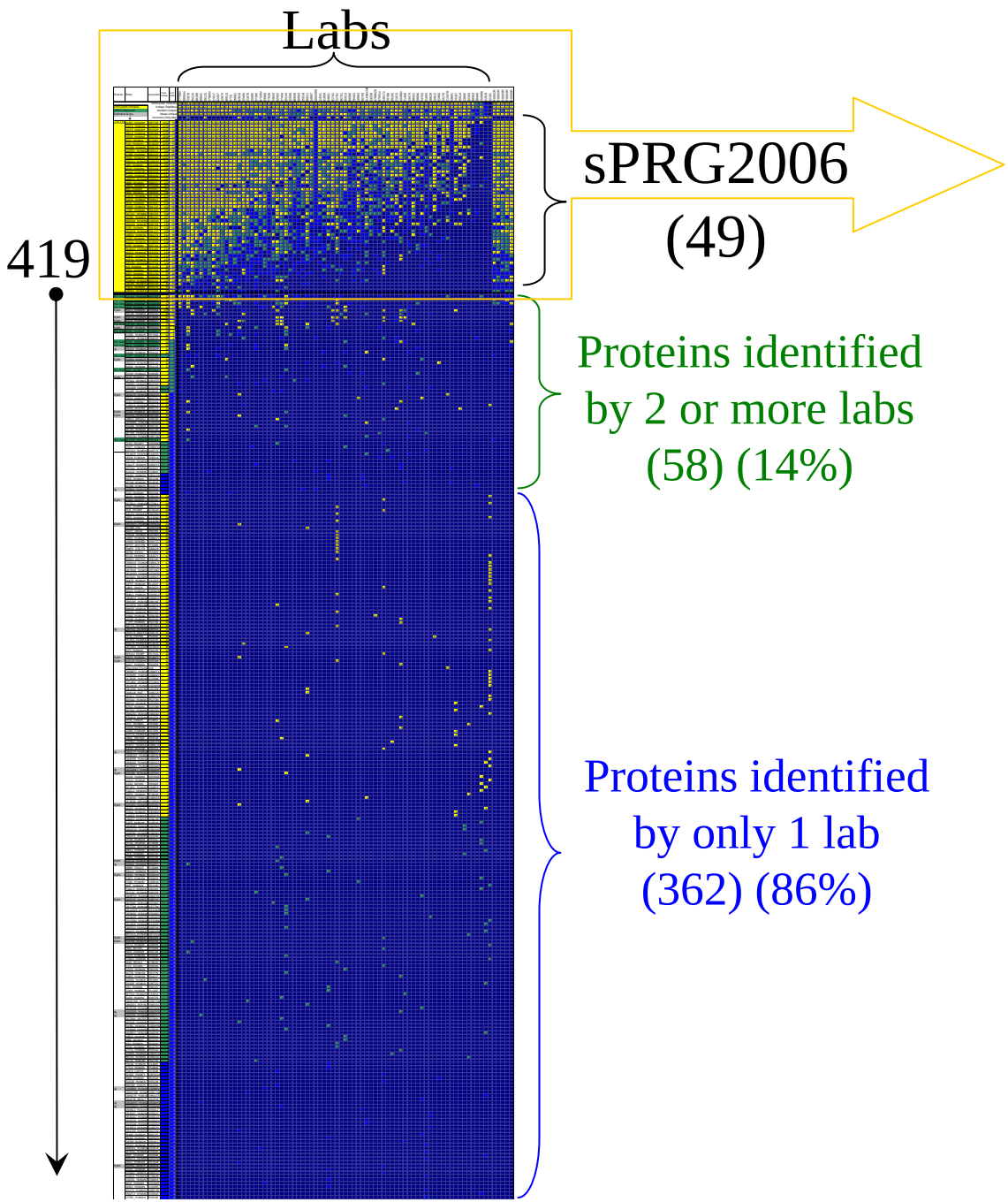
Proteomics Standards Research Group

sPRG2006 Protein Identification

Legend

- 1 - 2
- 3 - 5
- > 5

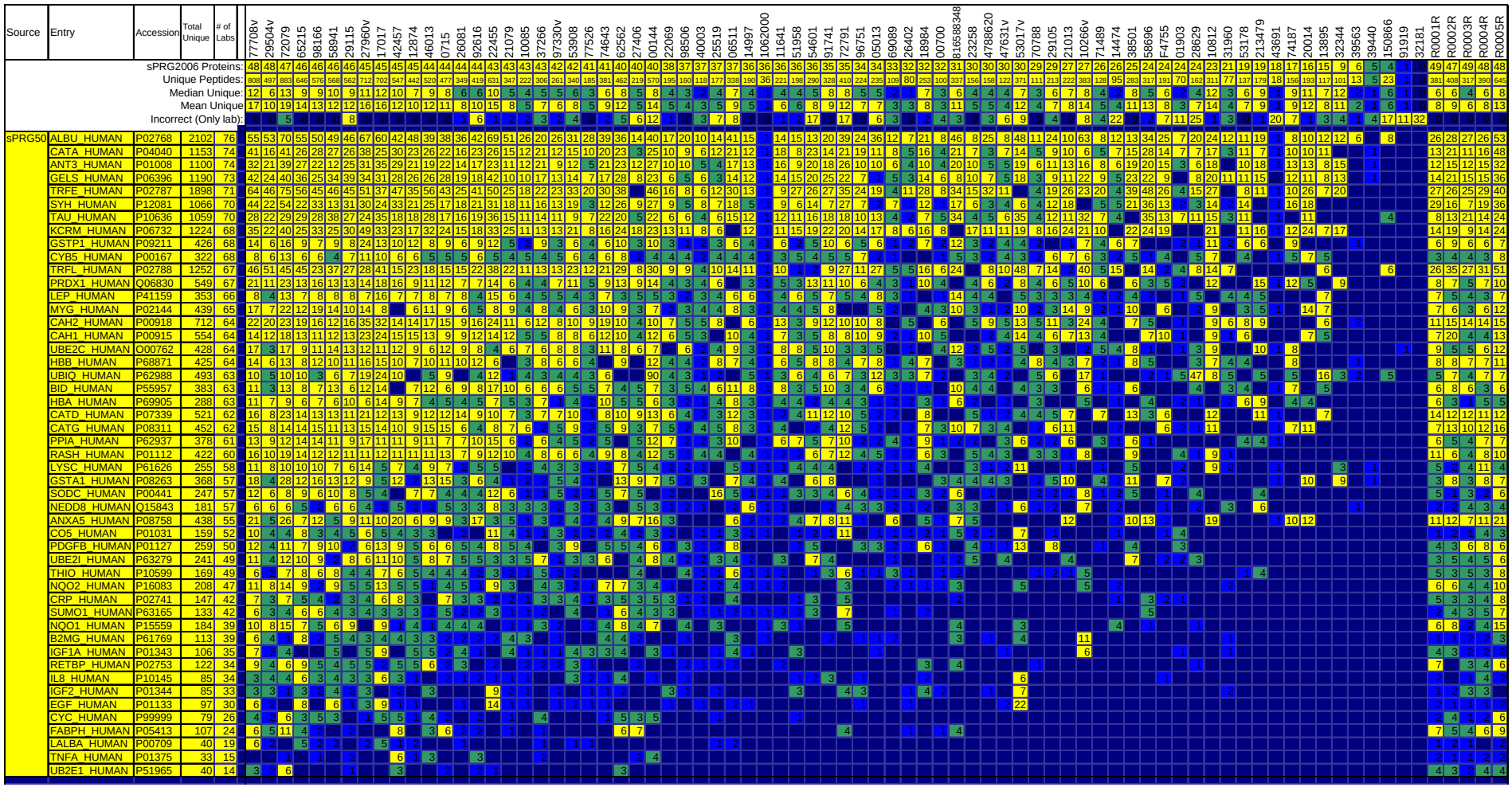
www.abrf.org/sprg



sPRG2006 Protein Identification

Proteomics Standards Research Group

Legend: 1 - 2 | 3 - 5 | > 5

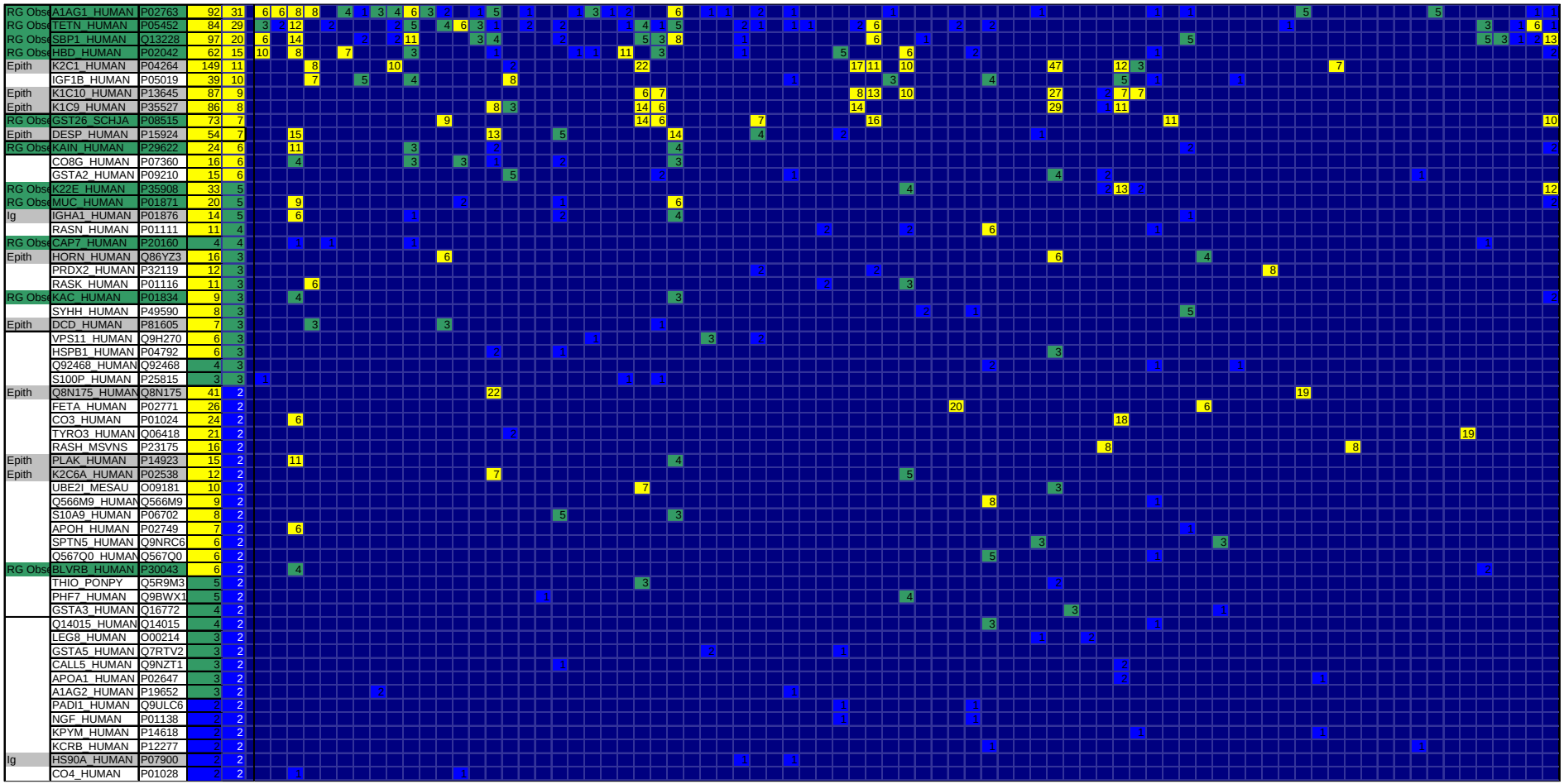


Full sPRG2006 HeatMap available at www.abrf.org/sprg

Proteins Identified by 2 or More Labs

Proteomics Standards Research Group

Legend: **1 - 2** **3 - 5** **> 5**



Full sPRG2006 HeatMap available at www.abrf.org/sprg

Proteomics Standards Research Group

Isoform Issue

IGF1A_HUMAN

MGKISSLPTQLFKCCFCDFLKVKMHTMSSSHLFYLALCLLTFTSSATAGPETLCGAELVDALQFVCG
DRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSARSVRAQRHTDMPKTQK
EVHLKNASRGSAGNKNYRM

IGF1B_HUMAN

MGKISSLPTQLFKCCFCDFLKVKMHTMSSSHLFYLALCLLTFTSSATAGPETLCGAELVDALQFVCG
DRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSARSVRAQRHTDMPKTQK
YQPPSTNKNTKSQRRKGWPKTHPGGEQKEGTEASLQIRGKKKEQRREIGSRNAECRGKKGK

IGF1A_HUMAN

GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA

IGF1B_HUMAN

GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA

Proteomics Standards Research Group

Molecular and Cellular Proteomics Workshop Paris - May 2005

Revised Guideline 7.

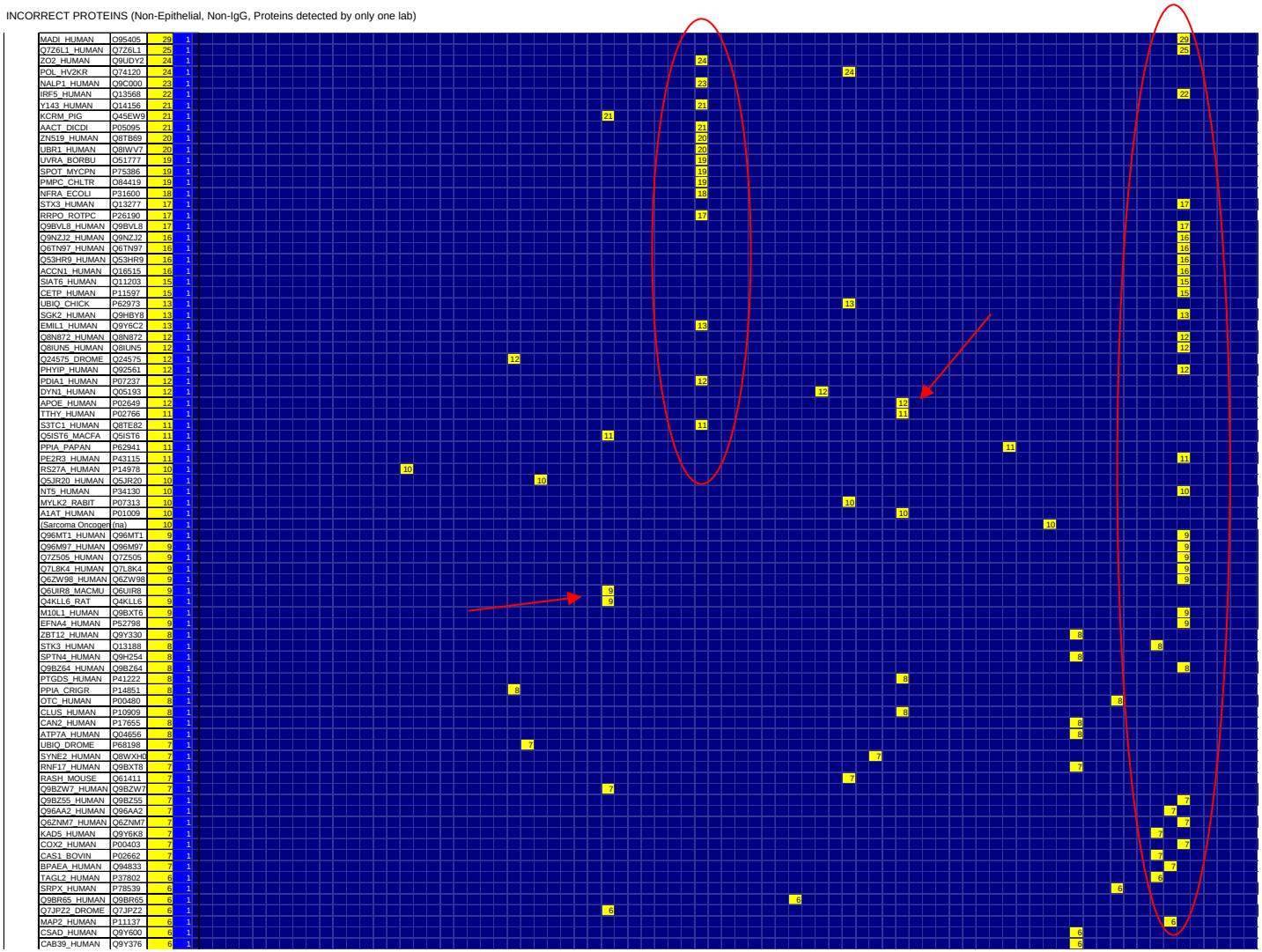
- Identical peptide sequences can occur as elements of multiple unique protein sequences due to biological variation.
- Shotgun proteomics (proteolytic digestion of protein mixtures) introduces the complication of loss of connectivity between peptides and their protein precursors.
- Assignment of peptide sequences results in two outcomes; *distinct* peptides that map to only one protein sequence or *shared* peptides that map to more than one protein sequence.
- Detection of shared peptides introduces an uncertainty bioinformatics redundancy versus physical redundancy.
- The apparent ambiguity in peptide assignment requires reporting of a protein group.

Sean Seymour - de novo sequencing/DB searching tutorial session - Tuesday 11 am

Incorrect IDs: High Peptide Count

Proteomics Standards Research Group

Legend: 1 - 2 | 3 - 5 | > 5



Protein Identification Accuracy

Proteomics Standards Research Group

sPRG								
R0001R	49					381	6	8
R0002R	47					408	6	9
R0003R	49					317	4	6
R0004R	48					390	6	8
R0005R	48					645	8	13

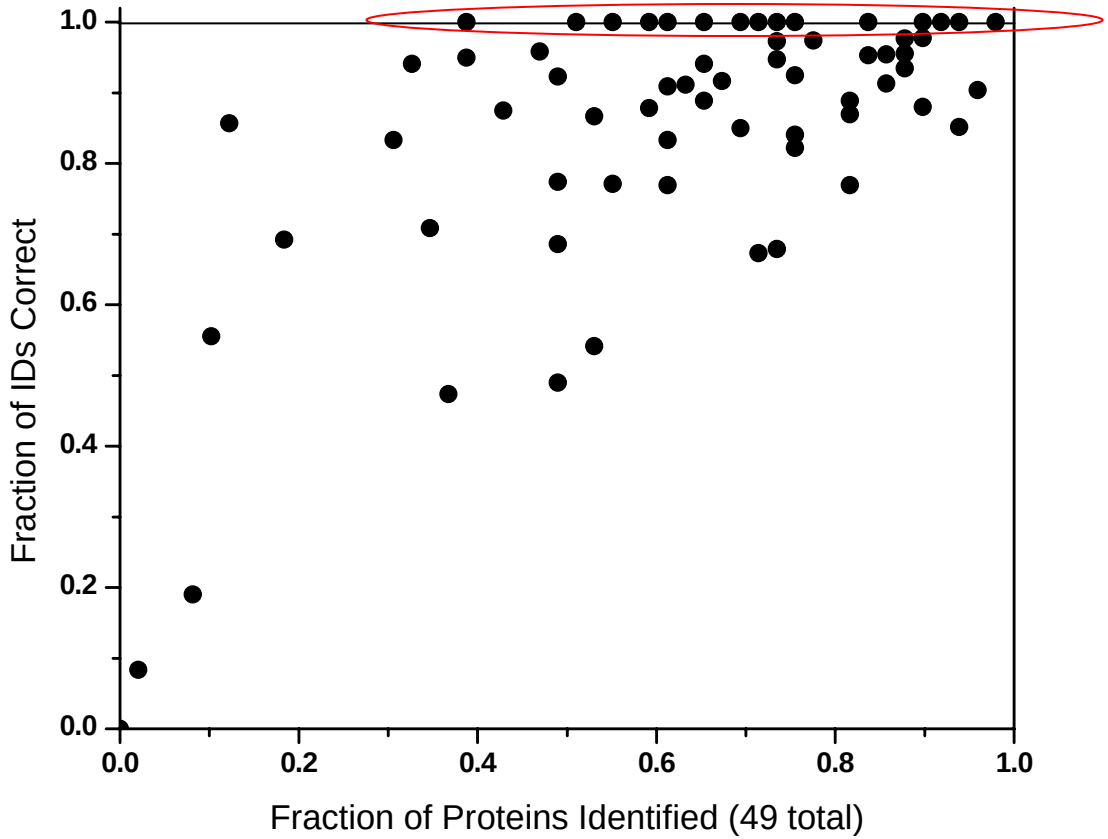
Total Correct / (Correct + Incorrect)

% Identified x Accuracy

	sPRG50 Identified					Unique Peptides		
Lab	Proteins	%	Incorrect	Accuracy	Overall	Peptides	Median	Mean
77708	48	98%	0	100%	98%	808	11.5	16.8
29504	48	98%	0	100%	98%	497	6.0	10.4
72079	47	96%	5	90%	87%	883	13.0	18.8
65215	46	94%	0	100%	94%	646	9.0	14.0
98166	46	94%	0	100%	94%	576	9.0	12.5
58941	46	94%	0	100%	94%	568	9.5	12.3
29115	46	94%	8	85%	80%	562	8.5	12.2
27960	45	92%	0	100%	92%	712	11.0	15.8
17017	45	92%	0	100%	92%	702	12.0	15.6
42457	45	92%	0	100%	92%	547	10.0	12.2
12874	45	92%	0	100%	92%	442	7.0	9.8
46013	44	90%	0	100%	90%	520	8.5	11.8
0715	44	90%	0	100%	90%	477	8.0	10.8
26081	44	90%	1	98%	88%	349	5.5	7.9
92616	44	90%	6	88%	79%	419	5.5	9.5
22455	43	88%	1	98%	86%	631	10.0	14.7
21079	43	88%	2	96%	84%	347	5.0	8.1
10085	43	88%	2	96%	84%	222	4.0	5.2
37266	43	88%	3	93%	82%	306	5.0	7.1
97330	42	86%	2	95%	82%	36	5.0	6.2
53908	42	86%	4	91%	78%	340	5.5	8.1
77526	41	84%	0	100%	84%	185	3.0	4.5
74643	41	84%	2	95%	80%	381	6.0	9.3
62562	40	82%	5	89%	73%	462	8.0	11.6
27406	40	82%	6	87%	71%	219	4.5	5.5
00144	40	82%	12	77%	63%	570	7.5	14.2
22069	38	78%	1	97%	76%	195	4.0	5.1
98506	37	76%	0	100%	76%	160	3.0	4.3
40003	37	76%	3	93%	70%	118	2.0	3.2
25519	37	76%	7	84%	63%	177	4.0	4.8
06511	37	76%	8	82%	62%	338	7.0	9.1
14997	36	73%	0	100%	73%	190	4.0	5.3
1062000	36	73%	0	100%	73%	36	1.0	1.0
11641	36	73%	1	97%	71%	221	4.0	6.1
51958	36	73%	2	95%	70%	198	4.0	5.5
54601	36	73%	17	68%	50%	290	5.0	8.1
91741	35	71%	0	100%	71%	328	8.0	9.4
72791	35	71%	17	67%	48%	410	8.0	11.7
96751	34	69%	0	100%	69%	224	4.5	6.6
05013	34	69%	6	85%	59%	235	4.5	6.9
69089	33	67%	3	92%	62%	109	2.0	3.3
26402	32	65%	0	100%	65%	80	1.0	2.5
18984	32	65%	2	94%	61%	253	7.0	7.9
00700	32	65%	4	89%	58%	100	2.5	3.1
8165883487	31	63%	3	91%	58%	337	6.0	10.9
23258	30	61%	0	100%	61%	156	4.0	5.2
4788620	30	61%	3	91%	56%	158	3.5	5.3
47631	30	61%	6	83%	51%	122	3.5	4.1
53017	30	61%	9	77%	47%	371	7.0	12.4
70788	29	59%	0	100%	59%	111	3.0	3.8
29105	29	59%	4	88%	52%	213	6.0	7.3
21013	27	55%	0	100%	55%	222	7.0	8.2
10266	27	55%	8	77%	43%	383	8.0	14.2
71489	26	53%	4	87%	46%	128	4.0	4.9
14474	26	53%	22	54%	29%	95	2.0	3.7
38501	25	51%	0	100%	51%	283	8.0	11.3
58696	24	49%	2	92%	45%	317	5.0	13.2
F4755	24	49%	7	77%	38%	191	6.0	8.0
01903	24	49%	11	69%	34%	70	2.0	2.9
28629	24	49%	25	49%	24%	162	3.5	6.7
10812	23	47%	1	96%	45%	311	12.0	13.5
31960	21	43%	3	88%	38%	77	3.0	3.7
53178	19	39%	0	100%	39%	137	6.0	7.2
213479	19	39%	1	95%	37%	179	9.0	9.4
43691	18	37%	20	47%	17%	18	1.0	1.0
74187	17	35%	7	71%	25%	156	9.0	9.2
20014	16	33%	1	94%	31%	193	11.0	12.1
13895	15	31%	3	83%	26%	117	7.0	7.8
32344	9	18%	4	69%	13%	101	12.0	11.2
39563	6	12%	1	86%	10%	13	1.5	2.2
39440	5	10%	4	56%	6%	5	1.0	1.0
150866	4	8%	17	19%	1.6%	23	5.5	5.8
91919	1	2%	11	8%	0.2%	1	1.0	1.0
32181	0	0%	32	0%	0%	0	0.0	0.0

Proteomics Standards Research Group

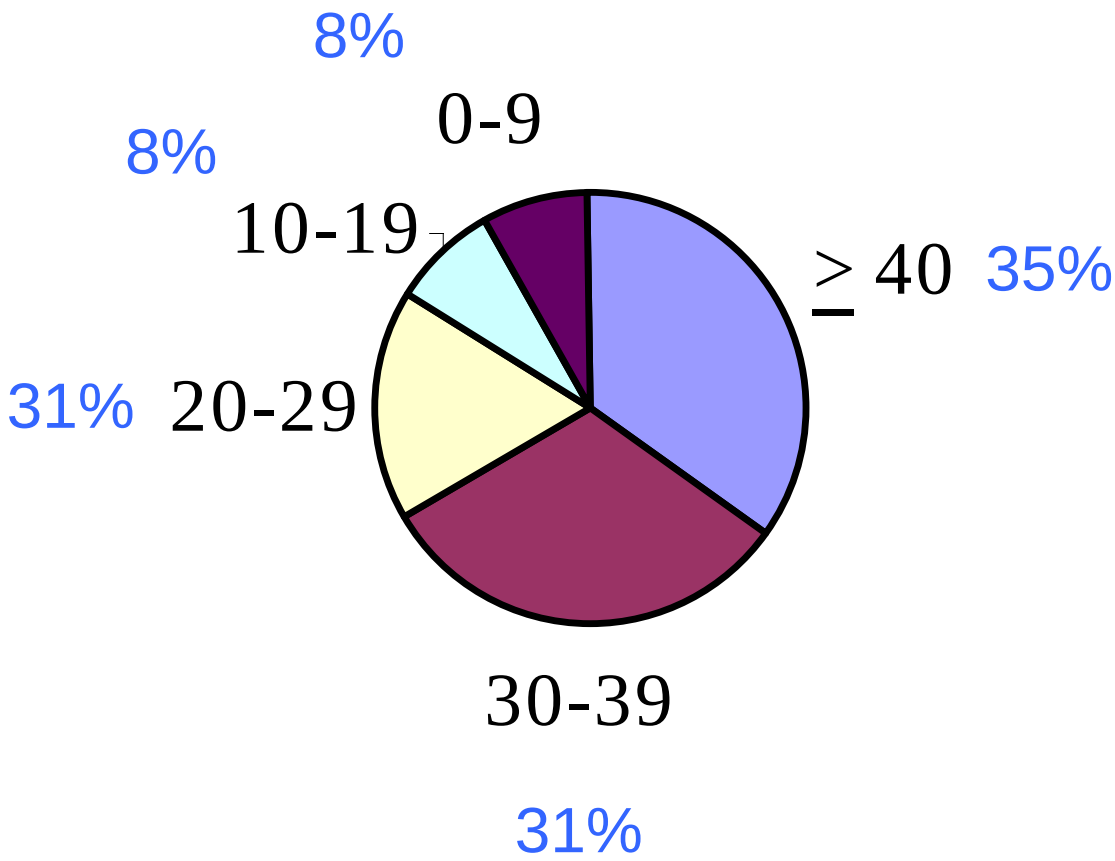
Protein Identification Performance for 74 Labs



The performance of each lab is represented by a point defined by the fraction of all known proteins identified and the fraction of all reported identifications that were correct. Those labs with no false identifications fall along the upper axis.

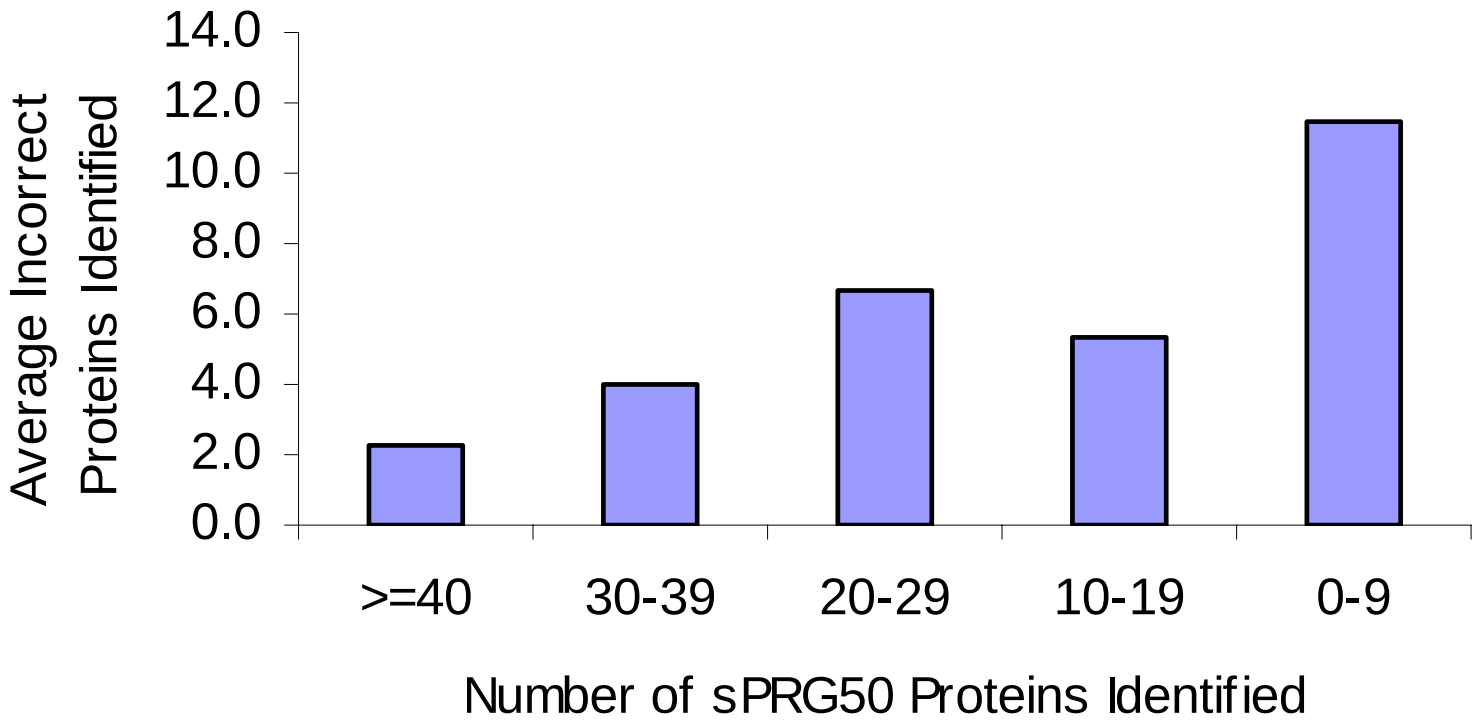
Proteomics Standards Research Group

Number of Proteins Identified by Respondents



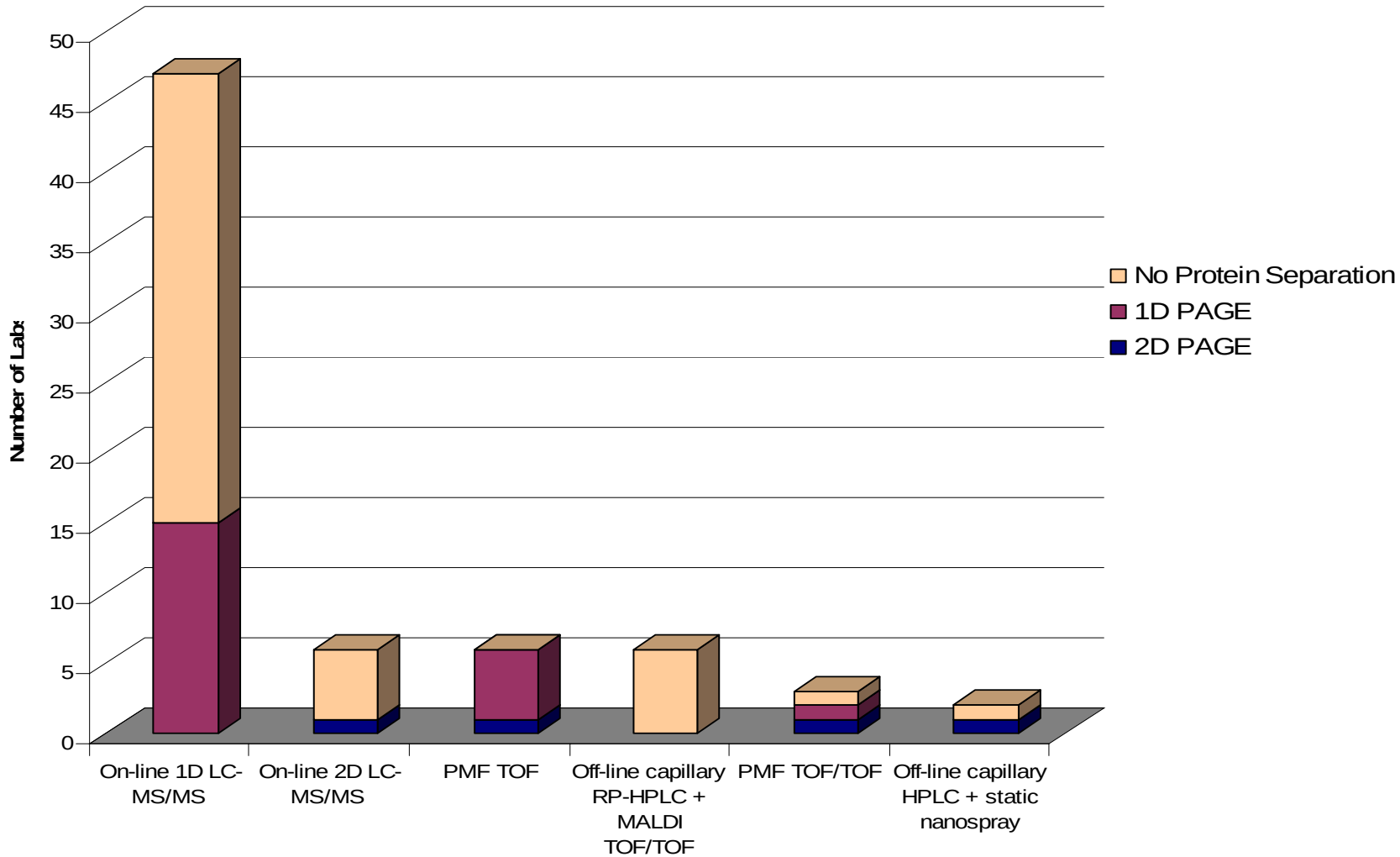
Proteomics Standards Research Group

Labs detecting fewer correct proteins tended to identify more incorrect proteins



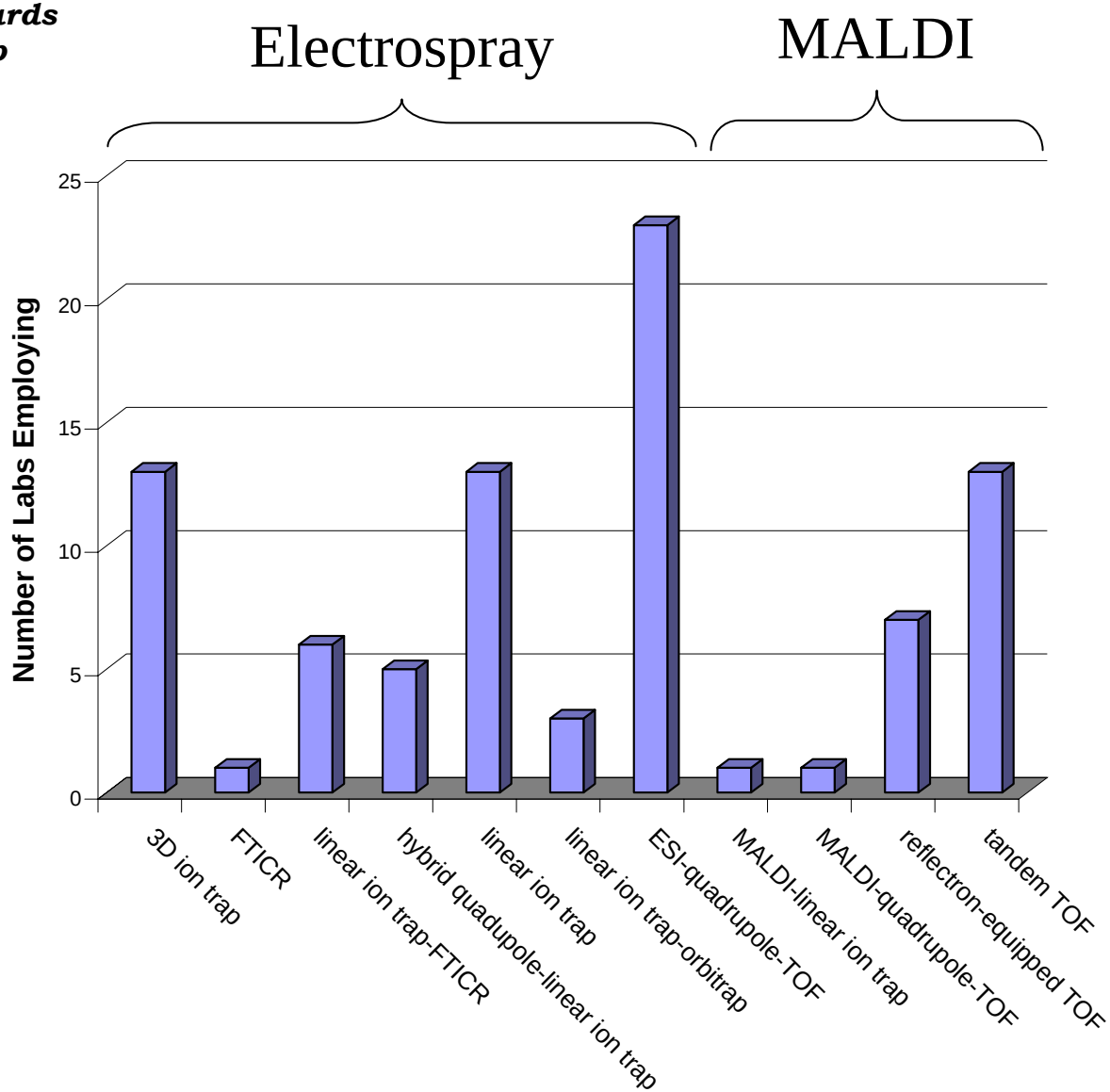
Primary Method of Analysis

Proteomics Standards Research Group



Proteomics Standards Research Group

Mass Spectrometers Used

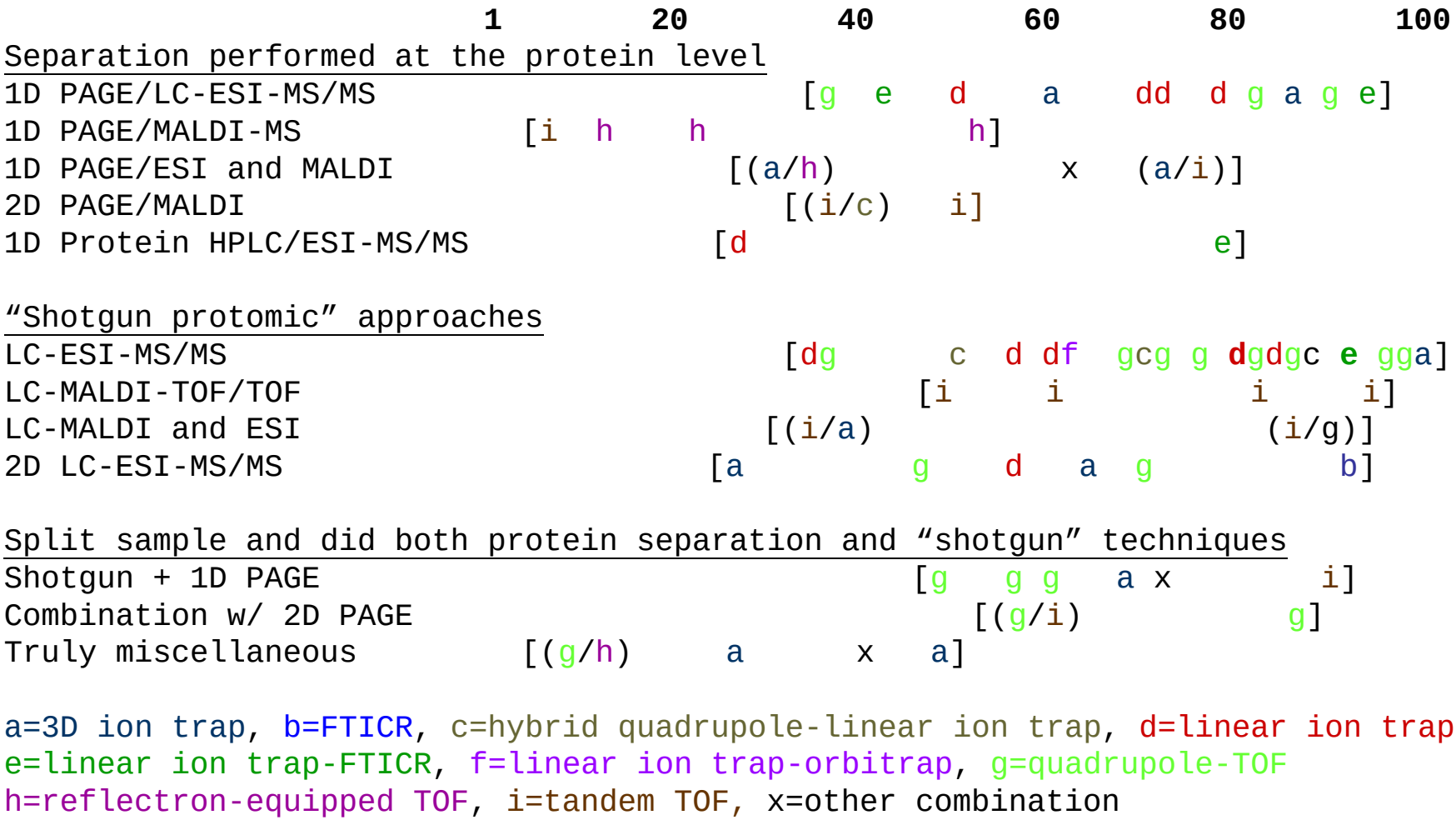


Identification Accuracy Using Different Experimental Approaches

Each data point represents the instrument(s) used and the accuracy achieved, where

Accuracy = total correct identifications / (correct identifications + incorrect identifications)

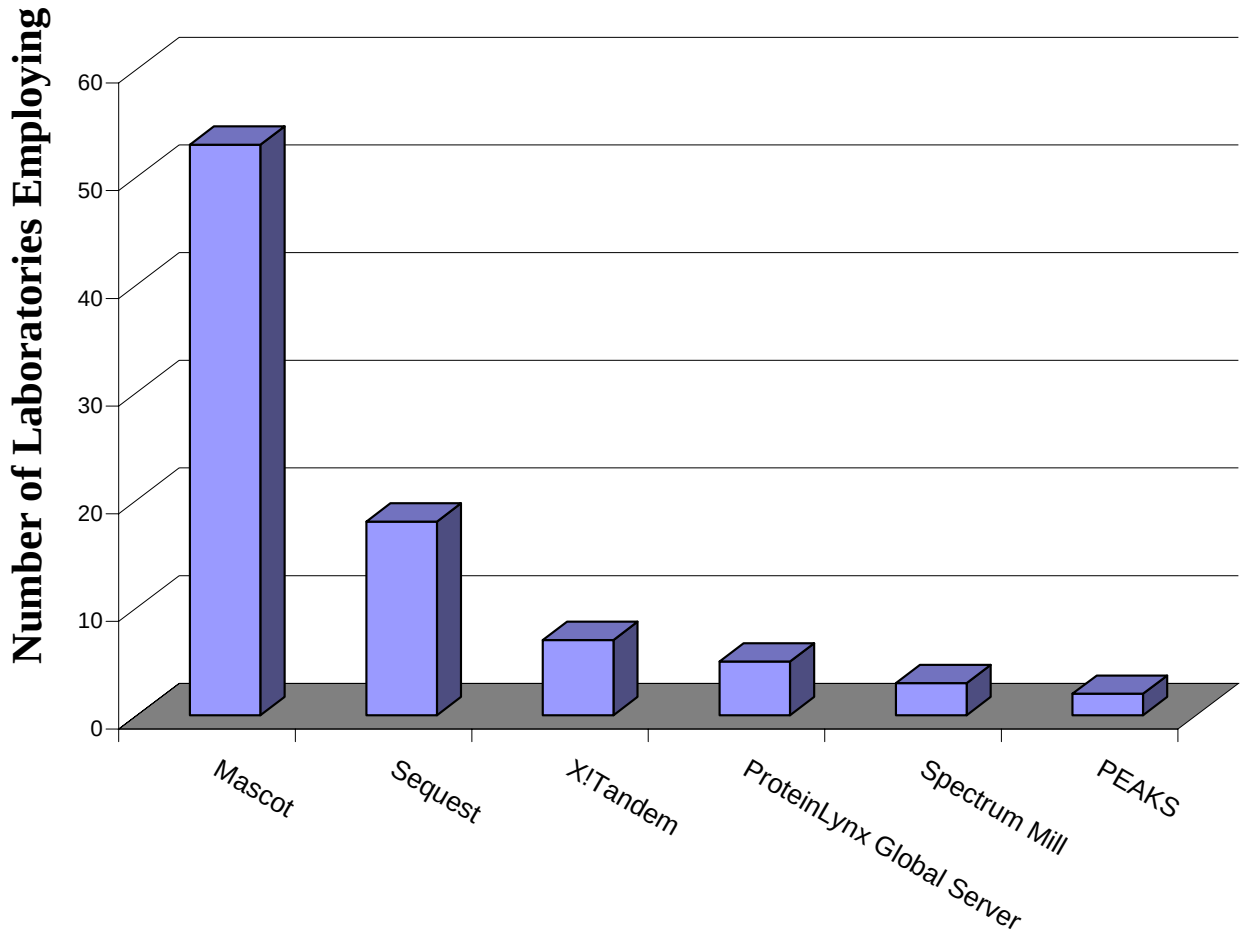
Brackets enclose range of scores for each method; slash indicates multiple instruments

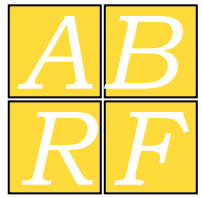


a=3D ion trap, b=FTICR, c=hybrid quadrupole-linear ion trap, d=linear ion trap
e=linear ion trap-FTICR, f=linear ion trap-orbitrap, g=quadrupole-TOF
h=reflectron-equipped TOF, i=tandem TOF, x=other combination

Proteomics Standards Research Group

Search Engines Used for Protein Identification from Tandem Mass Spectra

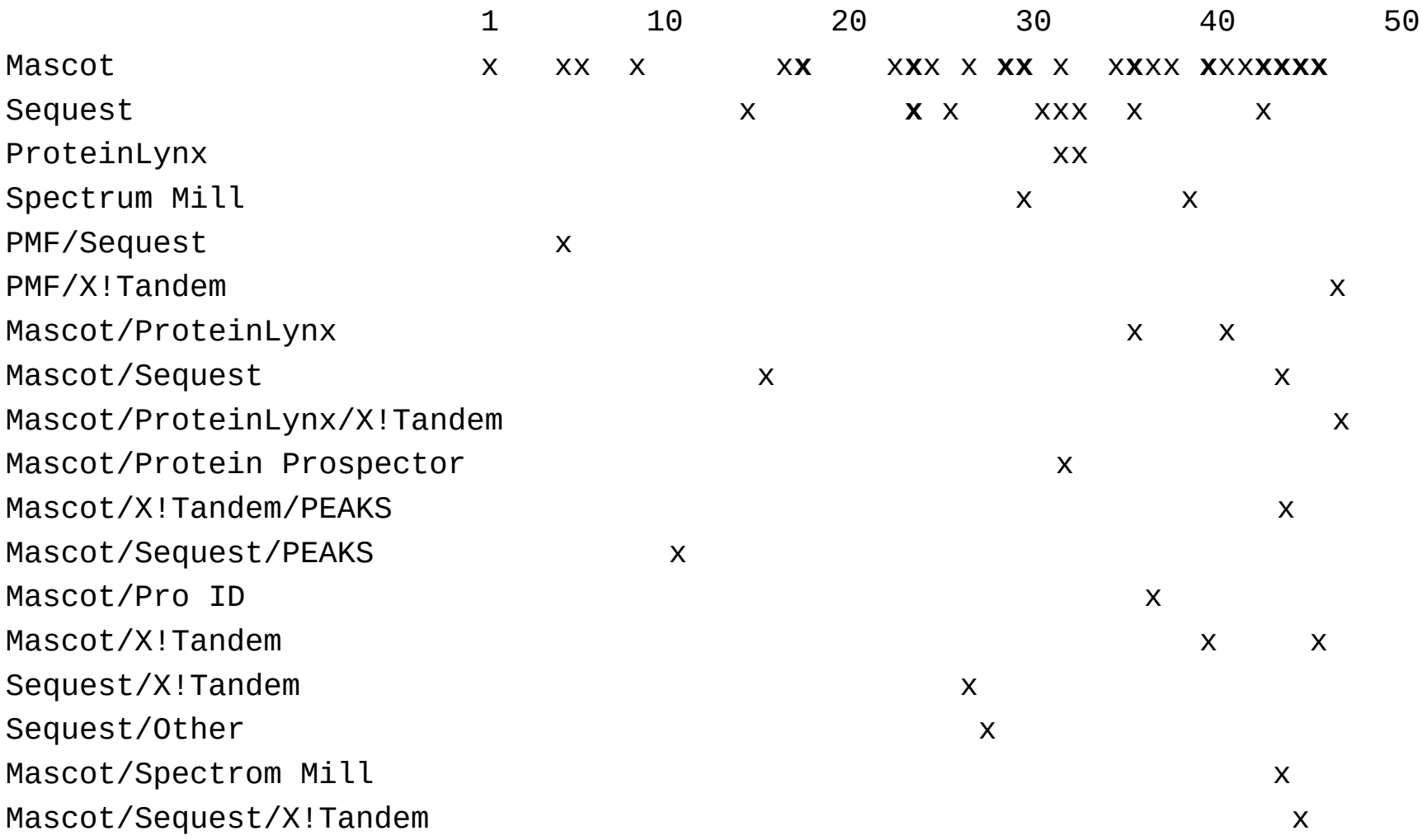




Proteomics Standards Research Group

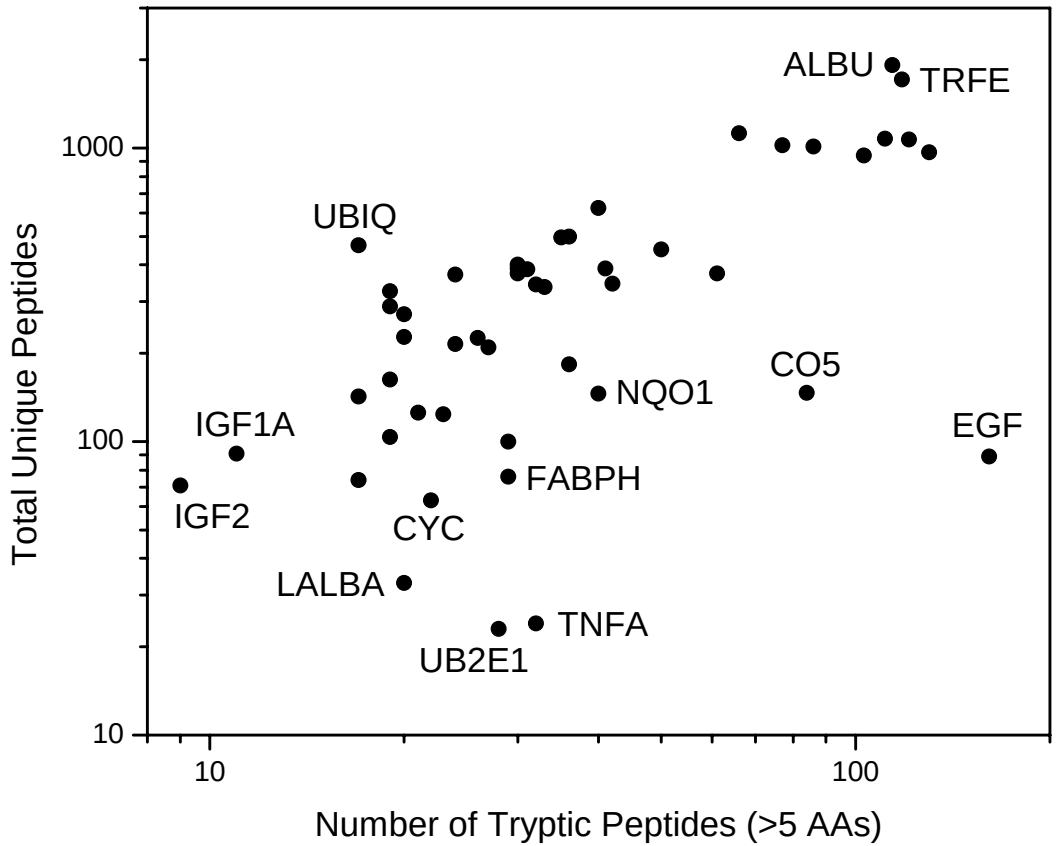
Number of Proteins Correctly Identified Using Each Search Engine/Combination of Engines

Each data point represents the results from an individual lab



Proteomics Standards Research Group

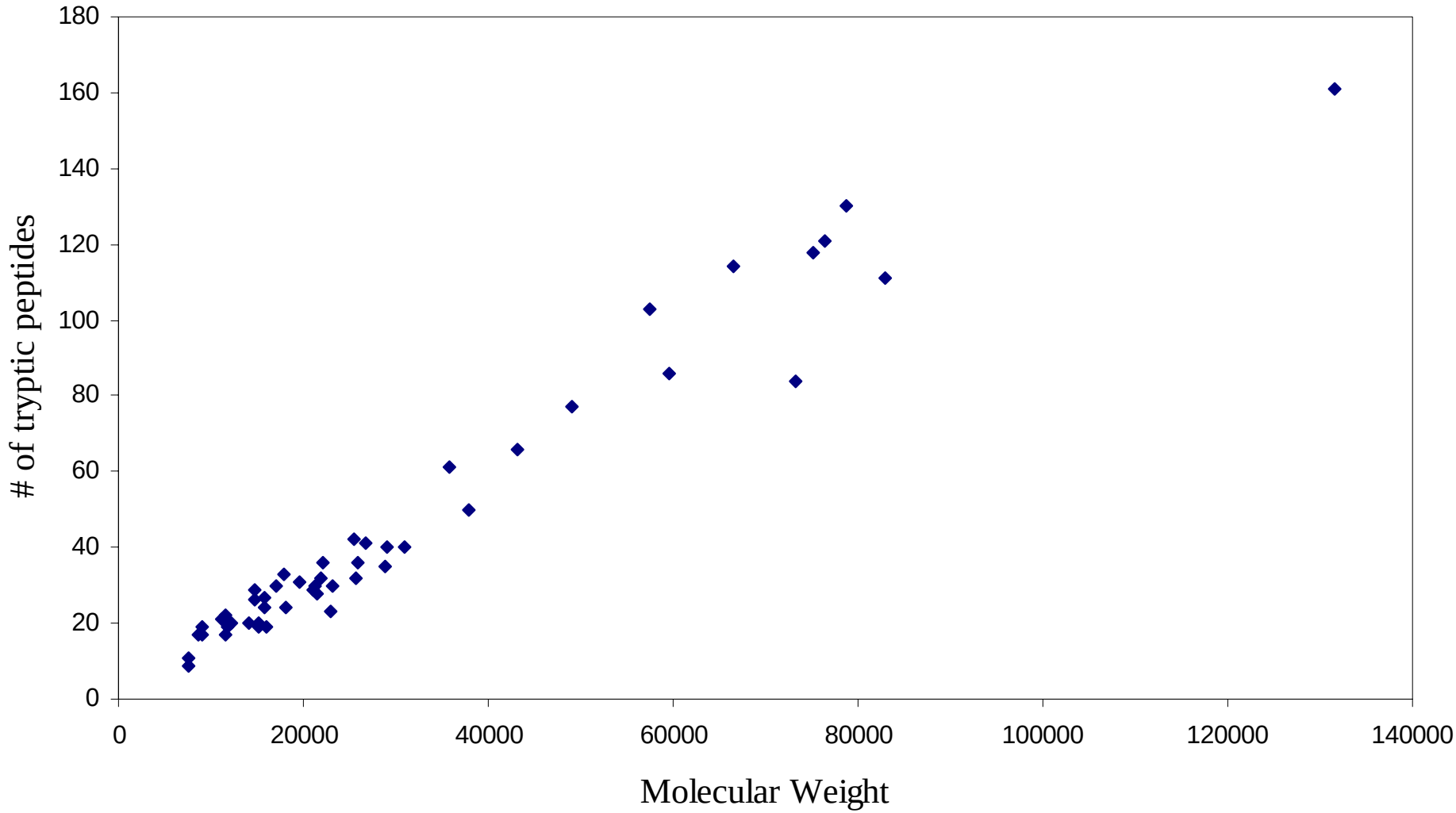
Comparison of Observed vs Expected Unique Tryptic Peptides



For each protein in the test mixture, the total number of reported unique peptides is compared to the maximum expected number of different simple tryptic peptides. Since an effort was made to prepare the proteins in eqimolar amounts (5 pmol), if all peptides were formed and detected with equal efficiency, a straight line would result. The origin of large dispersion is not clear, but outliers at the lower right are glycoproteins (CO5_HUMAN and EFG_HUMAN) while the highest yields are from secreted proteins (ALBU_HUMAN and TRFE_HUMAN). Other protein at the extremities are also labeled.

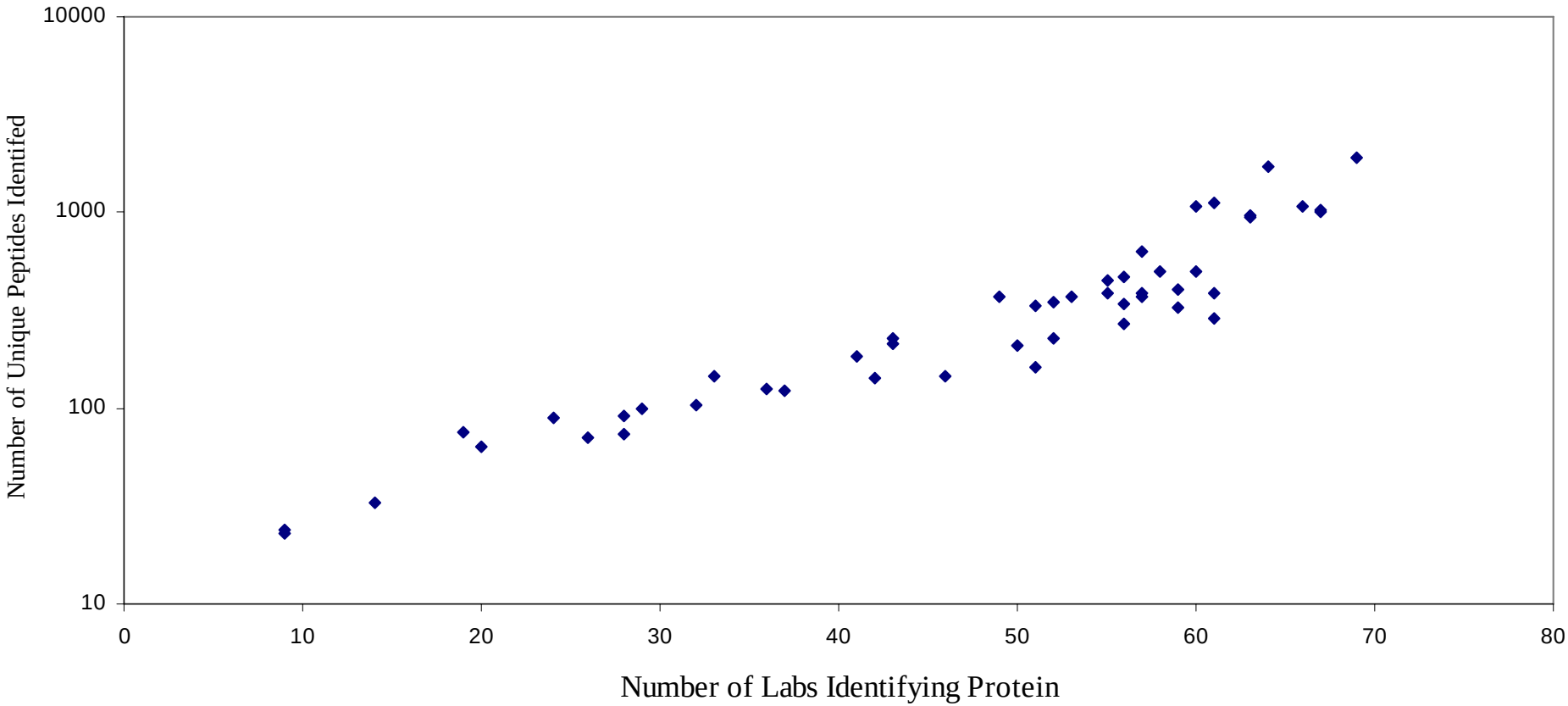
Proteomics Standards Research Group

Theoretical number of tryptic peptides vs molecular weight



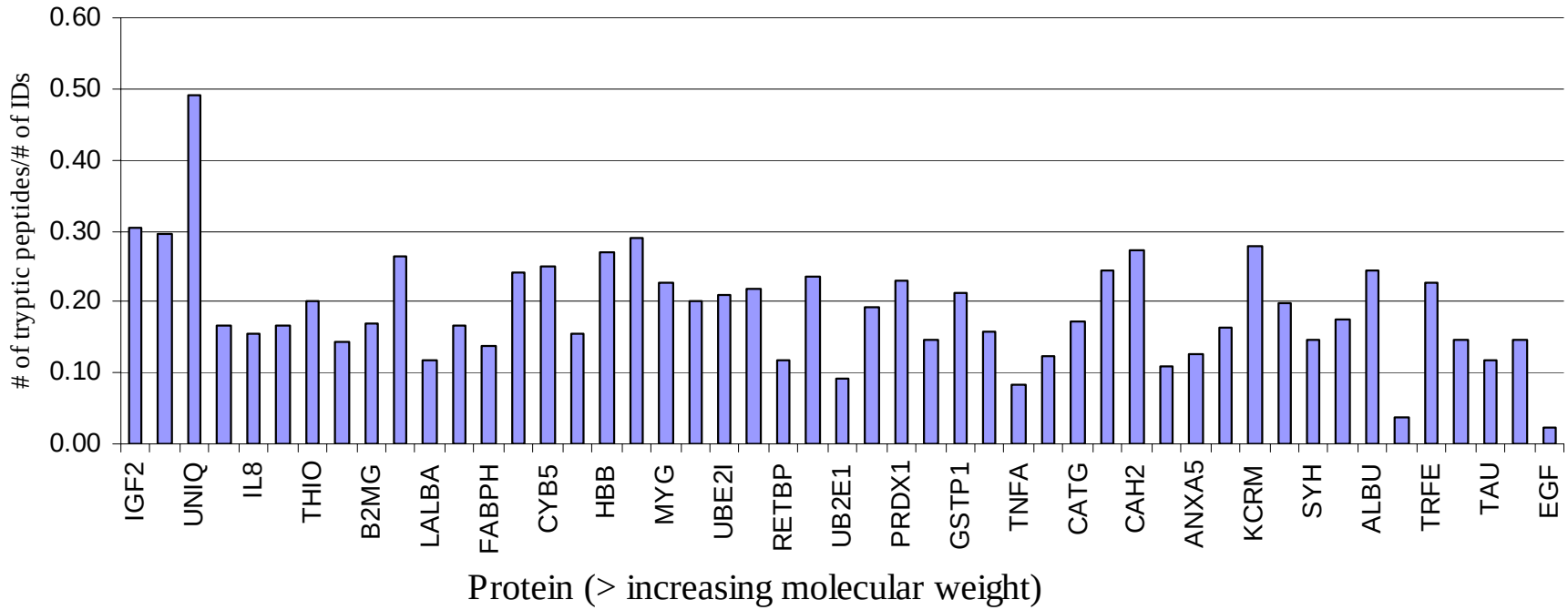
Proteomics Standards Research Group

Unique Peptides vs Successful IDs



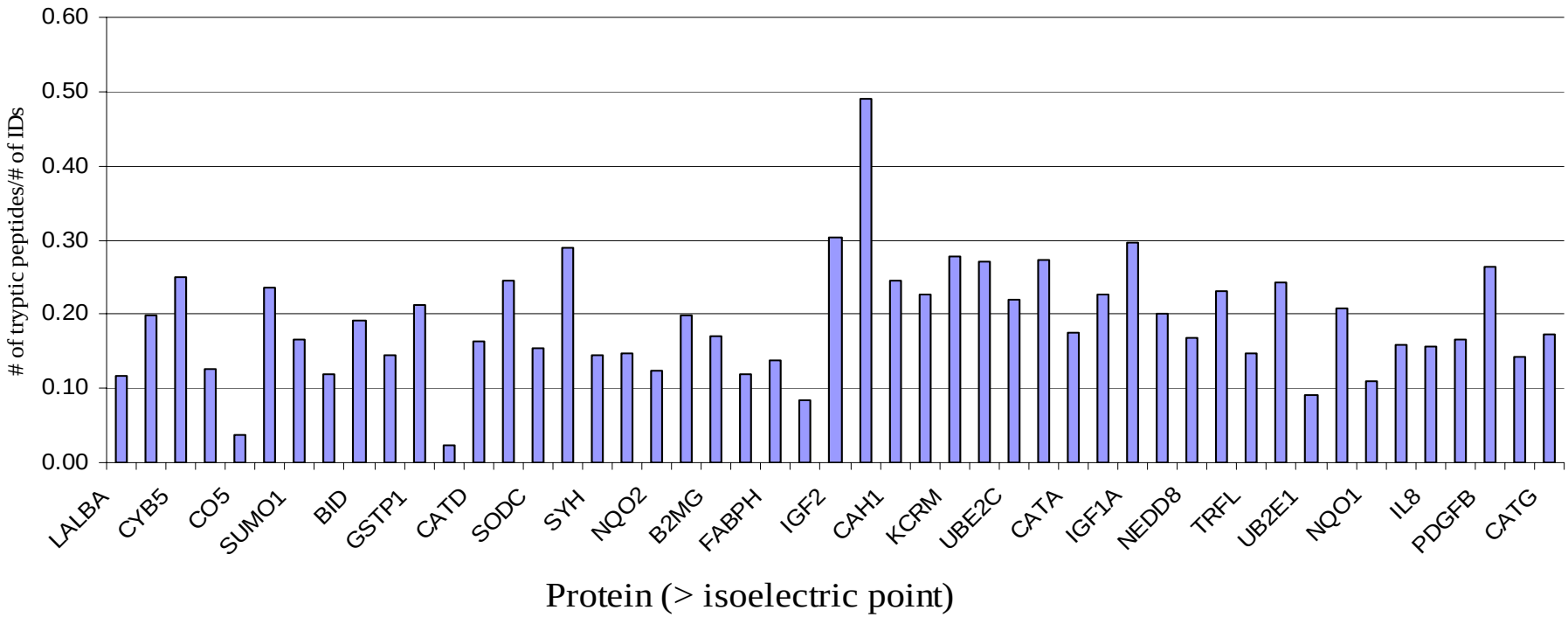
Proteomics Standards Research Group

Protein Sequence Coverage: Molecular Weight



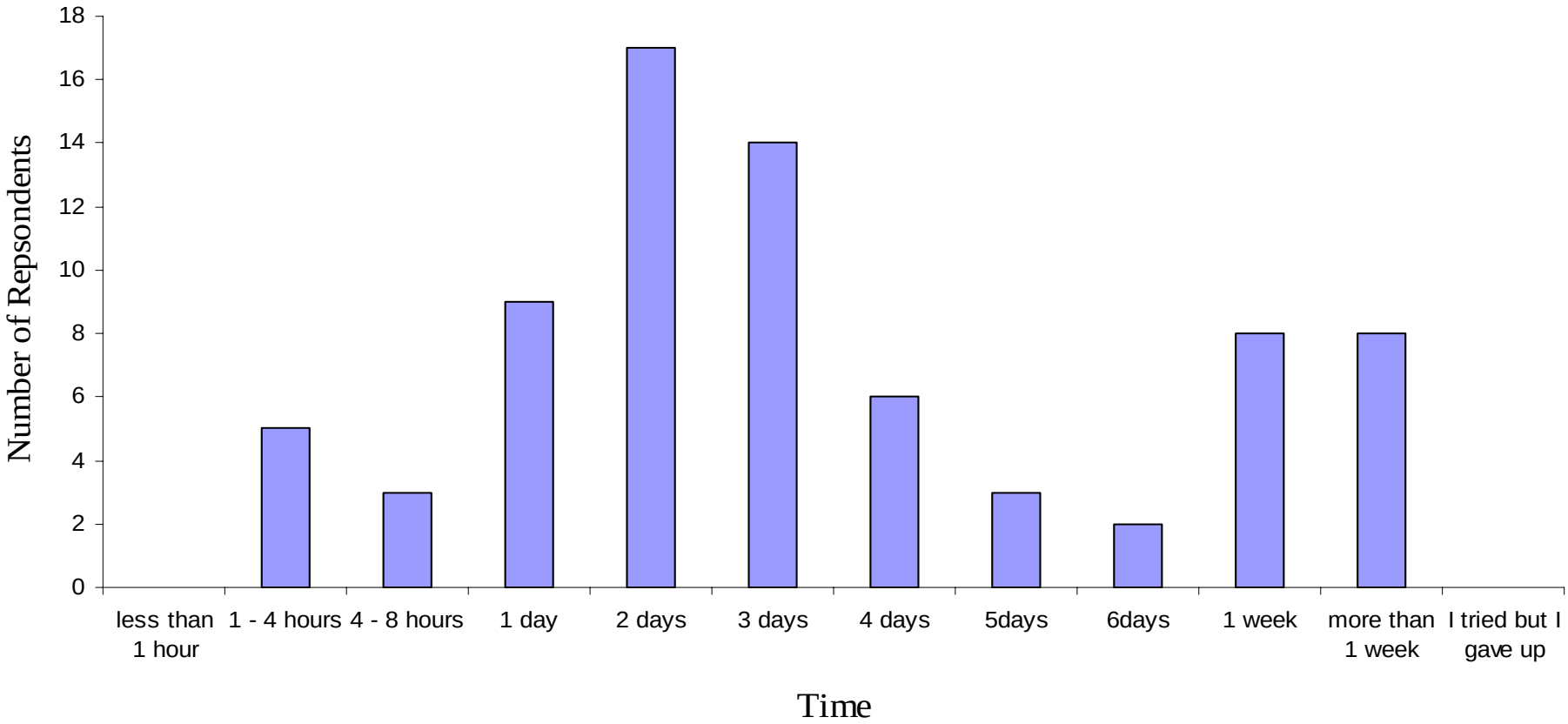
Proteomics Standards Research Group

Protein Sequence Coverage: Isoelectric Point



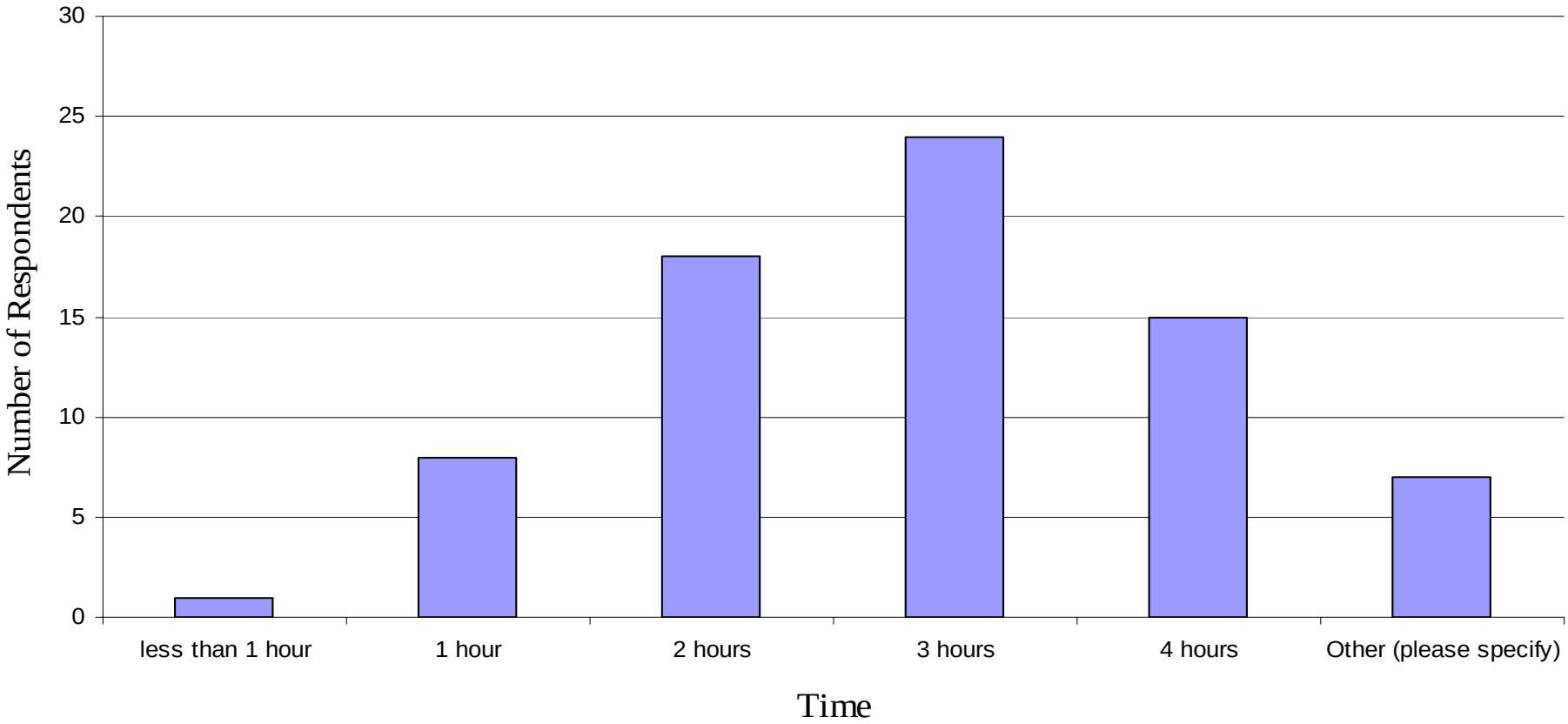
Proteomics Standards Research Group

Time to Prepare and Analyze



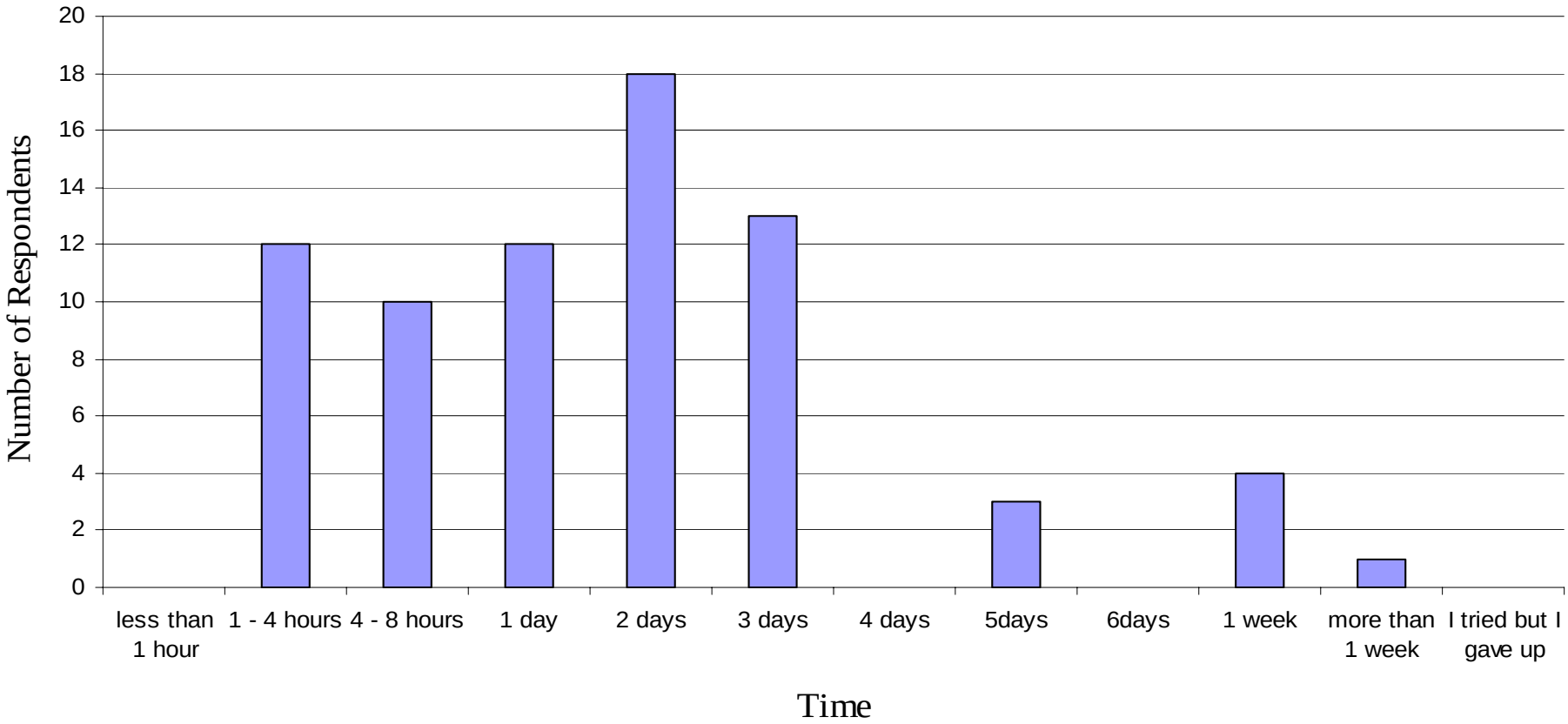
Proteomics Standards Research Group

Time to Complete Survey



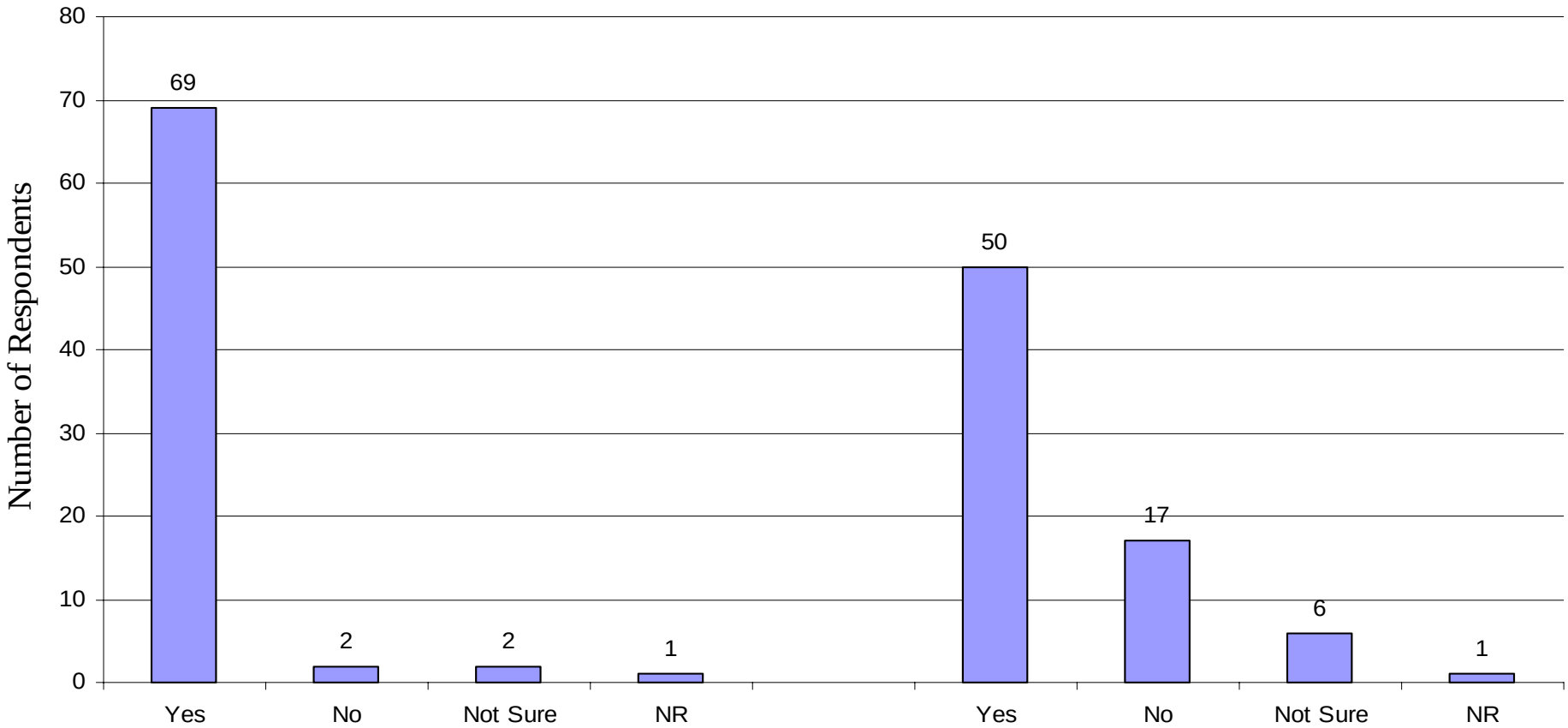
Proteomics Standards Research Group

Time for Data Analysis



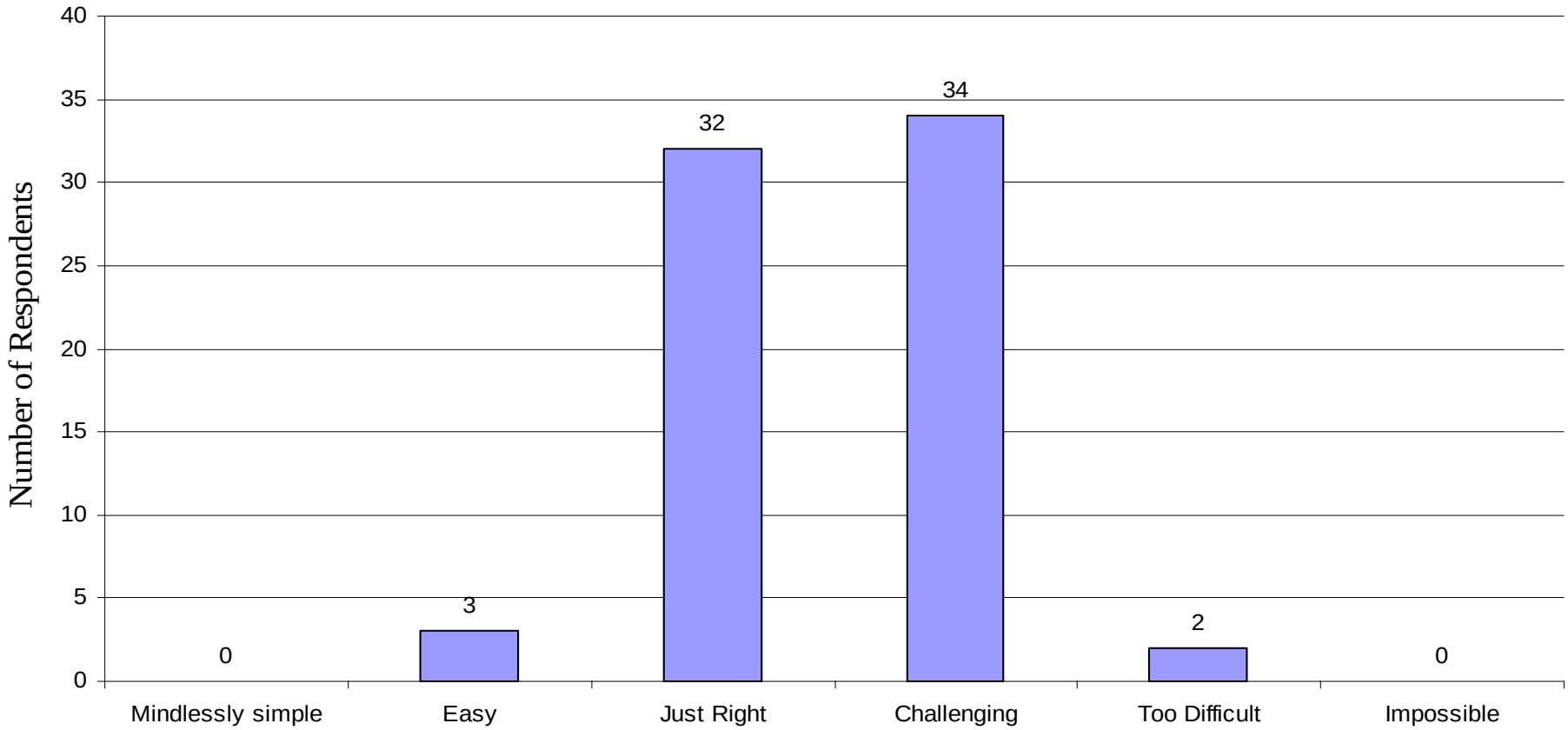
Proteomics Standards Research Group

Utility



Proteomics Standards Research Group

Level of Difficulty



Proteomics Standards Research Group

Conclusions

- A standard protein mixture has been developed that has broad usefulness for a variety of proteomics strategies.
- The fact that no single approach performed better than others may suggest that success is not so much method-dependent as possibly experience-, or technique-dependent (technique of the individual performing the analysis).
- Good results are achievable by a lab which does not necessarily have the latest instruments but optimizes variables within its control.

Conclusions (cont.)

- Many labs can reliably identify a large fraction of proteins at these concentrations with few false positive results.
- The nature of variability expected in a semi-complex protein mixture analysis will have to be dealt with in future standards design and analysis.
- We demonstrated in a tangible way, problems resulting from the myriad representations of proteins and lack of data representation standards.
- The study has led to a publicly available set of raw data files for analysis of a known protein mixture by a number of independent labs.

Proteomics Standards Research Group

Acknowledgements

Andrews Lab
Angela Walker
John Strahler
Jayson Falkner

Arnott Lab
Alexander James Arnott

Gawinowicz Lab
Susan Pollak

Lane Lab
John M. Neveu
Renee Robinson
Marcelo Cantu

Lilley Lab
Julie Howard
Renata Feret

Markey Lab
Jim Makusky
Ewa Jankowska-Stevens
Wenyao Shi

Martin Lab

Stein Lab
Lisa Kilpatrick
Jeri Roth
Sara Yang

Sigma-Aldrich
Graham Scott
George Lipscomb
Jennifer Williams
Michael Hadjisavas

Judy Boland
John Dapron
Malaika Durham
Jeffrey Turner

Steven Cockrill
Justin Wildsmith
Beth Geoppo
Kevin Ray
Mark Tizzard

Proteomics Standards Research Group

Proteomics Standards Research Group (sPRG) Members

Phil Andrews - University of Michigan

David Arnott - Genentech, Inc.

Mary Ann Gawinowicz - Columbia University

Jeffrey A. Kowalak - National Institutes of Health (Chair)

William S. Lane - Harvard University (EB Liaison)

Kathryn S. Lilley - University of Cambridge

Larry Martin - East West University

Stephen E. Stein - NIST

Proteomics Standards Research Group

THANK YOU

For participating in the ABRF 2006 Proteomics Standards Research Group Study