SCIENCE MEETS LIFE

MSqRob: analysis of label-free proteomics data in an R/Shiny environment

Ludger Goeminne

Emmy Van Quickelberghe

Lieven Clement

Kris Gevaert

11/01/2017

VIB

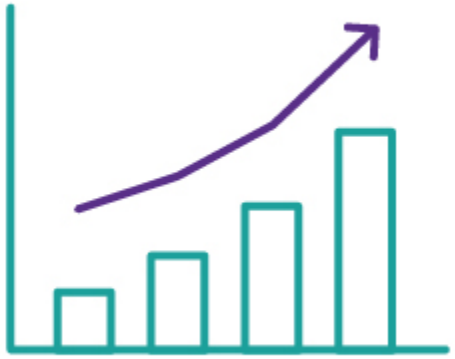
UNIVERSITEIT GENT

AGENTSCHAP INNOVEREN & ONDERNEMEN

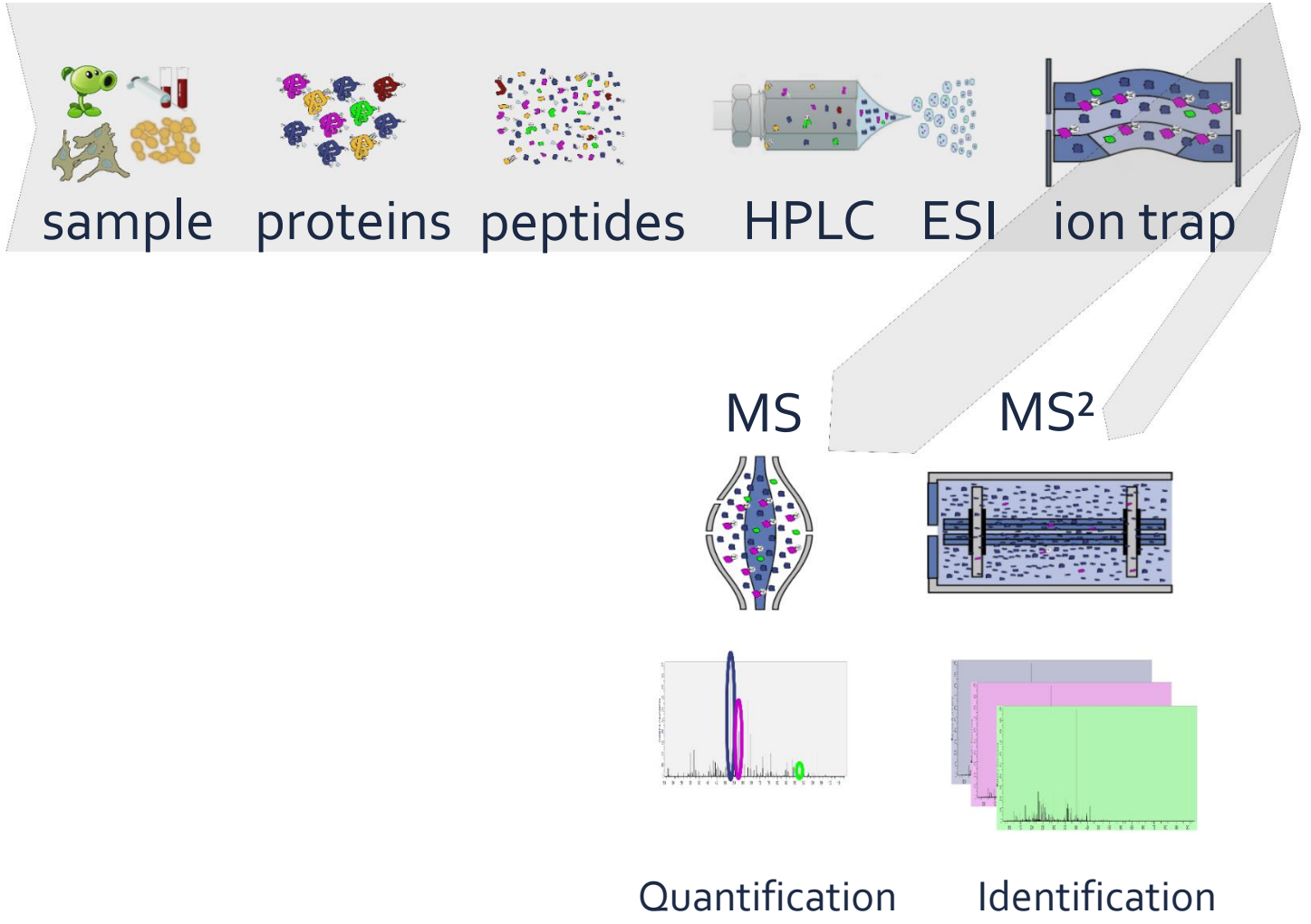
BIG N2N
Bioinformatics Institute Ghent
from nucleotides to networks

How often have you been stuck in the data analysis part?

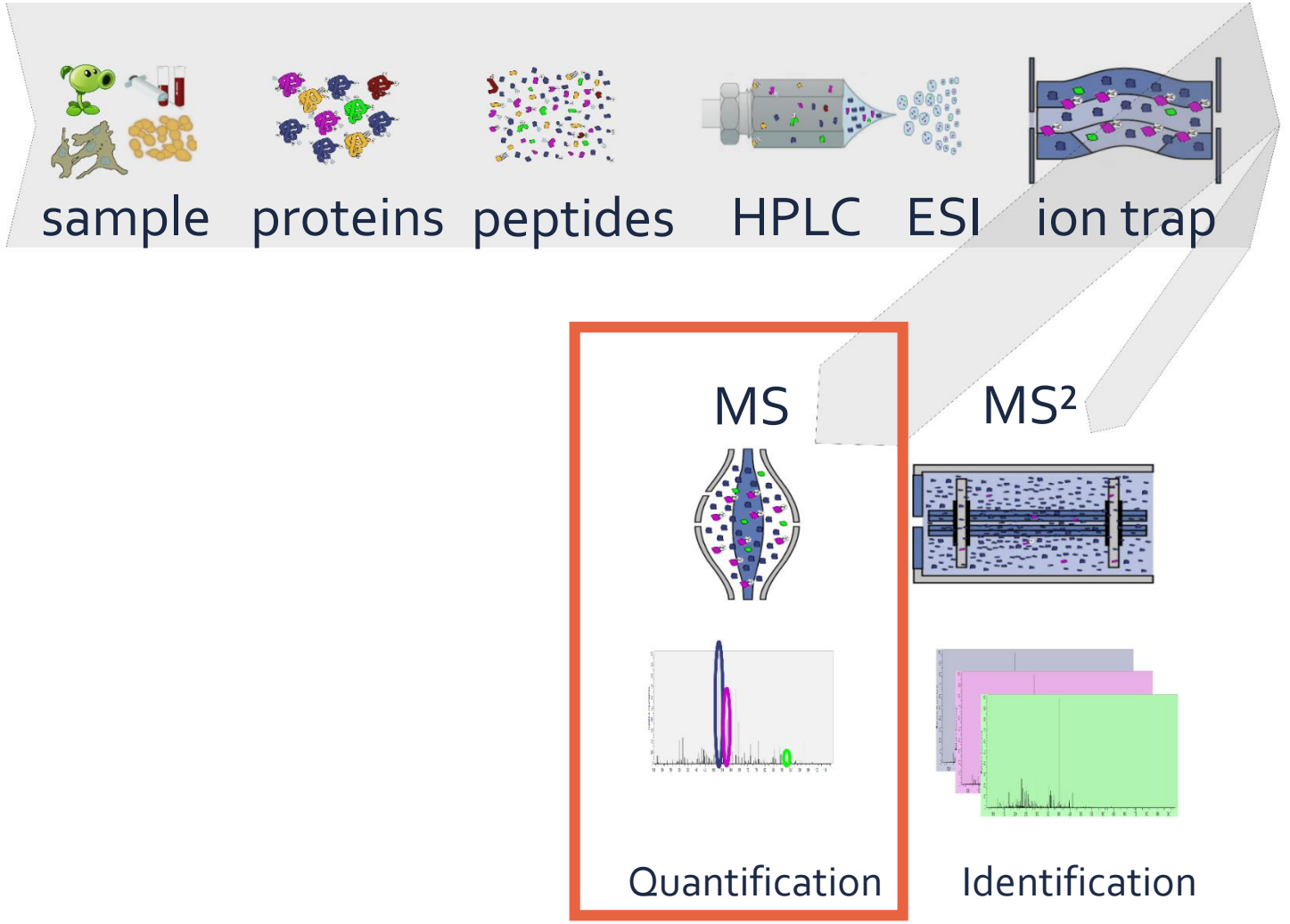
??????????



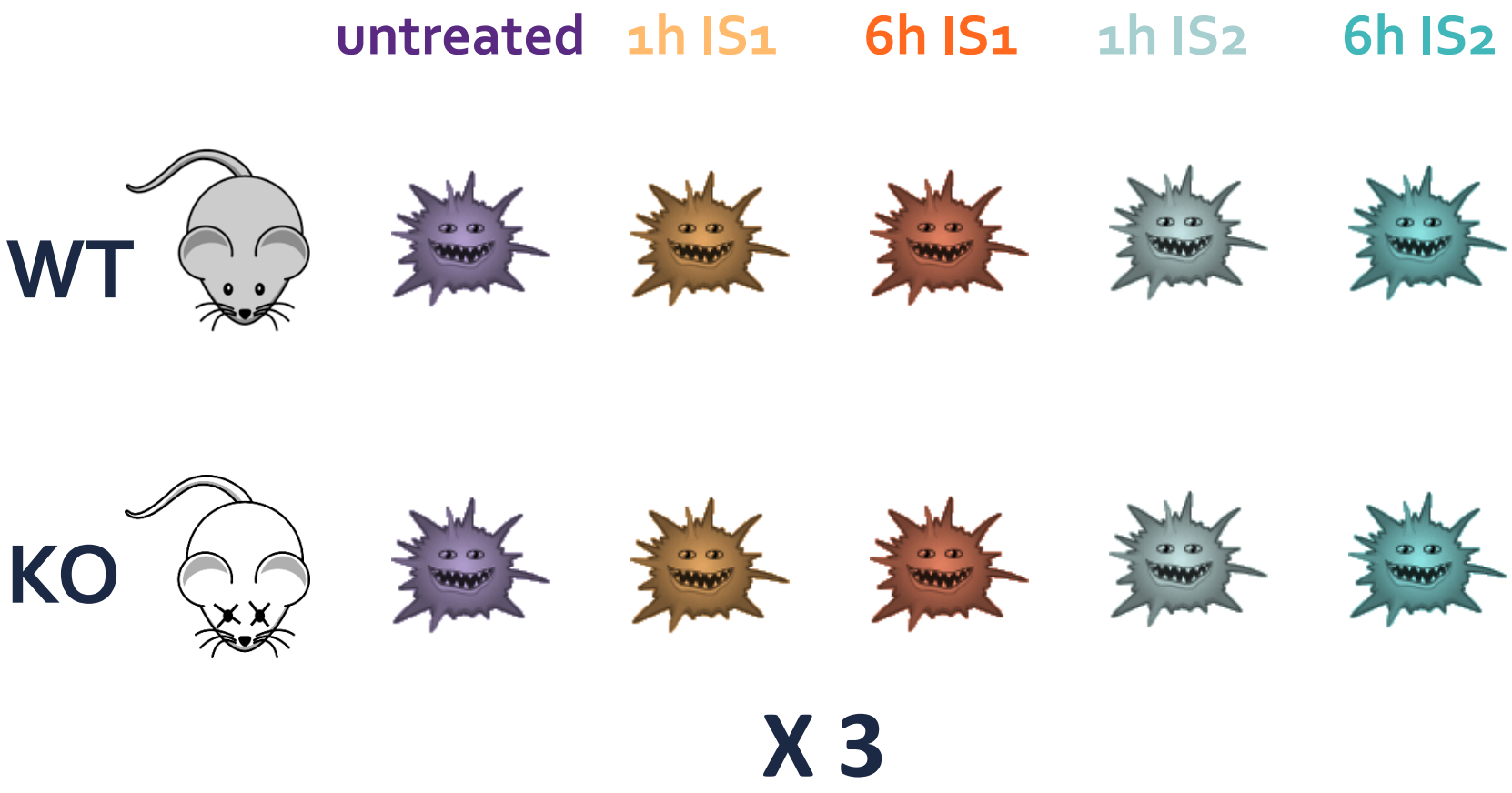
MS-based proteomics identifies many thousands of peptides



MS-based proteomics identifies many thousands of peptides



Repeated-measures designs are hard to analyze with existing tools



Many tools return uninteresting proteins because they cannot address complex designs

Perseus does only t-tests between two groups,

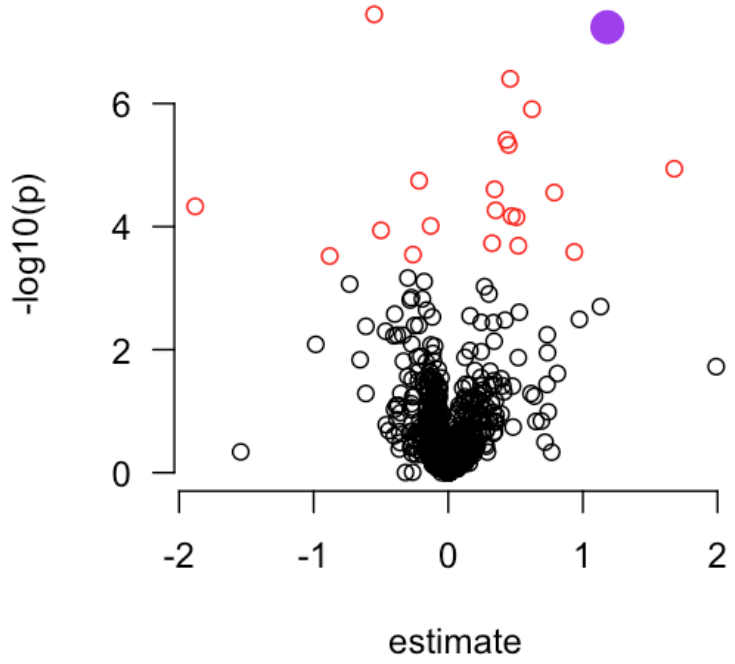
MSStats can account for repeated measures,
but only a few designs are possible
and it suffers from overfitting and outliers

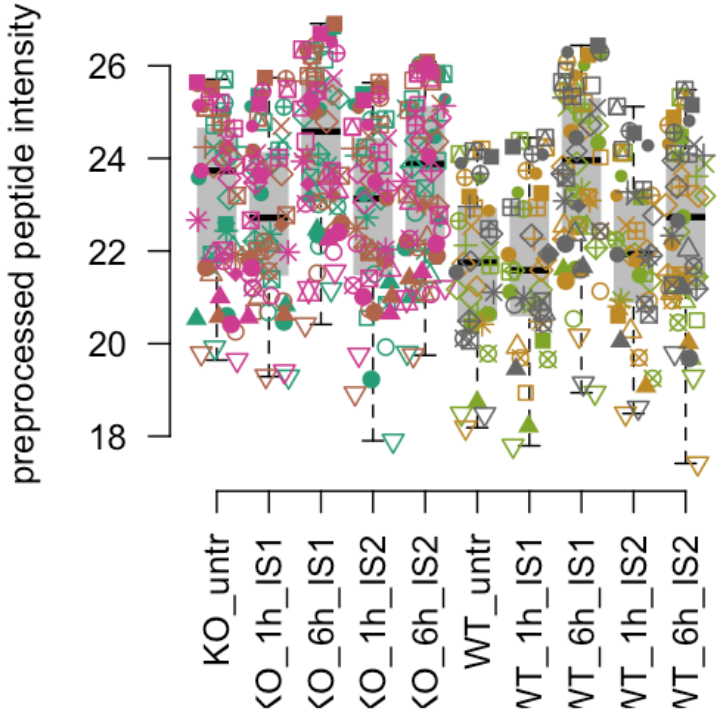
Statistics and proteomics join forces

Lieven Clement: statistics

Kris Gevaert: proteomics

MSqRob
can solve your data-analysis problems





Analysis of label-free proteomics data with MSqRob

Performance: why MSqRob works so well

Case study: a repeated measures design

Features: how you can use MSqRob

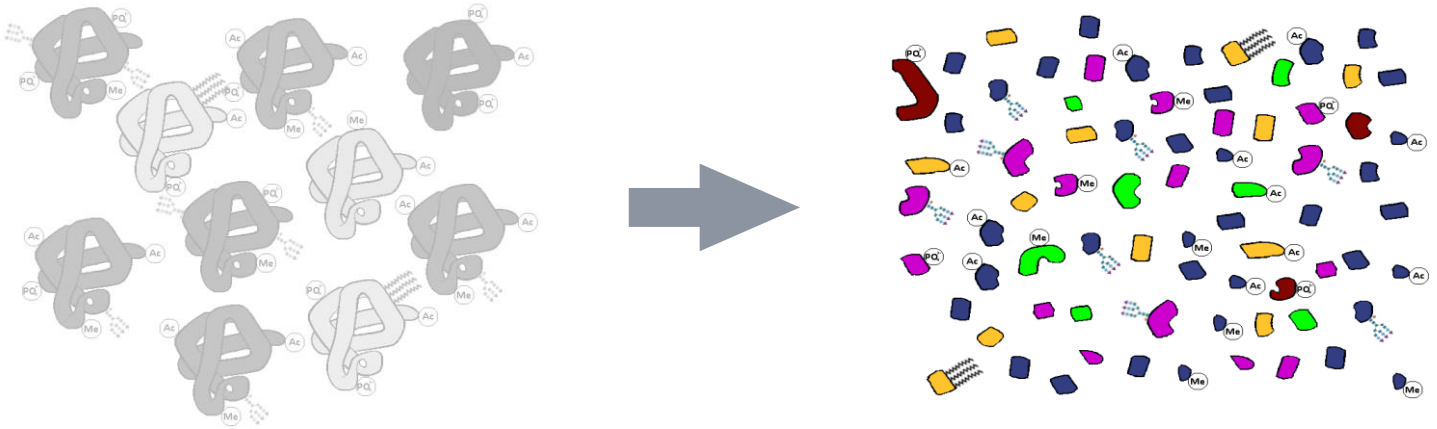
Analysis of label-free proteomics data with MSqRob

Performance: why MSqRob works so well

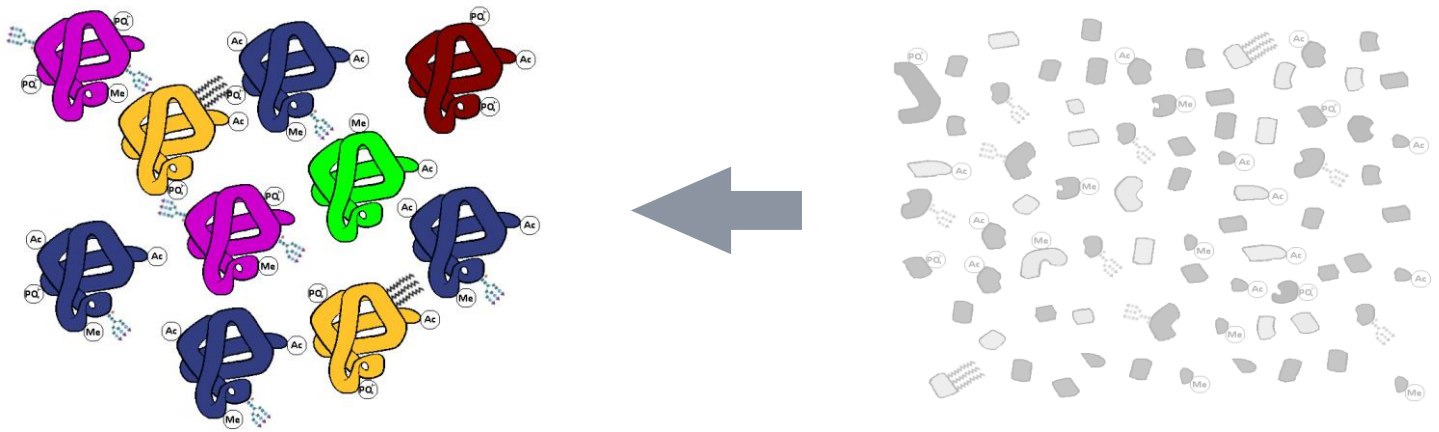
Case study: a repeated-measures design

Features: how you can use MSqRob

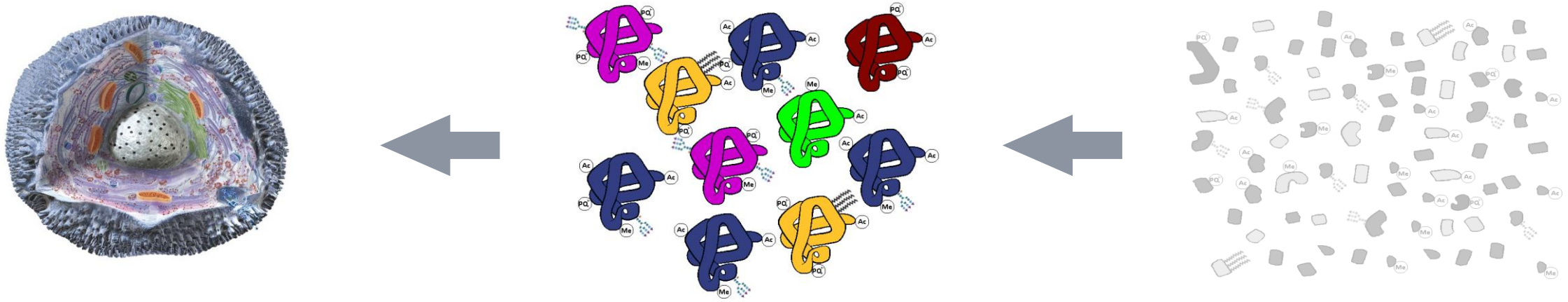
MS-based proteomics identifies peptides…



But we need protein-level information

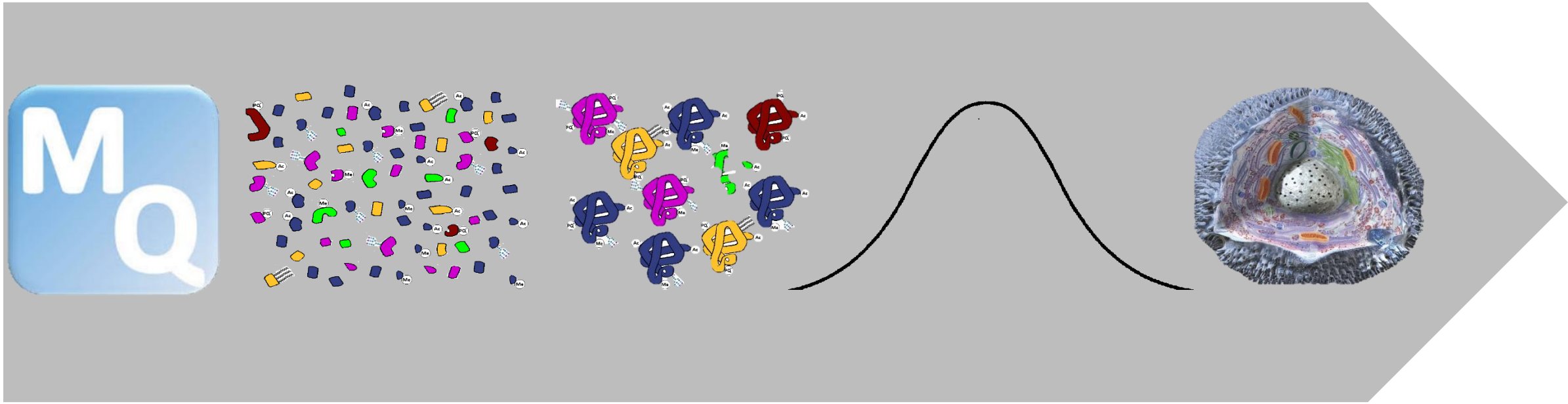


But we need protein-level information



Approach 1: summarization-based approaches

Peptide identification → **Estimate Protein intensities** → **Statistics** → **Results**

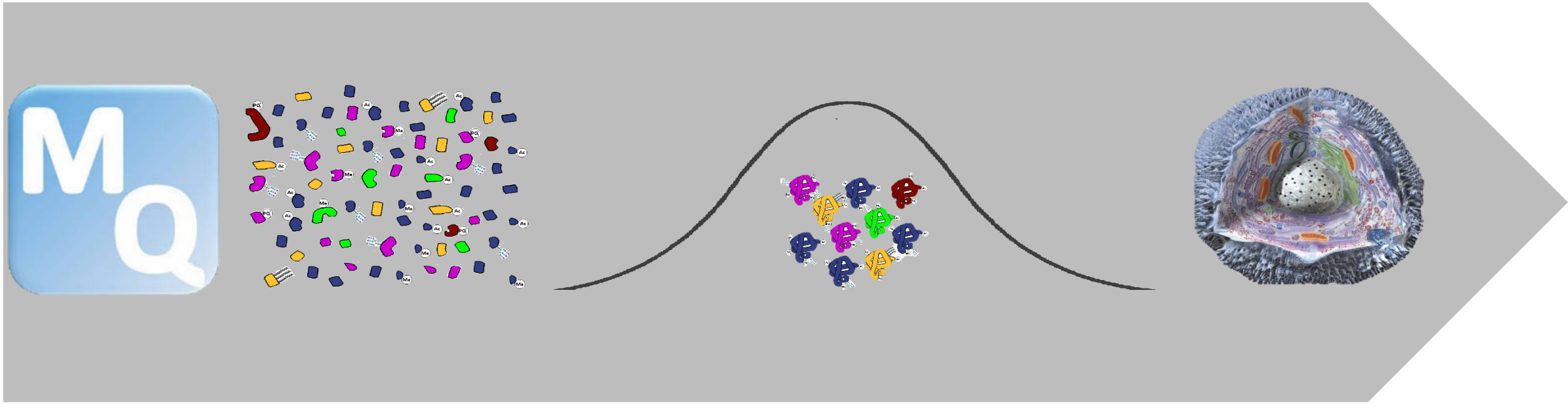


Approach 2: peptide-based models

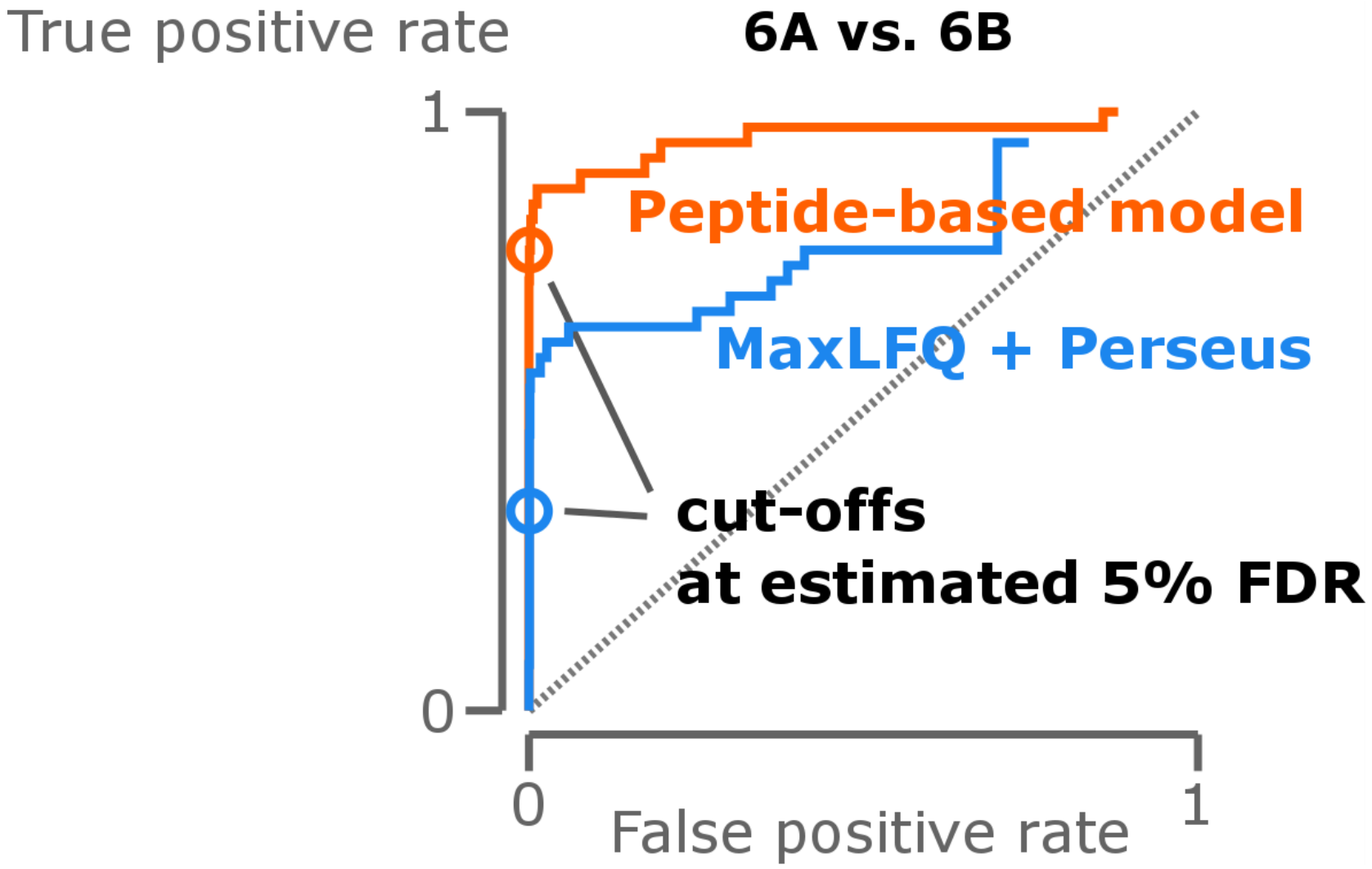
Peptide identification

Directly model peptide-intensities

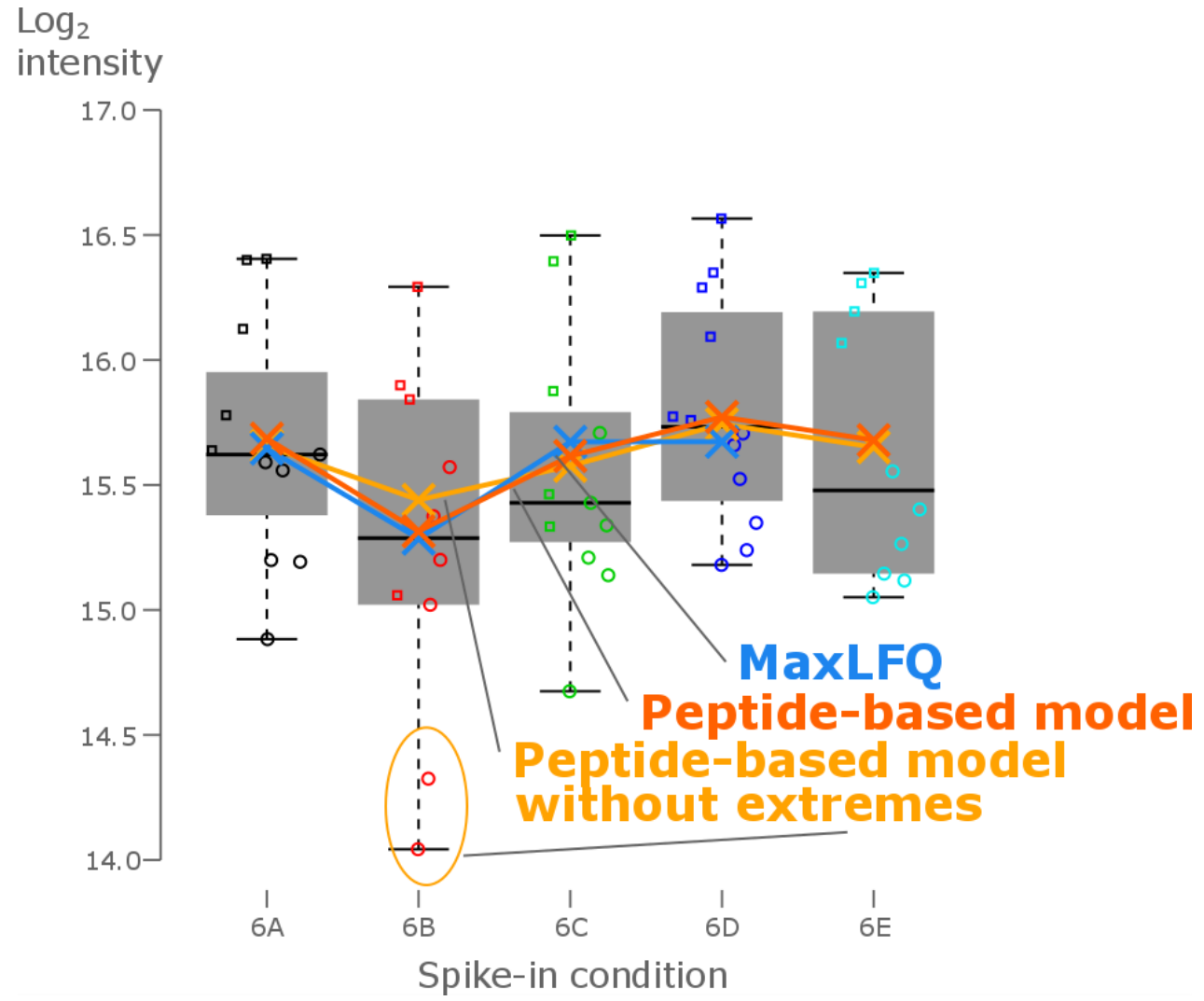
Results



Peptide-based models work better than summarization-based approaches



Existing methods suffer from overfitting, unstable variances and outliers



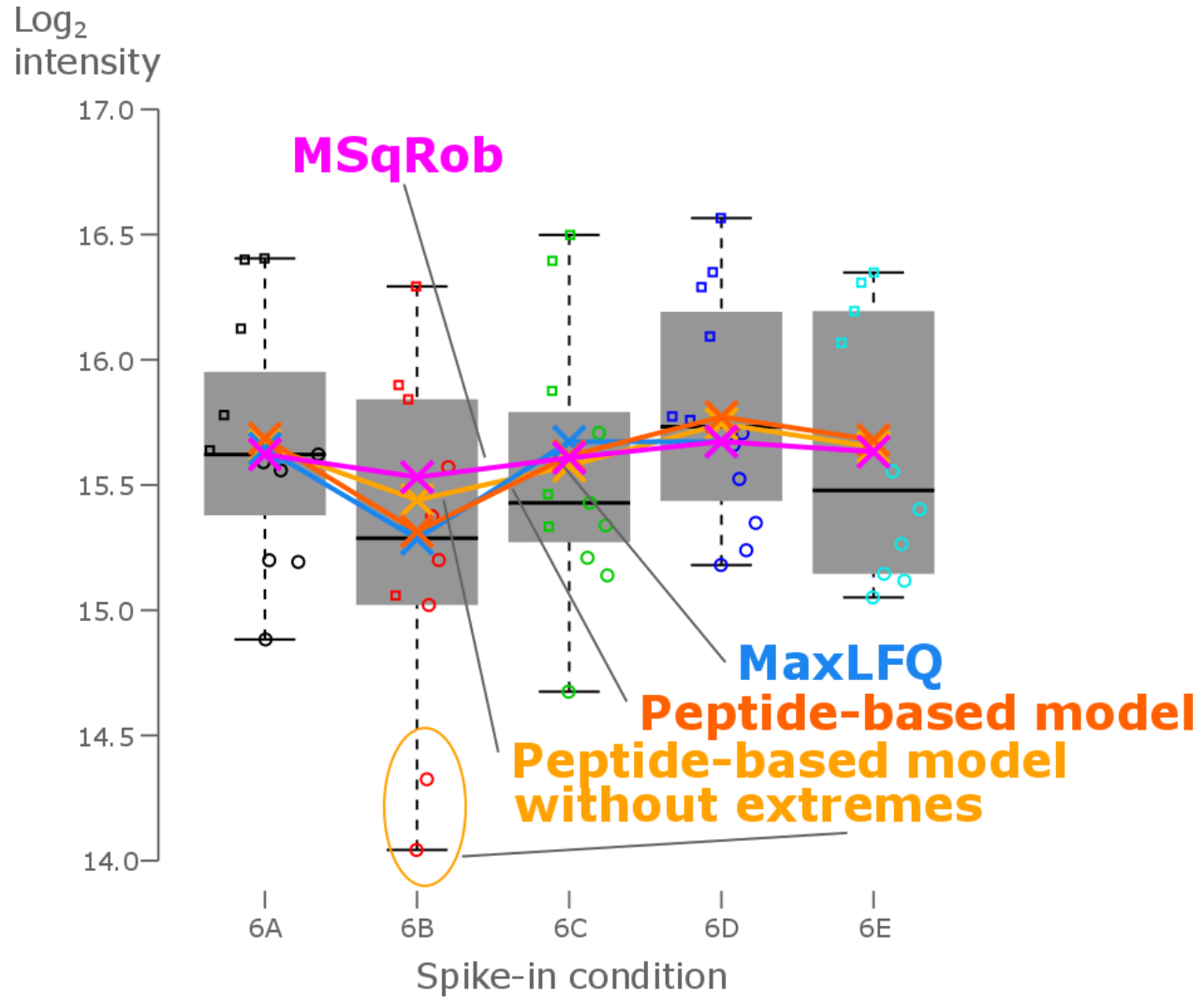
MSqRob adds 3 improvements to peptide-based models

Shrinkage estimation

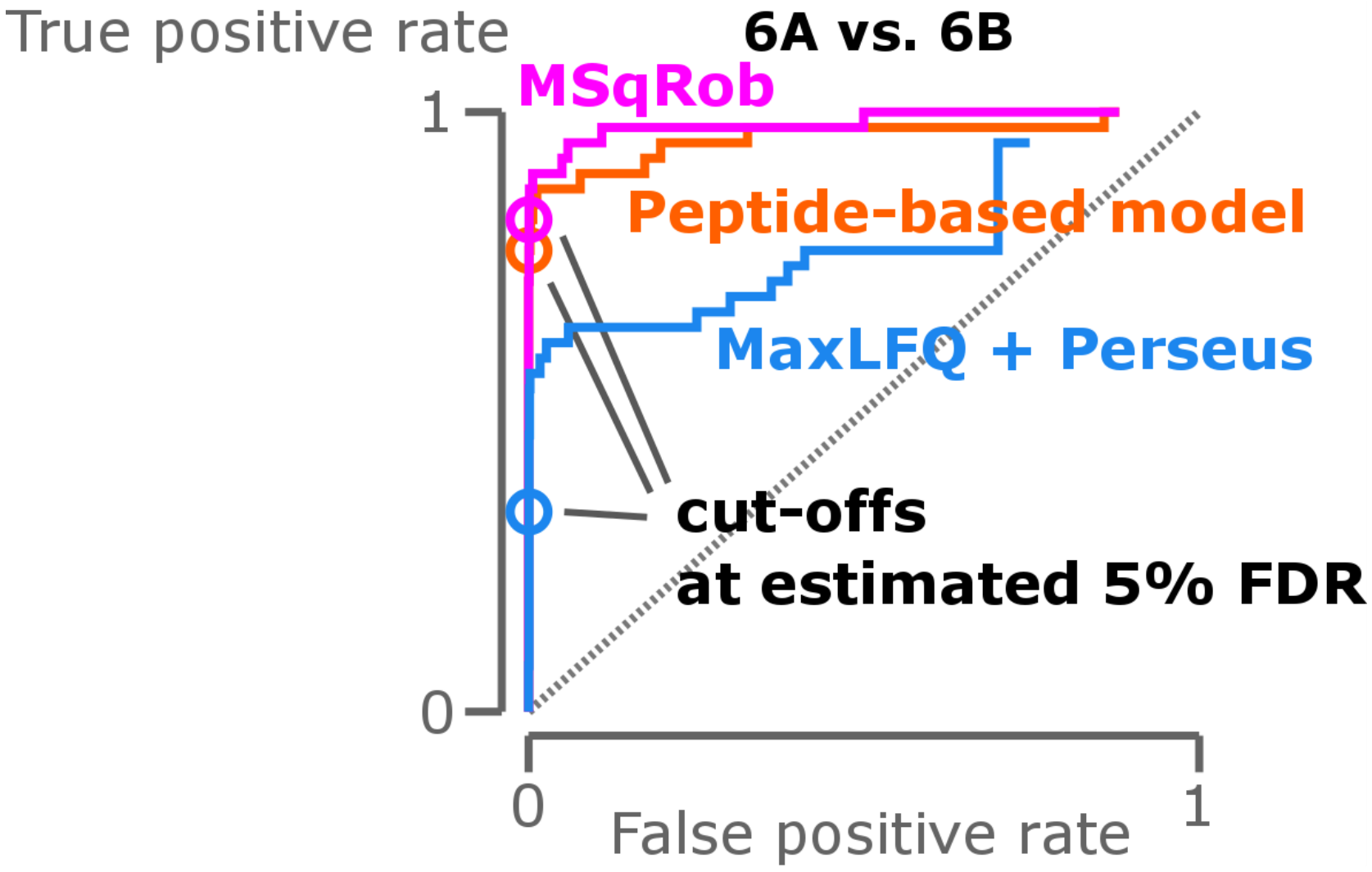
Borrowing information across proteins

Weighing down outliers

MSqRob provides more stable estimates



MSqRob outperforms other methods



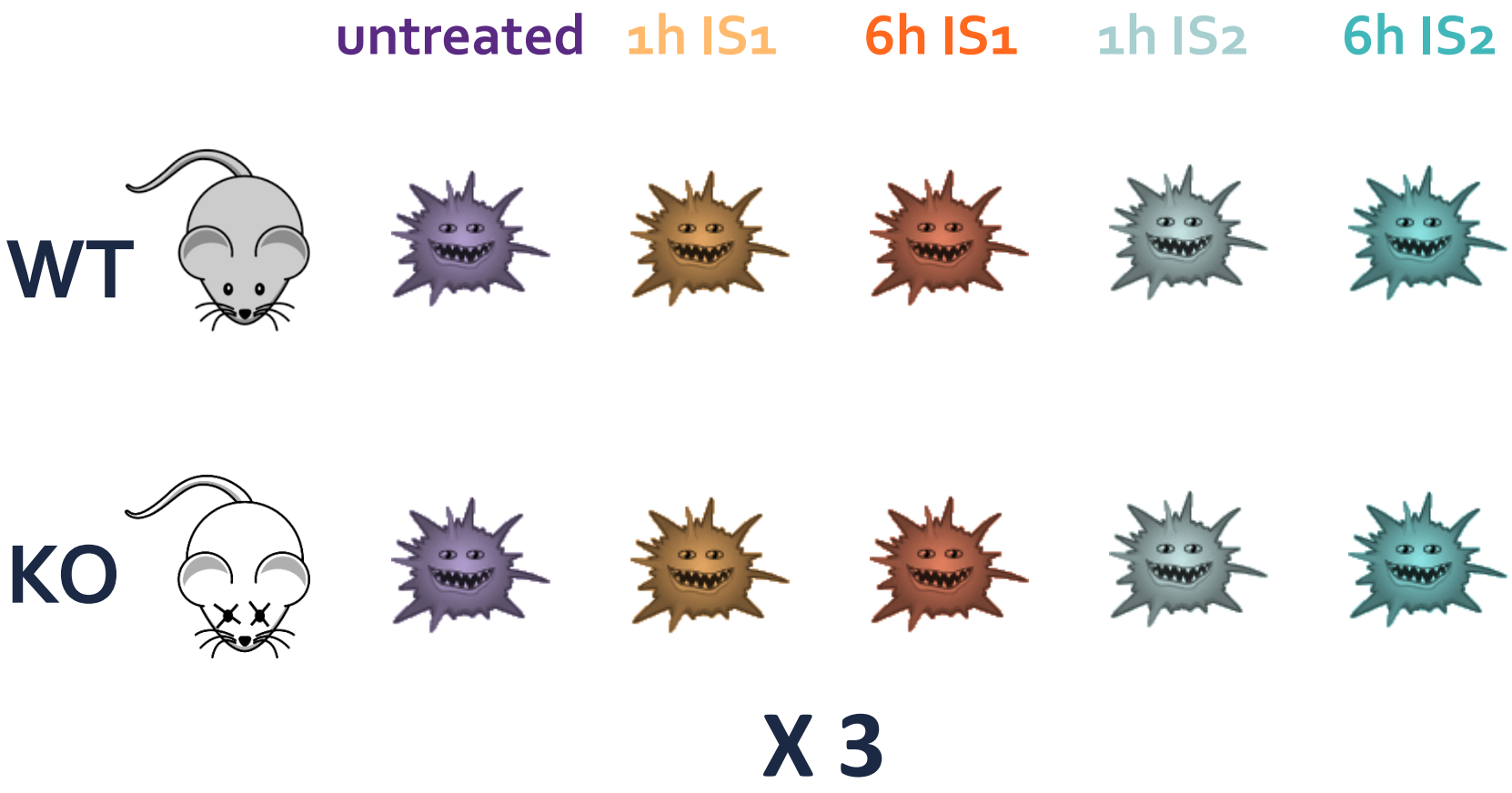
Analysis of label-free proteomics data with MSqRob

Performance: why MSqRob works so well

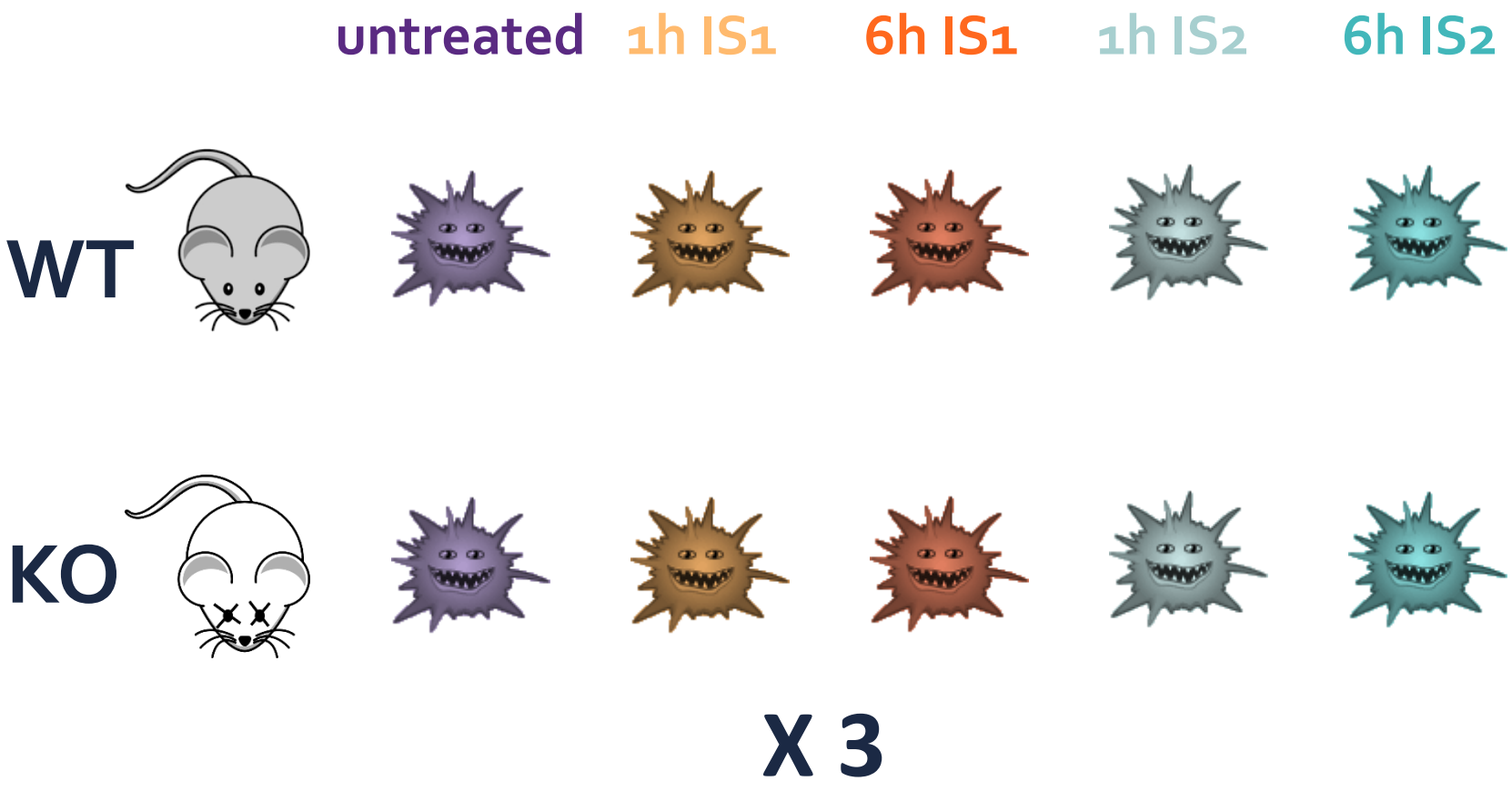
Case study: a repeated-measures design

Features: how you can use MSqRob

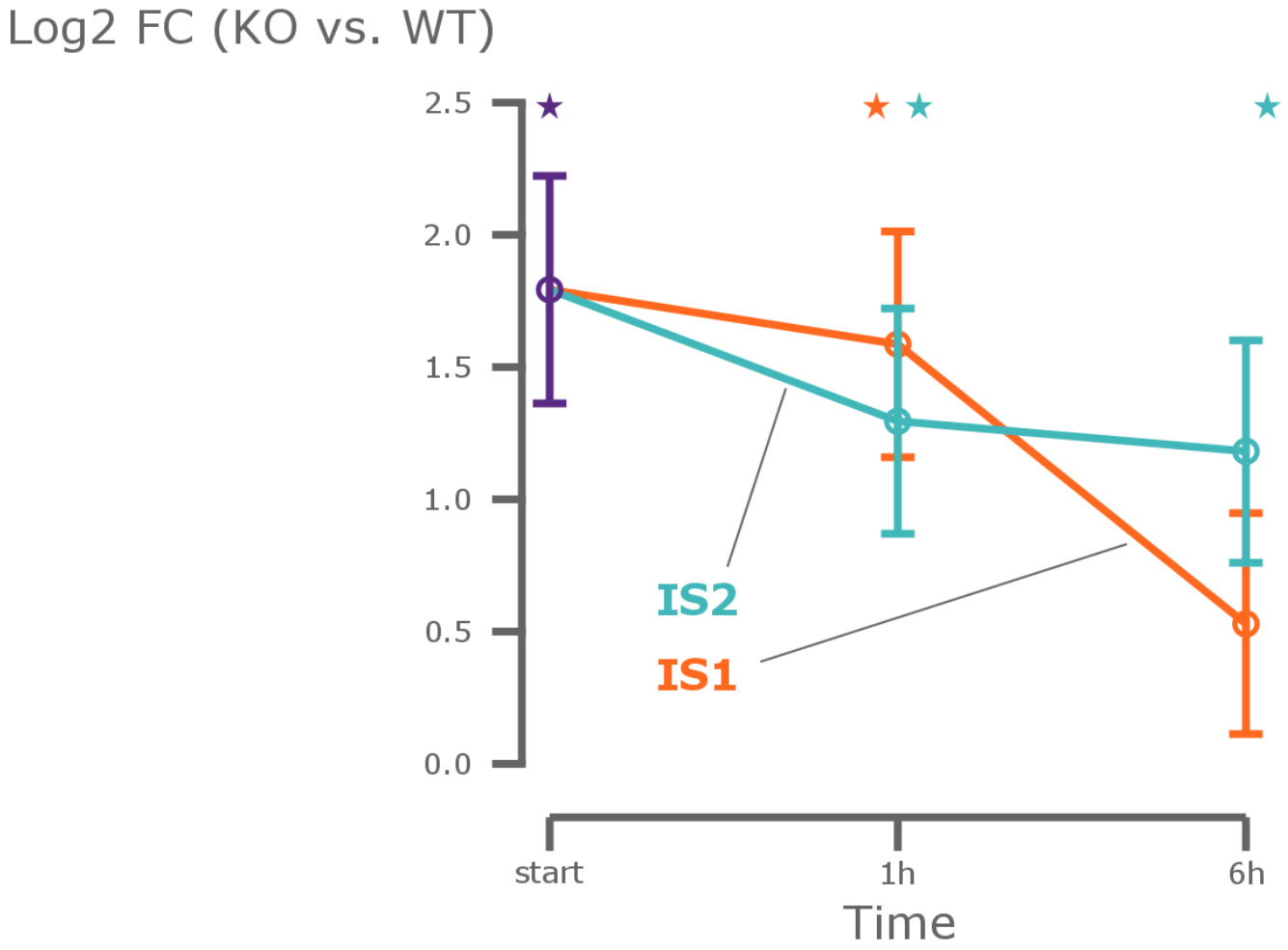
Repeated-measures designs are hard to analyze with existing tools



MSqRob handles repeated measures in a natural way



Significant upregulation in KO vs. WT, significant interaction between 6h and start



Analysis of label-free proteomics data with MSqRob

Performance: why MSqRob works so well

Case study: a repeated-measures design

Features: how you can use MSqRob

You only need MaxQuant's peptides.txt and your experimental annotation

Project Name

project

Specify the location where your output will be saved

Browse... /Users/lgoeminn/Doc

Upload complete

Specify the location of your experimental annotation file

Browse... annotation.xlsx

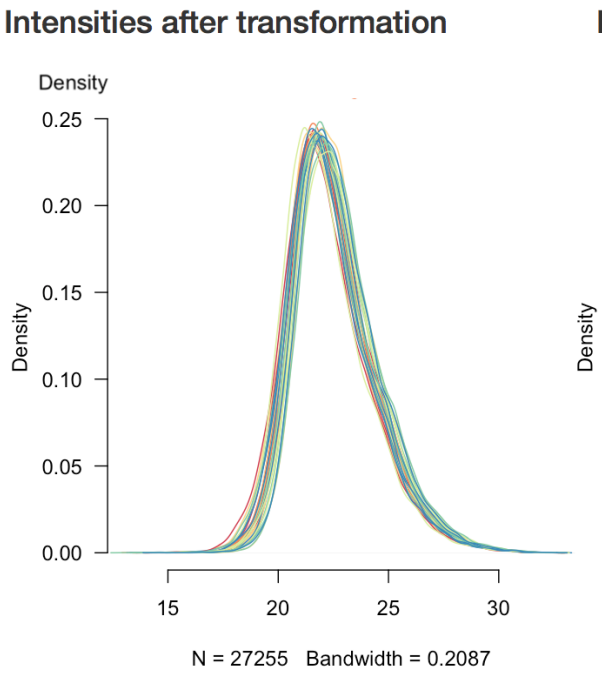
Upload complete

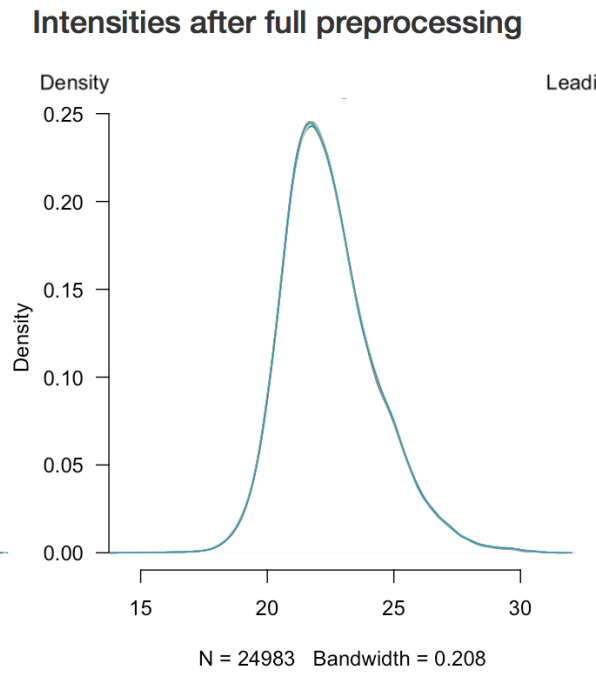
Specify the location of your peptides.txt file

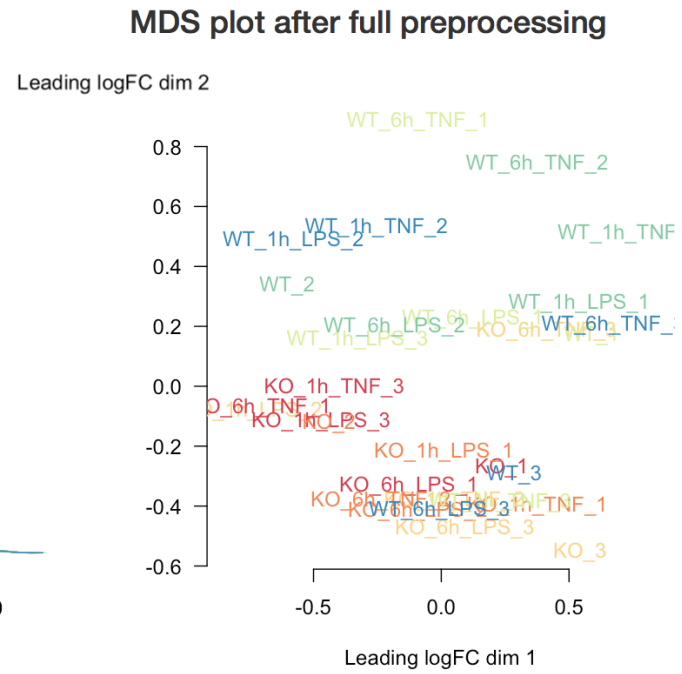
Browse... peptides.txt

Upload complete

Preprocess your data

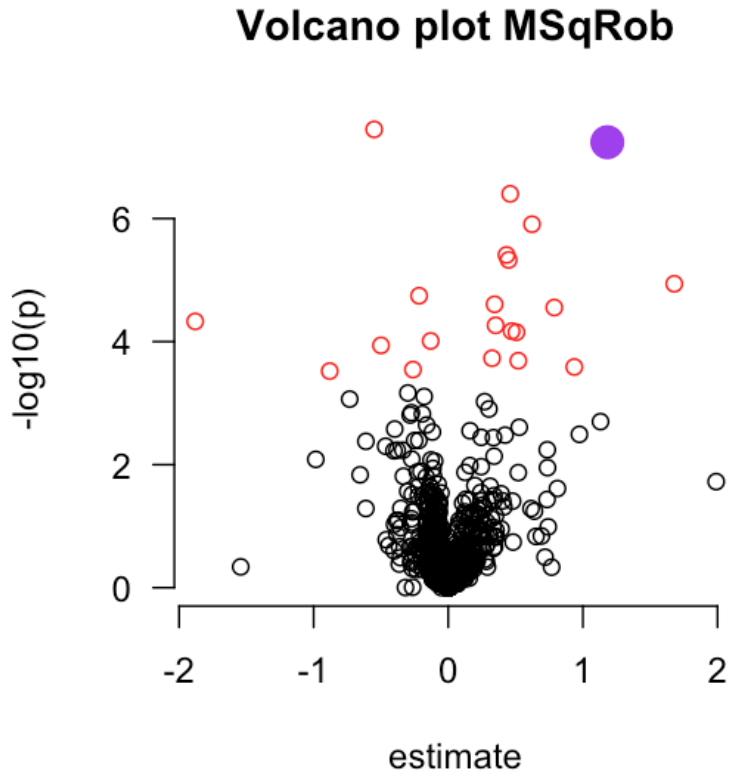




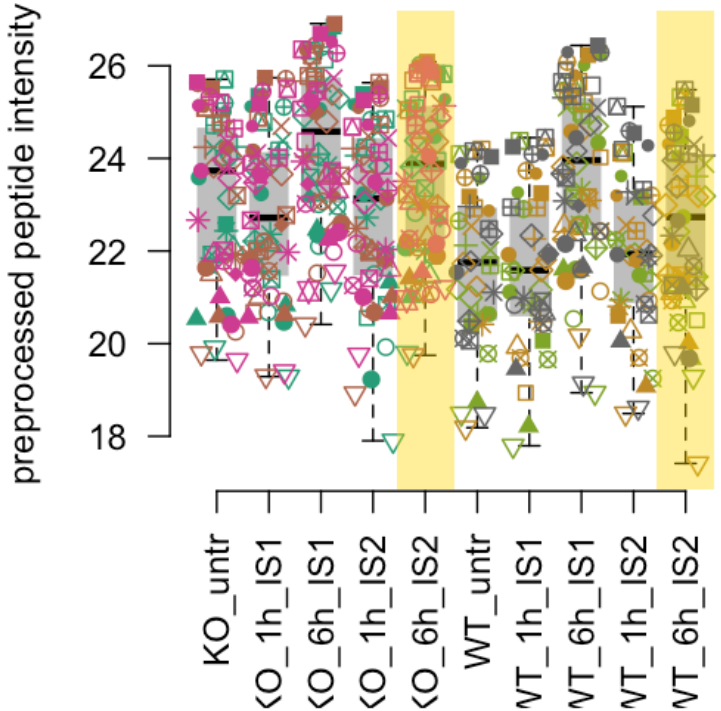


Inspect the results graphically

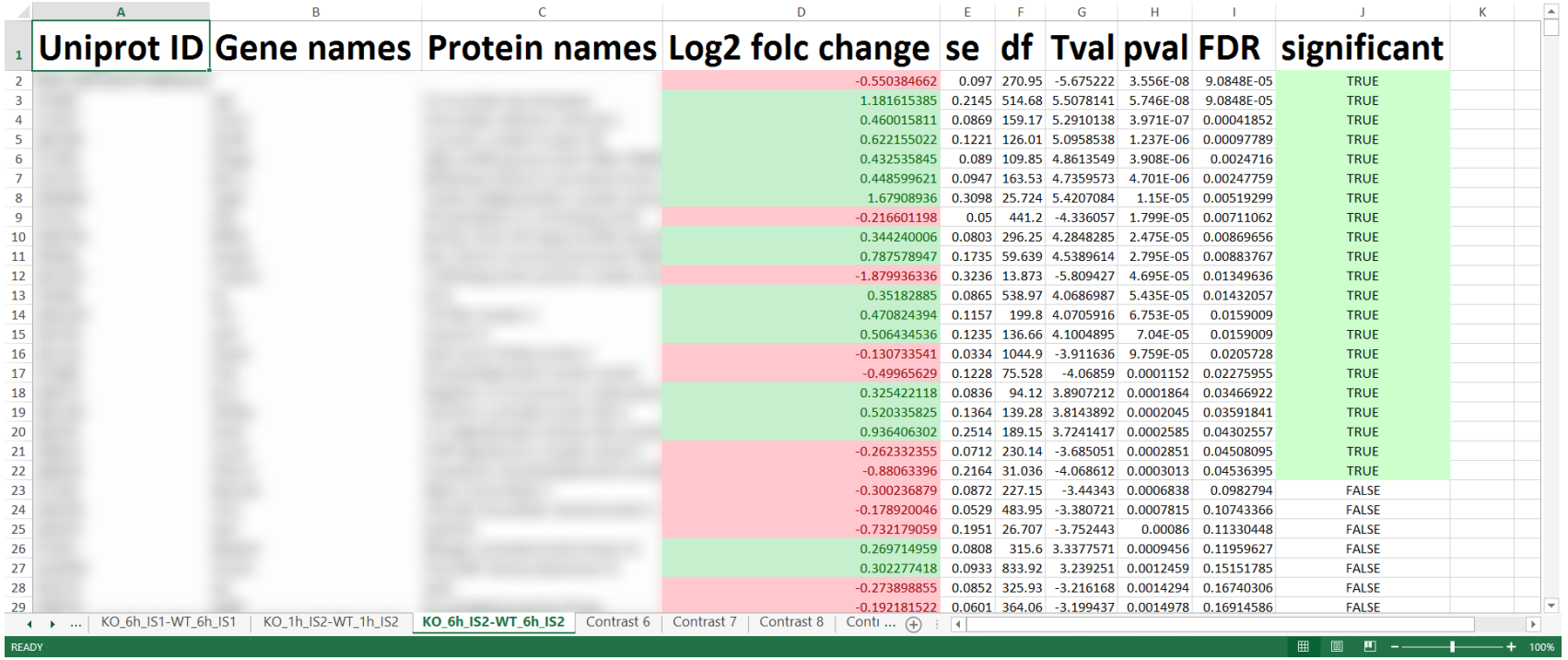
Volcano plot



Detail plot



Export the results to Excel



	A	B	C	D	E	F	G	H	I	J	K
1	Uniprot ID	Gene names	Protein names	Log2 folc change	se	df	Tval	pval	FDR	significant	
2				-0.550384662	0.097	270.95	-5.675222	3.556E-08	9.0848E-05	TRUE	
3				1.181615385	0.2145	514.68	5.5078141	5.746E-08	9.0848E-05	TRUE	
4				0.460015811	0.0869	159.17	5.2910138	3.971E-07	0.00041852	TRUE	
5				0.622155022	0.1221	126.01	5.0958538	1.237E-06	0.00097789	TRUE	
6				0.432535845	0.089	109.85	4.8613549	3.908E-06	0.0024716	TRUE	
7				0.448599621	0.0947	163.53	4.7359573	4.701E-06	0.00247759	TRUE	
8				1.67908936	0.3098	25.724	5.4207084	1.15E-05	0.00519299	TRUE	
9				-0.216601198	0.05	441.2	-4.336057	1.799E-05	0.00711062	TRUE	
10				0.344240006	0.0803	296.25	4.2848285	2.475E-05	0.00869656	TRUE	
11				0.787578947	0.1735	59.639	4.5389614	2.795E-05	0.00883767	TRUE	
12				-1.879936336	0.3236	13.873	-5.809427	4.695E-05	0.01349636	TRUE	
13				0.35182885	0.0865	538.97	4.0686987	5.435E-05	0.01432057	TRUE	
14				0.470824394	0.1157	199.8	4.0705916	6.753E-05	0.0159009	TRUE	
15				0.506434536	0.1235	136.66	4.1004895	7.04E-05	0.0159009	TRUE	
16				-0.130733541	0.0334	1044.9	-3.911636	9.759E-05	0.0205728	TRUE	
17				-0.49965629	0.1228	75.528	-4.06859	0.0001152	0.02275955	TRUE	
18				0.325422118	0.0836	94.12	3.8907212	0.0001864	0.03466922	TRUE	
19				0.520335825	0.1364	139.28	3.8143892	0.0002045	0.03591841	TRUE	
20				0.936406302	0.2514	189.15	3.7241417	0.0002585	0.04302557	TRUE	
21				-0.262332355	0.0712	230.14	-3.685051	0.0002851	0.04508095	TRUE	
22				-0.88063396	0.2164	31.036	-4.068612	0.0003013	0.04536395	TRUE	
23				-0.300236879	0.0872	227.15	-3.44343	0.0006838	0.0982794	FALSE	
24				-0.178920046	0.0529	483.95	-3.380721	0.0007815	0.10743366	FALSE	
25				-0.732179059	0.1951	26.707	-3.752443	0.00086	0.11330448	FALSE	
26				0.269714959	0.0808	315.6	3.3377571	0.0009456	0.11959627	FALSE	
27				0.302277418	0.0933	833.92	3.239251	0.0012459	0.15151785	FALSE	
28				-0.273898855	0.0852	325.93	-3.216168	0.0014294	0.16740306	FALSE	
29				-0.192181522	0.0601	364.06	-3.199437	0.0014978	0.16914586	FALSE	

KO_6h_IS1-WT_6h_IS1 | KO_1h_IS2-WT_1h_IS2 | KO_6h_IS2-WT_6h_IS2 | Contrast 6 | Contrast 7 | Contrast 8 | Conti ...

READY 100%

Experienced users:
run MSqRob in bash mode

CPTAC analysis

Ludger Goeminne, Kris Gevaert and Lieven Clement

2017-03-06

Process the CPTAC data with MSqRob

In this file, we show you how the CPTAC data should be processed. It can serve as a basis for your own data analysis. For each research hypothesis, we rank the proteins by significance and we make volcano plots. We also save the output to Excel. Settings can be saved optionally as well. Just adjust the parameters in the first block of code and knit the file. If you want to see an example of the same file adjusted for the *Francisella* dataset, check out https://github.com/statOmics/MSqRobData/blob/master/inst/extdata/Francisella/analysis_Francisella.Rmd. If you want to have this analysis with a lot of extra material and explanations and background on the experiments, check out our vignette at https://github.com/statOmics/MSqRob/blob/master/vignettes/MSqRob.Rmd.

```
library(MSqRob)
```

Import results in MSqRob GUI for manual inspection!

Download MSqRob from GitHub

https://github.com/statOmics/MSqRob

Use our tutorial paper!

Goeminne, L.J.E., Gevaert, K. and Clement, L. **Experimental design and data-analysis in label-free quantitative LC/MS proteomics: a tutorial with MSqRob**. Journal of Proteomics (minor revision).

Analysis of label-free proteomics data with MSqRob

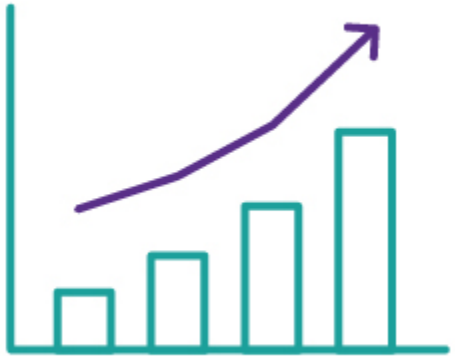
Performance: why MSqRob works so well

Case study: a repeated measures design

Features: how you can use MSqRob

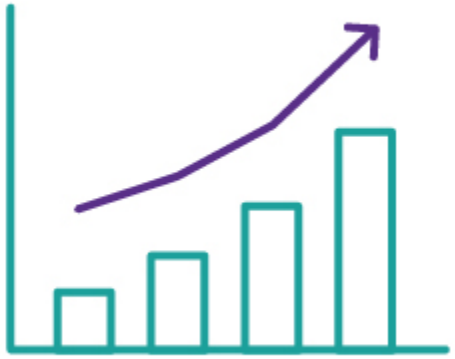
How often have you been stuck in the data analysis part?

??????????

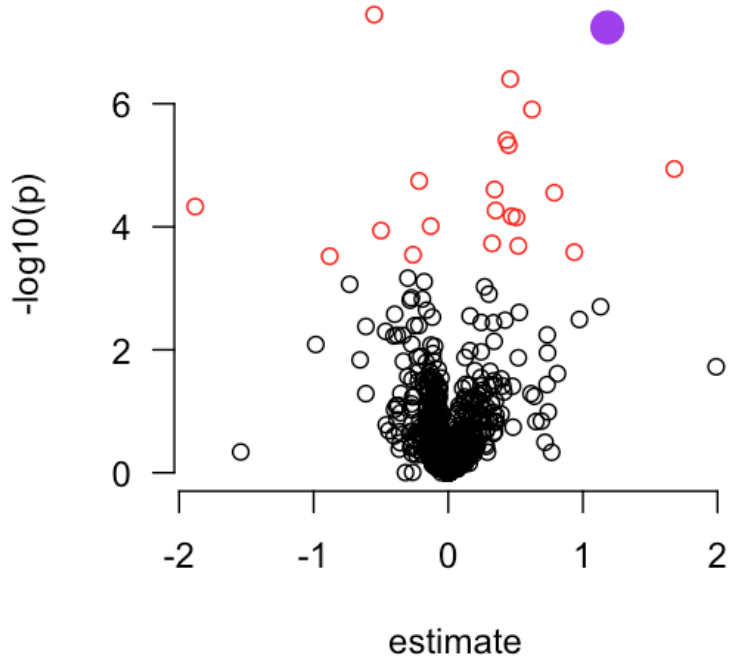


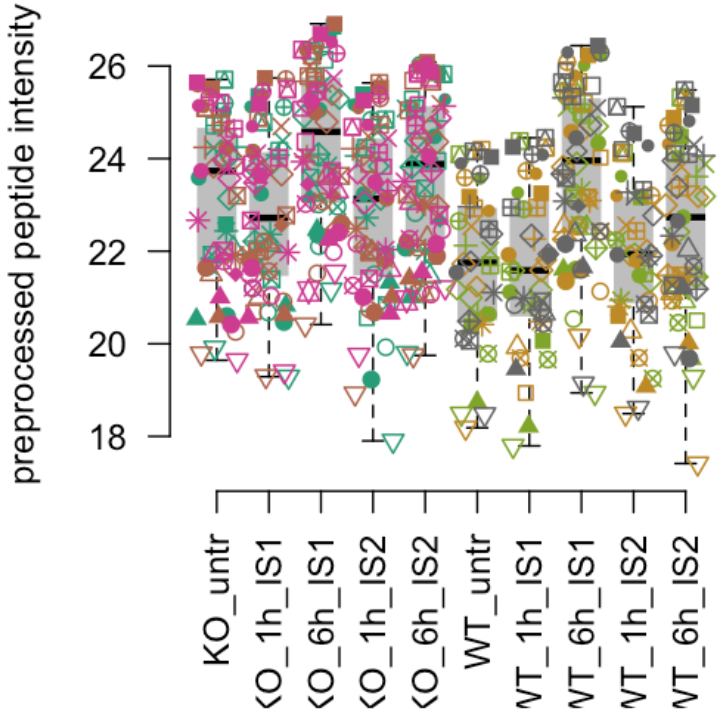
MSqRob can solve your problem

MSqRob



Give MSqRob a try!





ludger.goeminne@vib-ugent.be

https://github.com/statOmics/MSqRob

Lieven Clement

and lab members

FACULTY OF SCIENCES

VIB

UNIVERSITEIT GENT

Kris Gevaert

Emmy Van Quickelberghe

and other lab members

VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY

FACULTY OF MEDICINE AND HEALTH SCIENCES

BIG N2N
Bioinformatics Institute Ghent
from nucleotides to networks

AGENTSCHAP INNOVEREN & ONDERNEMEN

Vlaanderen is ondernemen