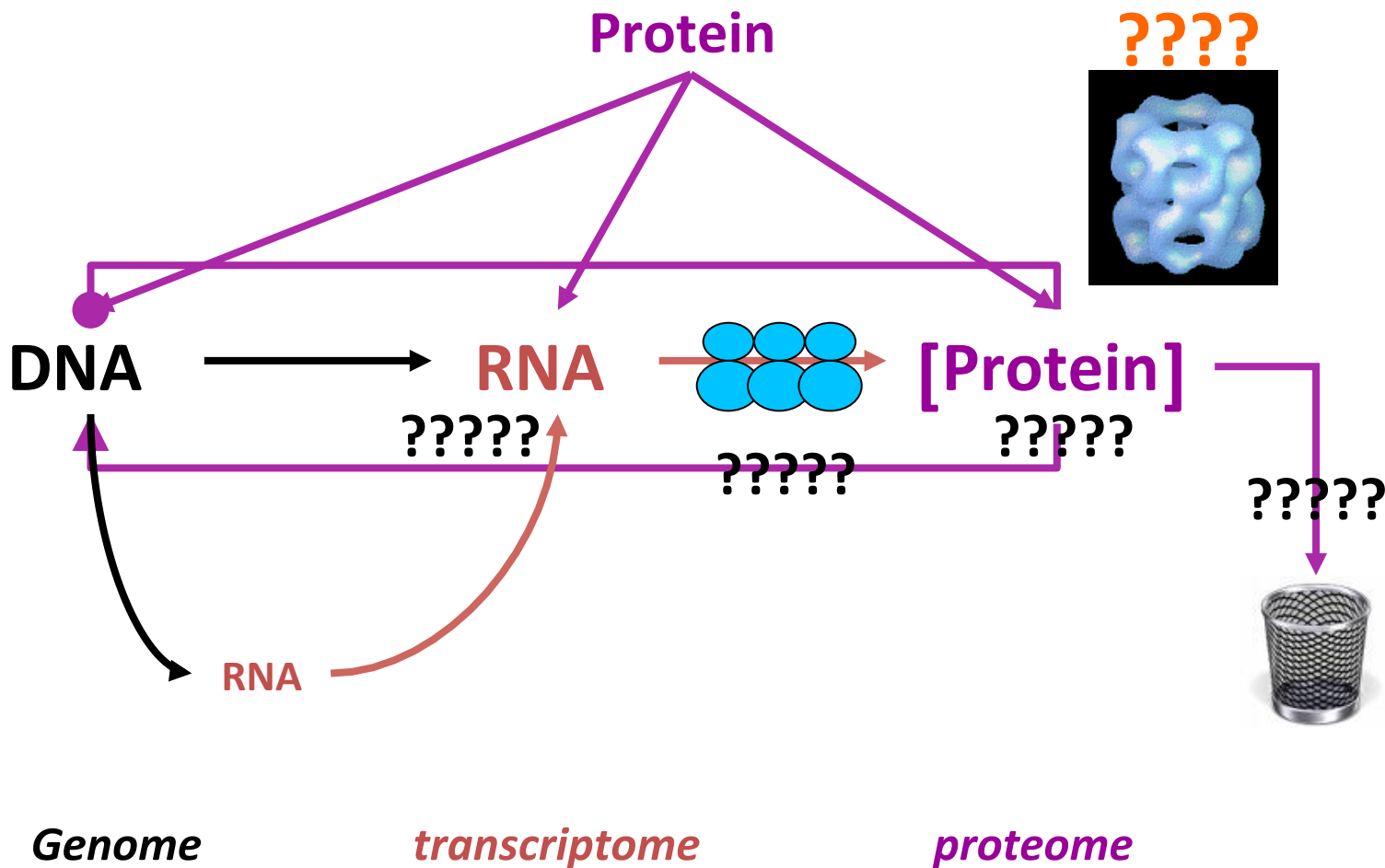


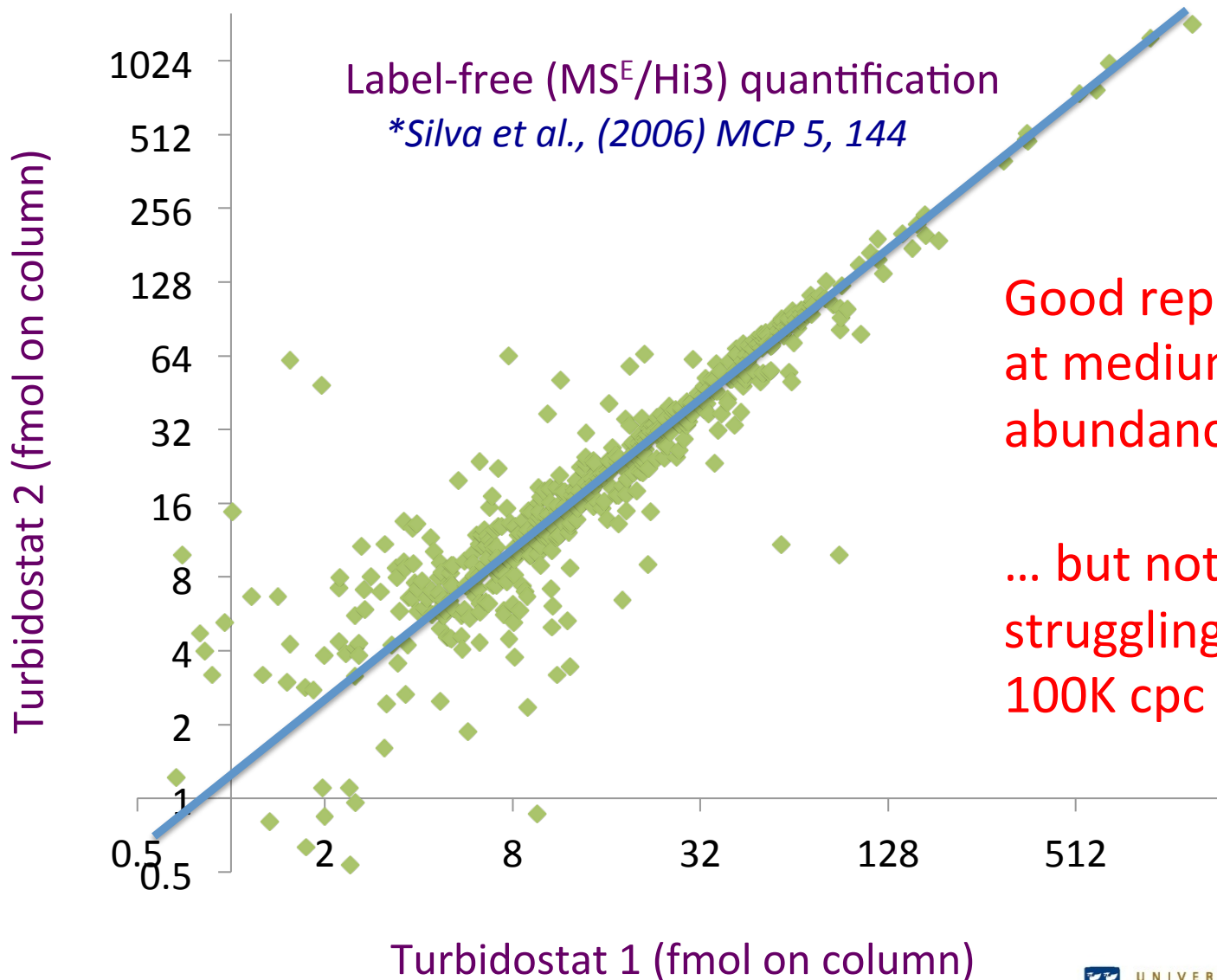
Absolute Protein Quantification in Yeast: pitfalls and potential

Simon Hubbard
Faculty of Life Sciences
University of Manchester

Life is ~~simple?~~ complicated



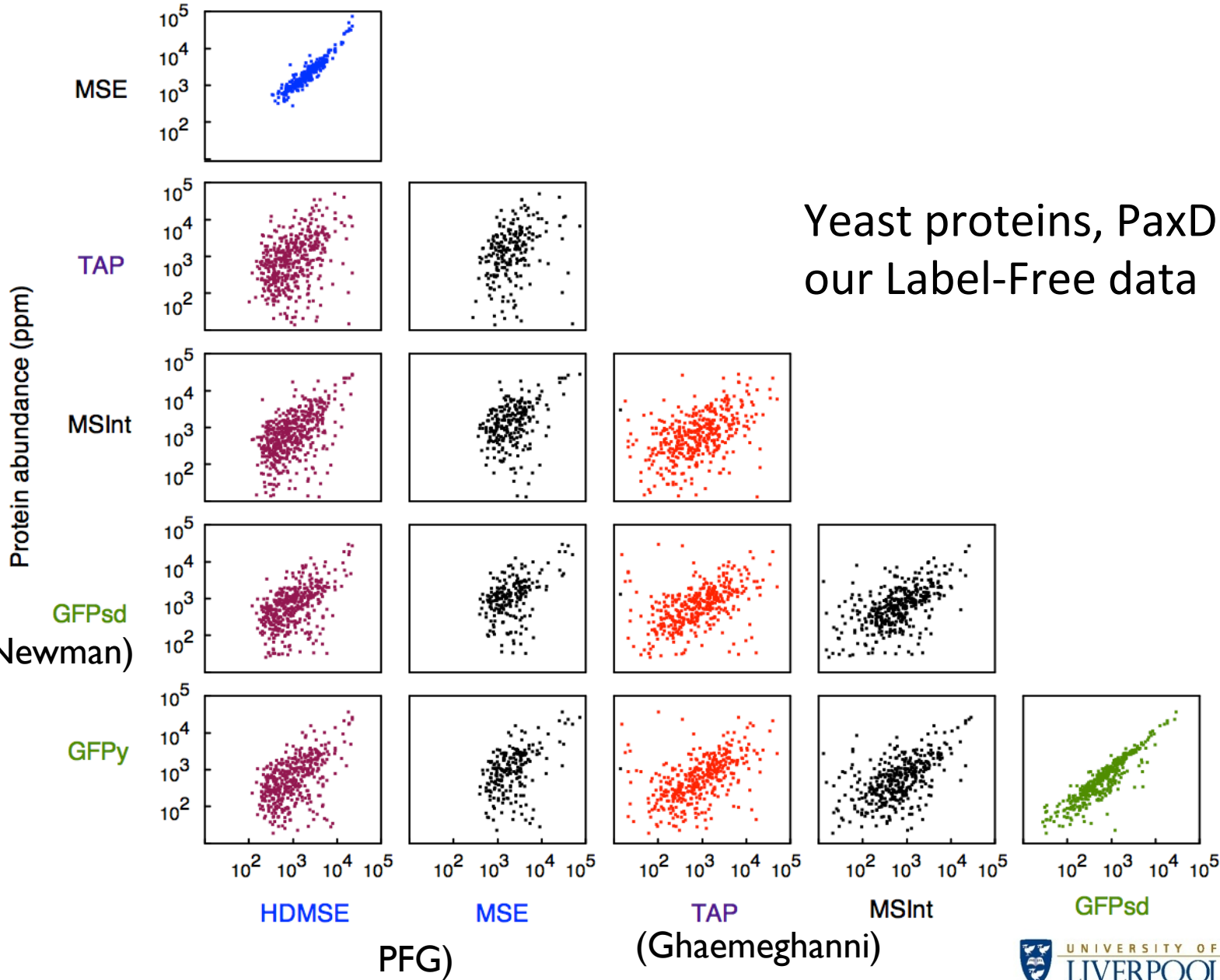
Quant is needed, but how to do it: Label free?



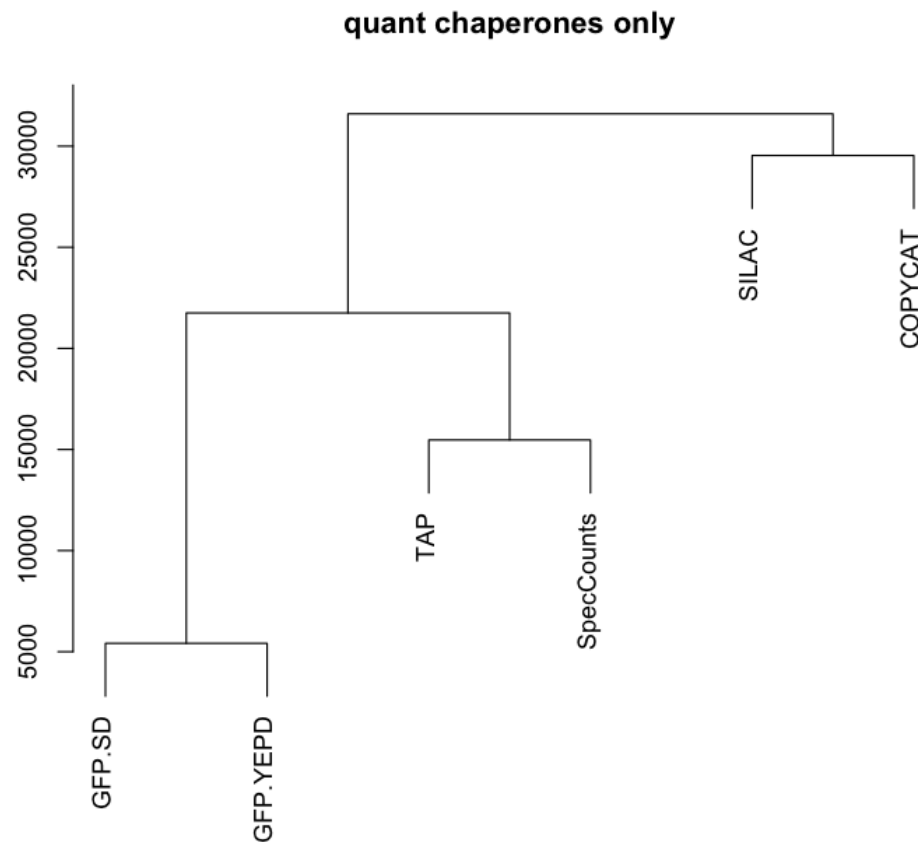
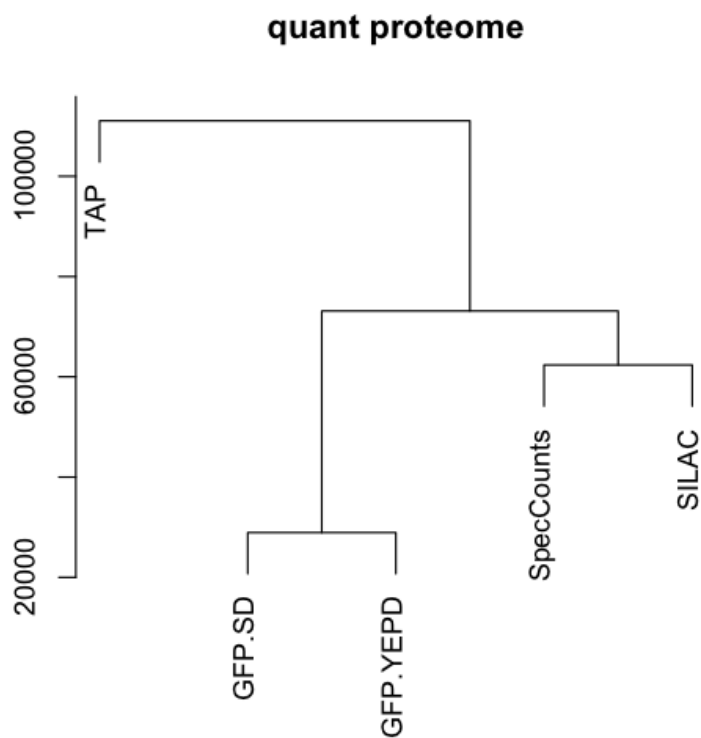
Good reproducibility
at medium/high
abundance ...

... but not “deep”,
struggling beyond
100K cpc in Hela Cells

Gold standards in quantification??

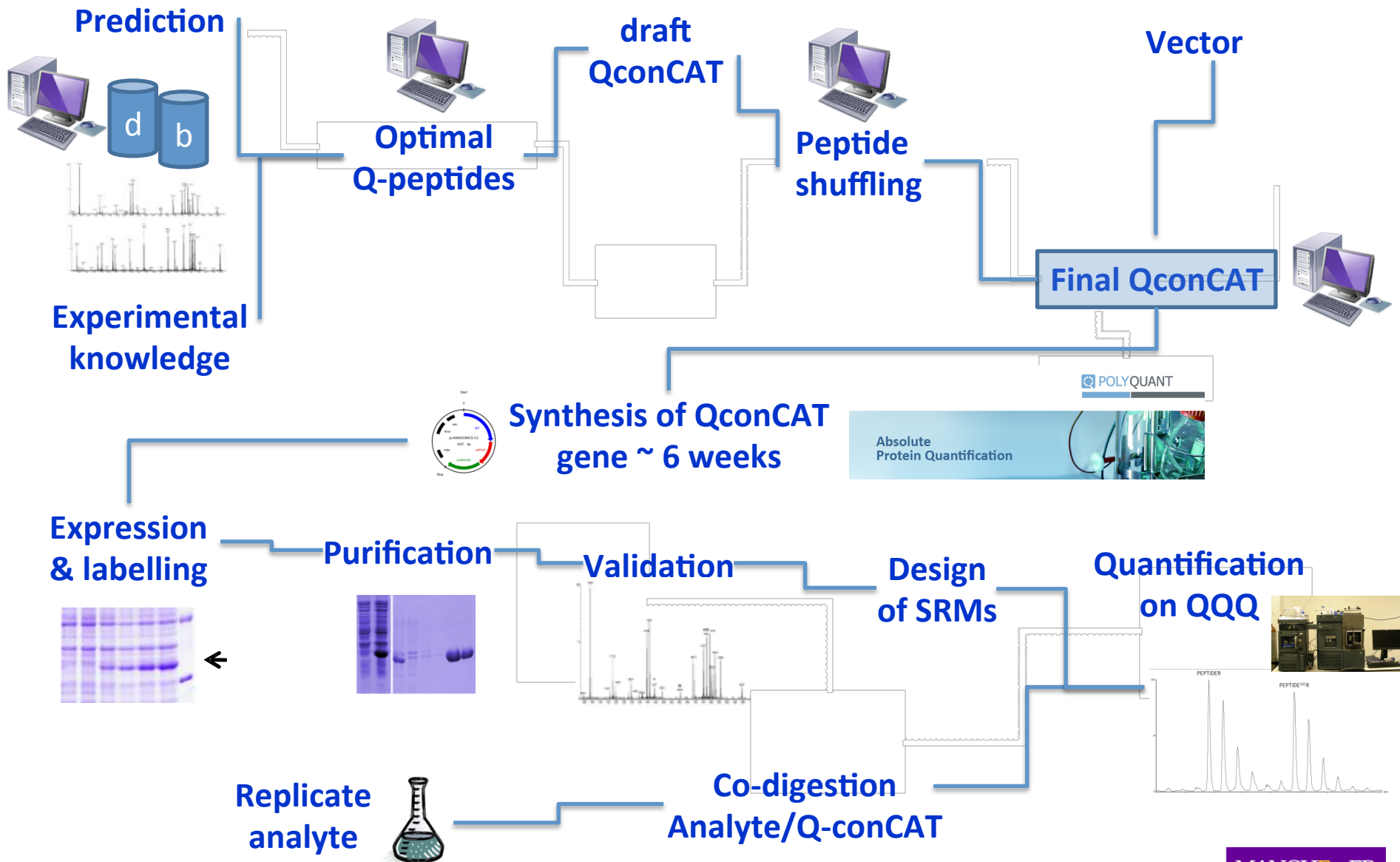


Quant is highly method dependent

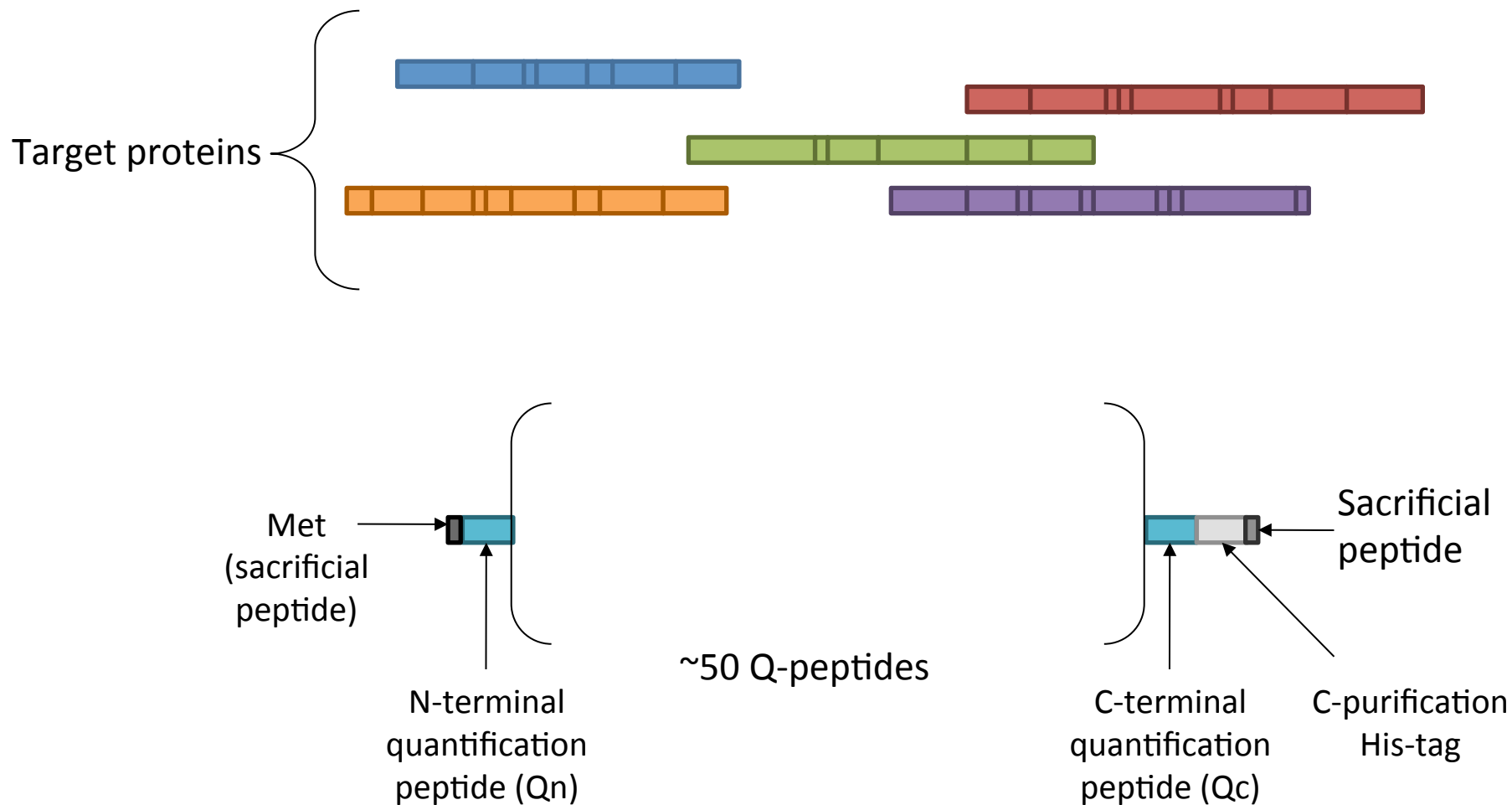


positive correlations, but ...

QconCAT workflow



Targeted proteomics: QconCAT Strategy



Beynon et al., Nature Methods. (2005) 2:587,
Pratt et al., Nature Protocols (2006) 1:1029
Rivers et al., MCP (2007) 6:1416

Choosing a good surrogate

- Reasons to be cheerful ...
 - Growing databases (e.g. PeptideAtlas)
 - Sequence unique
 - Proteotypic
 - **Flyability**
 - Quantotypic
 - **Stoichiometric with protein of interest**
- Fly in the ointment ...
 - Amino acid composition
 - PTMs
 - Missed cleavage
 - Isoforms & Paralogues

Isoformal proteomics is significant

8 P. Blakeley *et al.*

Proteomics 2010, 10, 1–14

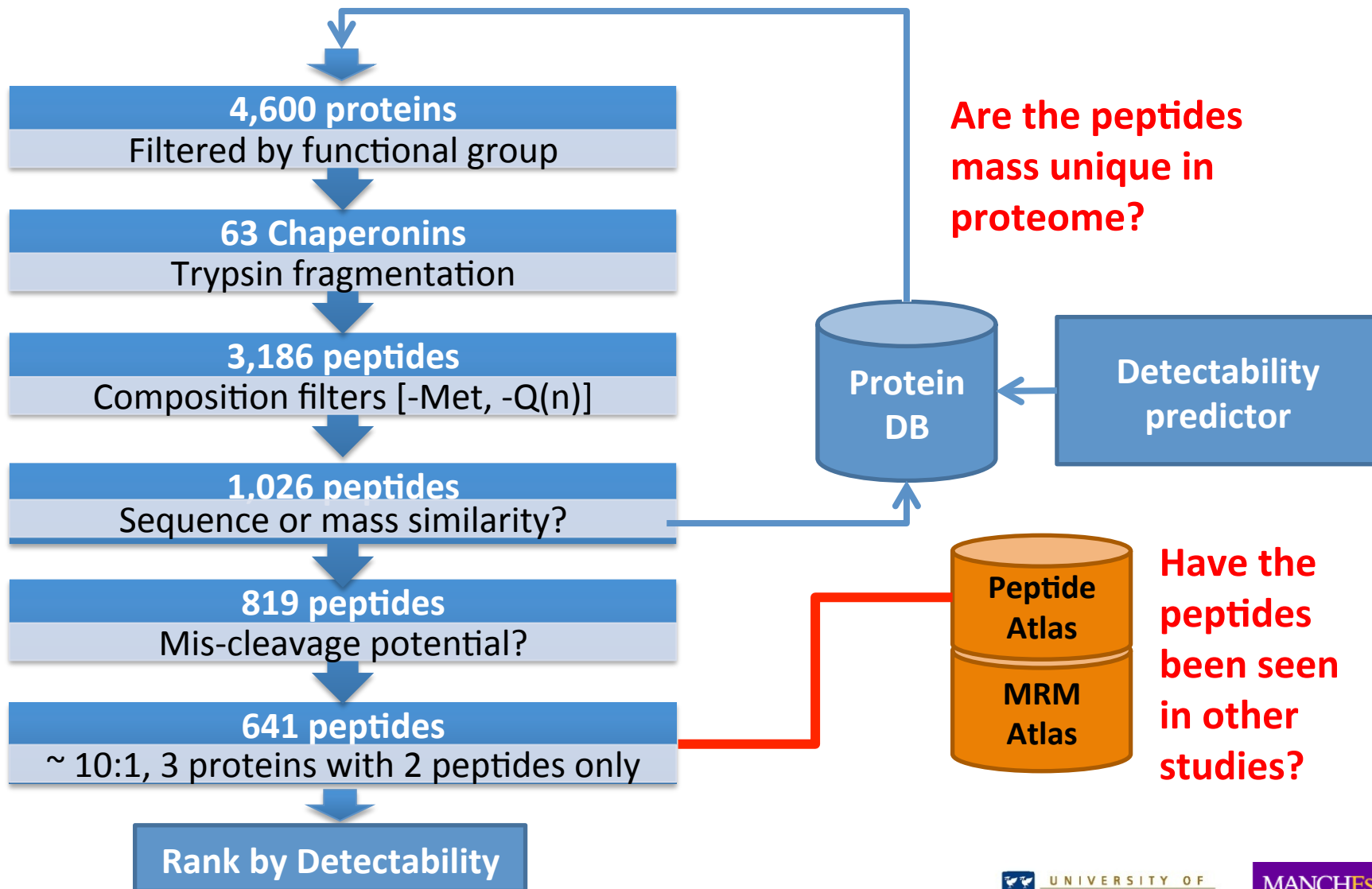
Table 3. Relative fractions of gene sets with associated peptide evidence

Species	Total number of genes	Genes with peptide evidence ^{a)}	Genes annotated with multiple non-redundant protein isoforms	AS genes with peptide evidence for specific alternative isoforms ^{b)}	AS genes with proteotypic peptides predicted for specific alternative isoforms ^{c)}
Human	22 839	7868 (34%)	9343	3,059 (33%)	8376 (90%)
Mouse	23 429	3096 (13%)	7280	747 (10%)	6874 (94%)
Chicken	16 736	3023 (18%)	3776	628 (17%)	3710 (98%)
<i>D. melanogaster</i>	14 039	907 (6%)	2079	211 (10%)	2020 (97%)
<i>C. elegans</i>	20 140	8256 (41%)	2337	1,216 (52%)	2130 (91%)

BIG GAP !

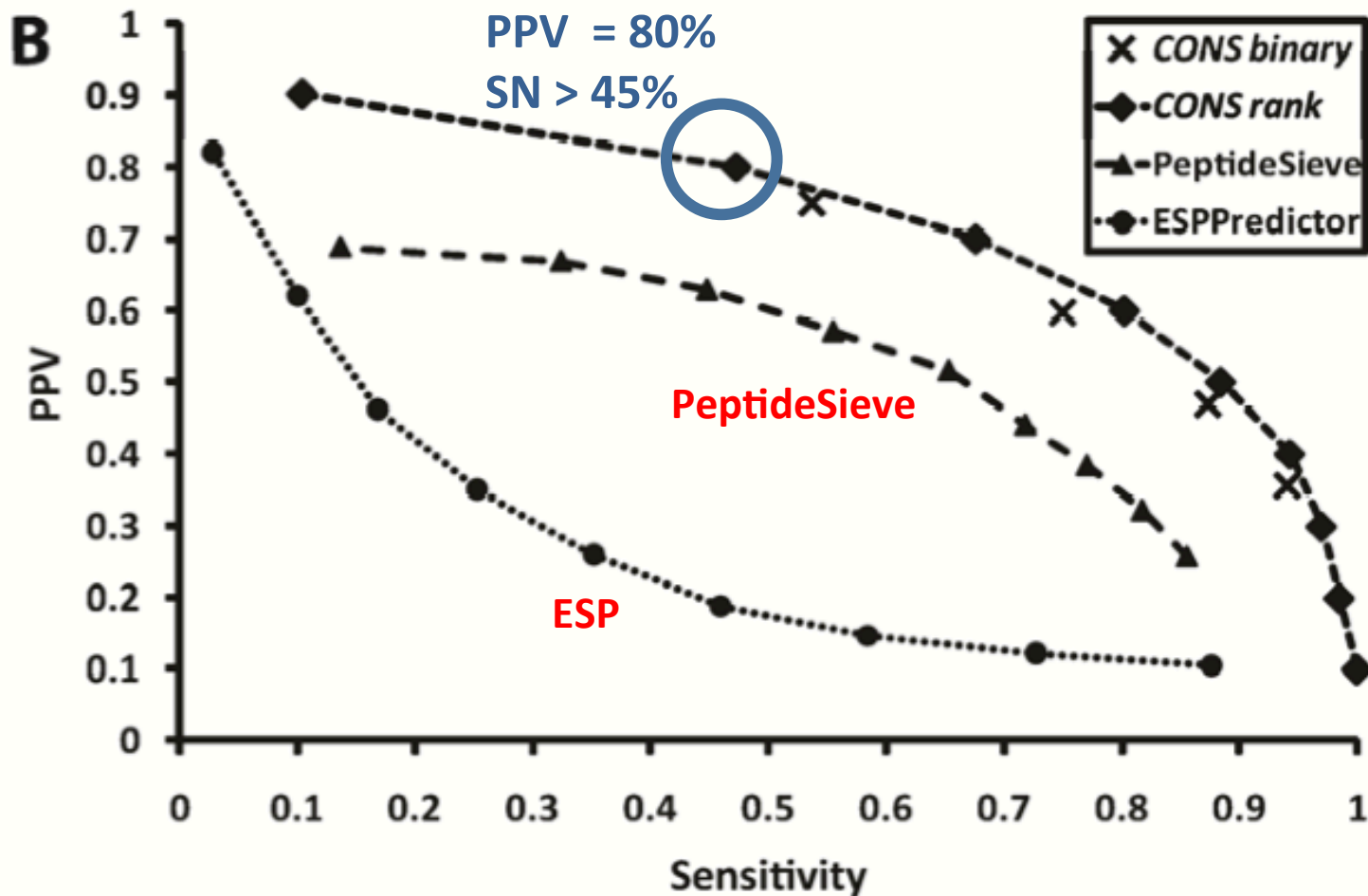
- Fully tryptic peptides only, up to and including 1 missed cleavage
- Non-conservative definition: any PSM which is not constitutive

COPYCAT design pipeline: chaperones example

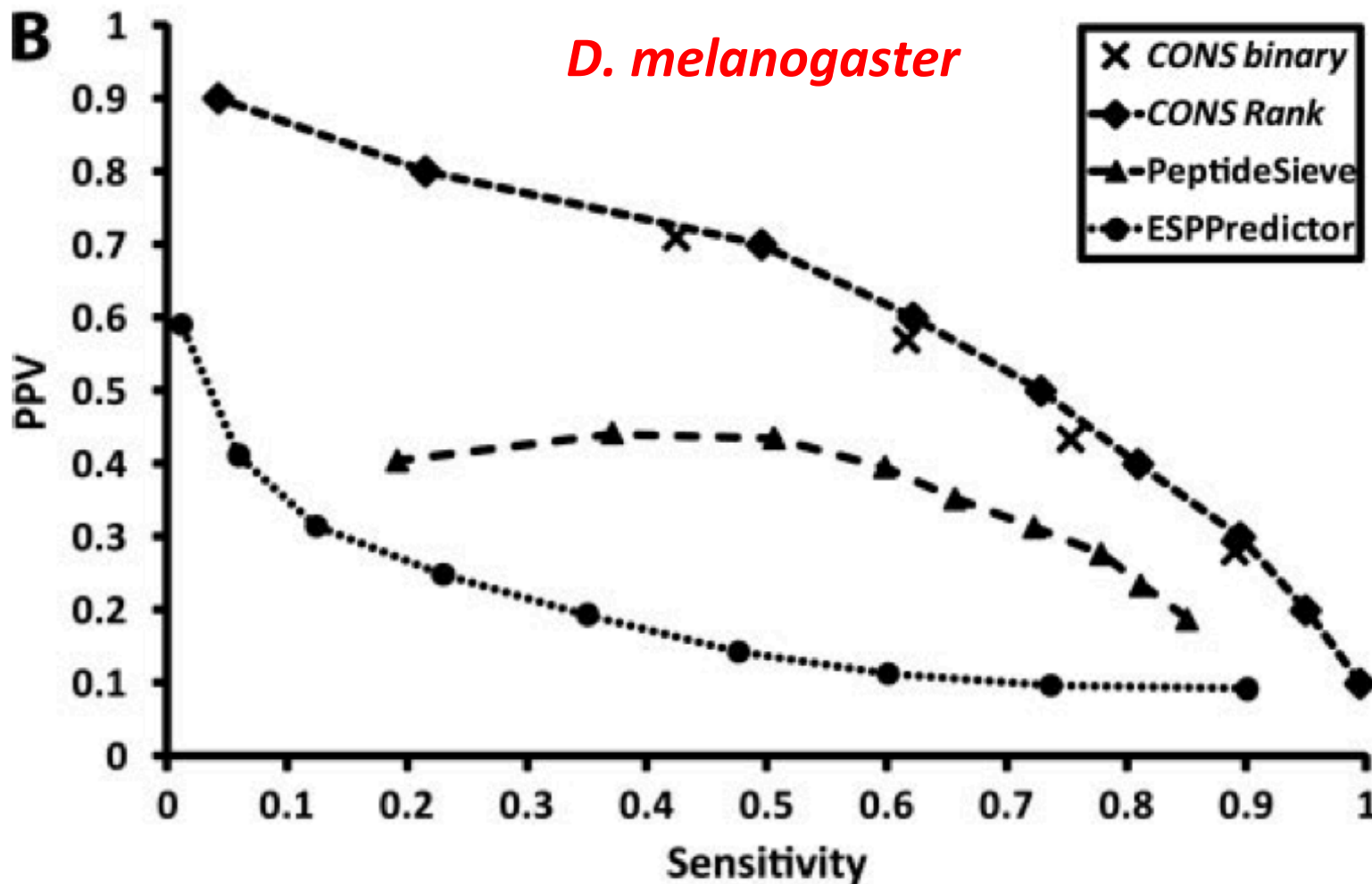


High performance prediction

Consensus Method (PeptideAtlas mc dataset) vs PeptideSieve & ESP



Algorithm generalises to other species



Candidate Peptides

- CONSeQuence
 - <http://king.smith.man.ac.uk/CONSeQuence>

Select protein fasta file (max 2MB): No file chosen
or paste protein sequence(s) in fasta format:

```
>sp|P49954|YEAST_NIT
MSASKILSQIKIVALVQLSGSSPDKMANLQRAATFIERAMKEQPDTKLVVLPECFNSPYSTDQFRKYSEVINPKPSTSVQFLSNLANKFK
IILVGGTIPELDPKTDKIYNTSIIIFNEDGKLIDKRRVHFLFDVDPNGISFHESETLSPGKSTTIDTKYKFGVGVGICYDMRFPPELAMLSA
RKGAFAMIYPSAFNTVTGPLHWHLLARSRAVDNQVYVMLCSPARNLQSSYHAYGHSIVVDPRGKIVAEAGEGEEIIYAEALDPEVIESFRQA
VELIKQRRFDVYSVDNAH
```

Select number of internal miscleavages: Select pr

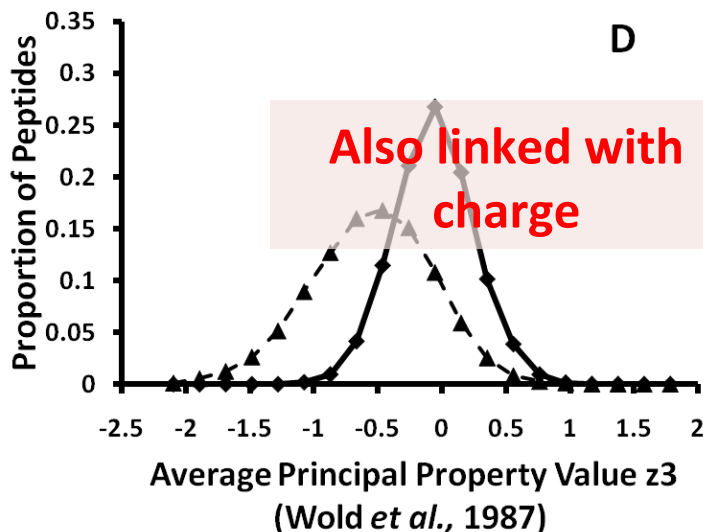
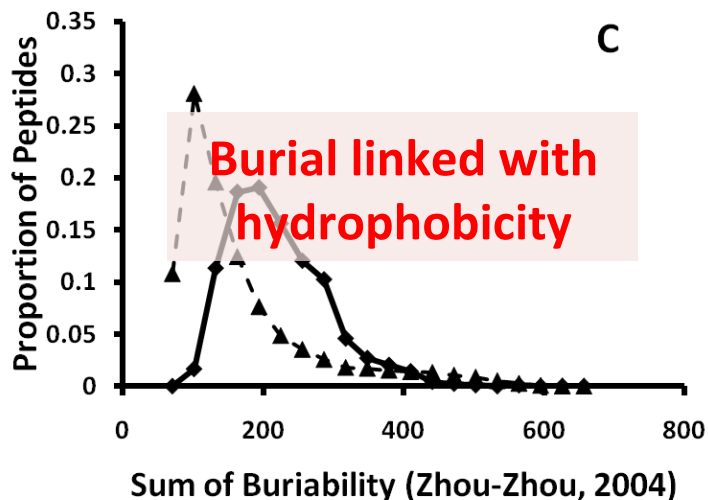
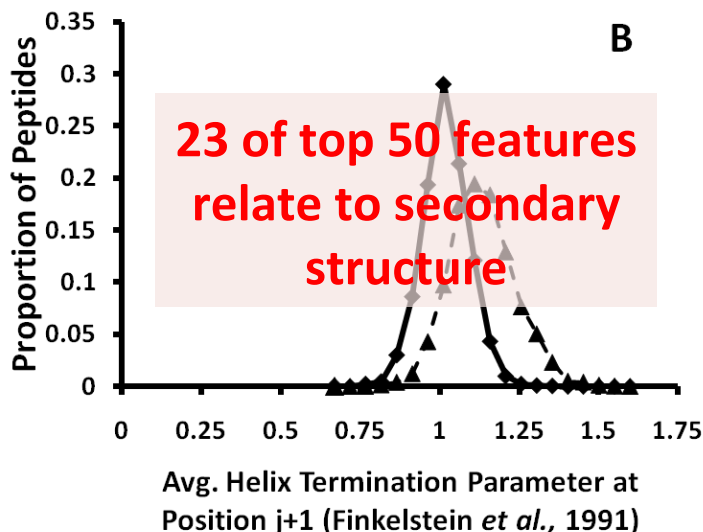
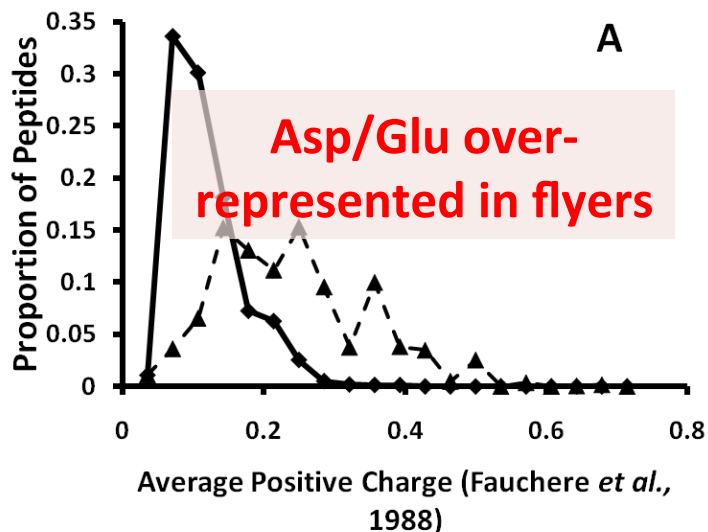
CONSeQuence+++

A total of 19 peptides available from protein.
[Download here](#)

This results page will be available for 72 hours

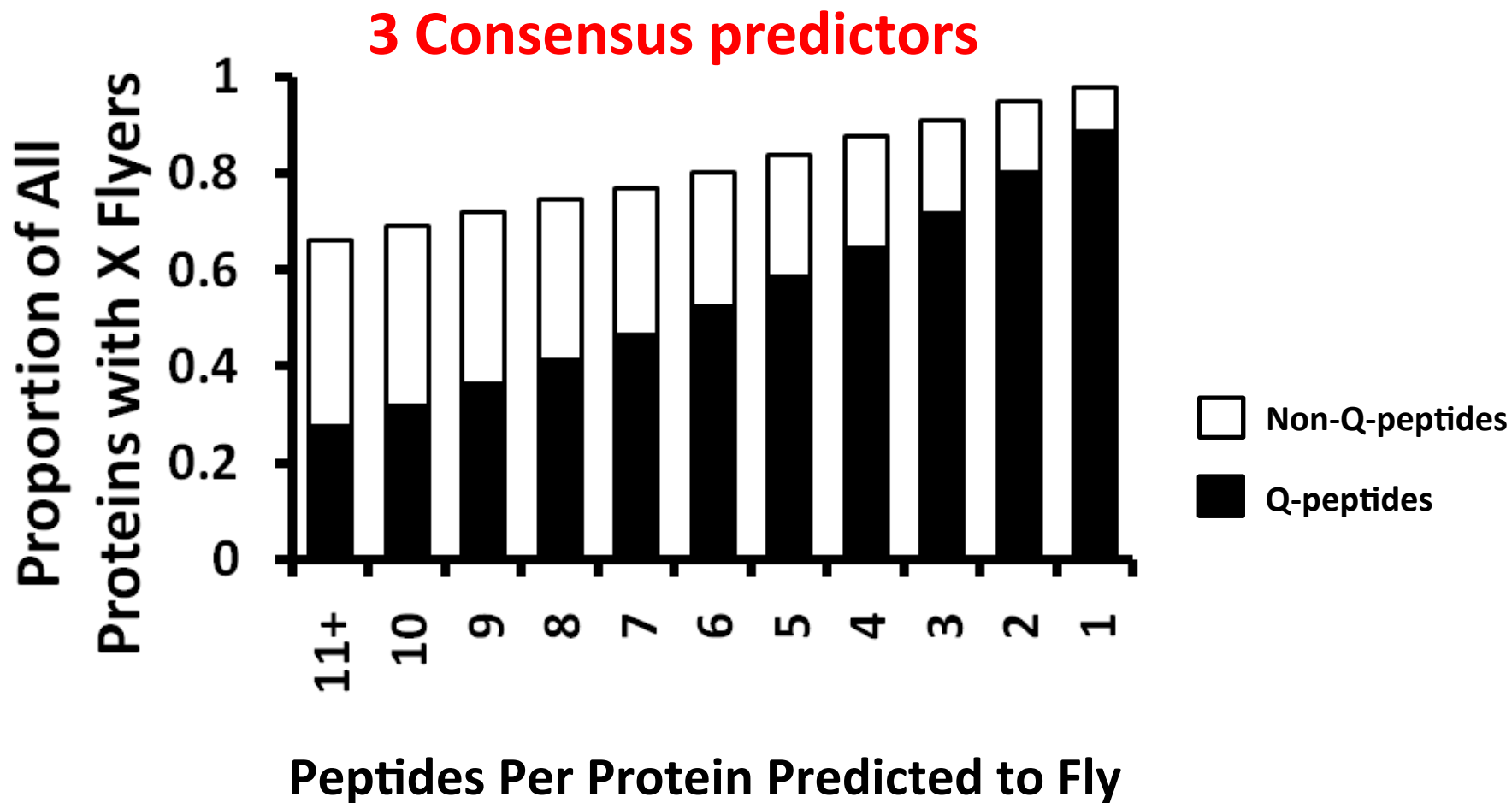
Protein	Peptide	IMC	Score
SP P49954 YEAST_NIT	VALVQLSGSSPDK	0	0.746
SP P49954 YEAST_NIT	MANLQR	0	0.210
SP P49954 YEAST_NIT	AATFIER	0	0.251
SP P49954 YEAST_NIT	EQPDTK	0	0.209
SP P49954 YEAST_NIT	LVVLPECFNSPYSTDQFR	0	0.780
SP P49954 YEAST_NIT	YSEVINPK	0	0.258
SP P49954 YEAST_NIT	EPSTSVQFLSNLANK	0	0.761
SP P49954 YEAST_NIT	IILVGGTIPELDPK	0	0.753
SP P49954 YEAST_NIT	IYNTSIIIFNEDGK	0	0.643

Properties of proteotypic "flyers"



◆ Flyers ▲ Non-Flyers

Do we have enough quantotypic peptides?



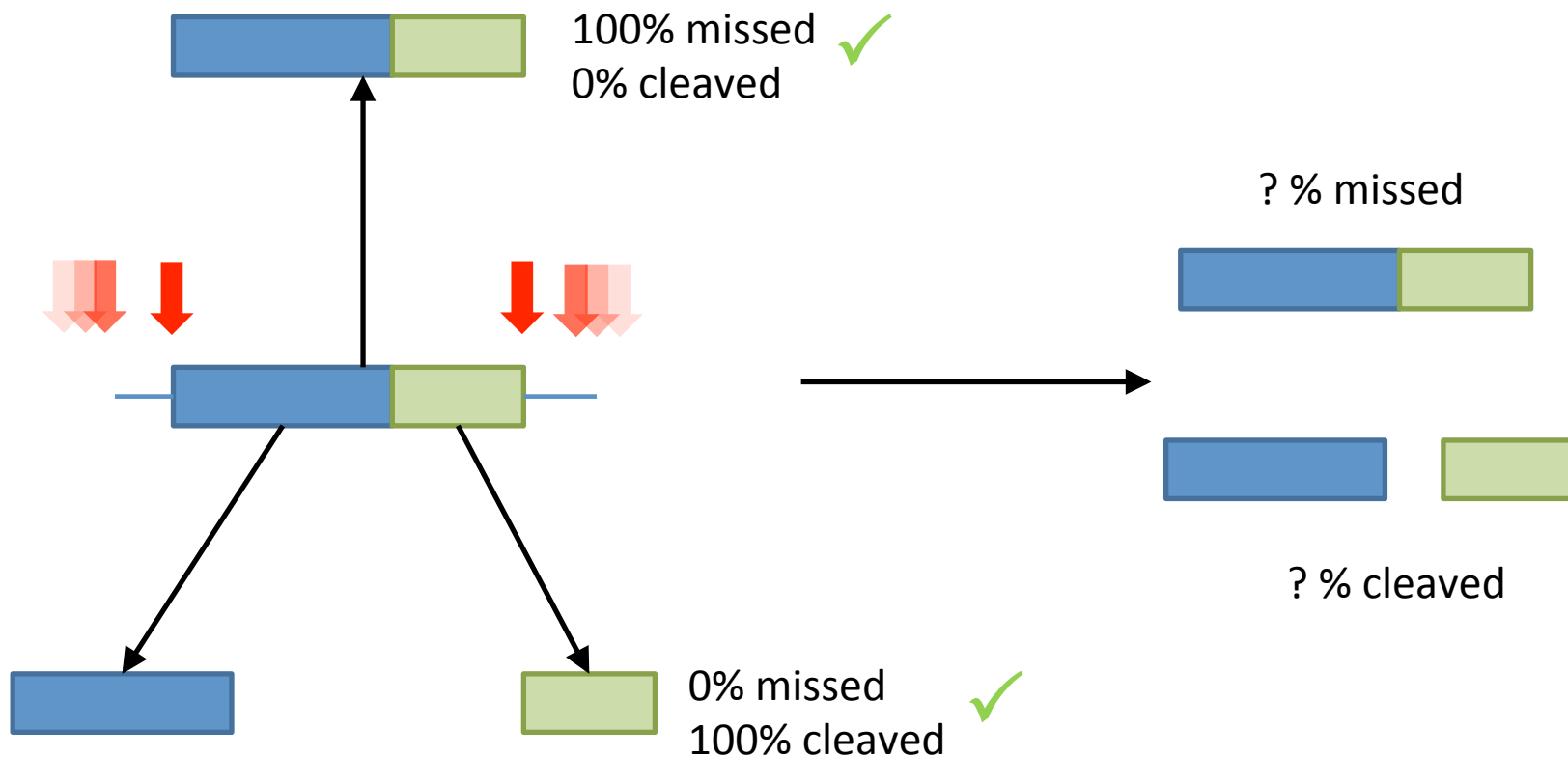
Missed Cleavages

- Dibasics are a problem
 - ~25% of available peptides in *S. cerevisiae*
- Internal missed cleavages*

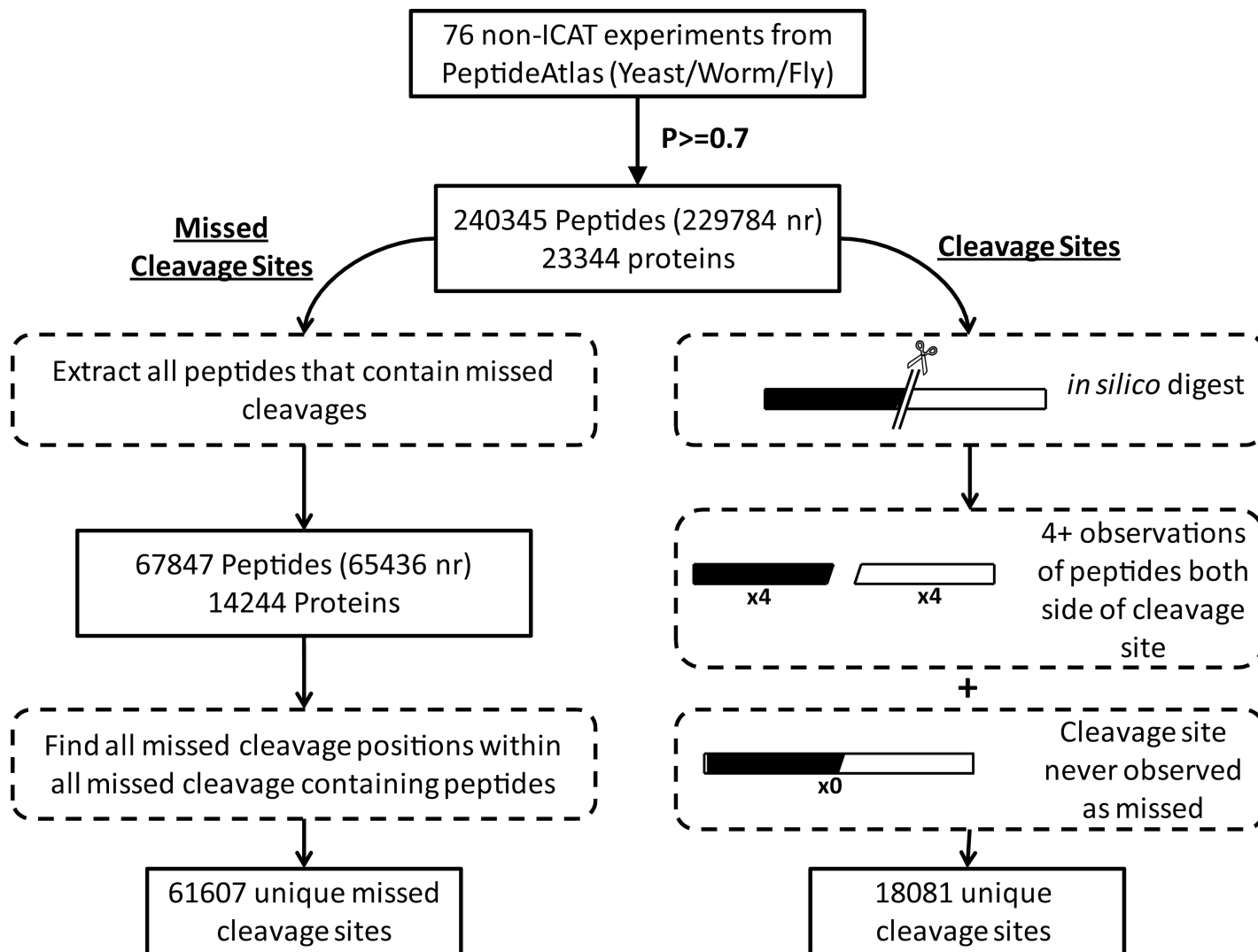
	<i>S. cerevisiae</i>	<i>C. elegans</i>	<i>D. melanogaster</i>	
Proteins	5124	8547	9673	
Proteins with mc peptides	4088	4799	5337	
Peptides	111119	57652	71574	
Peptides with no mc	77505	41644	53349	} mischief
Peptides with 1 mc	27417	13704	15412	
Peptides with 2 mc	5365	2300	2722	
Peptides with 3-6 mc	832	4	91	

*Peptide Atlas TPP \geq 0.7

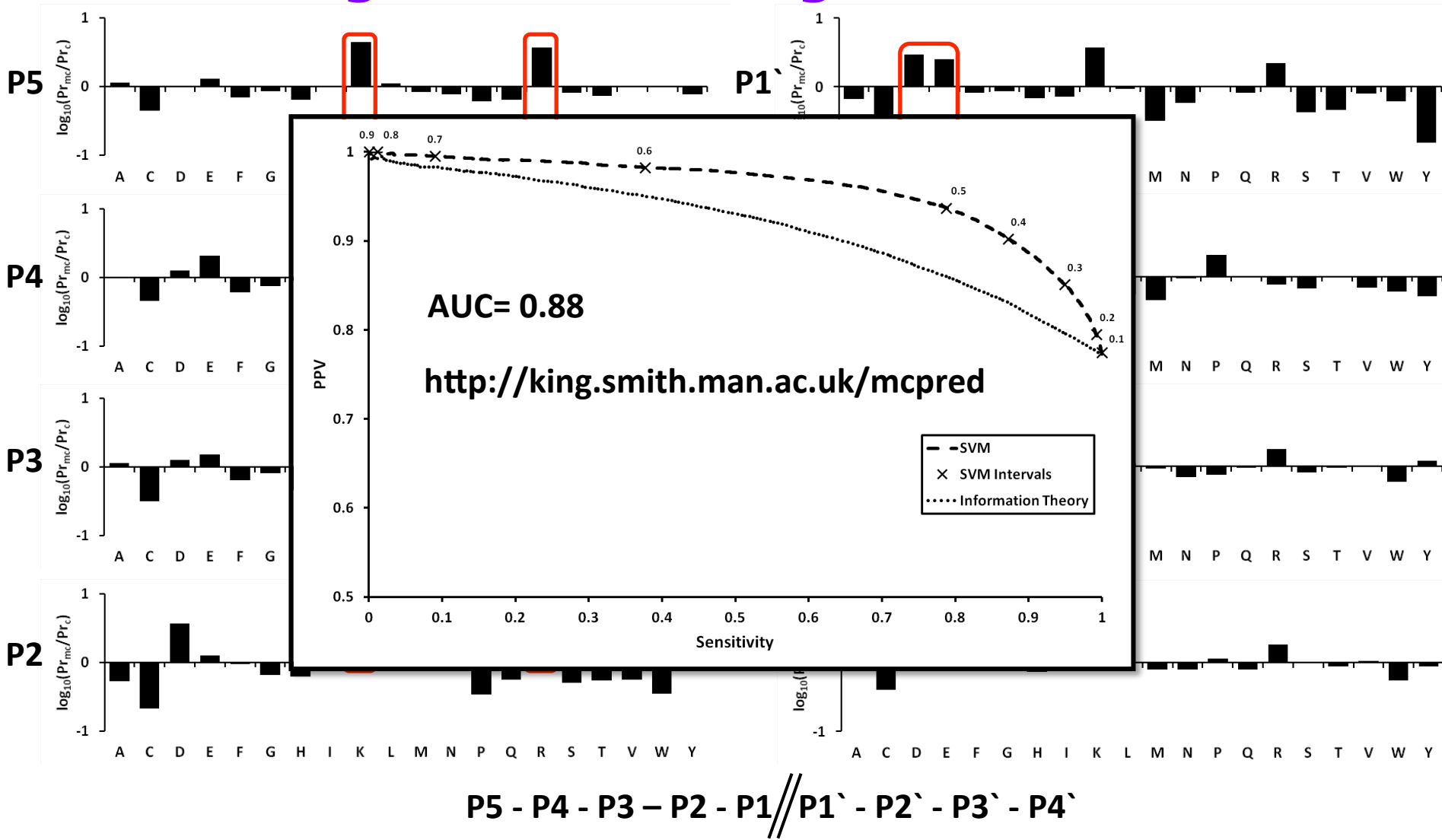
Missed cleavage problem



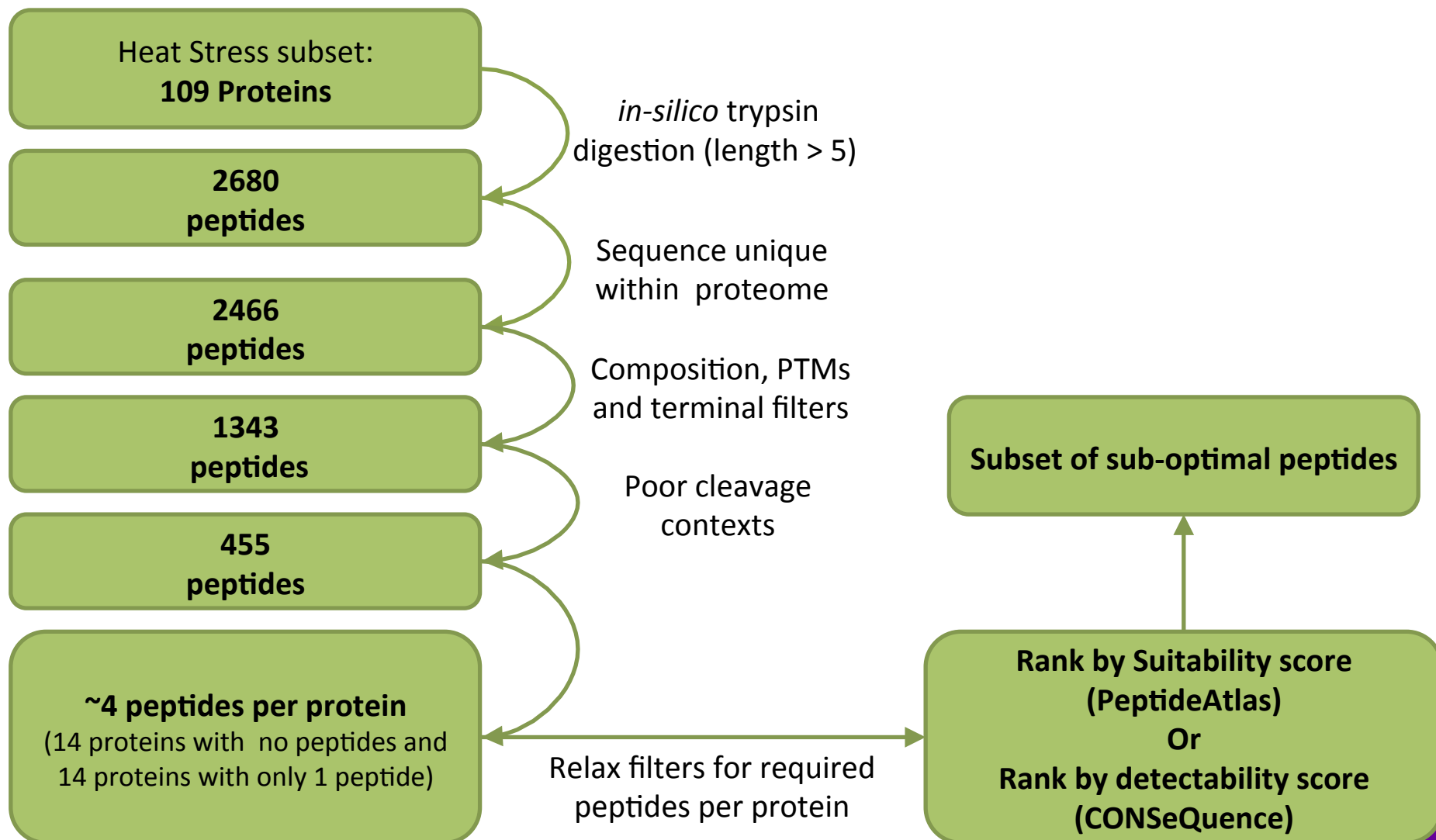
Missed cleavages are also important



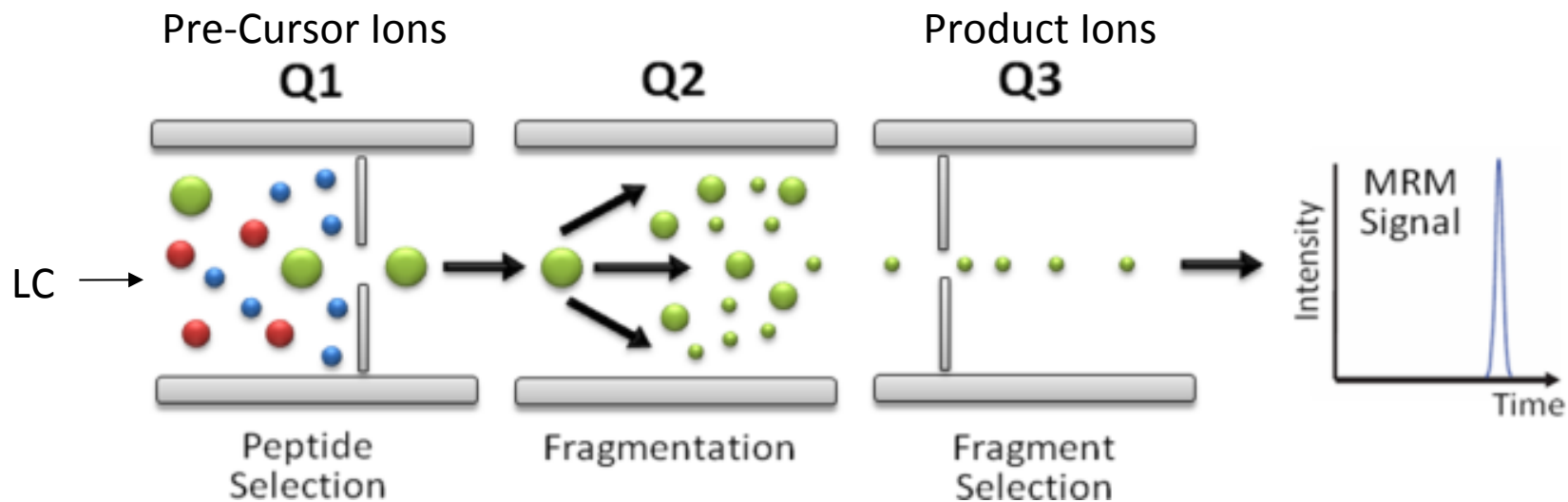
Predicting Missed Cleavages



Peptide Filtering



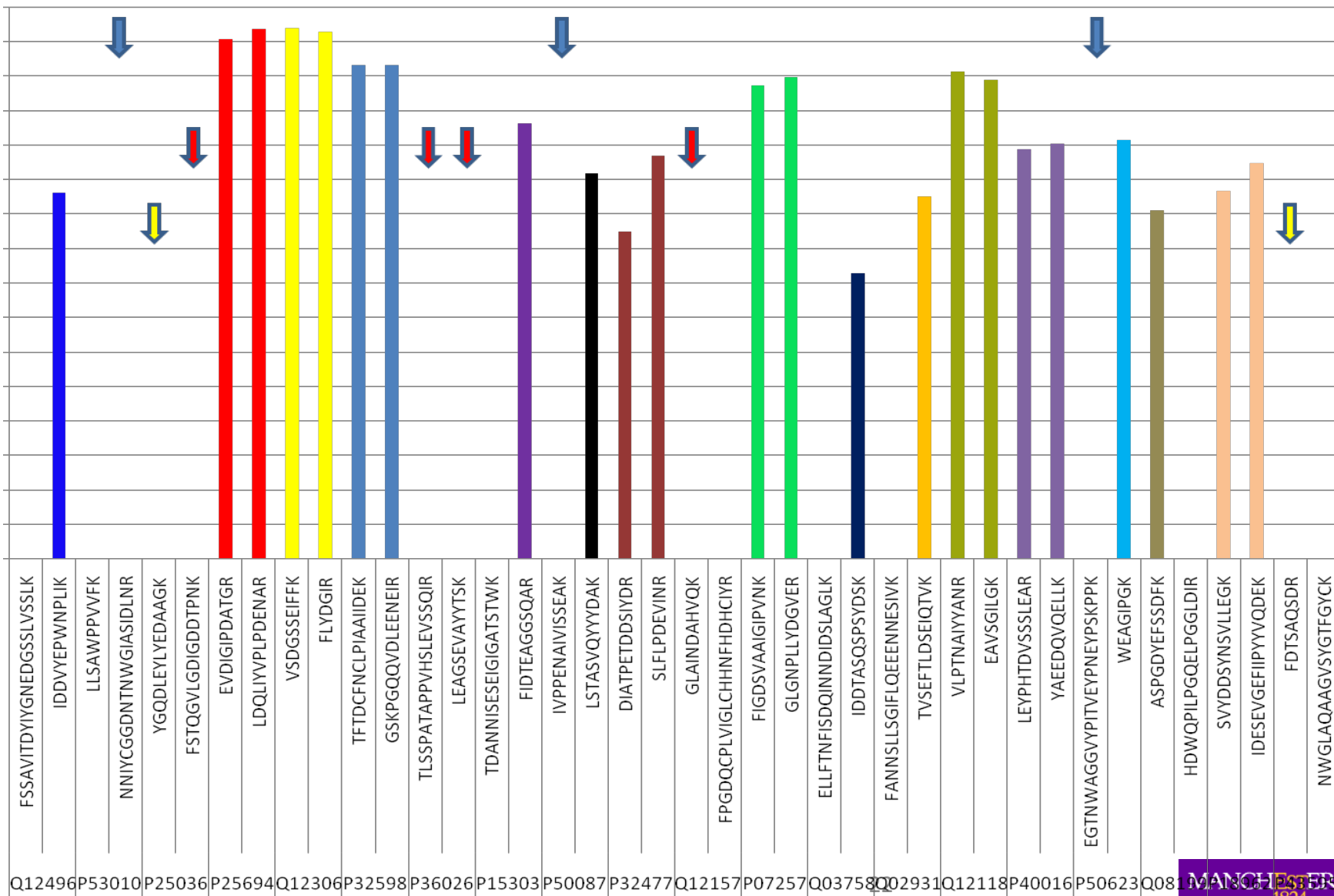
SRM Strategy



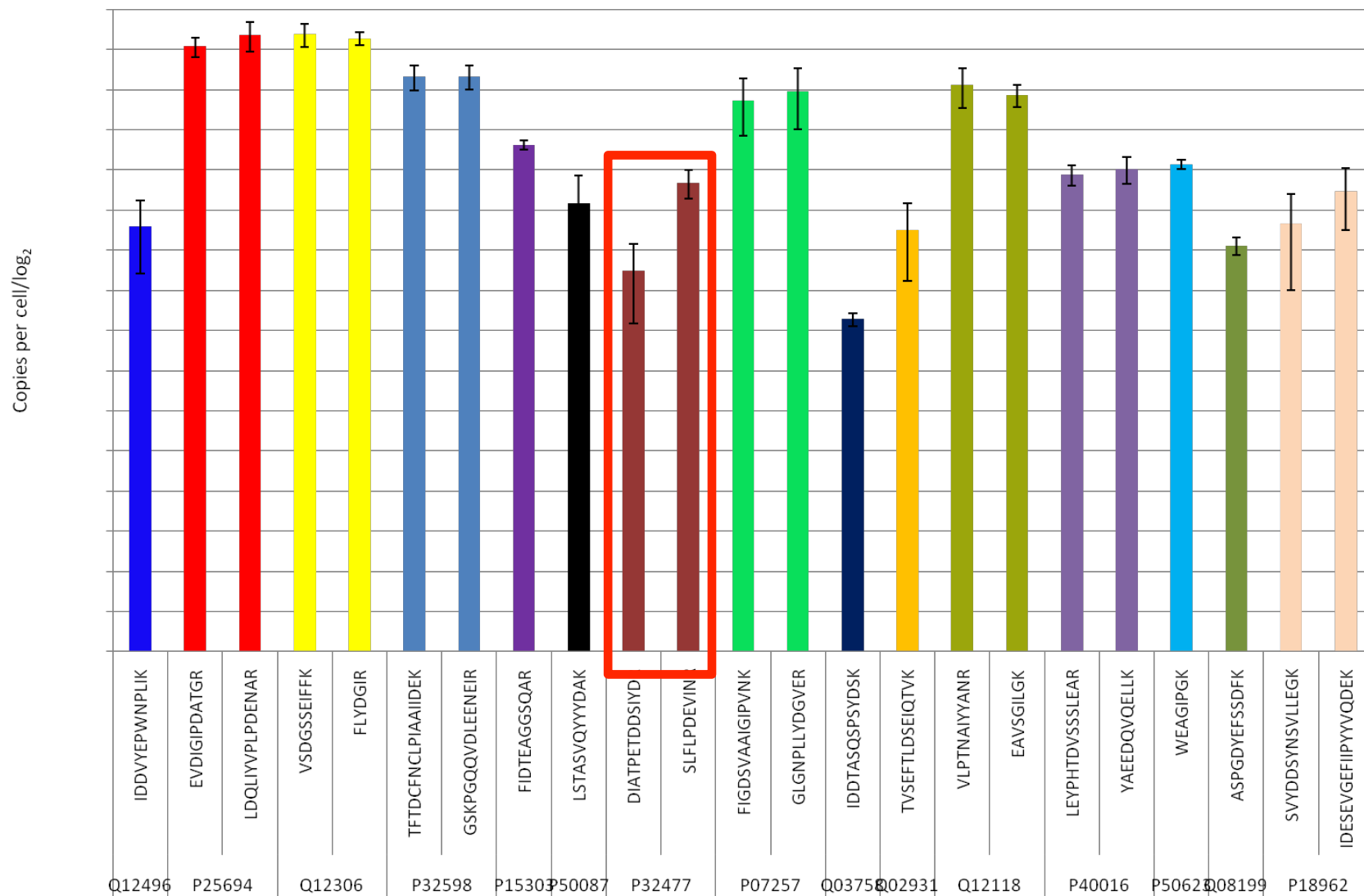
- High Sensitivity
- Absolute quantification achieved via XICs from SRMs
- Selecting best of 3 loadings of QconCAT

Typical Copycat quantification (Proteolysis)

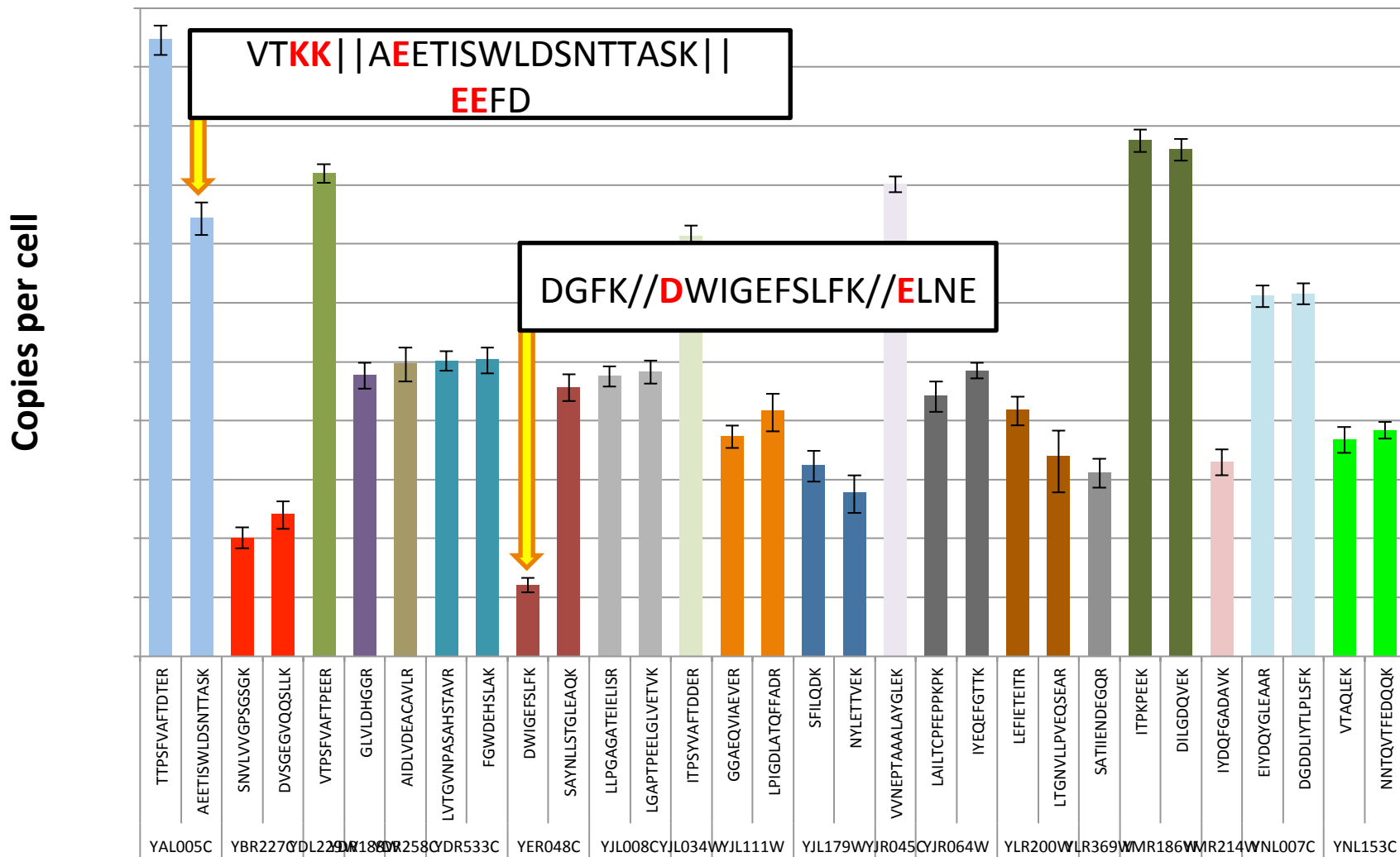
Copies per cell/ \log_2



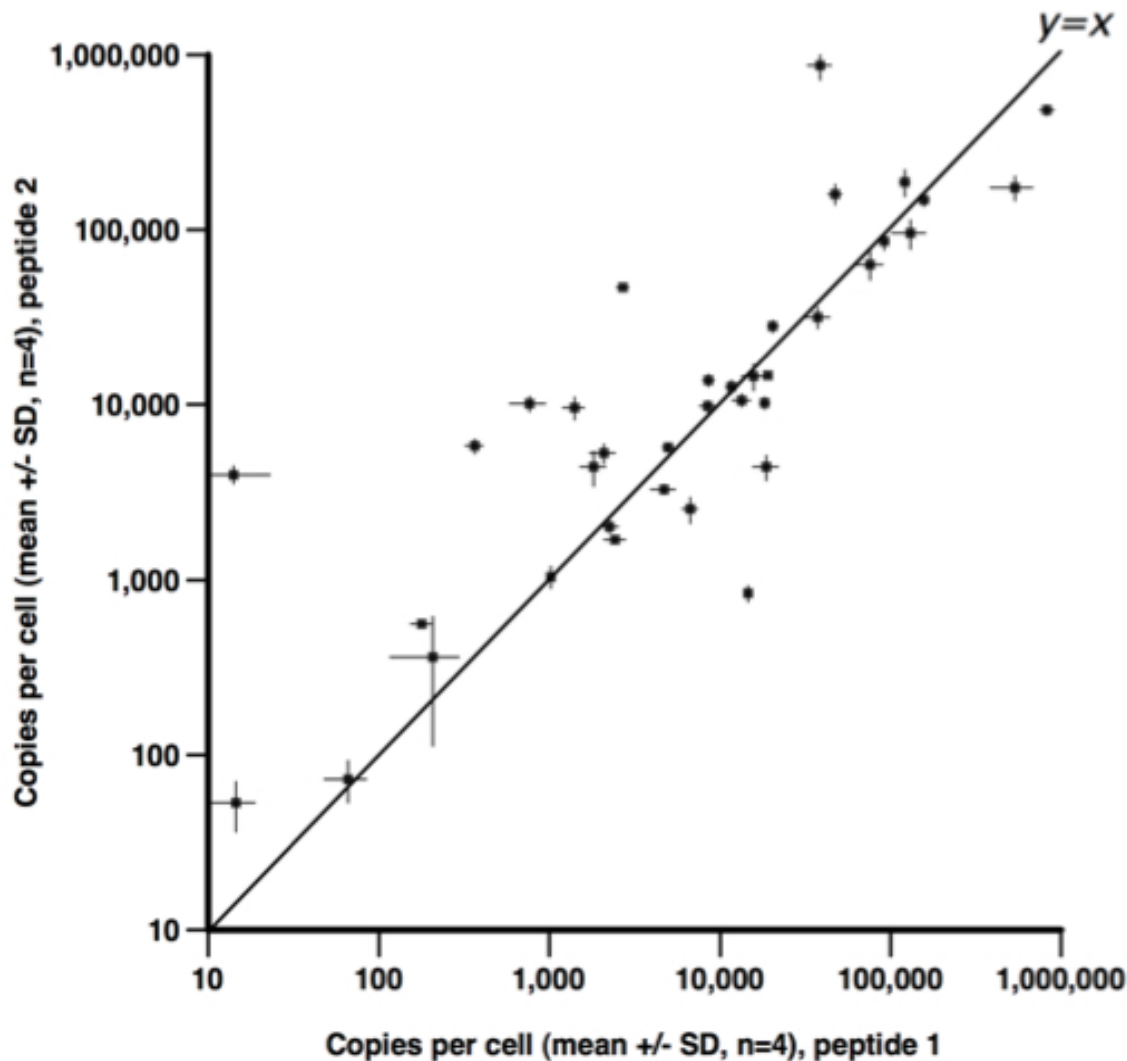
Type A peptides only



Chaperones CoPYcat



Reproducibility for 37 chaperone proteins

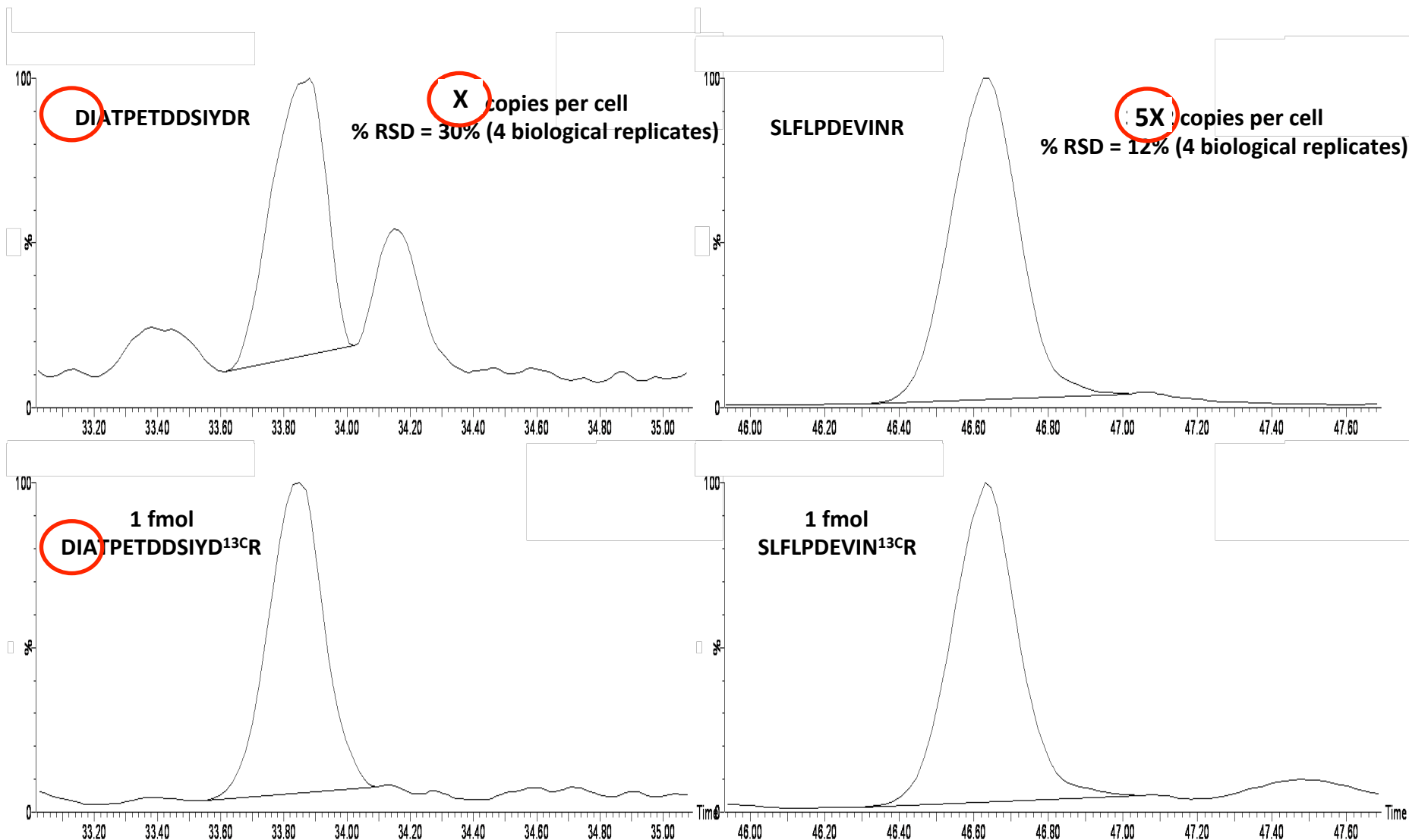


I want to be straight!

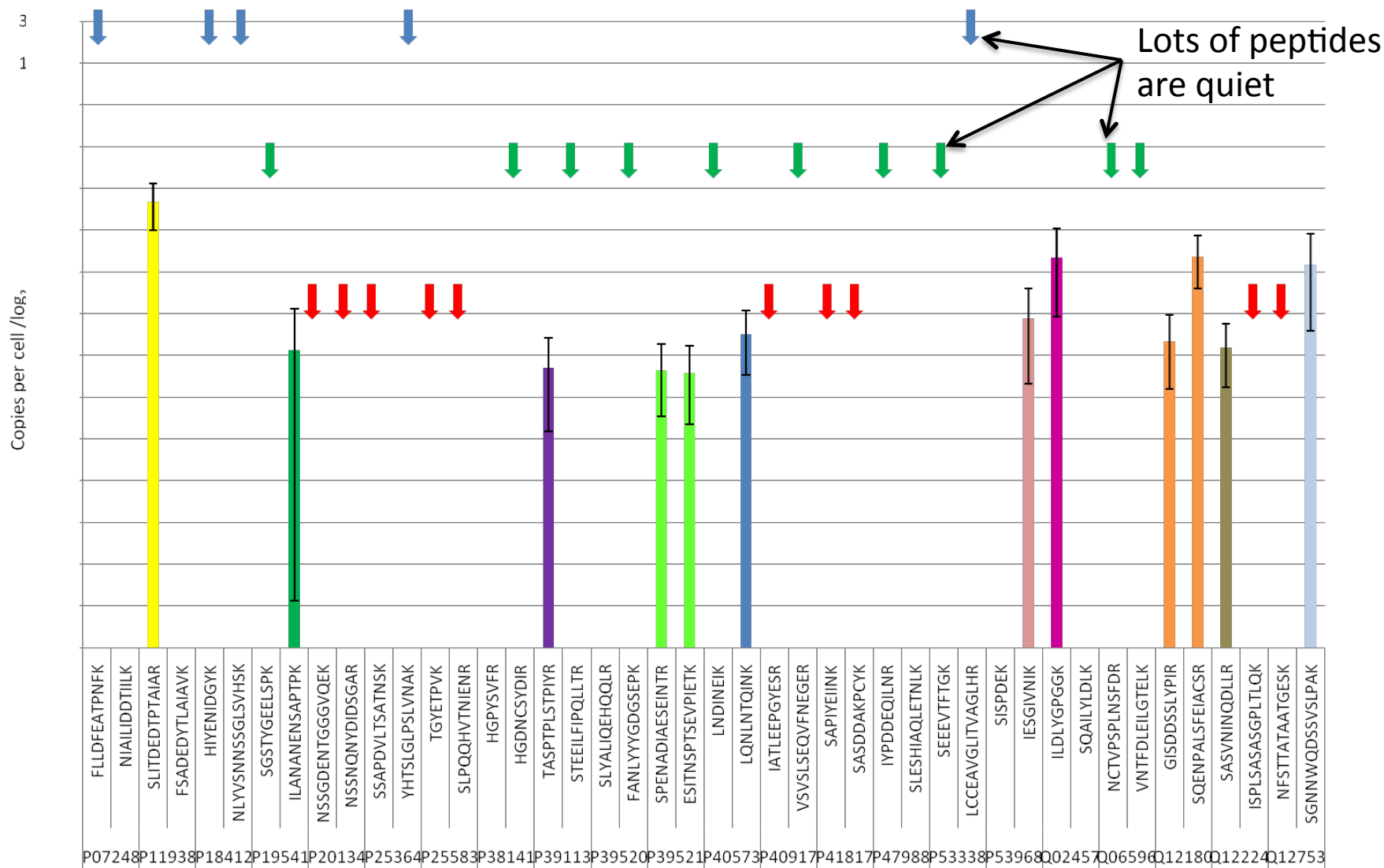
**Sibling peptides
compared over 4
biological replicates**

**Missed cleavage
explains many
“problem” cases**

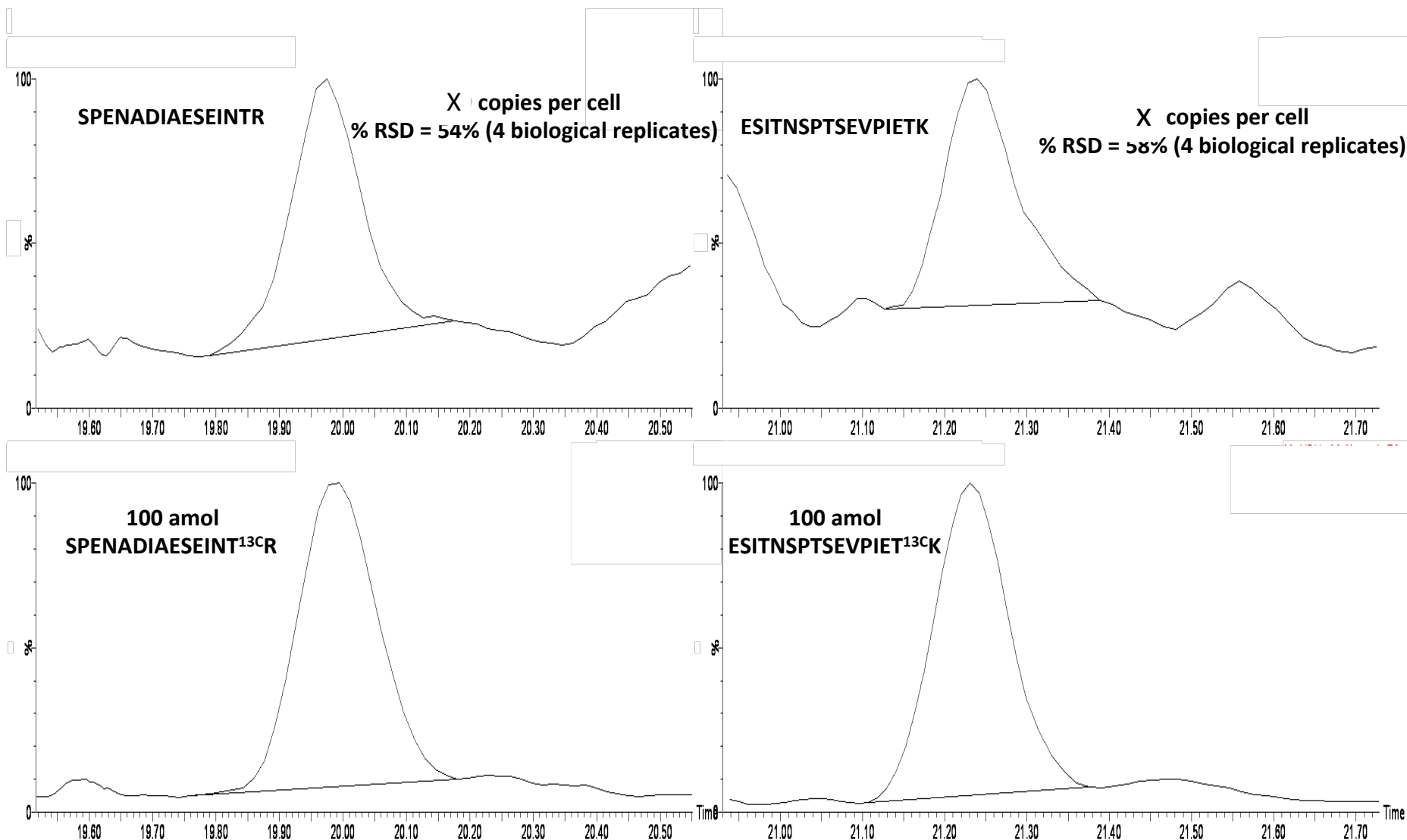
GSH1 (Proteolysis)




Transcription factor Copycat



FLH1 (Transcription factor)



Summary

- Absolute quant by direct means is within our grasp
 - Pipeline is “up and running”, 1000 protein mark soon
 - But **whole** proteome by direct means looks tough
- Methods do not agree 100% and have pros and cons
 - The “jury is out” on best approach 
- Targetted quant via SRM approaches is preferred solution
 - Able to cope with biological real world (isoforms, low abundance, PTMs)
 - Sub 100 cpc achievable

Acknowledgments



- **COPYCAT (Liverpool)**

Rob Beynon

Philip Brownridge

Vicky Harman

Lynn McLean

Deborah Simpson

Duncan Robertson

- **COPYCAT (Manchester)**

Simon Hubbard

Paul Sims

Simon Gaskell (now QMUL)

Claire Eyers

Chris Grant

Craig Lawless

Rae Watkins

Stephen Holman

Karin Lanthaler

Ronan O’Cualain



- **QconCAT (Liverpool)**

Jenny Rivers

Mary Doherty

Amy Claydon

MCISB - Kath Carroll *et al*



Hubbard group

David Wedge

Julian Selley

Paul Blakeley

Jenifer Siepen (now AZ)