

SCIENCE MEETS LIFE

# A Leap over the Hurdle in Label-Free Quantitative Proteomics

Uniting Spectral Counting and Peak-Intensity-Based Methods

Ludger Goeminne

18/09/2017

Promotors:

Lieven Clement

Kris Gevaert



# Which picture is the brightest?



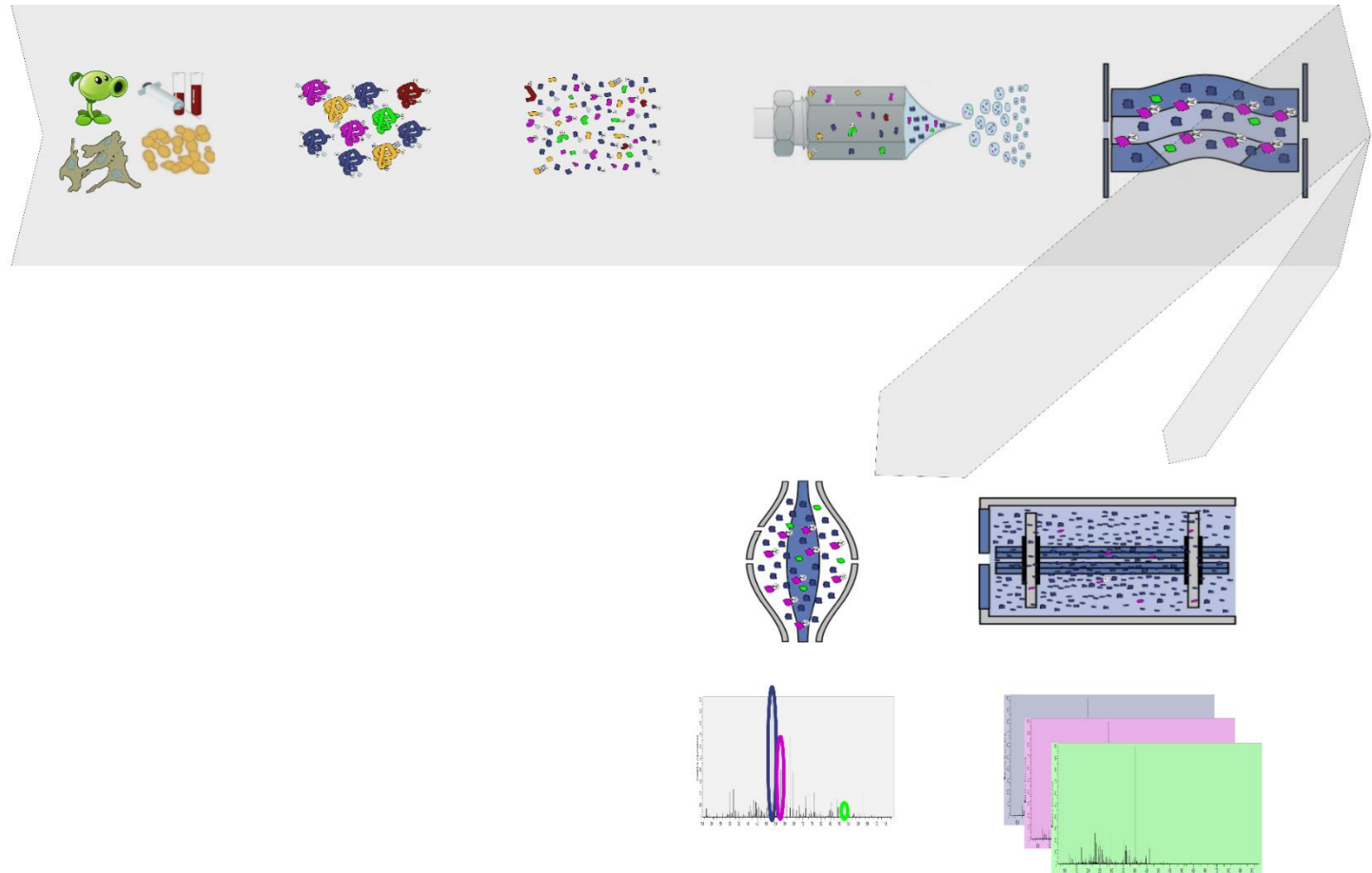
**What about now?**  
**Which one is the brightest?**



# And now?



# MS-based proteomics identifies many thousands of peptides



# How do we quantify proteins?



# How do we quantify proteins?

MSqRob: robust peptide-based model

$$y_{peptide} \sim peptide + treatment$$

peptide level

protein level

MSqRob      models peptide effects  
                 provides inference at protein level  
                 is robust against outliers

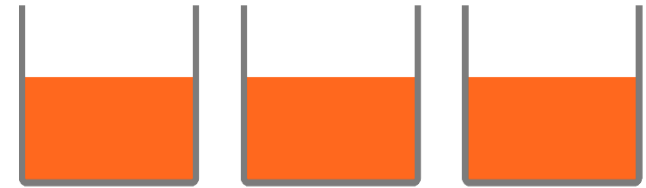
(Goeminne et al., 2016, MCP)

# How to detect proteins completely missing in one condition?

Protein X identified



Protein X not identified





# How to detect proteins completely missing in one condition?

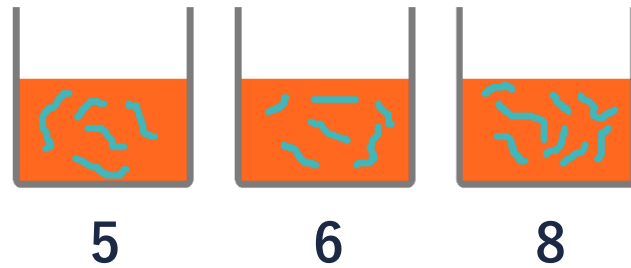
MSqRob: no estimate, theoretical fold change  $\infty$

⇒ No p-value, no FDR

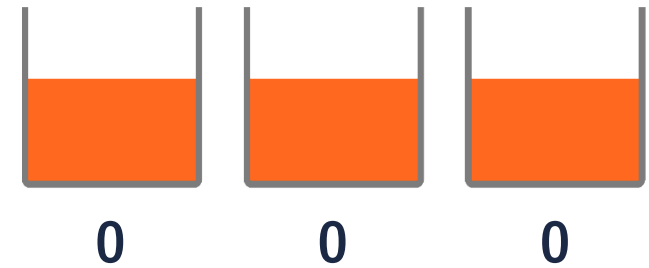
⇒ No differential abundance detected

# Peptide counting copes with zeros

Protein X identified



Protein X not identified

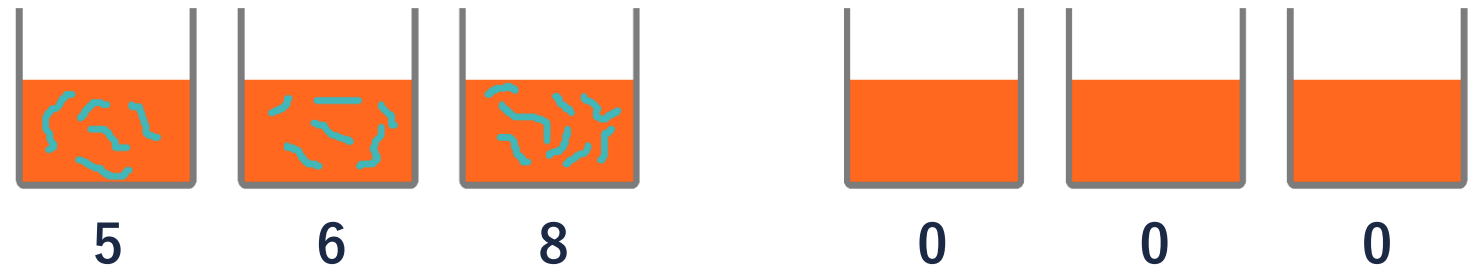


# Peptide counting copes with zeros

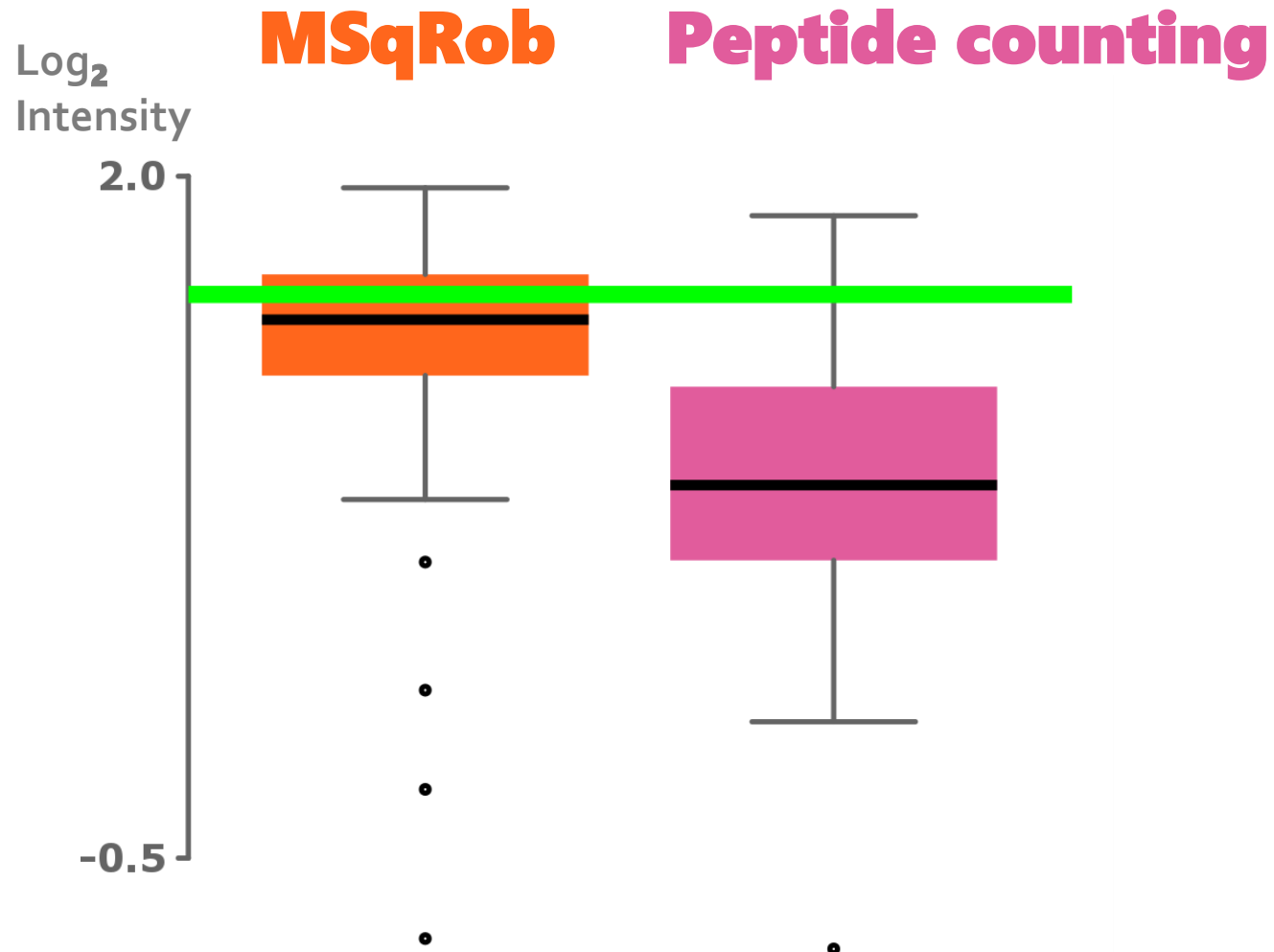
## Binomial regression

Assess probability to observe  $n_i$  peptides  
out of  $N$  identified peptides by random chance

=> p-value and FDR



# Peptide counting is less precise than intensity-based methods



# MSqRob is more precise

## Peptide counting finds missing in 1 condition

	Peptides absent	Peptides present
MSqRob	Missing	Stable intensity-based estimate
Peptide counting	Binomial estimate	Binomial estimate

# StatOmics/Proteomics: combine ideas



# Why combine both?

	Peptides absent	Peptides present
MSqRob	Missing	Stable intensity-based estimate
Peptide counting	Binomial estimate	Binomial estimate

# Why combine both?

MSqRob

Peptide counting

Peptides absent

Missing

Binomial estimate

Identification hurdle

Peptides present

Stable intensity-based estimate

Binomial estimate




# Why combine both?

## Combined statistic

	Peptides absent	Peptides present
MSqRob	<del>Mixing</del>	Stable intensity-based estimate
Peptide counting	Binomial estimate	Binomial estimate

# A hurdle model combines MSqRob's precision with the ability to detect complete absence

## Hurdle model

	Peptides absent	Peptides present
MSqRob	 Missing	Stable intensity-based estimate
Peptide counting	Binomial estimate	Binomial estimate

# The hurdle model combines precision with the ability to detect complete absence

Performance of individual vs combined statistics

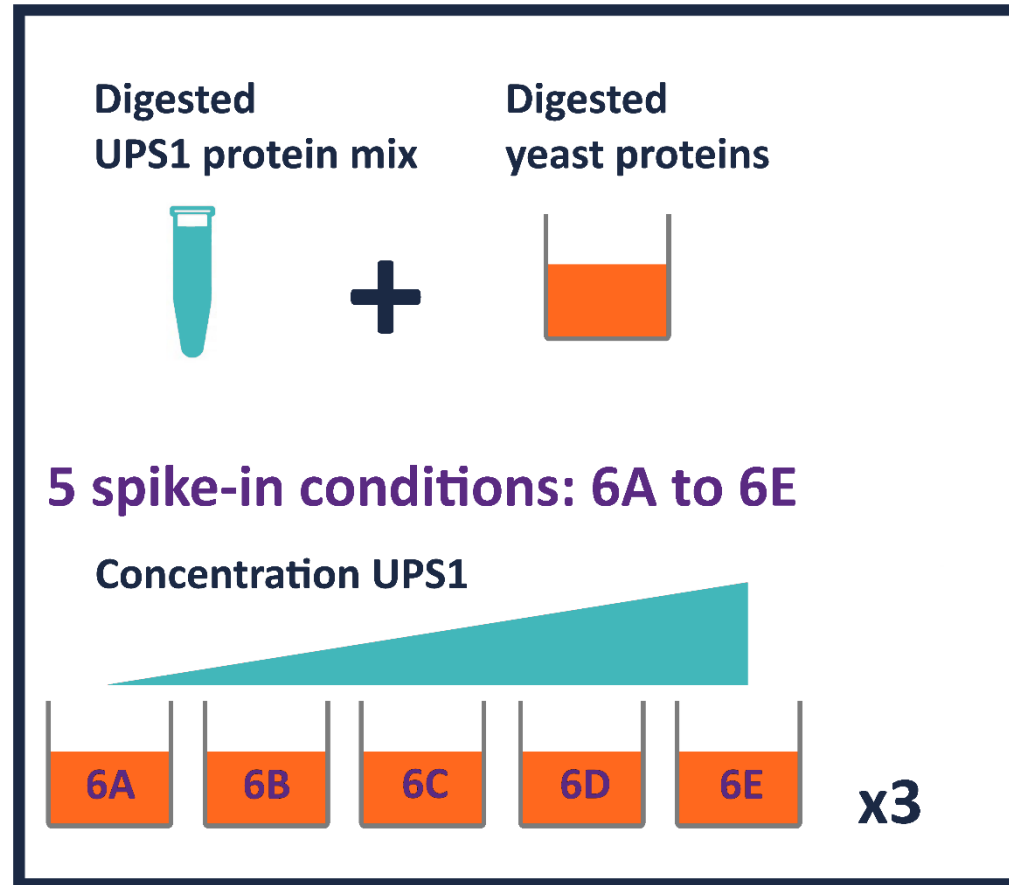
Inference in 2 stages

# The hurdle model combines precision with the ability to detect complete absence

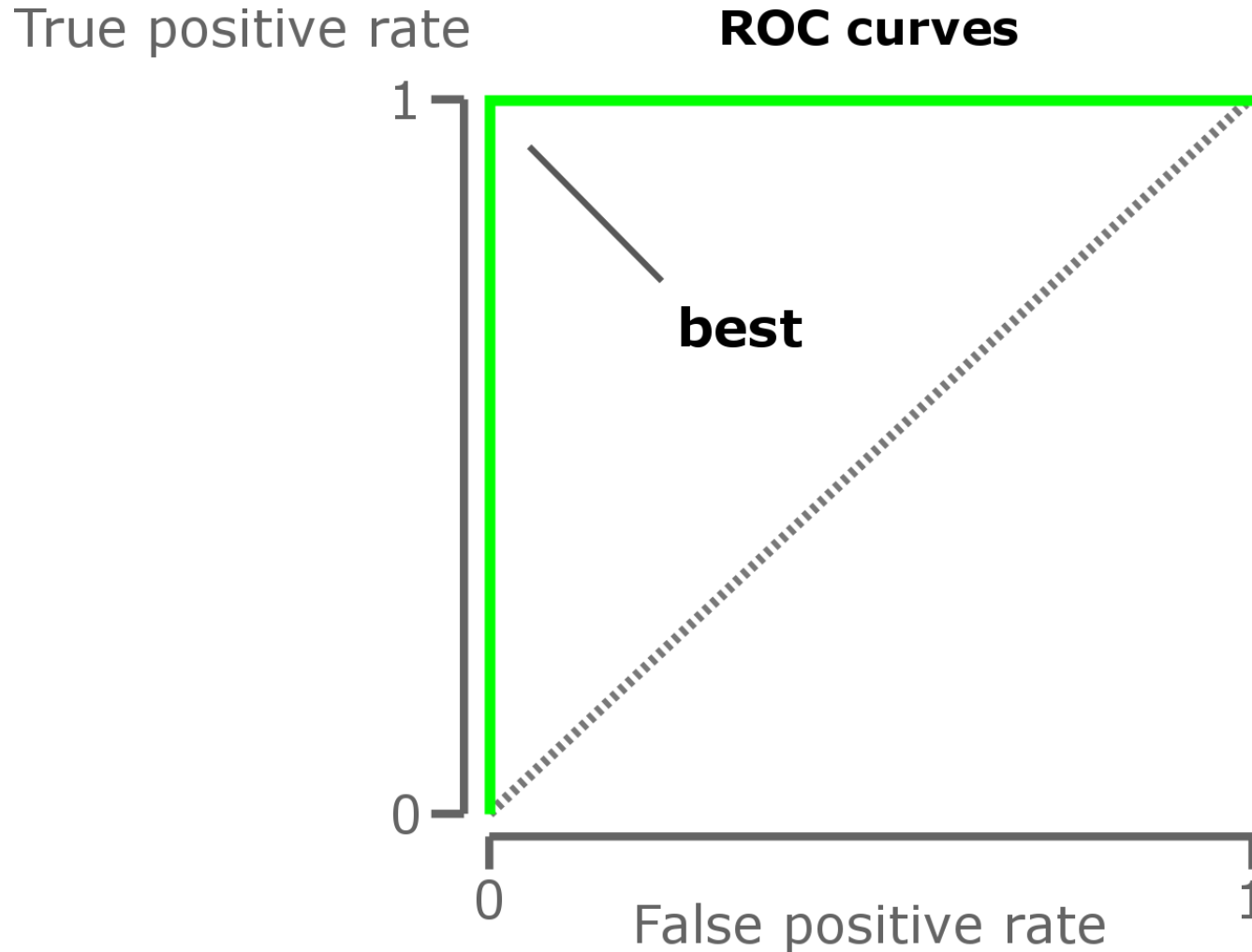
Performance of individual vs combined statistics

Inference in 2 stages

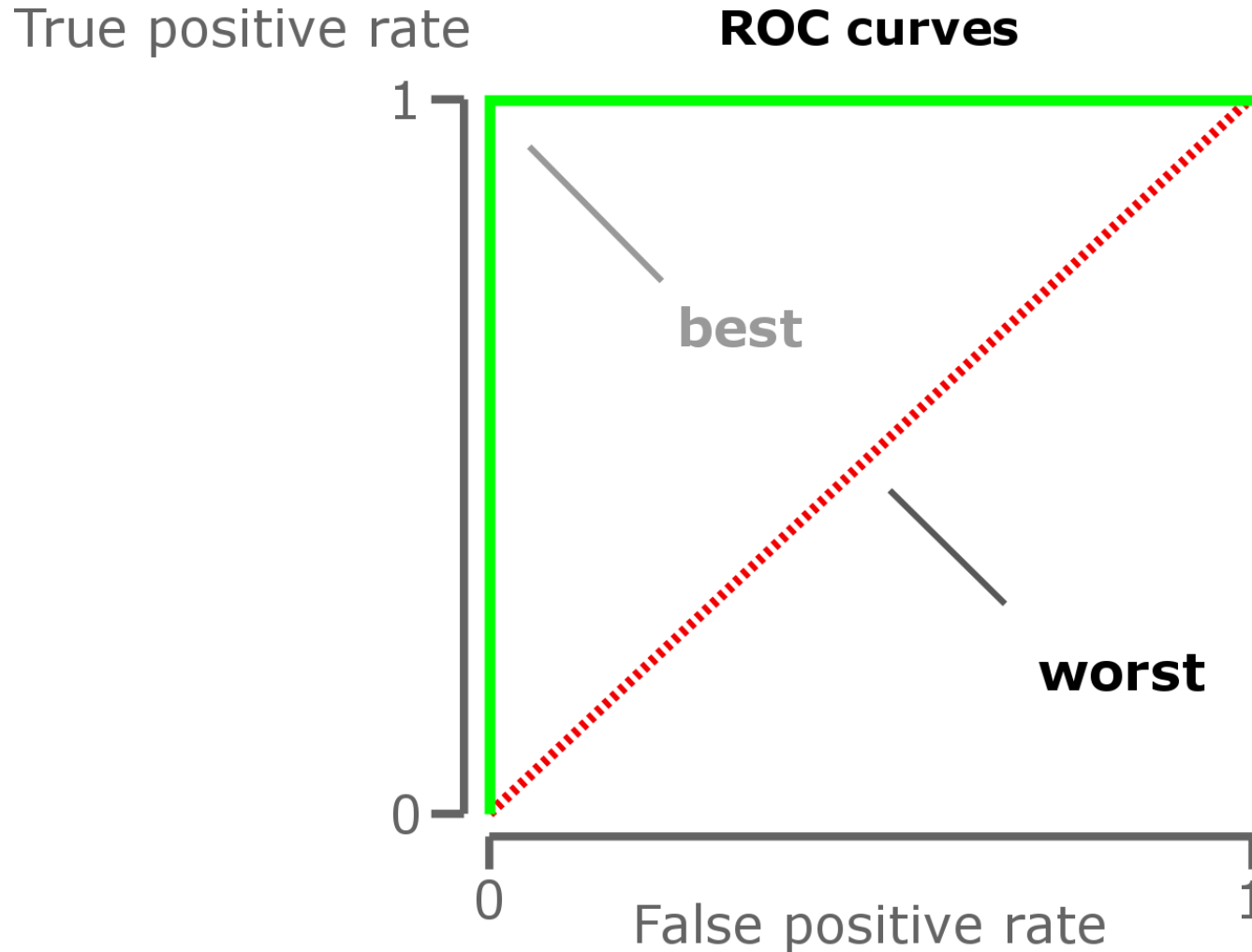
# We use a spike-in study to compare performances



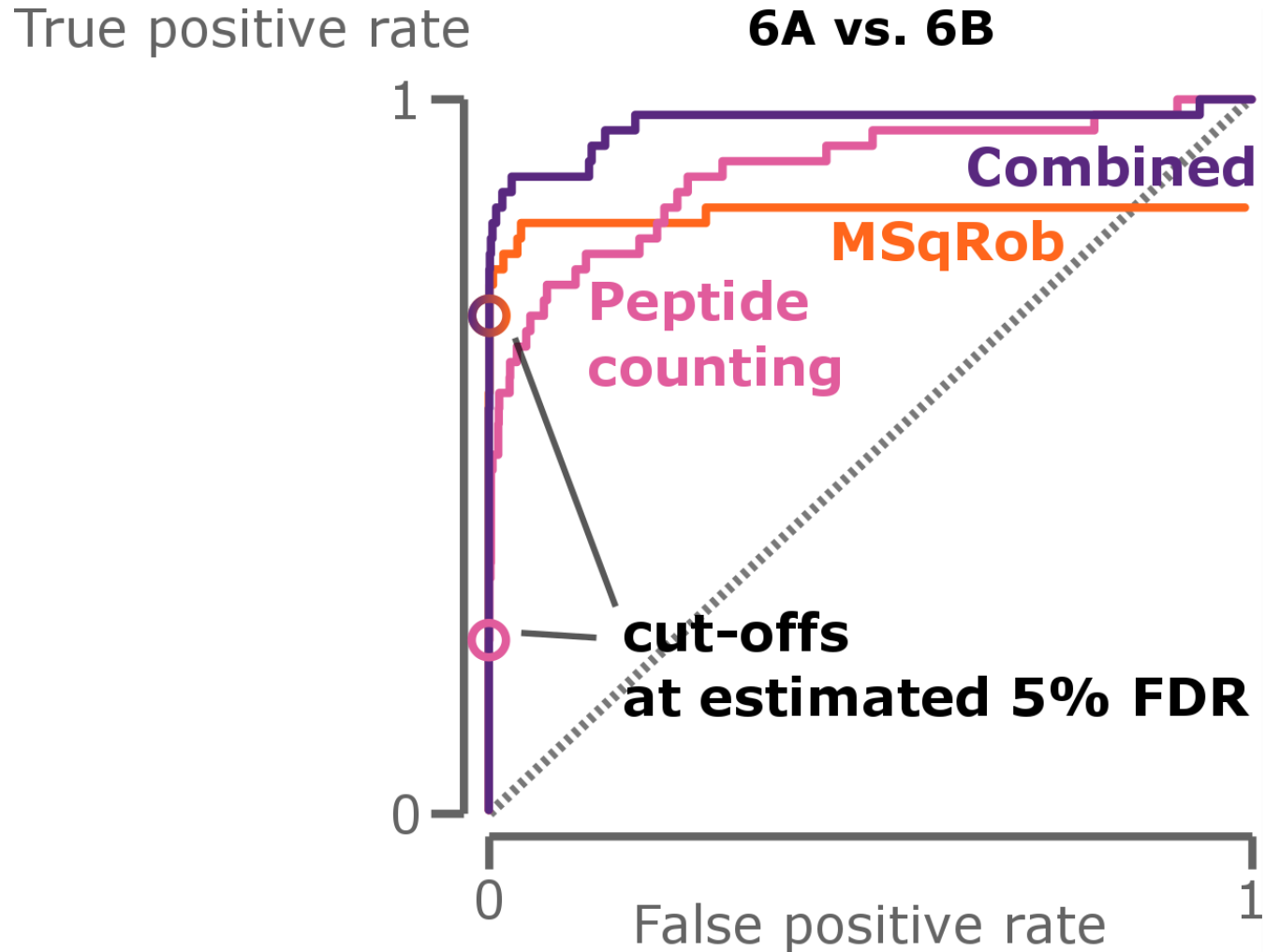
# The hurdle model outperforms MSqRob and the count-based method



# The hurdle model outperforms MSqRob and the count-based method

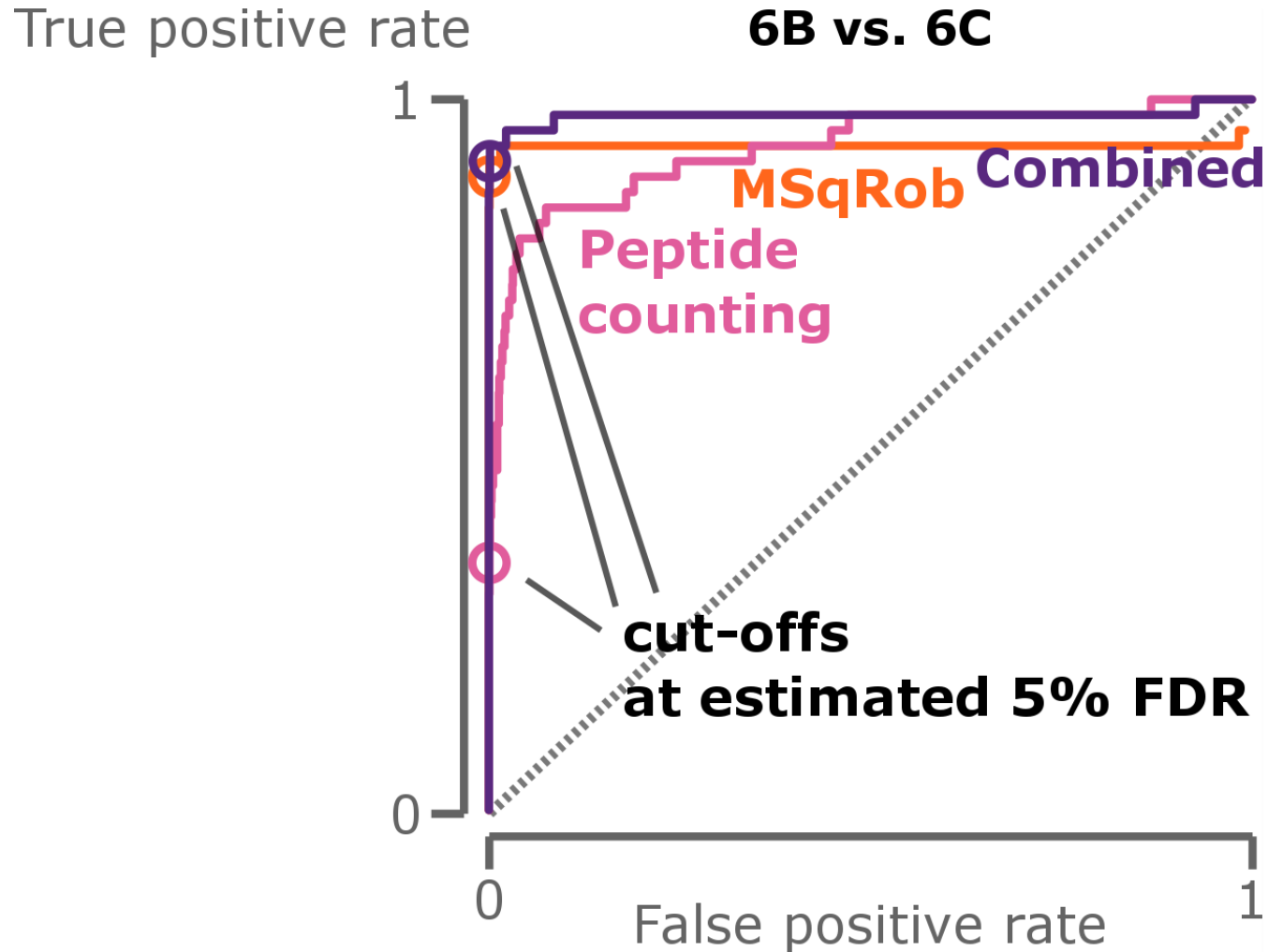


# The hurdle model outperforms MSqRob and the count-based method

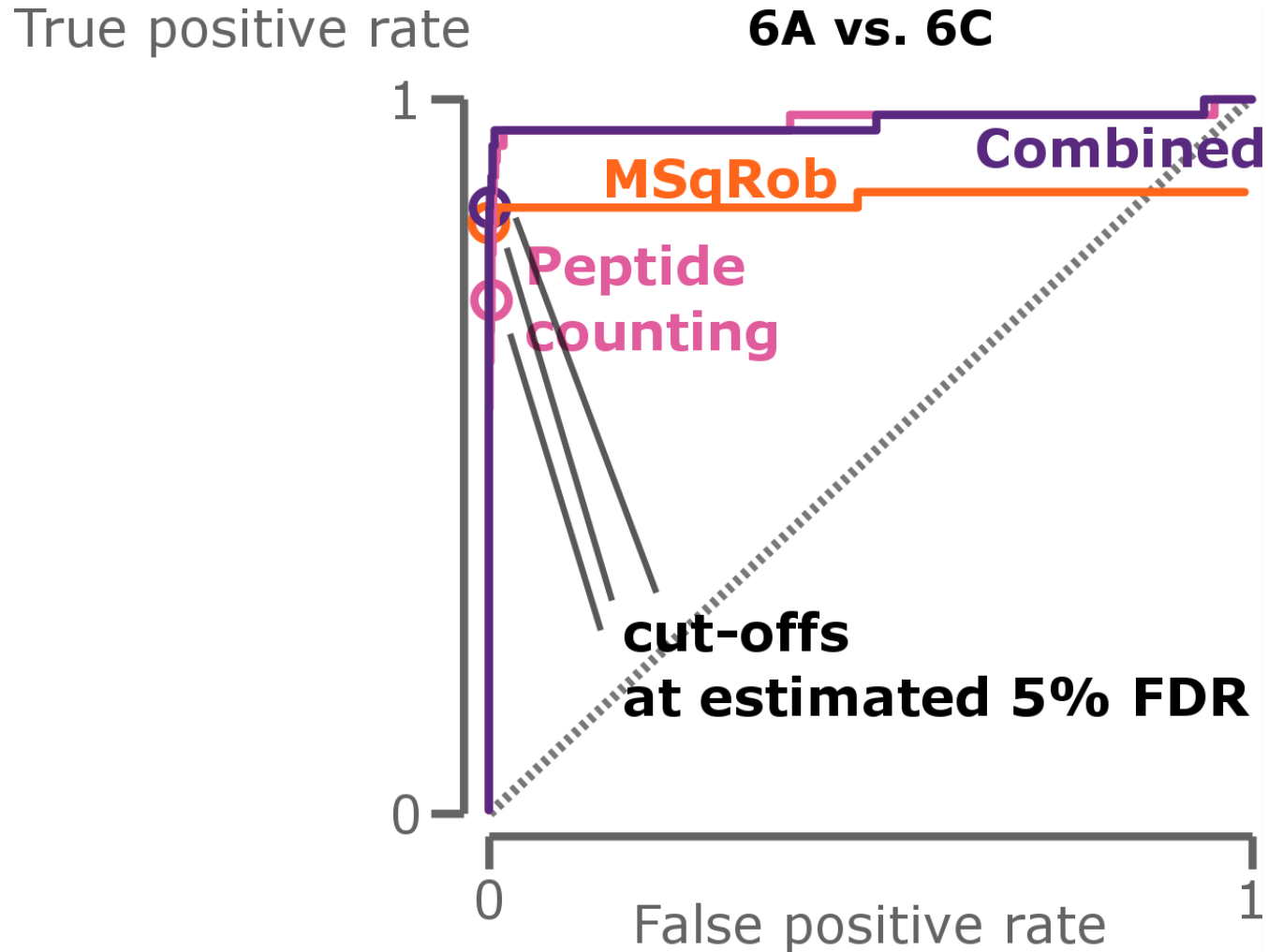




# The hurdle model outperforms MSqRob and the count-based method



# The hurdle model outperforms MSqRob and the count-based method



# The hurdle model in 2 stages answers 2 research questions

Performance of individual vs combined statistics

Inference in 2 stages

# The hurdle model in 2 stages answers 2 research questions

Performance of individual vs combined statistics

Inference in 2 stages

# The hurdle model in 2 stages answers 2 research questions

## Stage 1

Difference in intensities or number of peptides?

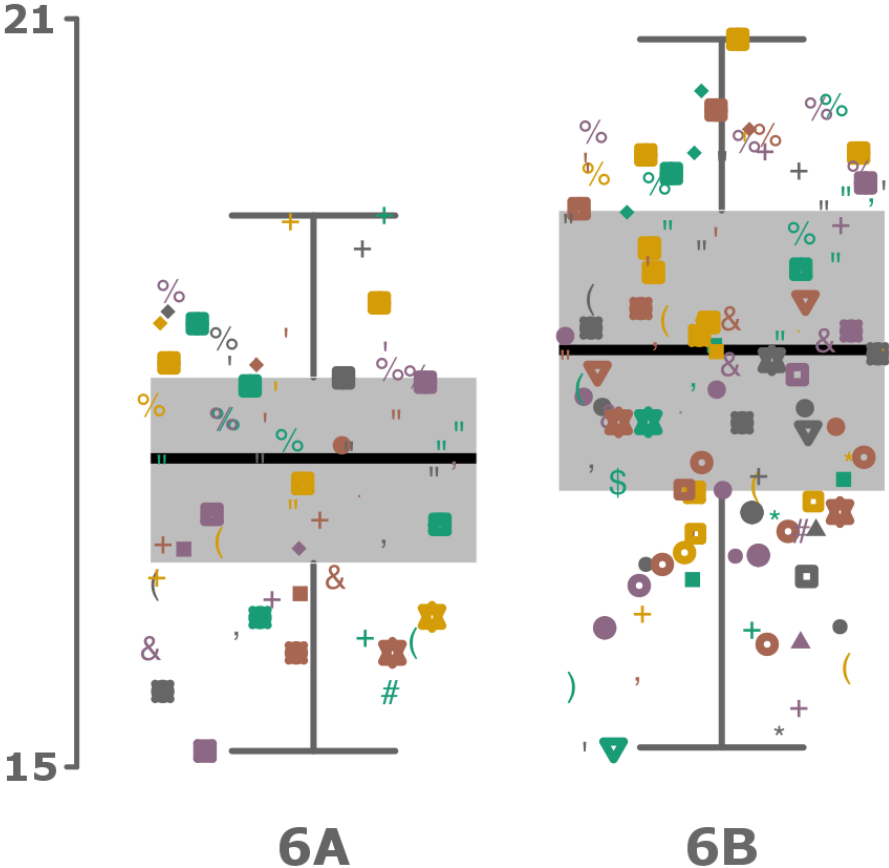
## Stage 2

- If present: test difference in intensities + difference in number of peptides
- If absent: only test difference in number of peptides

# Hurdle model detects the same proteins as MSqRob

Log2 intensity

## TAU HUMAN UPS



# Hurdle model detects proteins missing in one condition

Log2  
intensity

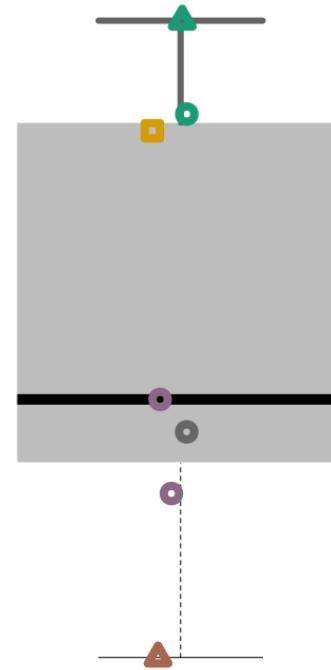
**UBIQ HUMAN UPS**

17.5

15.5

6B

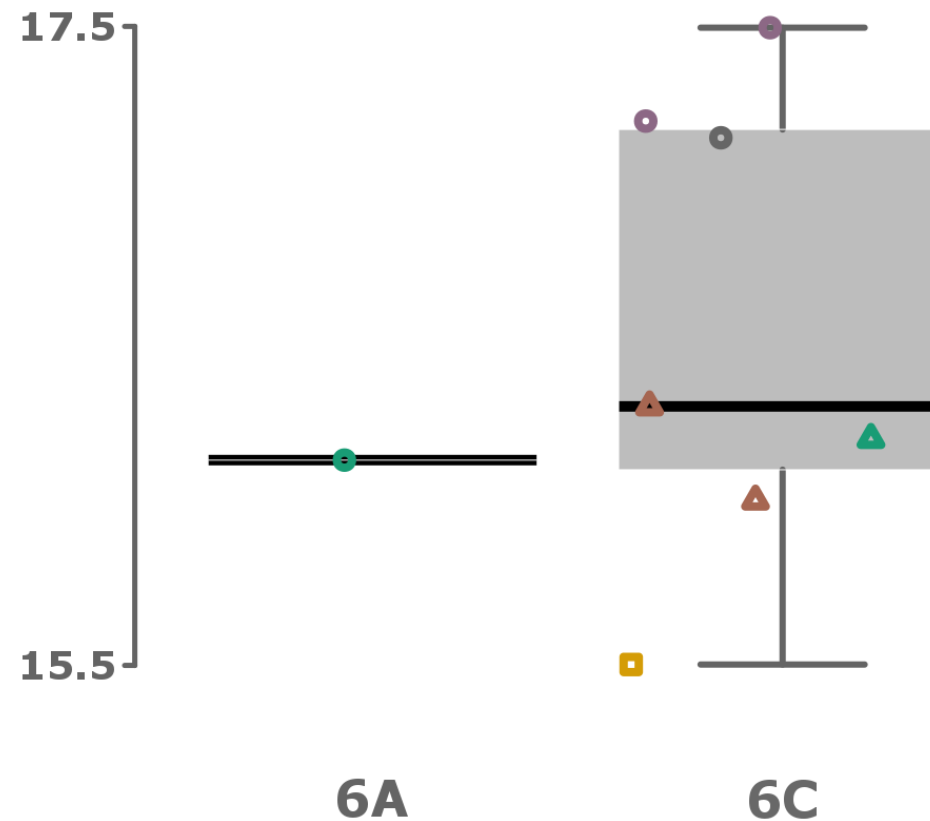
6C



# Hurdle model has increased power compared to MSqRob and binomial model

Log<sub>2</sub>  
intensity

UBIQ HUMAN UPS





# The hurdle model detects complete absence and answers 2 research questions

Performance of individual vs combined statistics

Inference in 2 stages





# [ludger.goeminne@vib-ugent.be](mailto:ludger.goeminne@vib-ugent.be)

<https://github.com/statOmics/MSqRob>

