



*Proteome Informatics*  
*Research Group*

# iPRG: Informatic Evaluation of Phosphopeptide Identification and Phosphosite Localization

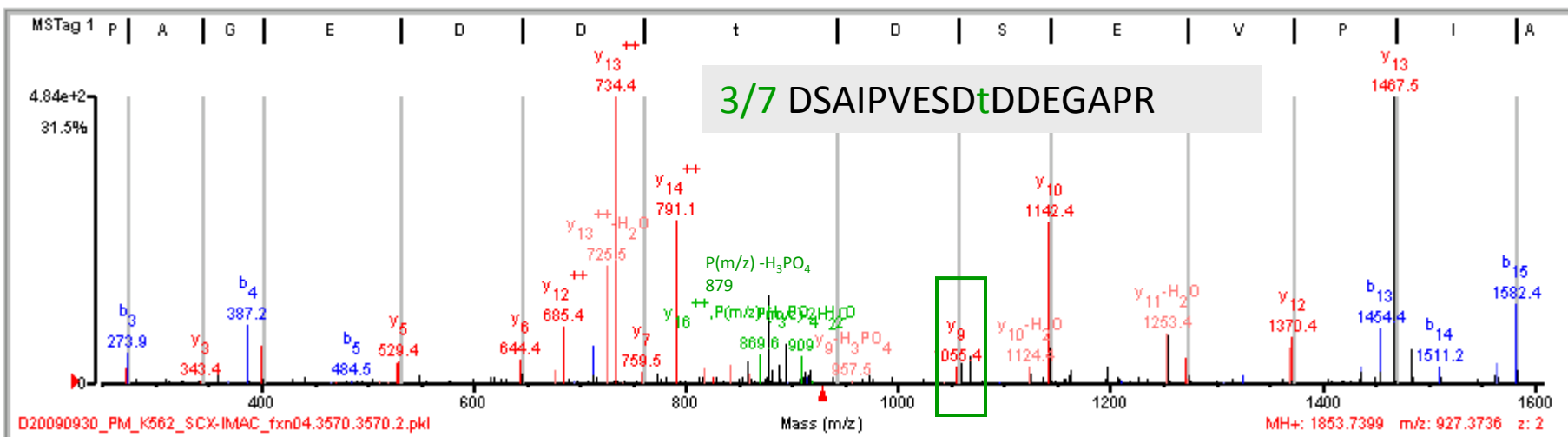
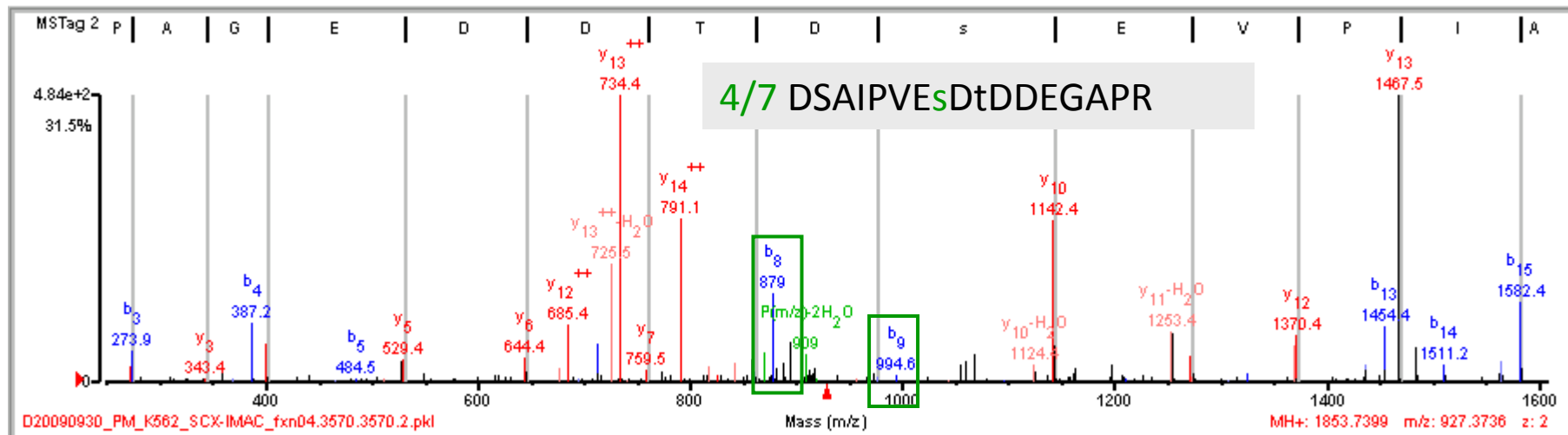
ABRF 2010, Sacramento, CA

March 22, 2010



# A Challenging Problem

Proteome Informatics  
Research Group



14/21 said can identify peptide but can not localize site



# Solution

- Not fun to do by hand!
- Software available that evaluates 'site-determining ions'
  - Generate per residue localization scores
  - Examples: Ascore, PTM Score (MaxQuant), pFind, PhosphoScore, etc.



# Study Goals

1. Evaluate the consistency of reporting phosphopeptide identifications and phosphosite localization across laboratories
2. Characterize the underlying reasons why result sets differ
3. Produce a benchmark phosphopeptide dataset, spectral library and analysis resource



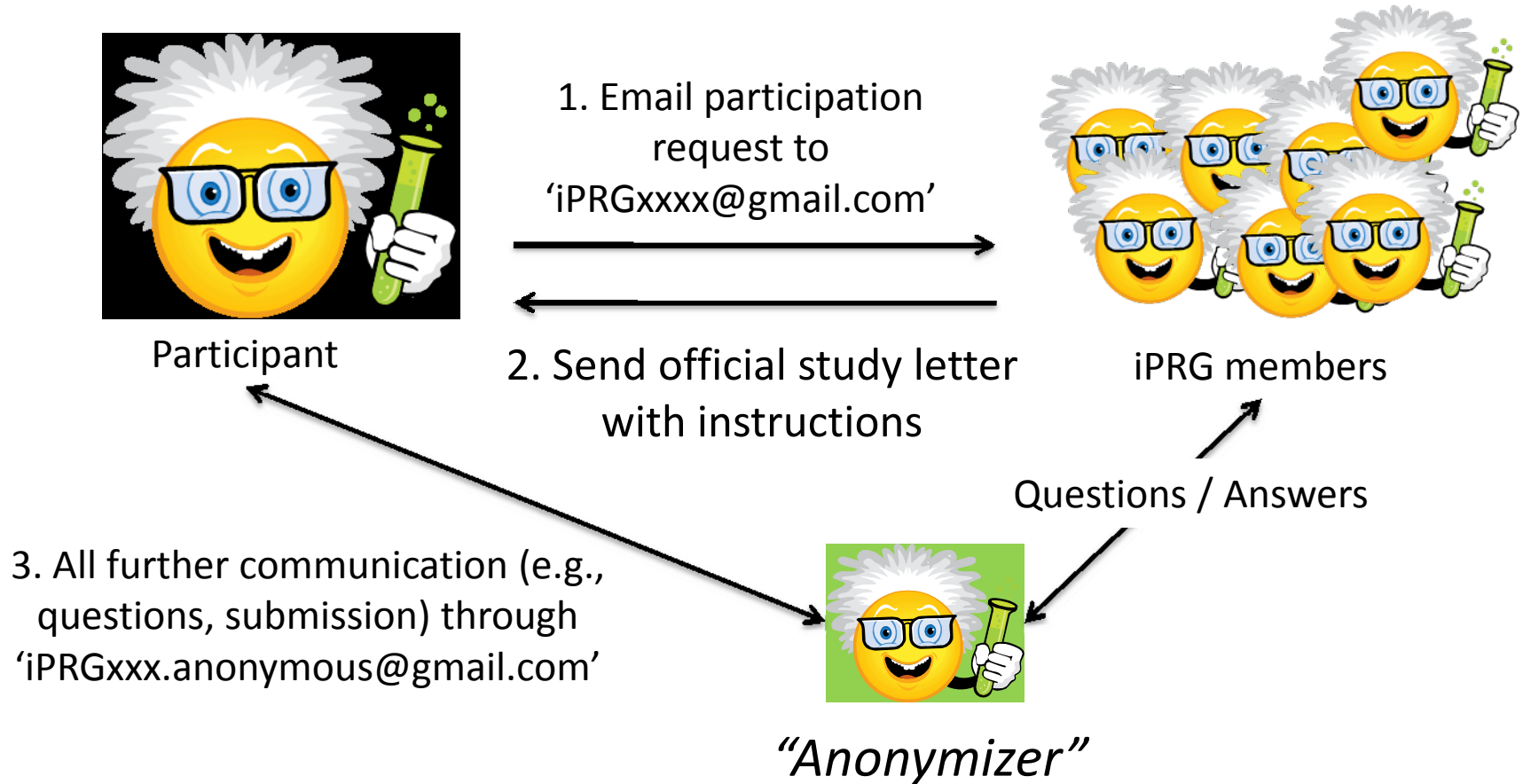
# Study Design

- Use a common dataset
- Use a common sequence database
- Allow participants to use the bioinformatic tools and methods of their choosing
- Use a common reporting template
- Fix the identification confidence (1% FDR)
- Require an indication of phosphosite ambiguity per spectrum
- Ignore protein inference – for now



# Soliciting Participants and Logistics

Study advertised on the ABRF website and listserv, Molecular and Cellular Proteomics blogsite, GenomeWeb and by direct invitation from iPRG members





# Study Materials and Instructions to Participants

- 1 Orbitrap XL dataset (3 files)
    - RAW, mzML, mzXML, MGF, pkl or dta – conversions by ProteoWizard
  - 1 FASTA file (SwissProt human seq's. v57.1)
  - 1 template (Excel)
  - 1 on-line survey (Survey Monkey)
1. Analyze the dataset
  2. Report the phosphopeptide spectrum matches in the provided template
  3. Complete an on-line survey
  4. Attach a 1-2 page description of your methodology



# Reporting Template

## ABRF iPRG 2010 Study Template: Phosphorylated Peptide Analysis

**Instructions:** Please fill in all REQUIRED fields. After deleting the example rows, create a new row for each *phosphopeptide* spectrum match. Multiple rows MAY be used to report ambiguous phosphosite localizations. Phosphorylated residues **MUST** be indicated in the 'Peptide Sequence' field, and results should be sorted by 'Peptide Identification Score' from most to least confident. Additional instructions can be found above each field header. **Results should be emailed to 'anonymous.IPRG2010@gmail.com' no later than Jan. 10, 2010.** Please make sure to fill out the REQUIRED survey [—————>](#)

### REQUIRED FIELDS

Name of data file (e.g., D20090930_PM_K562_SCX-IMAC_fxn03)	Identifiers should be unique scan numbers from data file but may also refer to a merged range of MS/MS scans (e.g., Scan:19, 2316.19.19.3.dta, 2316.19.19.3.pkl).	Precursor m/z as submitted to search engine	Precursor charge reported by search engine	Use lowercase s, t or y (e.g. SLsGSsPCPK) OR a trailing symbol (e.g. SLS#GS#PCPK) OR a string in parentheses (e.g. SLS(ph)GS(ph)PCPK) immediately following each phosphorylated residue. Only phosphorylation of S, T and Y will be compared; all other modifications (e.g., oxidized M) will be ignored. It will be assumed that all modifications indicated on S, T or Y are phosphorylations.	Protein identifier(s) from Fasta file. Use multiple values if peptide is found in multiple proteins, e.g., Q9NZ18; Q9UQ35. Protein inference will not be scored.	Total number of phosphorylations as evidenced by the precursor m/z and MS2 spectrum.	'Y' indicates this match is BETTER than the confidence threshold. 'N' indicates the match is WORSE. Please report BOTH types of identifications in your ranked list. Is this match above 1% FDR identification threshold (Y N)?	Indicate 'Y' if ALL phosphorylations have been confidently localized. 'N' if one or more have not. Are ALL phosphosites unambiguously localized (Y N)?	Phosphosite Localization Certainty	Peptide Identification Certainty	Peptide identification score reported by search engine (e.g., E-value, p-value, probability, Mascot score, etc.)
File	Spectrum Identifier	Precursor m/z	Precursor Charge	Peptide Sequence	Accession(s)	Num. Phosphosites	Peptide Identification Certainty	Localization Certainty	Peptide Identification Score		
D20090930_PM_K Scan:908		558.7576		2 qGsPVAAGAPAK	Q9NZ18	1	Y	Y	0.0002097		
D20090930_PM_K Scan:2017		710.82233		2 TsPDPSPVSAAPSK	Q13469	1	Y	N	45.41		
D20090930_PM_K Scan:683		692.28891		2 _APQTS(ph)S(ph)SPPPVR_	Q8IYB3	2	Y	N	30.09		
D20090930_PM_K Scan:4832		775.3548		2 SQtPPGVAtPPIPK	Q15648	2	Y	N	31.79		
D20090930_PM_K Scan:641		590.2127		2 SLsGSsPcPK	Q9UQ35	2	Y	N	0.0112023		
D20090930_PM_K Scan:641		590.2127		2 sLGSsPcPK	Q9UQ35	2	Y	N	0.0915611		





*Proteome Informatics*  
*Research Group*

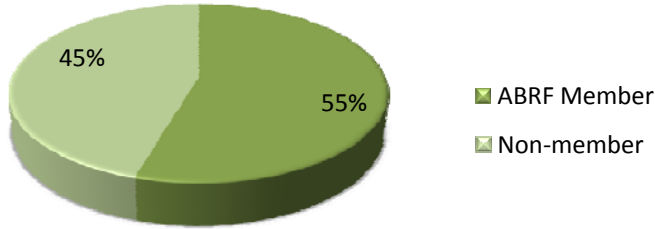
# THE PARTICIPANTS



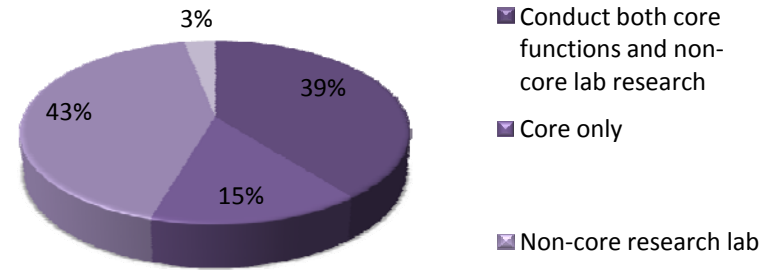
# • 59 requests / 32 submissions (54% return)

- 2 retractions
- + 7 iPRG members and 1 guest

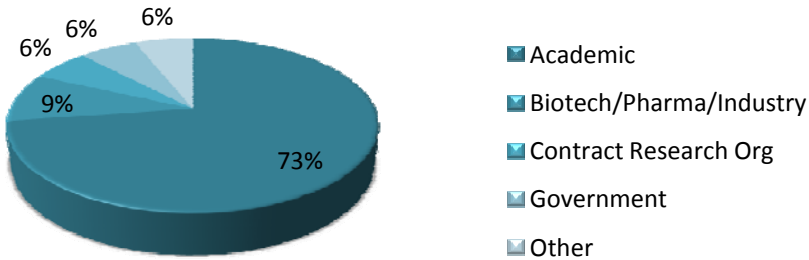
## Membership (n=33)



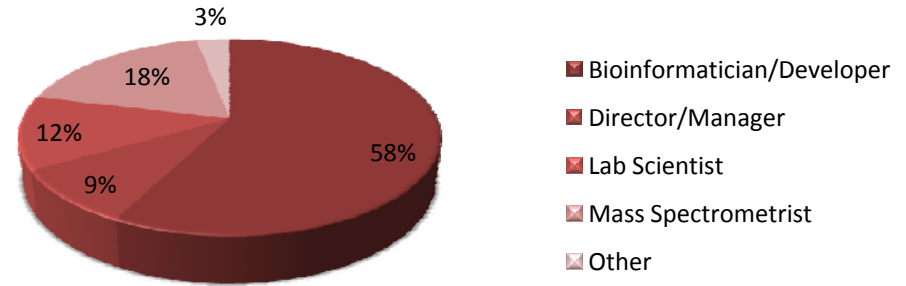
## Resource Lab Status



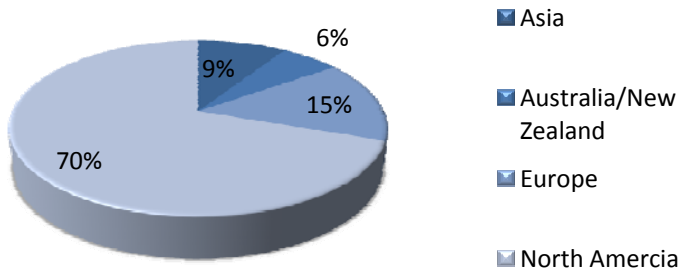
## Type of Lab



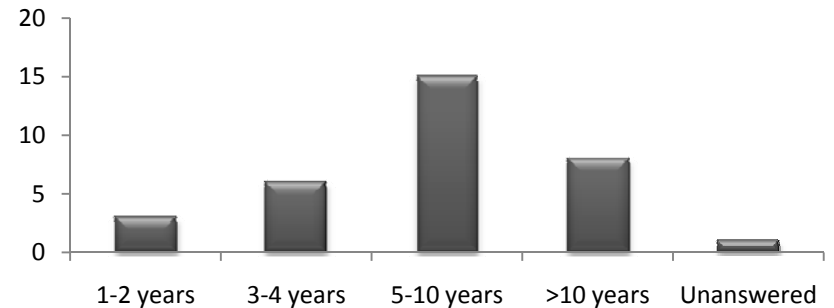
## Primary Job Function



## Location



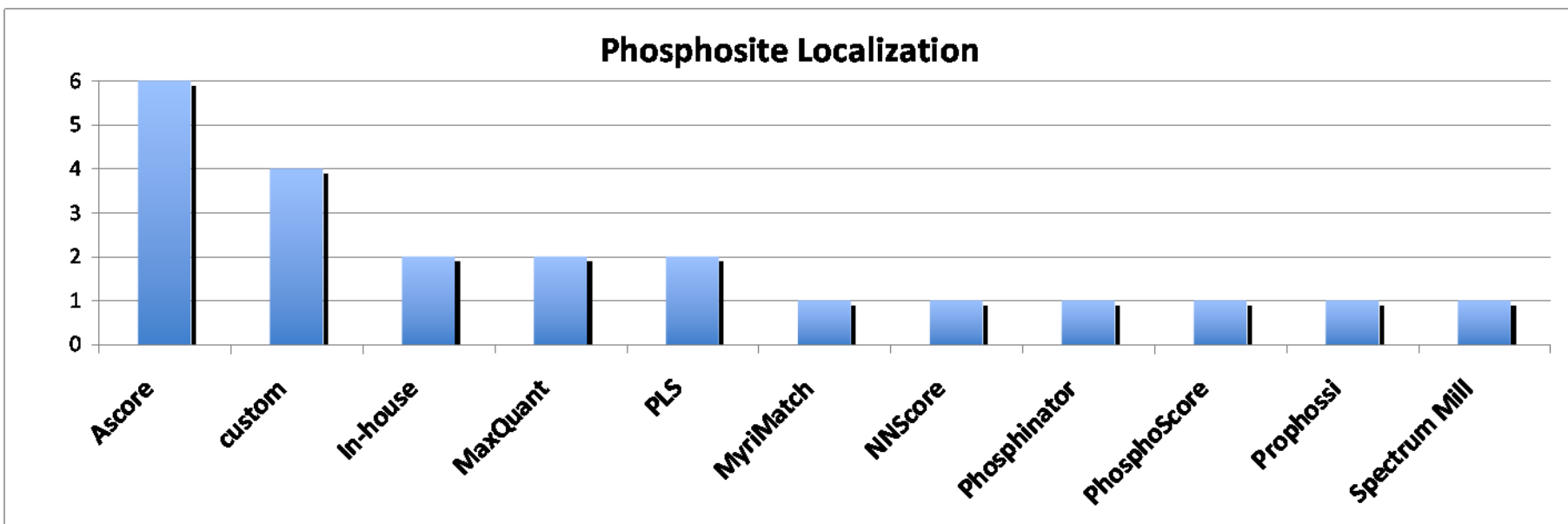
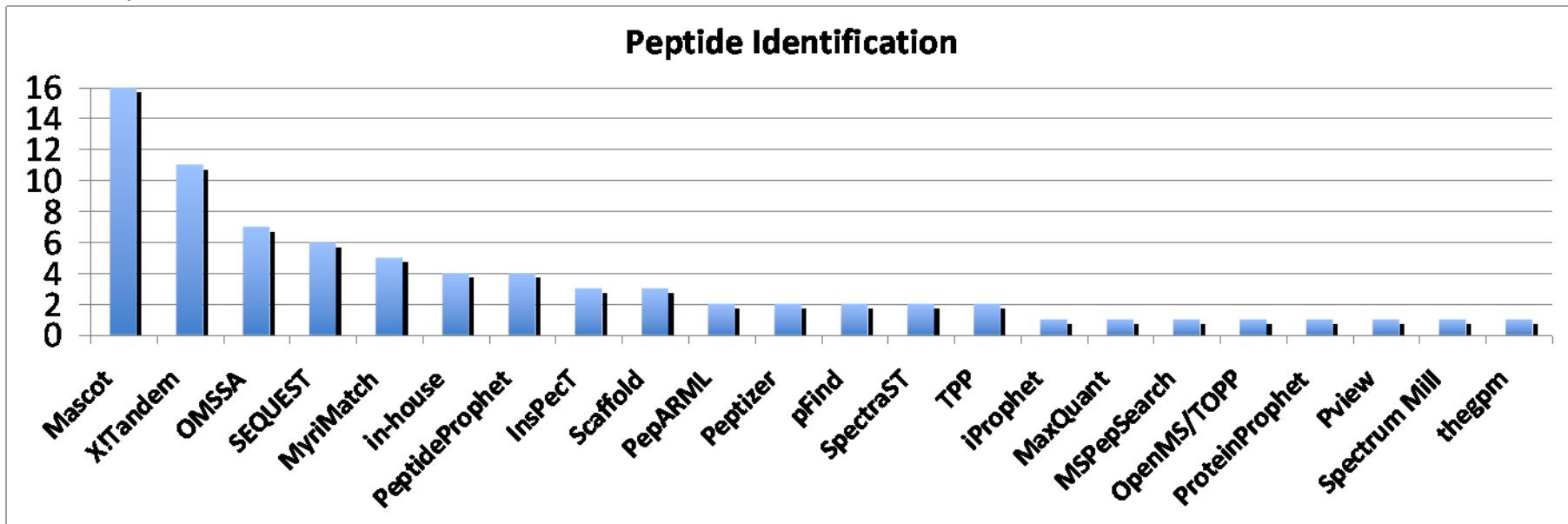
## Proteomics Experience





# Software Tools Used

Proteome Informatics  
Research Group



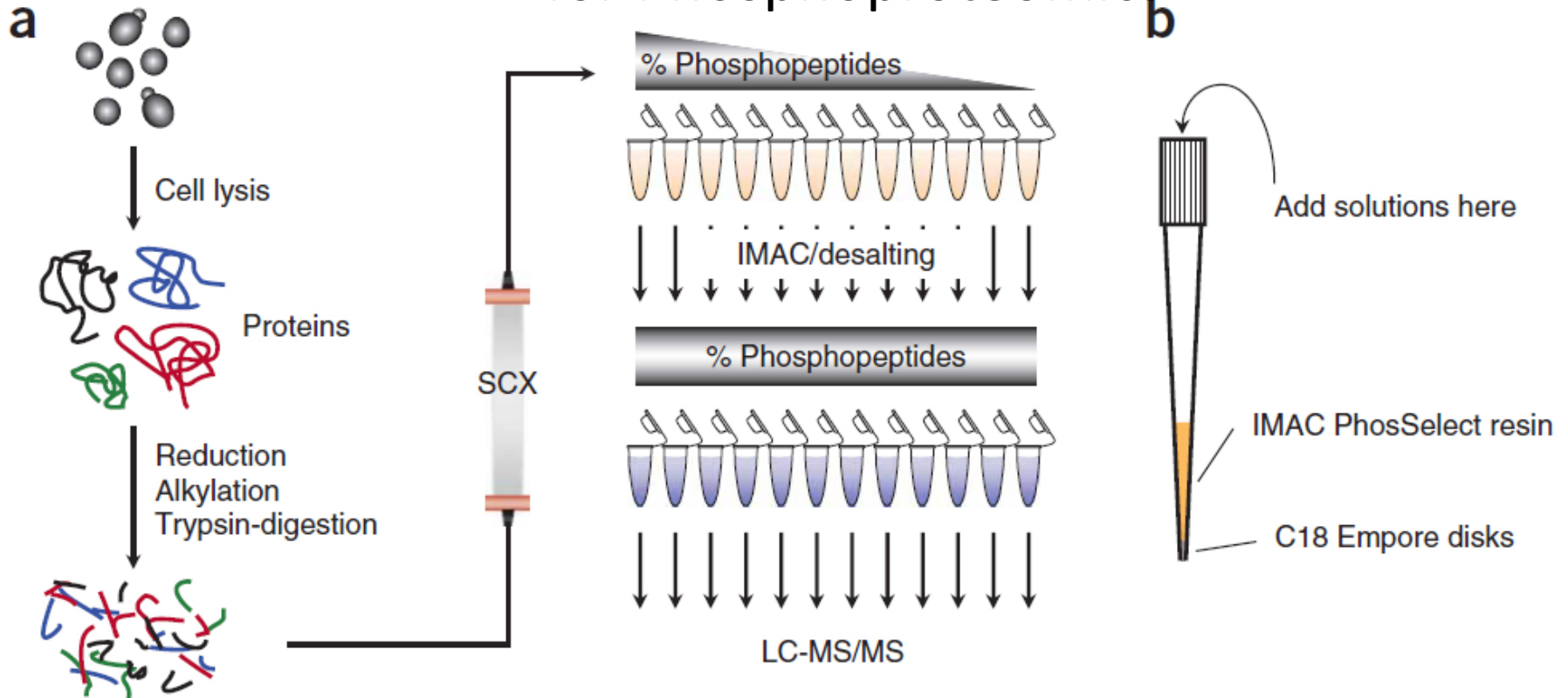


*Proteome Informatics*  
*Research Group*

# GENERATING A DATASET



# The SCX/IMAC Enrichment Approach for Phosphoproteomics



Sample: 7.5x10<sup>7</sup> human K562 human chronic myelogenous leukemia cells, 4mg lysate

Protocol: Villen, J, and Gygi, SP, Nat Prot, 2208, 3, 1630-1638.

Lysis: 8M urea, 75mM NaCl, 50 mM Tris pH 8.2, phosphatase inhibitors

SCX: PolyLC - Polysulfoethyl A 9.4 mm X 200mm, elute: 0-105mM KCl , 30% Acn .

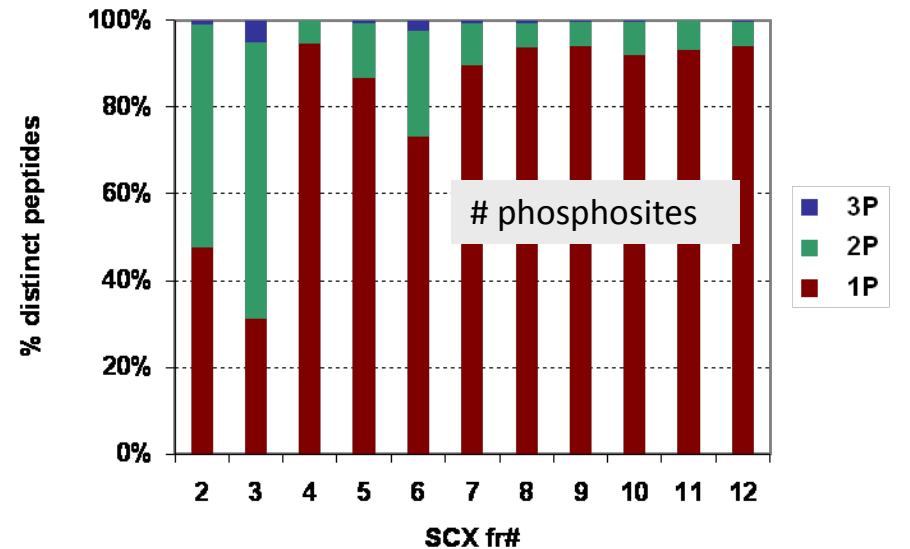
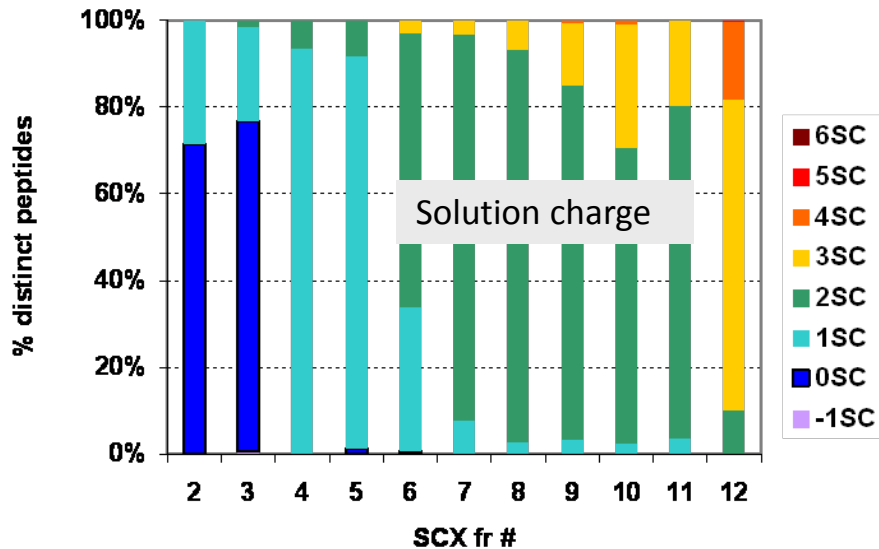
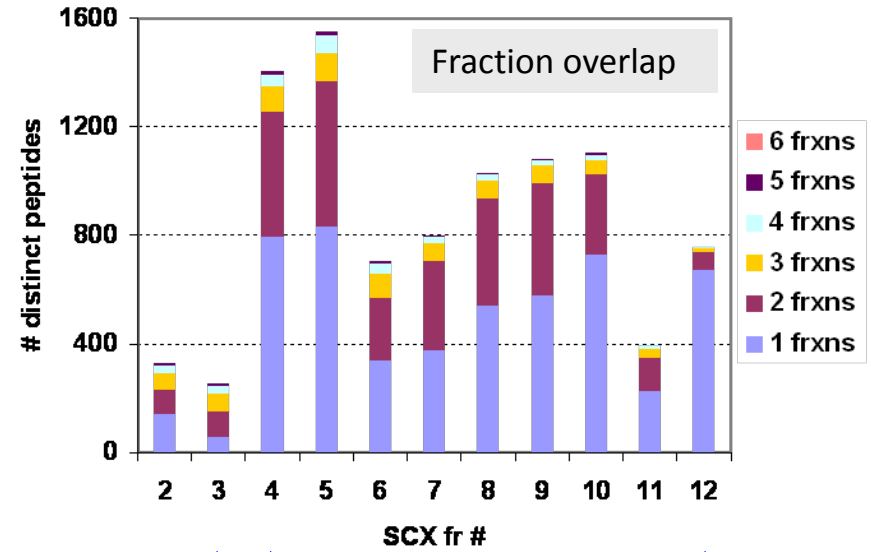
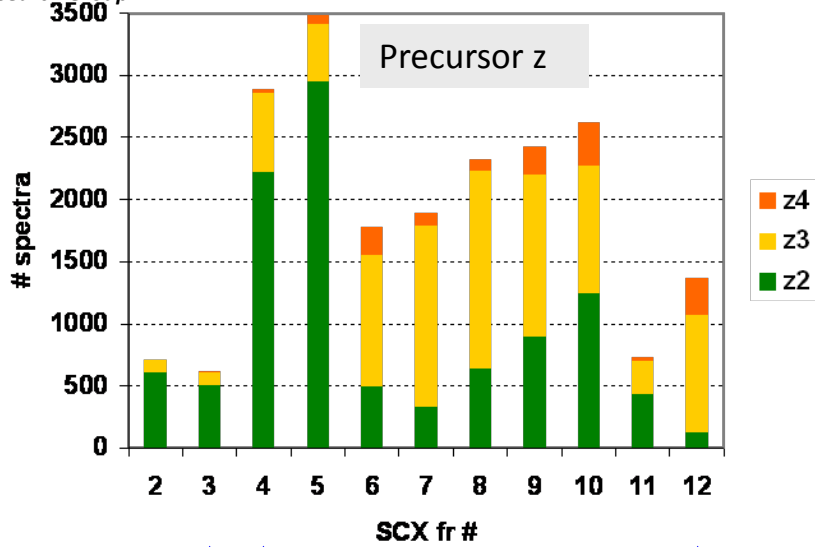
IMAC: Sigma - PhosSelect Fe IMAC beads, bind: 40% Acn, 0.1% formic acid, elute: 500 mM K<sub>2</sub>HPO<sub>4</sub> pH 7

MS/MS: Thermo Fisher Orbitrap XL, high-res MS1 scans in the Orbitrap (60k), Top-8 fragmented in LTQ, exclude +1 and precursors w/ unassigned charges, 20s exclusion time, precursor mass error +/- 10 ppm



# Preliminary Analysis of SCX Fractions and Dataset Selection

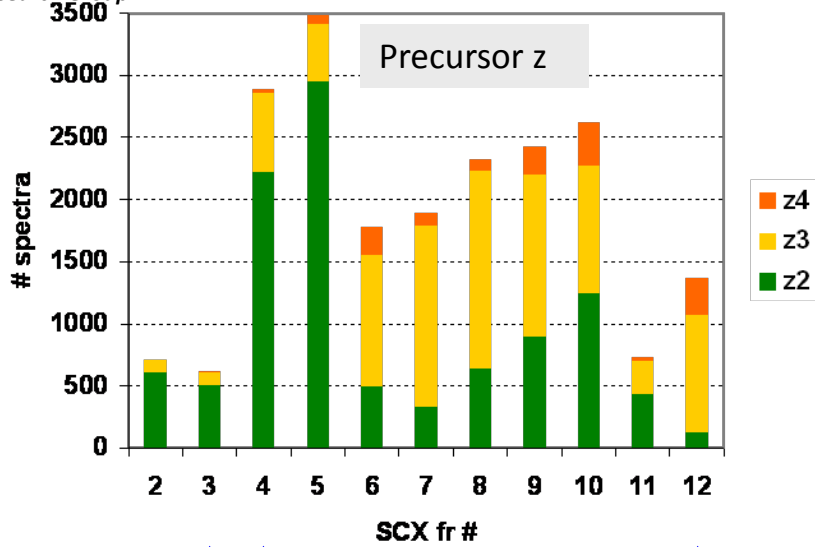
Proteome Informatics  
Research Group





# Preliminary Analysis of SCX Fractions and Dataset Selection

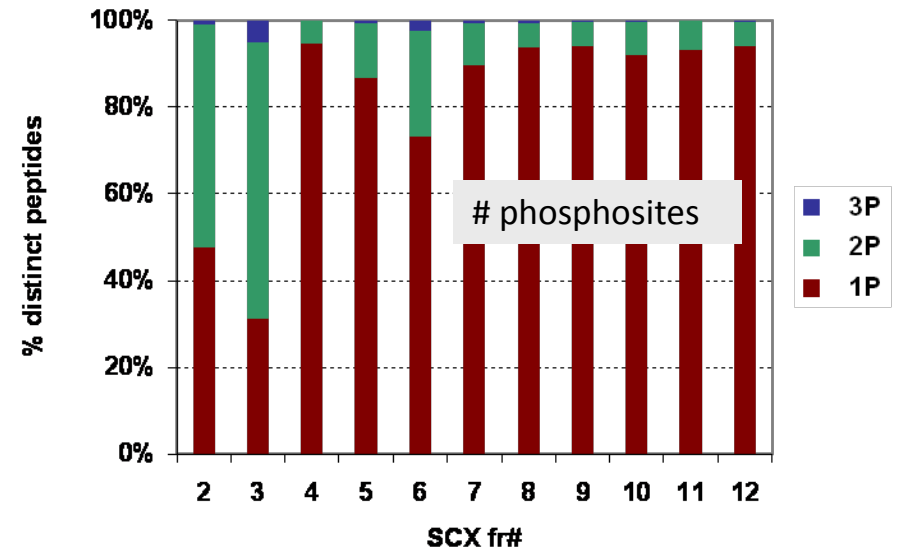
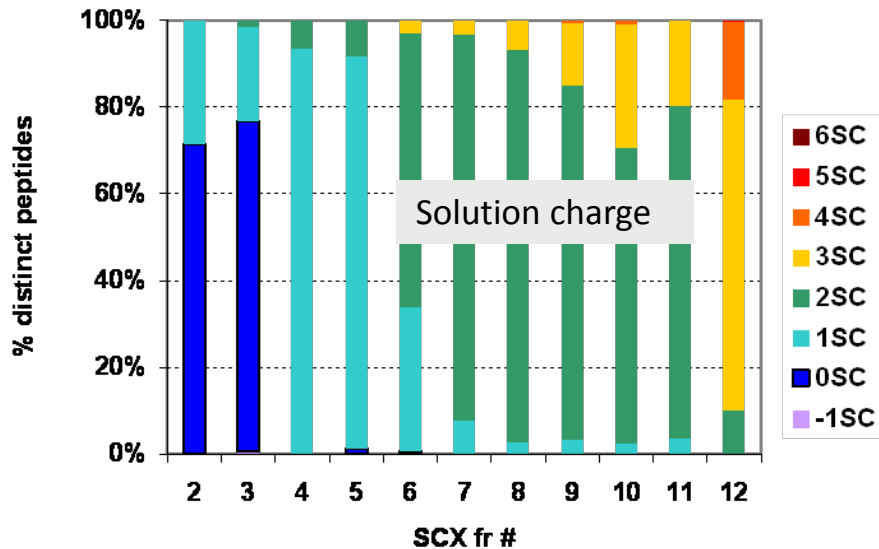
Proteome Informatics  
Research Group



Frxn 3: multi-phosphosites

Frxn 4: single phospho, single basic

Frxn 12: multi-basic residues (RHK)





*Proteome Informatics*  
*Research Group*

# PHOSPHOPEPTIDE IDENTIFICATION





# Notes on Analysis

- Identification and localization were analyzed separately
- All non-phosphopeptide IDs removed
- Reported confidence indicators were used as filters
  - Id = Is peptide spectrum match above 1% FDR? (Y|N)
  - Loc = Are **ALL** phosphosites unambiguously assigned? (Y|N)
- Mods indicated on S|T|Y are assumed phos, others ignored
  - Unique peptide comparison

SL<sup>sm(ox)</sup>DSQVPVYSPSIDLK → SLSMDSQVPVYSPSIDLK  
SL<sup>sm(ox)</sup>DSQVPVYSP<sup>s</sup>IDLK ↗

- Phosphopeptide comparison

SL<sup>sm(ox)</sup>DSQVPVYSPSIDLK → SL<sup>s</sup>MDSQVPVYSPSIDLK  
SL<sup>sm(ox)</sup>DSQVPVYSP<sup>s</sup>IDLK → SLSMDSQVPVYSP<sup>s</sup>IDLK



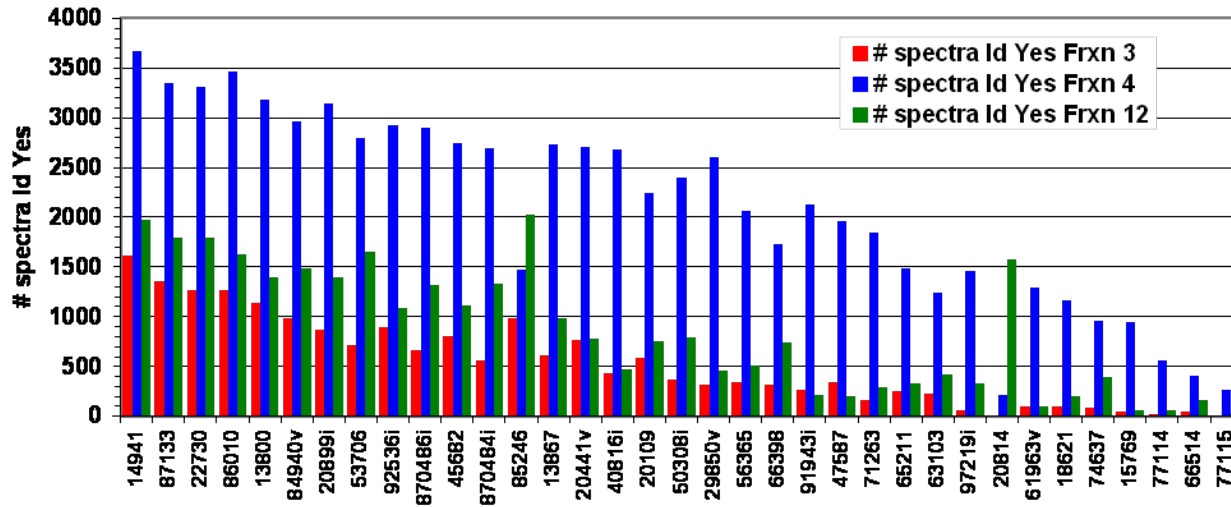


## Quotes From Participants (grading?)

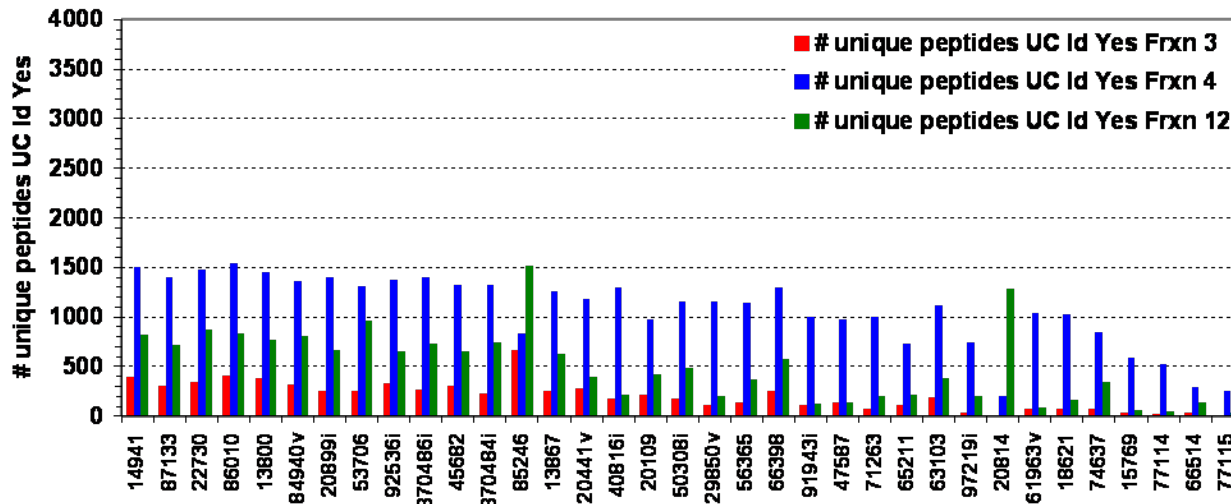
- “It is hard to see how the results of this analysis will be assessed. How will the accuracy of individual methods be determined? It is always easy to find more matches but less easy to determine whether they are credible.”
- “The most challenging part of this study is phosphate localization. But, the data in this study is not a gold standard. So, it's hard to judge which method works. .”
- “Well-designed study. I am looking forward to the results, and I am honestly wondering how the study designer comes up with a "model answer" (or maybe there won't be one?).”
- “How sure are the authors of phosphosite localisation? Have synthetic peptides been made to validate many of the peptides?”



# Relative Performance: Identification By Fraction



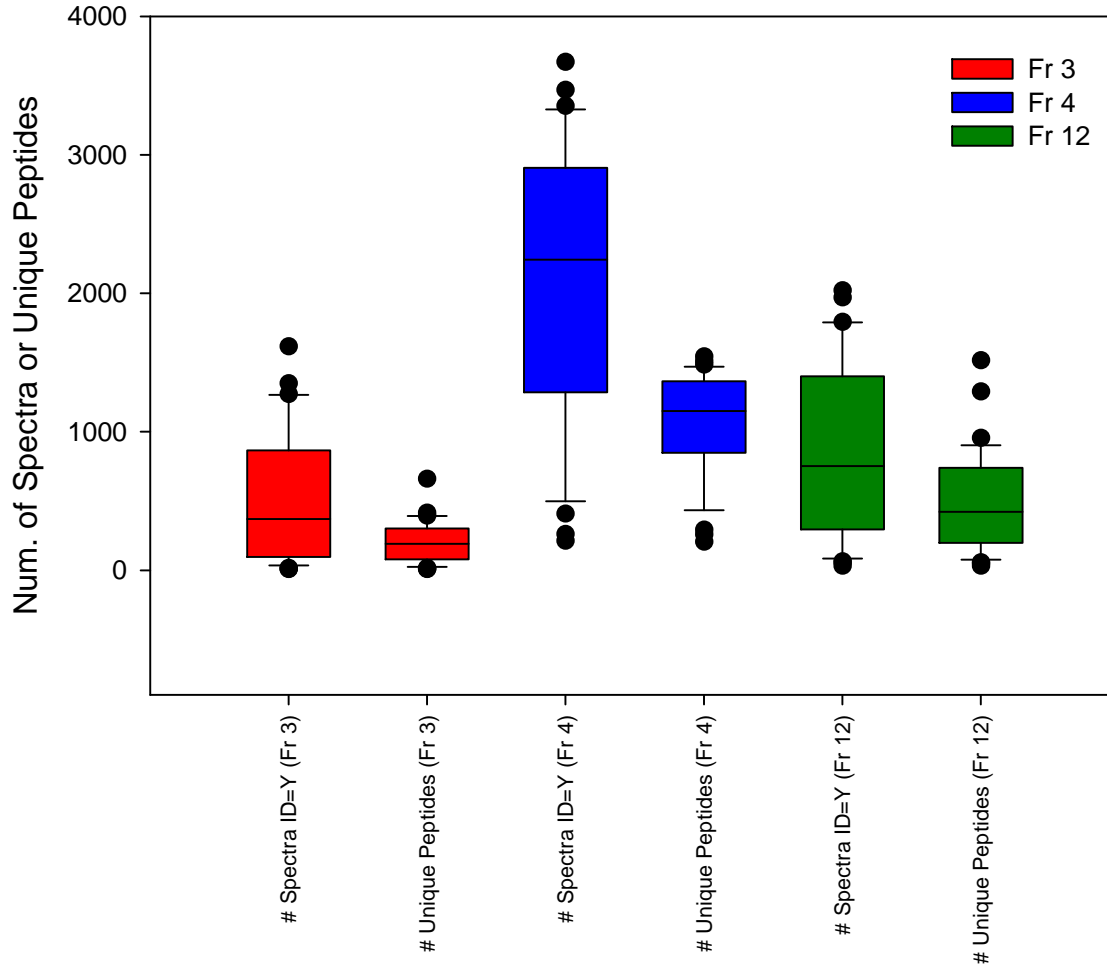
Performance was not equivalent across the 3 fractions for all participants.



Some participants saw more unique peptides than others.



# How Much Did Phosphopeptide Identifications (Spectra and Unique Peptides) Vary?



Unique Phosphopeptide CVs

Fr3=74%

Fr4=31%

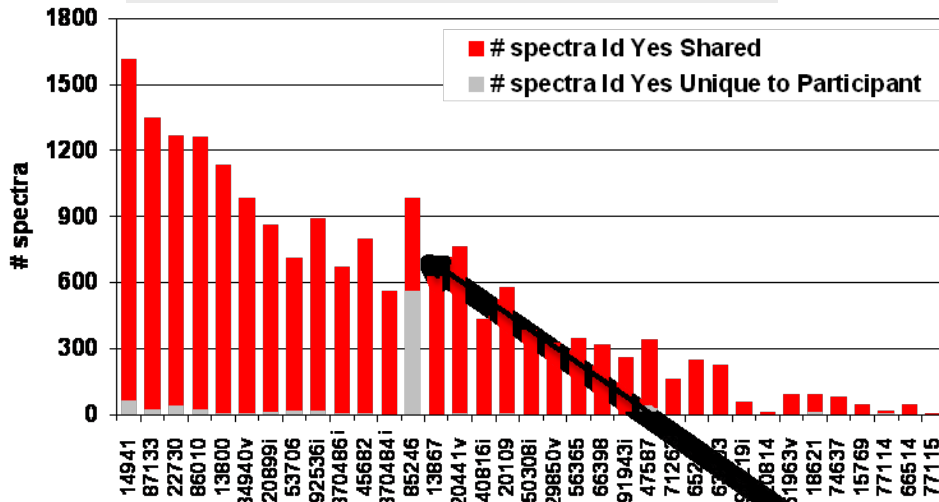
Fr12=84%\*



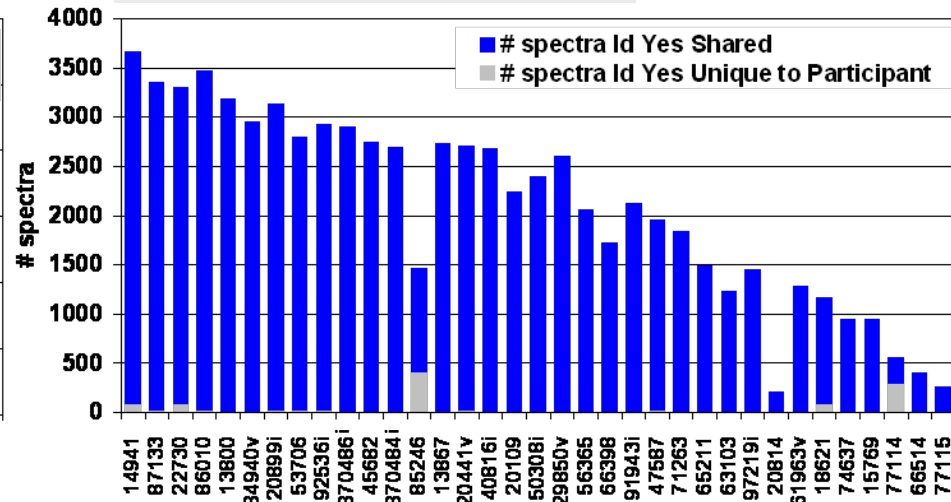
# One Participant Wonders

Proteome Informatics  
Research Group

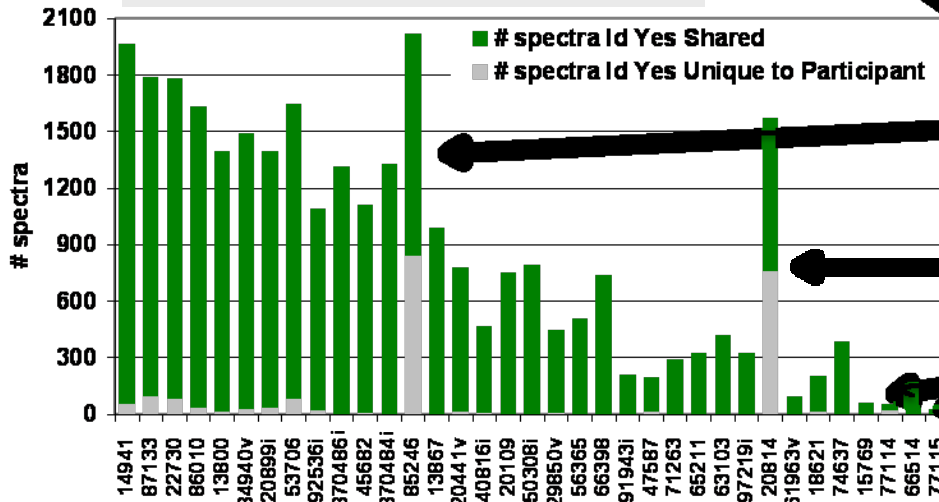
### Frnx 3 – most multiple phos per peptide



### Frnx 4 – most phosphopeptides



### Frnx 12 – highest precursor charges



Gray means – Number of spectra where < 2 people agreed on the Id

85246: 1205 spectra with 3-15 phosphosites, 624 spectra with 4-15

20814: ?, Frnx 12 >> Frnx 3,4

77114, 77115: merged multiple scans, so can't be compared with other 33

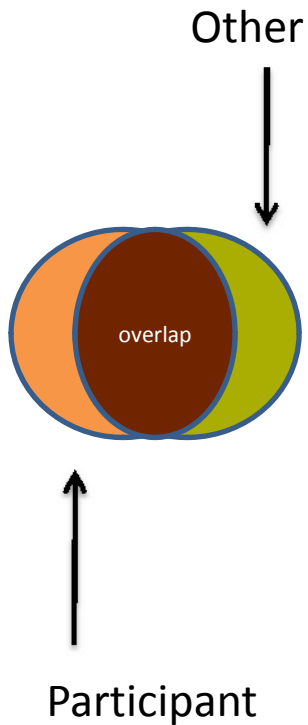


*Proteome Informatics*  
*Research Group*

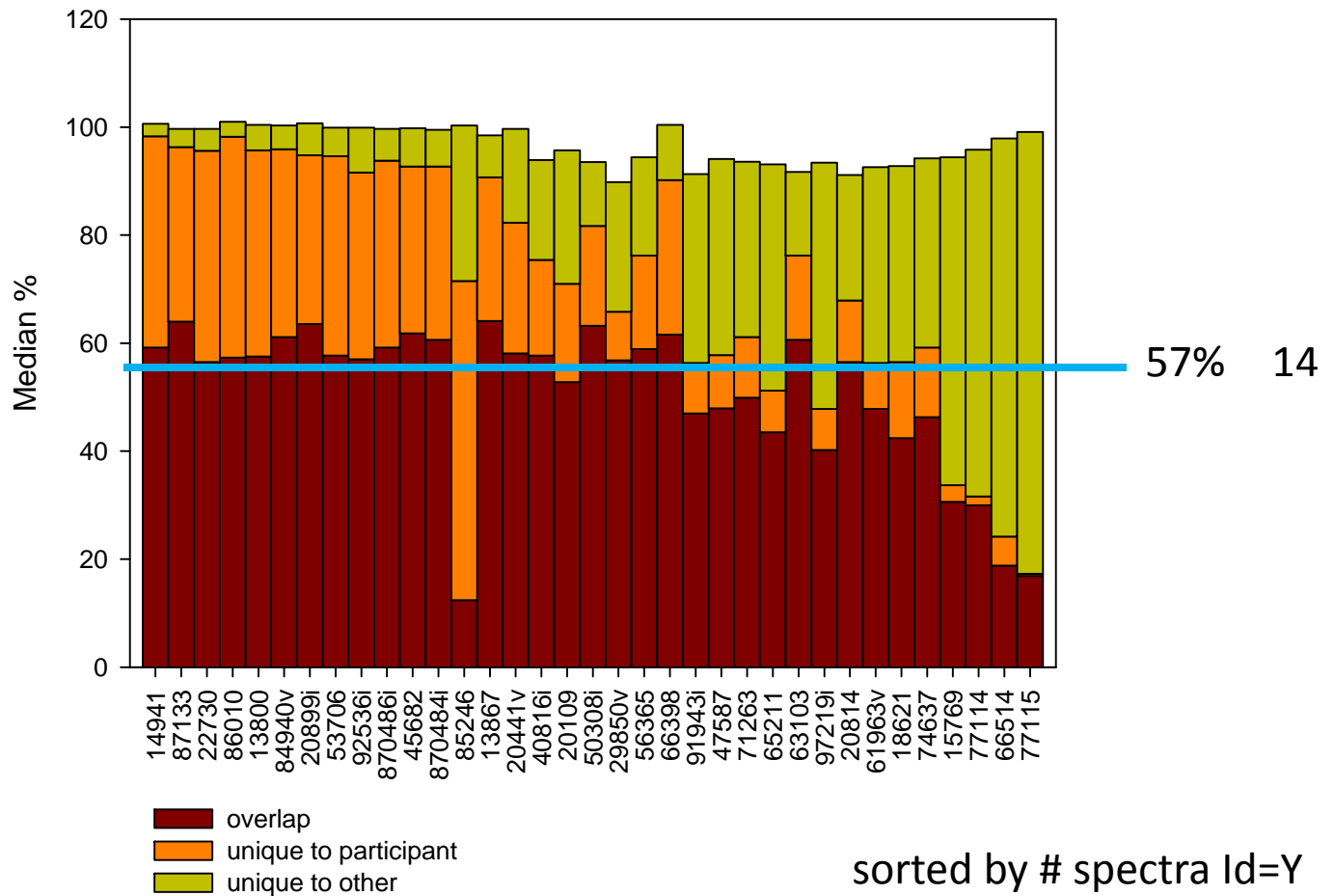
# **SIMILARITY OF RESULT SETS**



# On Average, How Similar Are the Sets of Confidently Identified Unique Peptides?



Pairwise Comparisons of Unique Peptides - All Fractions









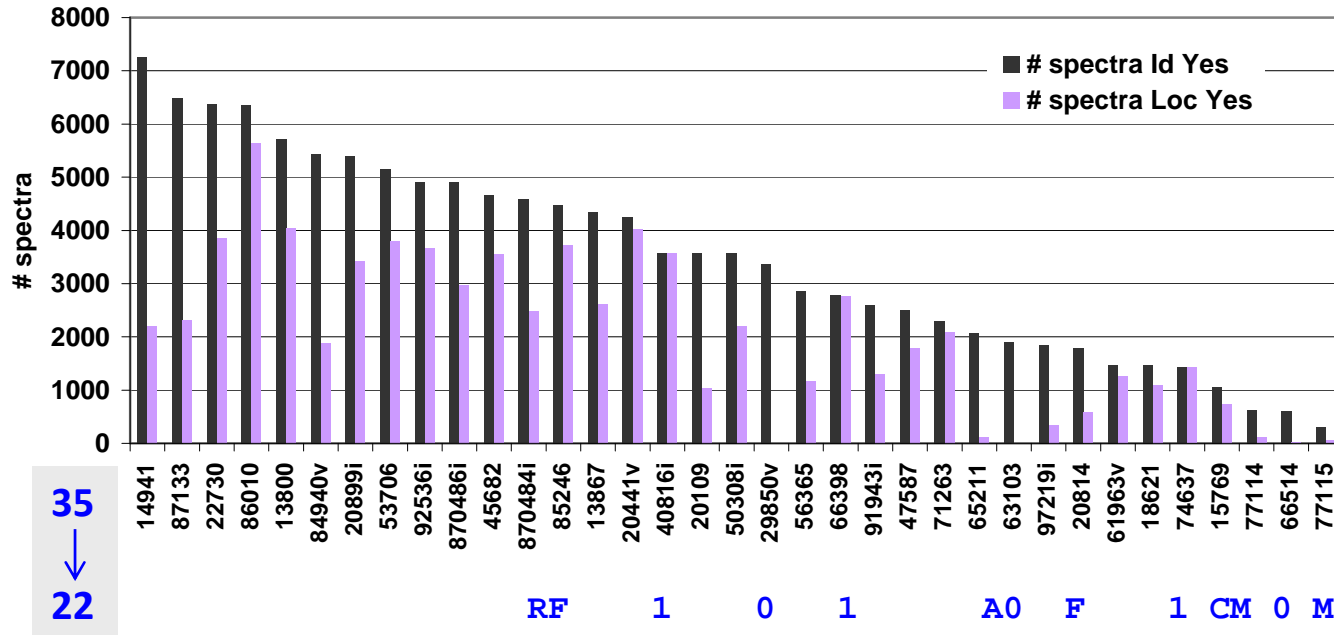
*Proteome Informatics*  
*Research Group*

# LOCALIZATION ANALYSIS



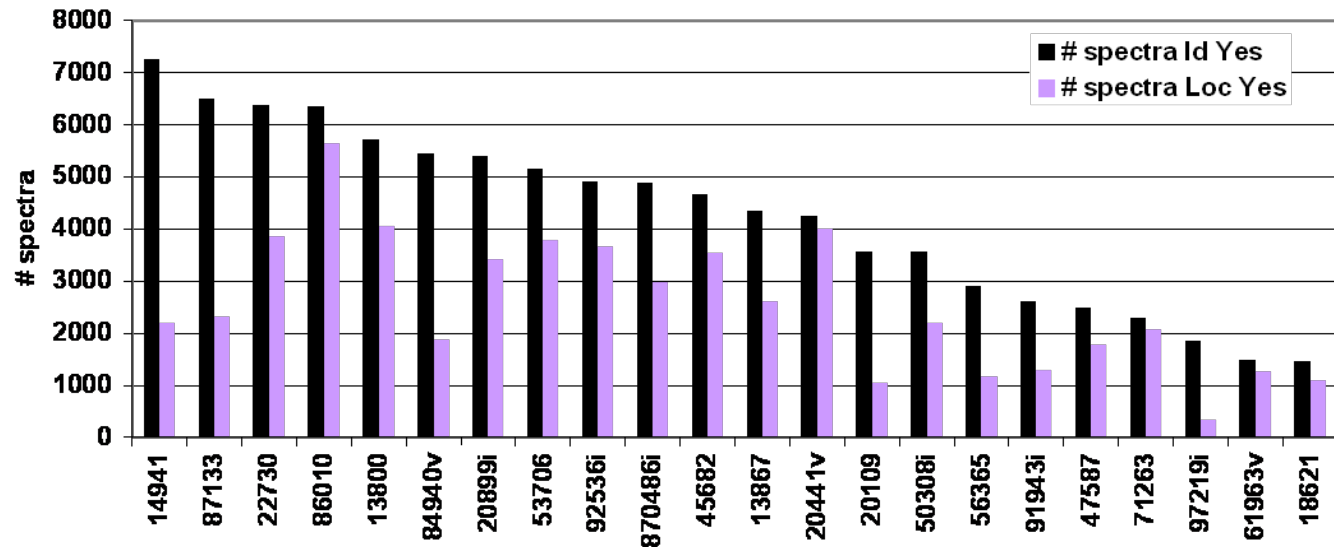
# Subset of Participants Used for Localization Analysis

Proteome Informatics  
Research Group



## Excluded

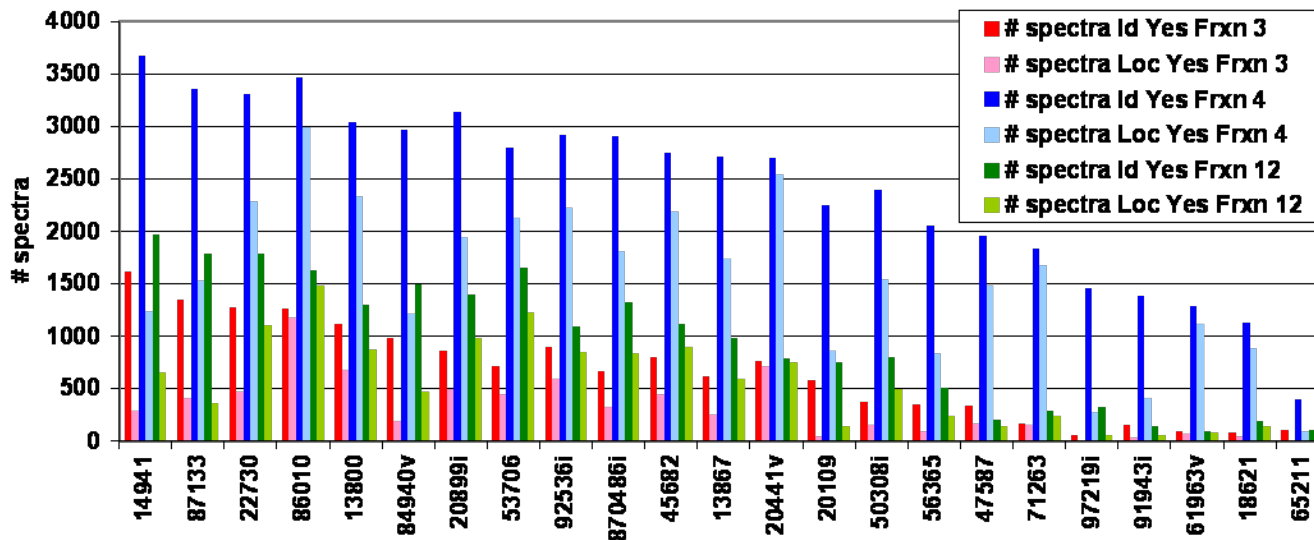
- 0 0% localization
- 1 100% localization
- F FDR - very high?
- R Replicate submission
- M Merged spectra
- C Categorization Errors
- A Y Loc only when no possible ambiguity



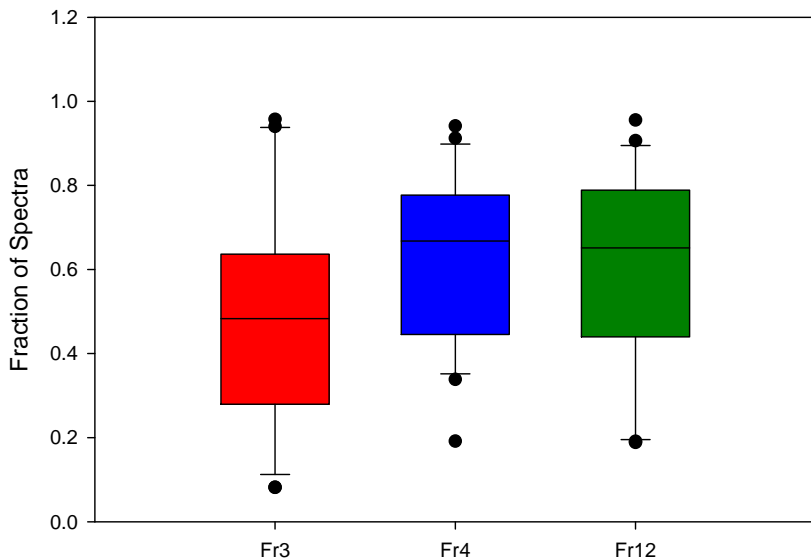


# Wide Range in Willingness to be Certain of Localization

Proteome Informatics  
Research Group



Fraction of Confidently Identified Spectra (Id=Y) Marked Fully Localized (Loc=Y)



Median Fraction of Confident Spectra Marked Loc=Y

Fr3 = 48%

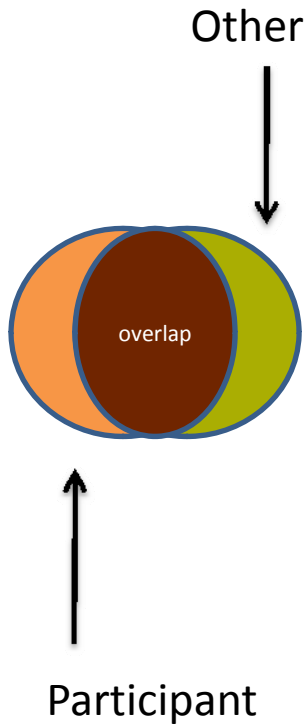
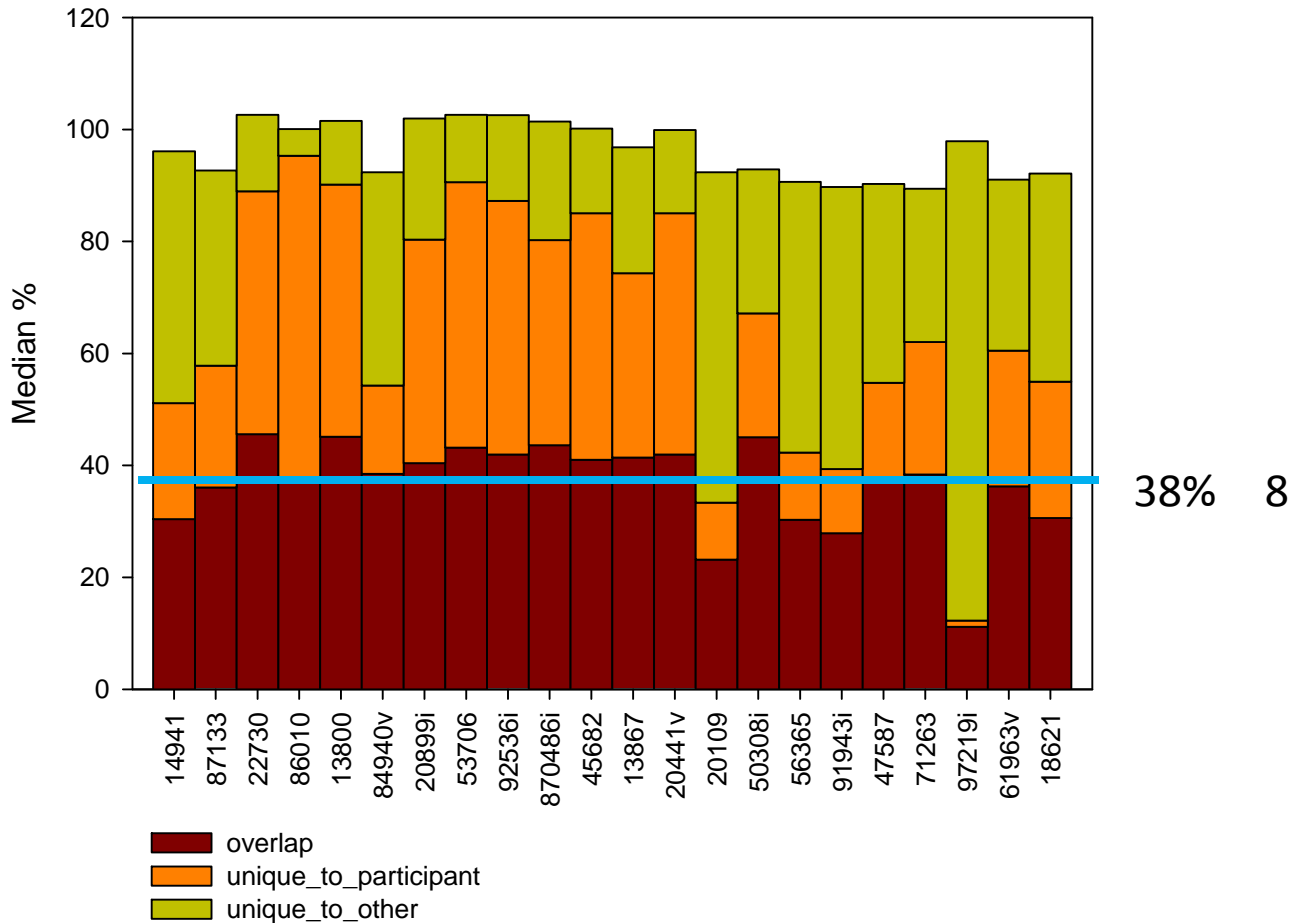
Fr4 = 67%

Fr12 = 65%



# On Average, How Similar Are the Sets of Confidently Identified and Localized Phosphopeptides (Id=Y, Loc=Y)?

### Pairwise Comparisons of Phosphopeptides - All Fractions



38% 8

n=22

sorted by # spectra Id=Y



# Phosphopeptide %Overlap

Proteome Informatics  
Research Group

By num. spectra Id=Y

	14941	87133	22730	86010	13800	84940v	20899i	53706	92536i	870486i	45682	13867	20441v	20109	50308i	56365	91943i	47587	71263	97219i	61963v	18621
14941		35.8	31.8	28.5	33.7	31	35.2	29.8	31.3	36.6	32.2	33.8	26.6	19.5	33.4	26.8	20.9	27.2	27.3	16.1	24	24
87133	35.8		37.7	33.5	39.4	32.5	35.9	36.7	35.4	38.1	39.5	39.4	34.8	25.2	41.6	28.1	30.8	38.4	38	13.3	36.2	30.5
22730	31.8	37.7		51	58.3	47.8	45.8	56.8	53.1	51.5	50.2	51.5	45.4	23.4	54.4	31.8	27.3	38.5	38.9	8.7	38	30.7
86010	28.5	33.5	51		66.4	33.1	44.1	52.6	51.6	45.5	47.9	40.2	46.6	18.2	39.2	24.7	22.6	30.4	34.9	6.4	31.8	24.7
13800	33.7	39.4	58.3	66.4		41.6	49.1	55.4	58.8	52.7	52.6	47.7	49.2	23.2	49.9	30.9	29.1	37.9	42.5	8.5	37.9	31
84940v	31	32.5	47.8	33.1	41.6		40.6	40.5	39.8	44.9	39.1	50.8	33.5	19.3	48.9	45.2	27.4	37.9	35.1	11.1	33.6	29.7
20899i	35.2	35.9	45.8	44.1	49.1	40.6		45.9	45.9	48.6	49.4	44.3	40.2	20.3	45.1	31.2	25.5	34	35.6	10.4	31.8	29.1
53706	29.8	36.7	56.8	52.6	55.4	40.5	45.9		50.8	46.2	50.2	48.1	45.9	22.3	48.4	29.9	25.4	36.8	38.1	8	36.6	28.8
92536i	31.3	35.4	53.1	51.6	58.8	39.8	45.9	50.8		49	45.7	44.5	44.1	22.4	45	29.8	27.3	36.7	37.9	7.7	34.9	28.7
870486i	36.6	38.1	51.5	45.5	52.7	44.9	48.6	46.2	49		45.9	46.9	42.3	23.1	50.1	32	28.5	39.6	39.6	10.7	36.3	30.7
45682	32.2	39.5	50.2	47.9	52.6	39.1	49.4	50.2	45.7	45.9		44.8	42.5	20.4	45.4	29.4	25.4	36.6	37.6	8.9	37.1	32.8
13867	33.8	39.4	51.5	40.2	47.7	50.8	44.3	48.1	44.5	46.9	44.8		41.6	24.8	54.3	36.5	28.9	40.3	41.2	11.3	38.6	31.8
20441v	26.6	34.8	45.4	46.6	49.2	33.5	40.2	45.9	44.1	42.3	42.5	41.6		26.3	43.6	28	28.4	38.5	42.6	8.8	47.3	29.6
20109	19.5	25.2	23.4	18.2	23.2	19.3	20.3	22.3	22.4	23.1	20.4	24.8	26.3		28.6	22	25.1	30.9	31.8	14.2	32.6	20
50308i	33.4	41.6	54.4	39.2	49.9	48.9	45.1	48.4	45	50.1	45.4	54.3	43.6	28.6		38.6	37.5	46.1	48.4	13.4	44.2	36.8
56365	26.8	28.1	31.8	24.7	30.9	45.2	31.2	29.9	29.8	32	29.4	36.5	28	22	38.6		26	34	38.7	13.6	30.7	26.9
91943i	20.9	30.8	27.3	22.6	29.1	27.4	25.5	25.4	27.3	28.5	25.4	28.9	28.4	25.1	37.5	26		39.4	39.8	10.5	36.1	31.1
47587	27.2	38.4	38.5	30.4	37.9	37.9	34	36.8	36.7	39.6	36.6	40.3	38.5	30.9	46.1	34	39.4		46.7	12.7	47.3	35.4
71263	27.3	38	38.9	34.9	42.5	35.1	35.6	38.1	37.9	39.6	37.6	41.2	42.6	31.8	48.4	38.7	39.8	46.7		12.2	45.2	37.5
97219i	16.1	13.3	8.7	6.4	8.5	11.1	10.4	8	7.7	10.7	8.9	11.3	8.8	14.2	13.4	13.6	10.5	12.7	12.2		12.3	11.2
61963v	24	36.2	38	31.8	37.9	33.6	31.8	36.6	34.9	36.3	37.1	38.6	47.3	32.6	44.2	30.7	36.1	47.3	45.2	12.3		33.4
18621	24	30.5	30.7	24.7	31	29.7	29.1	28.8	28.7	30.7	32.8	31.8	29.6	20	36.8	26.9	31.1	35.4	37.5	11.2	33.4	

## Descending # of total spectra Id=Y

By median %overlap

	22730	13800	50308i	870486i	53706	92536i	20441v	13867	45682	20899i	84940v	71263	47587	86010	61963v	87133	18621	14941	56365	91943i	20109	97219i
22730		58.3	54.4	51.5	56.8	53.1	45.4	51.5	50.2	45.8	47.8	38.9	38.5	51	38	37.7	30.7	31.8	31.8	27.3	23.4	8.7
13800	58.3		49.9	52.7	55.4	58.8	49.2	47.7	52.6	49.1	41.6	42.5	37.9	66.4	37.9	39.4	31	33.7	30.9	29.1	23.2	8.5
50308i	54.4	49.9		50.1	48.4	45	43.6	54.3	45.4	45.1	48.9	48.4	46.1	39.2	44.2	41.6	36.8	33.4	38.6	37.5	28.6	13.4
870486i	51.5	52.7	50.1		46.2	49	42.3	46.9	45.9	48.6	44.9	39.6	39.6	45.5	36.3	38.1	30.7	36.6	32	28.5	23.1	10.7
53706	56.8	55.4	48.4	46.2		50.8	45.9	48.1	50.2	45.9	40.5	38.1	36.8	52.6	36.6	36.7	28.8	29.8	29.9	25.4	22.3	8
92536i	53.1	58.8	45	49	50.8		44.1	44.5	45.7	45.9	39.8	37.9	36.7	51.6	34.9	35.4	28.7	31.3	29.8	27.3	22.4	7.7
20441v	45.4	49.2	43.6	42.3	45.9	44.1		41.6	42.5	40.2	33.5	42.6	38.5	46.6	47.3	34.8	29.6	26.6	28	28.4	26.3	8.8
13867	51.5	47.7	54.3	46.9	48.1	44.5	41.6		44.8	44.3	50.8	41.2	40.3	40.2	38.6	39.4	31.8	33.8	36.5	28.9	24.8	11.3
45682	50.2	52.6	45.4	45.9	50.2	45.7	42.5	44.8		49.4	39.1	37.6	36.6	47.9	37.1	39.5	32.8	32.2	29.4	25.4	20.4	8.9
20899i	45.8	49.1	45.1	48.6	45.9	45.9	40.2	44.3	49.4		40.6	35.6	34	44.1	31.8	35.9	29.1	35.2	31.2	25.5	20.3	10.4
84940v	47.8	41.6	48.9	44.9	40.5	39.8	33.5	50.8	39.1	40.6		35.1	37.9	33.1	33.6	32.5	29.7	31	45.2	27.4	19.3	11.1
71263	38.9	42.5	48.4	39.6	38.1	37.9	42.6	41.2	37.6	35.6	35.1		46.7	34.9	45.2	38	37.5	27.3	38.7	39.8	31.8	12.2
47587	38.5	37.9	46.1	39.6	36.8	36.7	38.5	40.3	36.6	34	37.9	46.7		30.4	47.3	38.4	35.4	27.2	34	39.4	30.9	12.7
86010	51	66.4	39.2	45.5	52.6	51.6	46.6	40.2	47.9	44.1	33.1	34.9	30.4		31.8	33.5	24.7	28.5	24.7	22.6	18.2	6.4
61963v	38	37.9	44.2	36.3	36.6	34.9	47.3	38.6	37.1	31.8	33.6	45.2	47.3	31.8		36.2	33.4	24	30.7	36.1	32.6	12.3
87133	37.7	39.4	41.6	38.1	36.7	35.4	34.8	39.4	39.5	35.9	32.5	38	38.4	33.5	36.2		30.5	35.8	28.1	30.8	25.2	13.3
18621	30.7	31	36.8	30.7	28.8	28.7	29.6	31.8	32.8	29.1	29.7	37.5	35.4	24.7	33.4	30.5		24	26.9	31.1	20	11.2
14941	31.8	33.7	33.4	36.6	29.8	31.3	26.6	33.8	32.2	35.2	31	27.3	27.2	28.5	24	35.8	24		26.8	20.9	19.5	16.1
56365	31.8	30.9	38.6	32	29.9	29.8	28	36.5	29.4	31.2	45.2	38.7	34	24.7	30.7	28.1	26.9	26.8		26	22	13.6
91943i	27.3	29.1	37.5	28.5	25.4	27.3	28.4	28.9	25.4	25.5	27.4	39.8	39.4	22.6	36.1	30.8	31.1	20.9	26		25.1	10.5
20109	23.4	23.2	28.6	23.1	22.3	22.4	26.3	24.8	20.4	20.3	19.3	31.8	30.9	18.2	32.6	25.2	20	19.5	22	25.1		14.2
97219i	8.7	8.5	13.4	10.7	8	7.7	8.8	11.3	8.9	10.4	11.1	12.2	12.7	6.4	12.3	13.3	11.2	16.1	13.6	10.5	14.2	

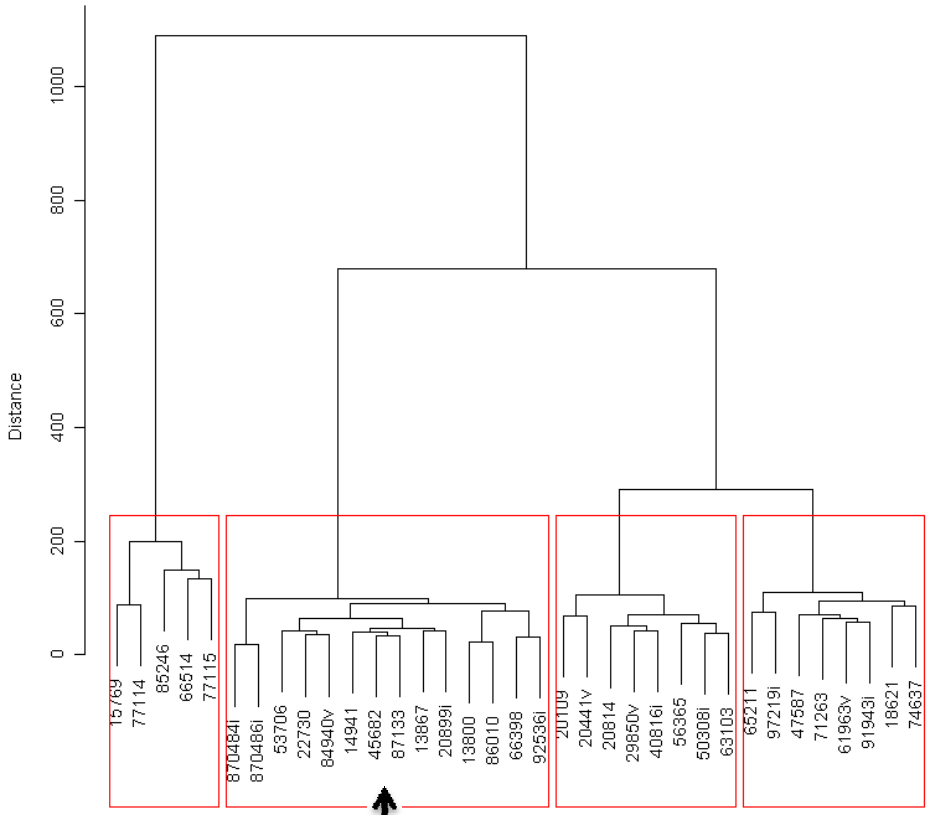
## Descending similarity (median % overlap)

n=22



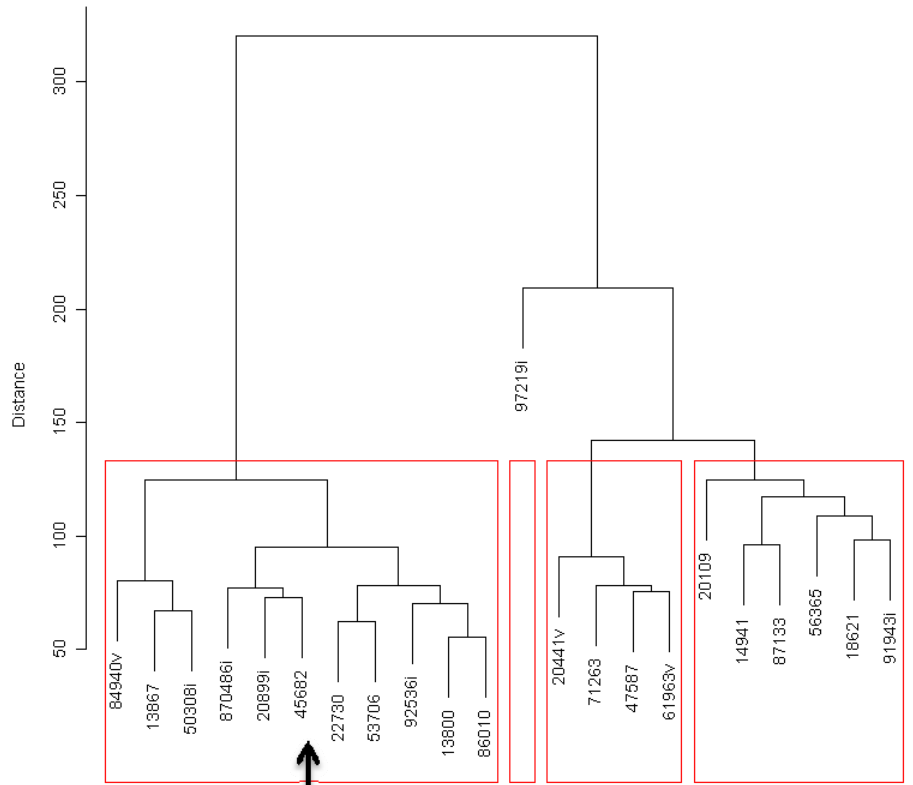
# Relatedness of Participants by Overlap

Clusters by Unique Peptide Overlap



n=35

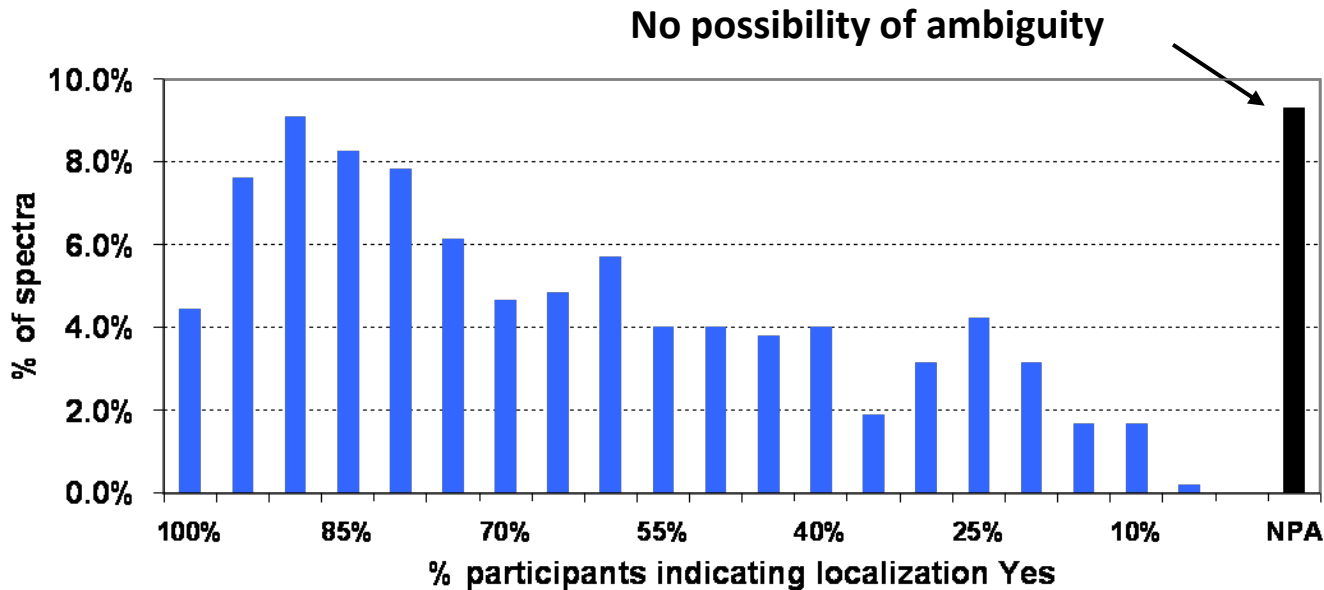
Clusters by PhosphoPeptide Overlap



n=22



# If Participants Agree on the Identity, Do They Also Agree Site Localization Can be Certain?



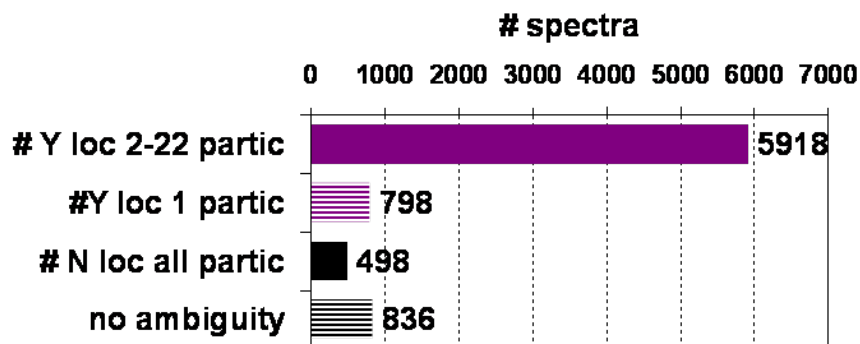
Frxn 4  
Subset of  
472 spectra  
for which  
20/22 participants  
all agree on  
Identity



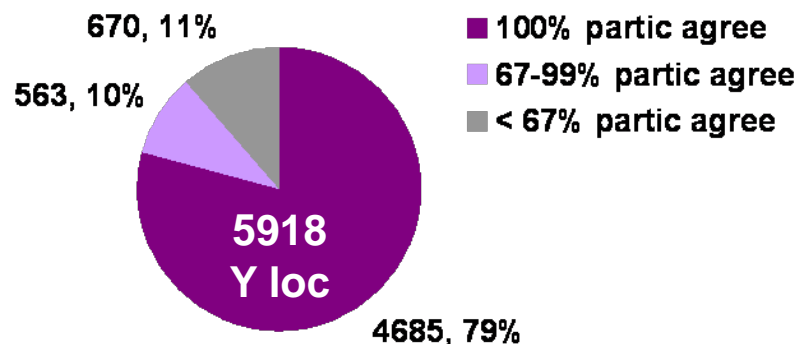


# What Fraction of the Time Do They Agree On Localization(s)?

8050 spectra with  $\geq 2/22$  Id Yes (Frnx 3, 4, 12)



5918/8050 spectra with  $\geq 2/22$  Loc Yes  
and Site Ambiguity Possible



For all of the participants that agree on identity when

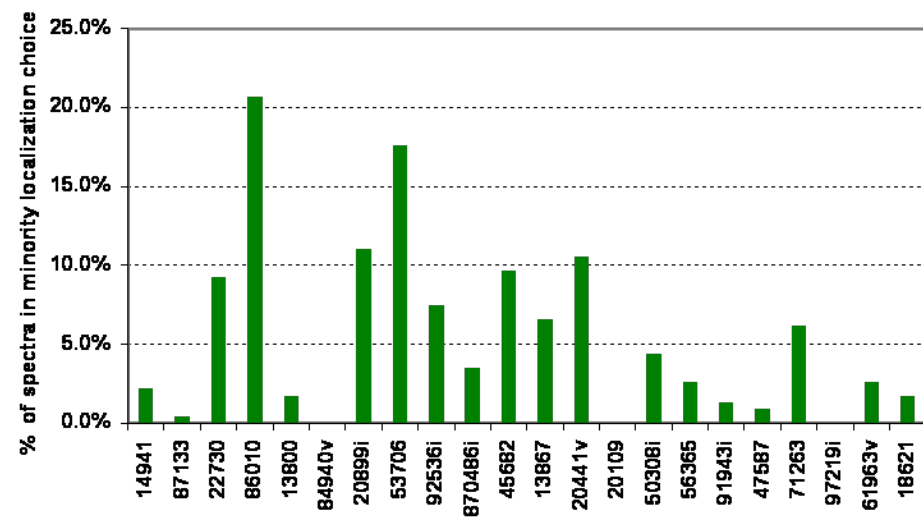
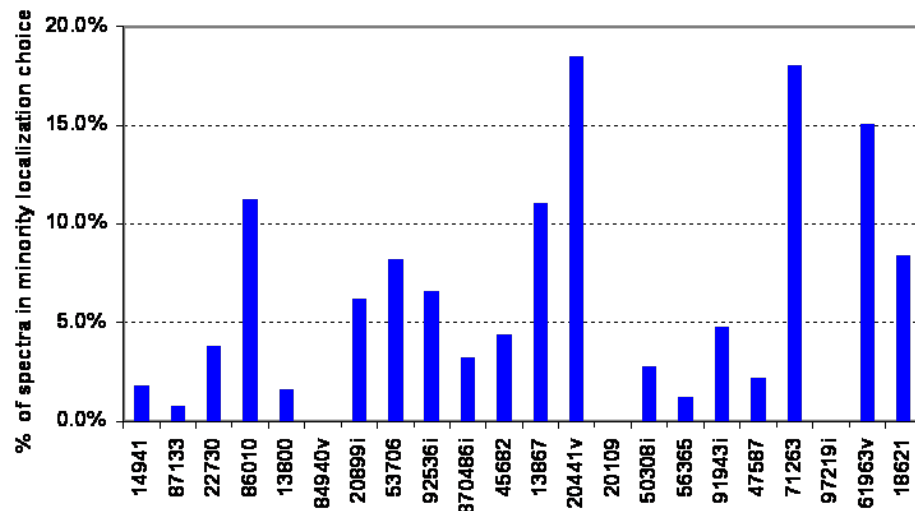
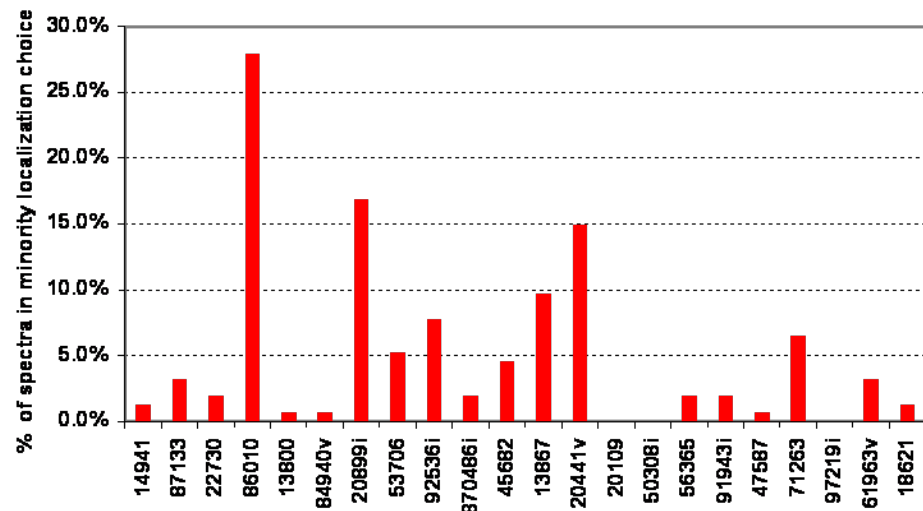
- site ambiguity is possible ( $\#S, T, Y > \#$  phos)
- $\geq 2$  participants mark Loc=Y

➤ For 79% (4,685 of 5,918) of the spectra, all participants who mark Loc=Y unanimously agree on the localization of the phosphosites



# Which Participants are More Likely to Disagree on Localization?

Proteome Informatics  
Research Group



## # Spectra with Loc Agreement 50.1-99.9%

Frxn 3: 154

Frxn 4: 498

Frxn 12: 227

The participants who are the most willing to localize are more likely to disagree with the majority view.



## Quotes From Participants

- “This study has started a dialogue of accurate phosphate identification and localization. Perhaps, the results of this study will point out the inadequacies of current informatics methods in identification and localization of phosphates.”
- “Great study choice! I learned a lot about available software, limitations, etc!”



*Proteome Informatics*  
*Research Group*

# DATA RESOURCE

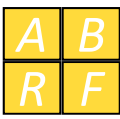


# Resource for Inspecting Peptide Id Certainty Overlaps - Frxn 4

Proteome Informatics  
Research Group

- YY:** Y – identification      Y – localization  
**YN:** Y – identification      N – localization  
**NS:** N – identification, but top sequence same as consensus  
**ND:** N – identification, and top sequence different than consensus

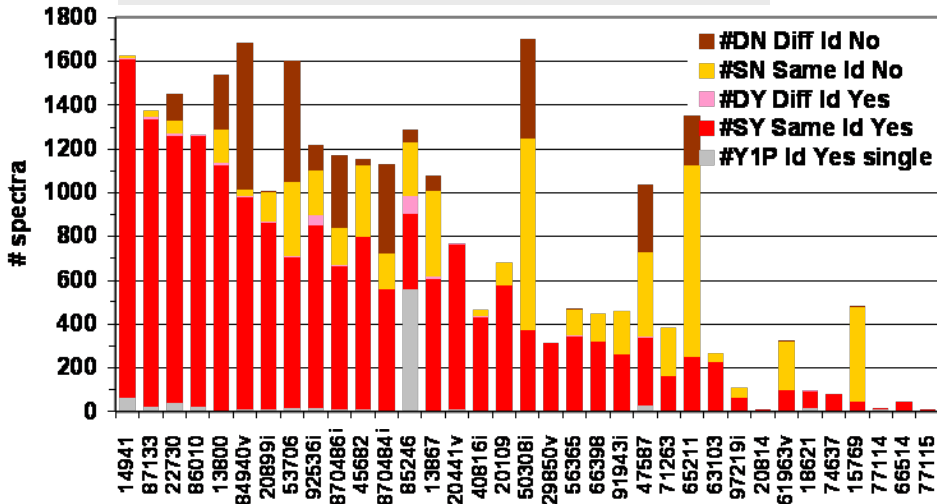
1	spectrum	bestSequence	nTerm	m/z	z	#sty	14941	87133	22730	86010	13800	84940v	20899i	53706	92536i	870486i	45682	13867	20441v	20109	50308i
91	9610	ADEDLIFRLEGVDGGQSPR	Acetyl	732.671	3	1	YY			YY	YY	ND	YY	ND	YY	YY		ND	YY		ND
92	4042	SDNGELEDKPPAPPVR	Acetyl	921.915	2	1	YY			YY	YY	ND			YY	YY					ND
93	9618	ADEDLIFRLEGVDGGQSPR	Acetyl	1098.5008	2	1	YY			YY	YY	ND		ND	YY	YY					ND
94	5115	ASGSPGPGEGSAGGEKR	Acetyl	889.0653	3	1	YN		ND	YY	YY	ND	YN	ND	YY	YN		ND	YY		ND
95	4671	ADHSFSDGVPSDSVEAAK	Acetyl	970.8986	2	1	YN		ND	YY	YY	ND	YN	ND	YY	YN	ND				ND
1790	3055	AGDLLEDSPK		562.749	2	1	YY	YY	YY	YY	YY	YN	YY	YY	YY	YY	NS	YY	YY	YY	YY
1791	1939	QGSPVAAGAPAK		558.7596	2	1	YY	YY		YY	YY	ND	YY			YY	YY	YY	YY	NS	YY
1792	7388	PAGPLLMGAPAGESPQPK		1059.8584	3	2	YY	YY	YN	YY	YY	YN	YY	YY	NS	YY	YY	YY	YY		NS
1793	5784	MVQLSPPALAAPAPGR		872.8923	2	1	YY	YY	ND		NS	ND		NS		YY	YY	YY			ND
1794	4633	IPSPNILK		481.26	2	1	YY	YY	YY	YY	NS	ND	YY	YY	YY	YY	YY	YY	YY	YY	YY
1795	4559	LPLVPESPR		544.2808	2	1	YY	YY	YY	YY	YY	YN	YY	YY	YY	YY	YY	YY	YY	YY	YY
1826	9516	VLNTGSDVEEAVADALK		905.9274	2	1	YY	YN	YY	YY	YY	YY	YY	YY	YY	YN	YY	YY	YY	YY	YY
2251	6134	SLSAASCAPISLPAASR		905.431	2	1	YN	YN	YN	YN	YN	YN	YN	YN	YN	YN	YY	YN		YN	YN
2252	8405	SSSPAPADIAQTVGEDLR		982.95	2	1	YN	YN	YY	YN	YY	YN	YN	YY	YN	YN	NS	NS		YN	NS
2253	6259	IGVPSATEIIK		604.3192	2	1			ND			ND	YY	NS	YY	YN	NS	NS	YY	YN	YN
2254	4298	VLGTSPEAIDSAENR		820.3205	2	1	YN	YN			NS	ND		NS		YN	YY	NS		NS	ND
2905	7666	SVVGTPAYLAPEVLR		826.4246	2	1	YN	YN	YN	YY	YY	YN	YY	YY	YY	NS	YY	YN	YY		NS
2906	7325	IPSGTTPPTIPPDETFGGR		957.4463	3	1	YN	YN	YY	YY	YY	YY	YN	NS	YY	NS	YY	YY	YY	NS	NS
2907	5060	SILDDSFK		502.721	2	1	YN		YY	YY	YY	ND	YY	YY	YY	NS	NS	NS			NS
2908	606	ATSPEAGGGGALK		626.7827	2	1	YY	YY	YY	YY	ND	YN		NS	NS	NS	YY	NS	YY	YN	NS
2909	704	LVGATATSPPPK		653.3251	2	1	YN	YY	YN	YY	YY	YN	YN	YY	YY	NS	YY	YN	YY	YN	NS
2910	3640	QMSFDLTK		533.2198	2	1	YY	YY	YY	YY	ND	YY	YY		YY	NS	ND	NS			NS
2911	866	SITSPSTSTTK		643.7967897	2	1		YN	YY	YY		YY	YY	NS	YY	NS	YY	YY			NS
2912	6783	LSPPAGLFTSAR		648.8214	2	1	YY		YY	YY	YY	ND	YY		YY	NS	NS	NS			YY
3217	9632	DSTGTVISDVEEDAGELSR		758.6606	3	1	YN	YY	YY	YY	YY	YN	YY	YY	YY	ND	YY	YY	YY		YN
3218	7040	DVTNFTVGGFAPMSPR		597.9312	3	1	YY	YY	YY	YY	YN	YY	YY	YY		ND	YY	YY	YY	YN	YY
3219	2369	SESPPLSDPK		617.2757	2	1	YN	YN	YY	YY	YY	YY	YY	YY	YY	ND	NS	YY	YY		YY
3220	3531	GEFSASPMK		581.745	2	1	YY	YY	YY	YY	NS	YY	YY	YY	YY	ND	NS	YY	YY	NS	NS
3221	9669	AGYGCPLLSDLTSPVPR		812.0666	3	1	YY	YY	YY	YY	YY	YN	NS	YY	YN	ND	YY	YY	YY	YN	YN
3222	1981	LSPGESAYQK		580.2558	2	1	YY	YY		YY	YY	ND		YY		ND	YY	YY			NS
3223	5699	CRSPGMLEPLGSSR		805.3468	2	1	YY	YY		YY	YY	ND		ND		ND	YY	YY	YY	YN	ND
3224	10271	ATENDIYNFFSPLNPVR		692.9849667	3	1		YY	YY	YY	YY	YY		YY	YY	ND	YY	NS		NS	NS



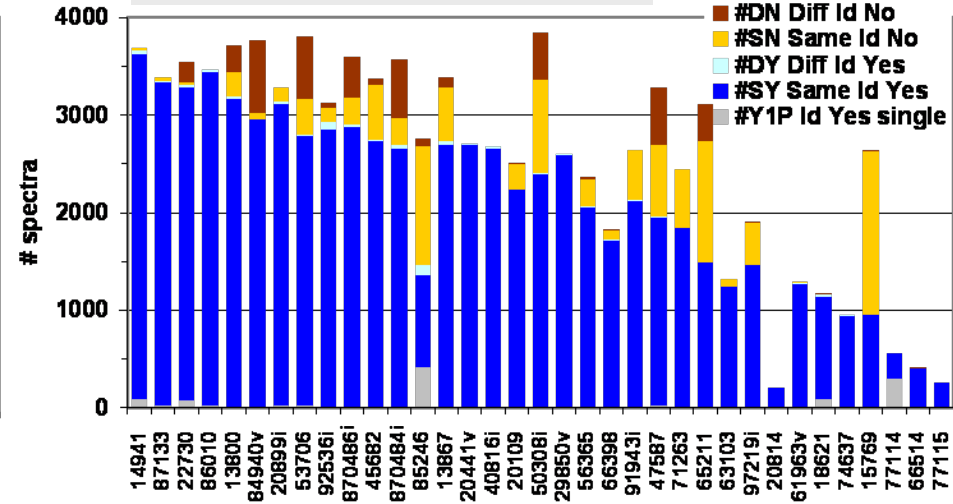
# Room for Improvement in ID Certainty Thresholds

Proteome Informatics  
Research Group

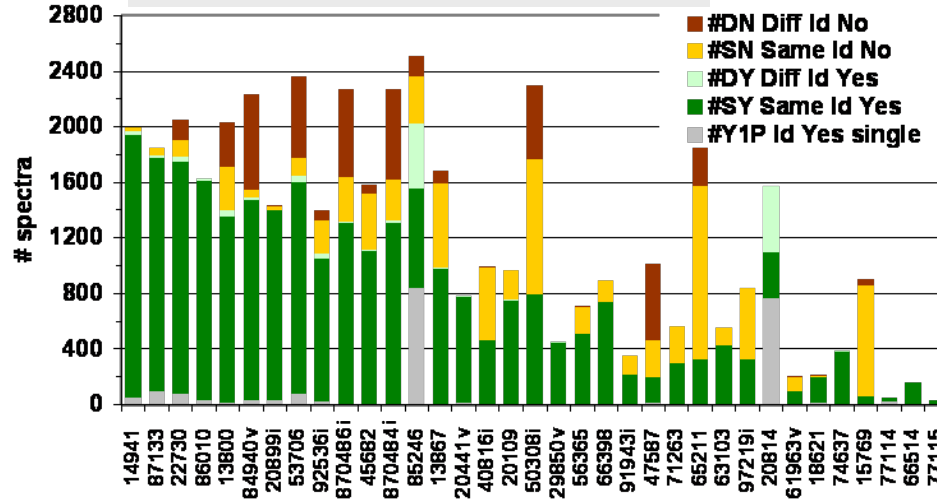
### Frxn 3 – most multiple phos per peptide



### Frxn 4 – most phosphopeptides



### Frxn 12 – highest precursor charges





# Preliminary Conclusions

1. Wide range of spectra marked confidently identified
2. Wide range of spectra marked confidently localized
3. Lack of a uniform method for calculating and reporting ambiguity made it hard to compare results from some participants (13 of 35 were only partially included)
4. Some participants succeeded without localization software but most at least used some measure of ambiguity
5. Typically, very few identifications were unique to any one participant
6. Unique peptide assignments were roughly 57% identical (n=35)
7. Confidently localized phosphopeptides assignments were roughly 38% identical (n=22)
8. Participants that performed well often shared the highest similarity with other participants
9. Participants did not hesitate to mark peptides with ambiguous phosphosite localizations (57% of identified spectra on average)
10. If all of the participants agree on the identification, phosphosite ambiguity is possible, and that localization is possible, for 79% of the spectra, participants unanimously agree on the localization(s)



# Selected Survey Responses

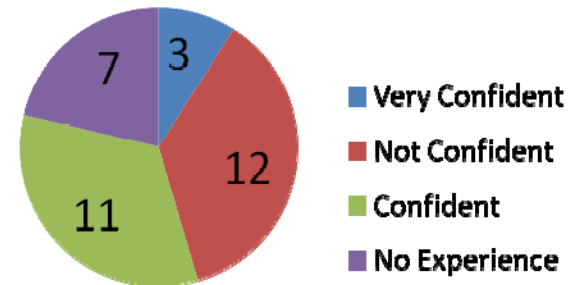
- Do you think this type of study is useful?
  - Yes 33 (100%)
- How difficult do you think this study was?
  - Easy 4
  - Challenging 17
  - Just right 12
- Based on this study, would you consider participating in future ABRF studies?
  - Yes 33 (100%)
- Have you participated in previous ABRF studies?
  - No 14 (42%)
  - Yes 19 (58%)



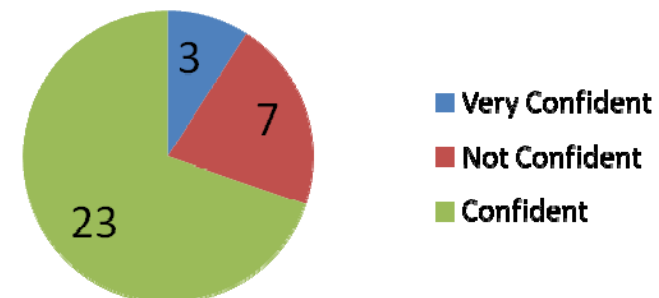


## Survey cont.

- Before this study, how confident were you of your ability to identify and rank phosphopeptides including assessing phosphorylation site localization?



- Now, after completing the study, how confident are you of your ability to identify and rank phosphopeptides including assessing specific phosphorylation sites?





# iPRG Membership

- Manor Askenazi - Dana-Farber Cancer Institute
- Karl Clauser - Broad Institute of MIT and Harvard
- Lennart Martens (incoming chair) - Ghent University, Belgium
- W. Hayes McDonald - Vanderbilt University
- Paul A Rudnick (outgoing chair) – NIST
  
- Karen Meyer-Arendt (outgoing member) - University of Colorado
- \*Brian C. Searle (outgoing member) - Proteome Software, Inc.
- **\*William S. Lane (outgoing member) - Harvard University**
- **\*Jeffrey A Kowalak (EB Liaison) (outgoing) – NIMH**
  
- Eric Deutsch (incoming member) – Institute for Systems Biology
- Nuno Bandiera (incoming member) – UCSD
- Robert Chalkley (incoming member) – UCSF

\*Founding member



# Acknowledgements

- Phillip Mertins, The Broad Institute
  - All wet lab work and an analysis
- Steve Gygi, Harvard Medical School
  - Test datasets
- Matthew Chambers, Vanderbilt University Medical Center
  - Data format conversions (ProteoWizard)
- Steve Stein and Yuri Mirokhin, NIST
  - A K562 phosphopeptide spectral library
- Renee Robinson, Harvard University
  - “The Anonymizer”



*Proteome Informatics*  
*Research Group*

# Thank You, Participants!

13800

13867

14941

15769

18621

20109

20441

20814

22730

29850

45682

47587

53706

56365

61963

63103

65211

66398

66514

71263

74637

77114

77115

84940

85246

86010

87133

66398

66514

71263

74637

77114

77115

84940

85246

86010

87133



*Proteome Informatics*  
*Research Group*

**END**