Use of external protein standards to approach absolute quantification in yeast extracts

Aaron MILLAN-OROPEZA
13/07/2018

Context of study: European project CHASSY

Consortium between nine academy and industry partners aiming to develop yeast as microbial factories using **systems and synthetic biology** (omics) to produce high-added value products of cosmetic and nutraceutic sectors.

http://chassy.eu/

Context of study: European project CHASSY

Consortium between nine academy and industry partners aiming to develop yeast as microbial factories using **systems and synthetic biology** (omics) to produce high-added value products of cosmetic and nutraceutic sectors.
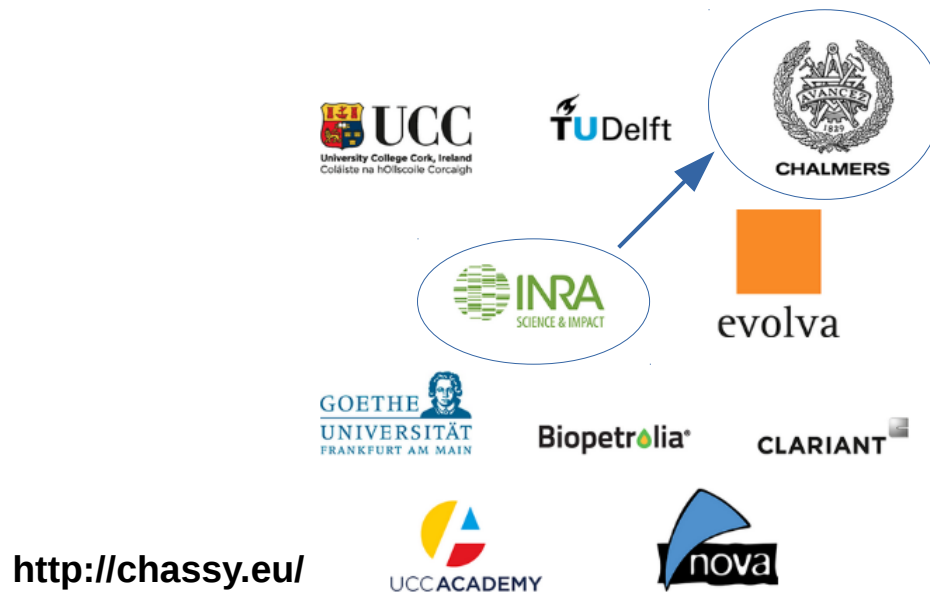
Ivan Domenzain
PhS student at Chalmer UT
in charge of modeling

http://chassy.eu/

Need of absolute quantification in large proteomic datasets with reliable accuracy for **modeling at a genome-scale level**

Relative and absolute quantification by MS

Quantification in proteomics relies on the ability to detect small changes in protein and peptide abundance in response to different conditions.

Relative quantification compares the levels of a specific protein in different conditions with results being expressed in relative units.

Absolute quantification is the determination of the exact amount or mass concentration of a protein (ex. fmol).

Wasinger *et al. (*2013) Int J Proteomics 180605.

Universal Proteomic Standars (UPS) for "absolute" quantification

UPS2 is a mixture of 48 human proteins distributed in 6 orders of magnitude well-known to be difficult to analyse.

- 50 pmoles (10^{-12}) – 500 amoles (10^{-18})
- 6 groups (different quantity) of 8 proteins (different molecular weight)
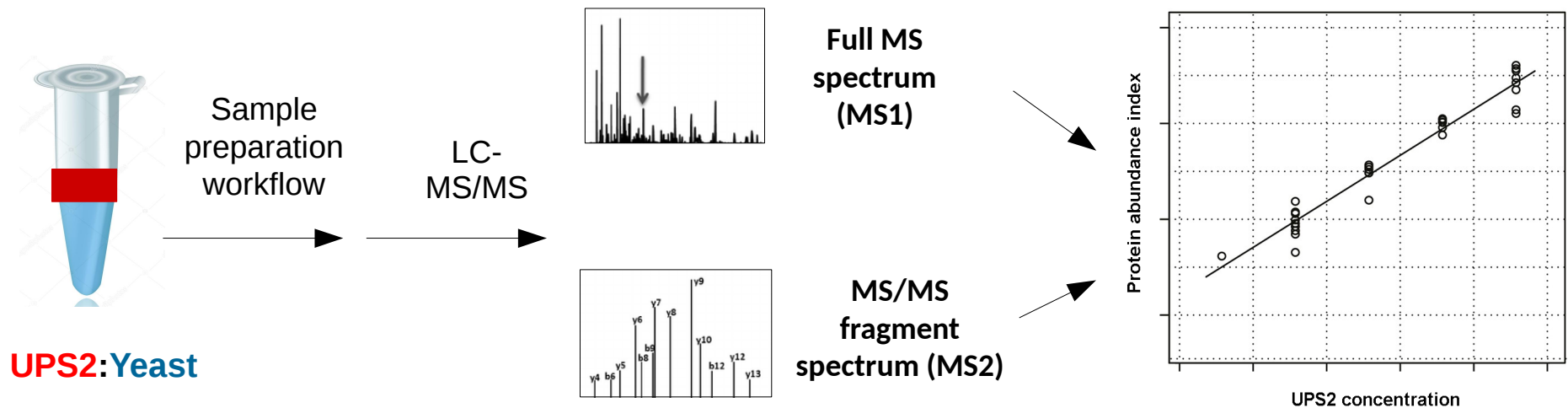
Universal Proteomic Standars (UPS) for "absolute" quantification

UPS2 is a mixture of 48 human proteins distributed in 6 orders of magnitude well-known to be difficult to analyze.

- 50 pmoles (10^{-12}) – 500 amoles (10^{-18})
- 6 groups (different quantity) of 8 proteins (different molecular weight)
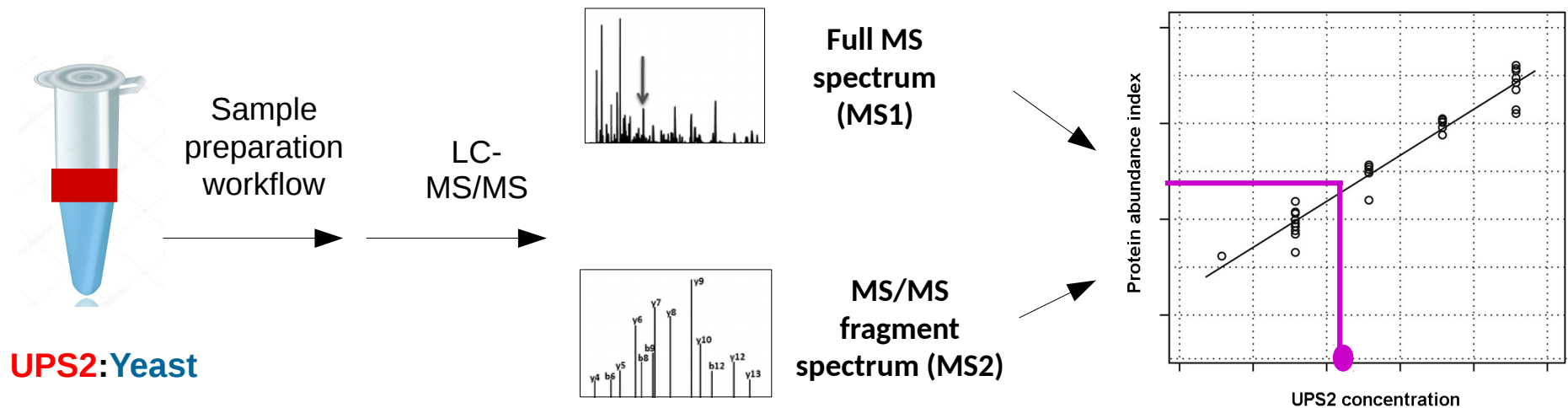
The regression curve between protein **quantification indexes** (MS1 or MS2) and **UPS proteins** concentrations will be used to calculate the **absolute values** of yeast to generate suitable datasets for genome-scale modeling

Not all the UPS2 proteins can be detected

☹ **High costs** of standard 10 µg / 580-750 €

☹ Limited to a small number of samples per experiment due to high amounts of UPS2 spiked
  (**1-4 µg UPS2** per sample)

☹ A maximum of **4 orders of magnitude (32 proteins)** have been detected
  (Soufi et al. 2015, Front Microbiol)

☹ Ambiguity of details concerning the spiking amounts

☹ First tests showed difficulty to detect UPS proteins
  in complex mixtures (14 proteins, 2 orders of magnitude)

Not all the UPS2 proteins can be detected

☹ **High costs** of standard 10 μg / 580-750 €

☹ Limited to a small number of samples per experiment due to high amounts of UPS2 spiked
(**1-4 μg UPS2** per sample)

☹ A maximum of **4 orders of magnitude (32 proteins)** have been detected
(Soufi et al. 2015, Front Microbiol)

☹ Ambiguity of details concerning the spiking amounts

☹ First tests showed issues to detect UPS proteins
in complex mixtures (**14 proteins, 2 orders of magnitude**)

Log10 (NSAF) vs Log10 (fmol)

28 UPS proteins
$R^2 = 0.727$
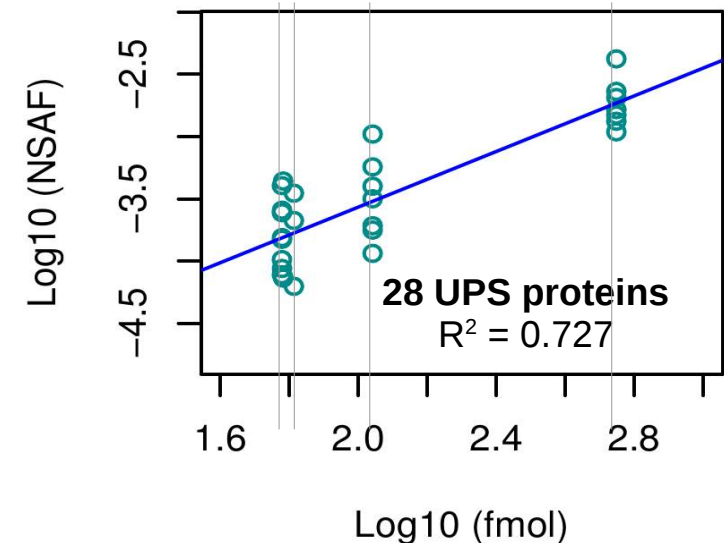
By adding **UPS1**, standard with the **same 48 proteins** at equimolar concentrations, we were able to generate 2 new orders of magnitude in a complex sample

Sample preparation : spiked samples

UPS1 (6 μg)   UPS2 (10.6 μg)   Yeast sample x 15 samples
(5 conditions x 3 replicates)

Reduction, Alkylation, Digestion (Lys-C and Trypsin, 1:30)

127 ng UPS1 → 1 : 6 (UPS:yeast)
382 ng UPS2 → 1 : 2 (UPS:yeast)
750 ng Yeast

1 : 1.5 (UPS:yeast)

Sample preparation : spiked samples

UPS1 (6 µg)   UPS2 (10.6 µg)   Yeast sample x 15 samples
(5 conditions x 3 replicates)

Each yeast sample was **spiked** with UPS1 and UPS2 after digestion

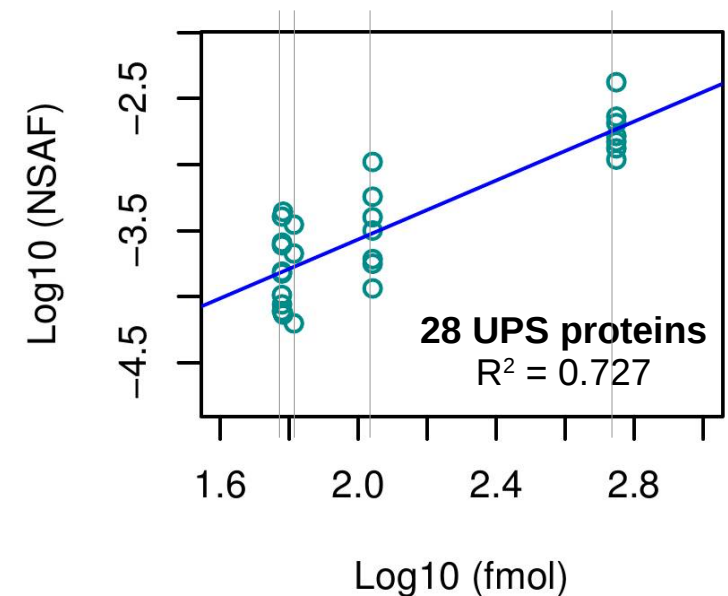Reduction, Alkylation, Digestion (Lys-C and Trypsin, 1:30)

LC-MS/MS

Orbitrap Fusion™ Lumos™ Tribrid™
50-cm column
216 min run

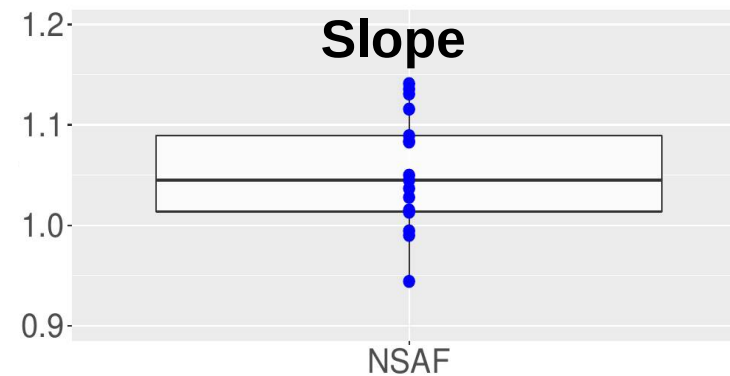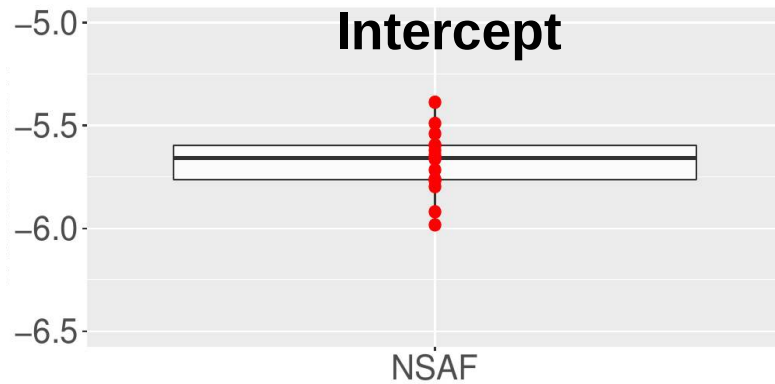**127 ng UPS1** → 1 : 6 (UPS:yeast)
382 ng UPS2 → 1 : 2 (UPS:yeast)
750 ng Yeast

1 : 1.5 (UPS:yeast)
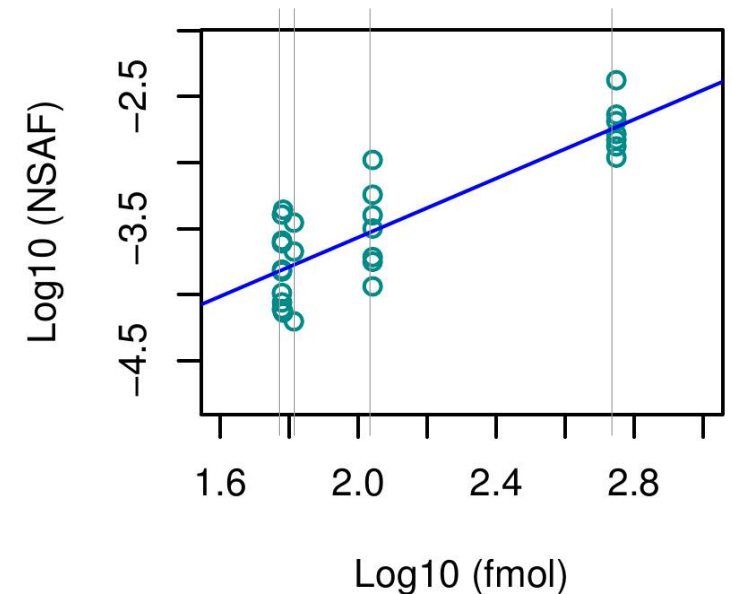
28 UPS proteins
$R^2 = 0.727$

Log10 (NSAF)

Log10 (fmol)

Spiked samples showed similar linear regressions

Despite the dispersion of proteins on each order of magnitude, the slopes and intercepts were consistent in the 15 samples
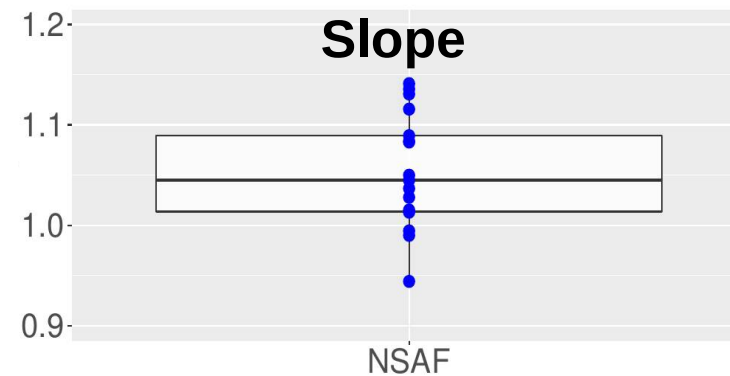
n = 15 samples

Spiked samples showed similar linear regressions
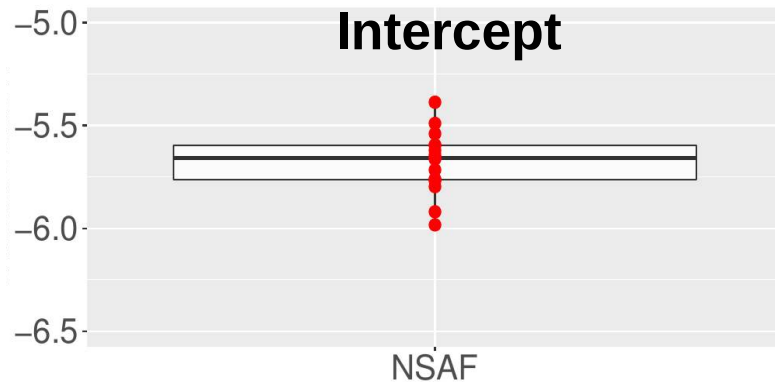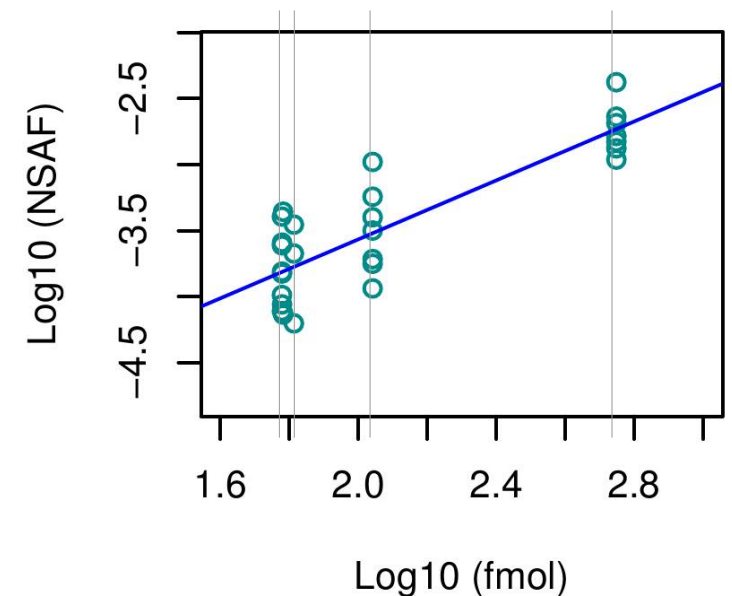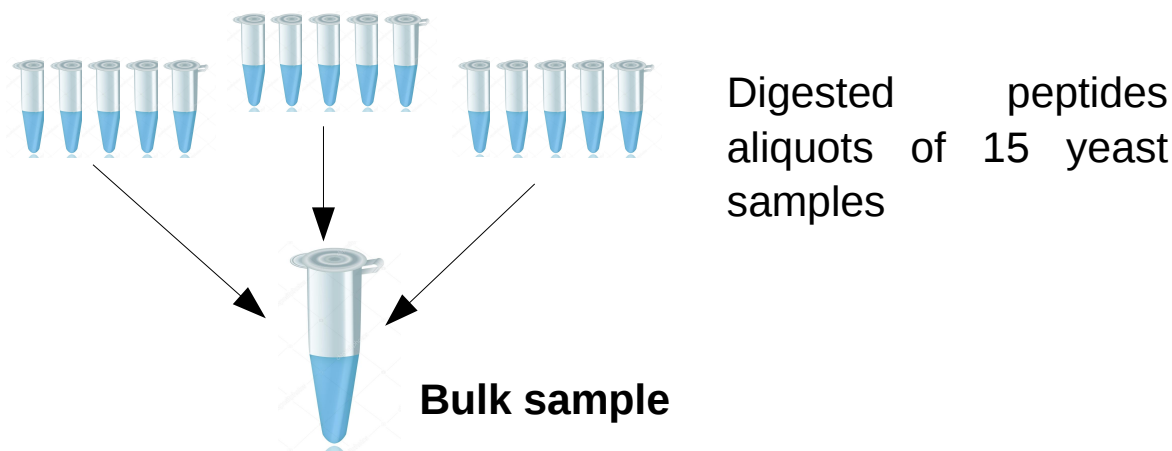

Despite the dispersion of proteins on each order of magnitude, the slopes and intercepts were consistent in the 15 samples

n = 15 samples

Can we use a **representative sample** during the proteomics analysis ?

Preparation of representative sample: bulk

Digested peptides aliquots of 15 yeast samples

Bulk sample

127 ng UPS1 → 1 : 6 (UPS:yeast)
382 ng UPS2 → 1 : 2 (UPS:yeast)
750 ng Bulk

1 : 1.5 (UPS:yeast)

LC-MS/MS

Orbitrap Fusion™ Lumos™ Tribrid™
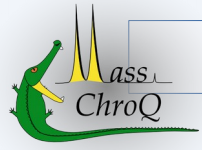50-cm column
216 min run

The costs of UPS standards are reduced using a bulk sample

Bioinformatic workflow

X!Tandempipeline

2739 yeast proteins
31 UPS proteins

MassChroQ

Data cleaning

Select **UPS** proteins

(5) IBAQ, (6) Top3

Linear regression

XIC

Select **UPS** proteins

**(1) PAI, (2) emPAI,
(3) Spectral Count,
(4) NSAF**

Linear regression

Abundance inference on
yeast proteins

Spectral count

Bioinformatic workflow

X!Tandempipeline

2739 yeast proteins
31 UPS proteins

Extracted Ion current-based data

MassChroQ

↓

Data cleaning

↓

Select **UPS** proteins

↓

(5) IBAQ, (6) Top3

↓

Linear regression

XIC

Spectral count-based data

Select **UPS** proteins

↓

**(1) PAI, (2) emPAI,
(3) Spectral Count,
(4) NSAF**

↓

Linear regression

↓

Abundance inference on yeast proteins
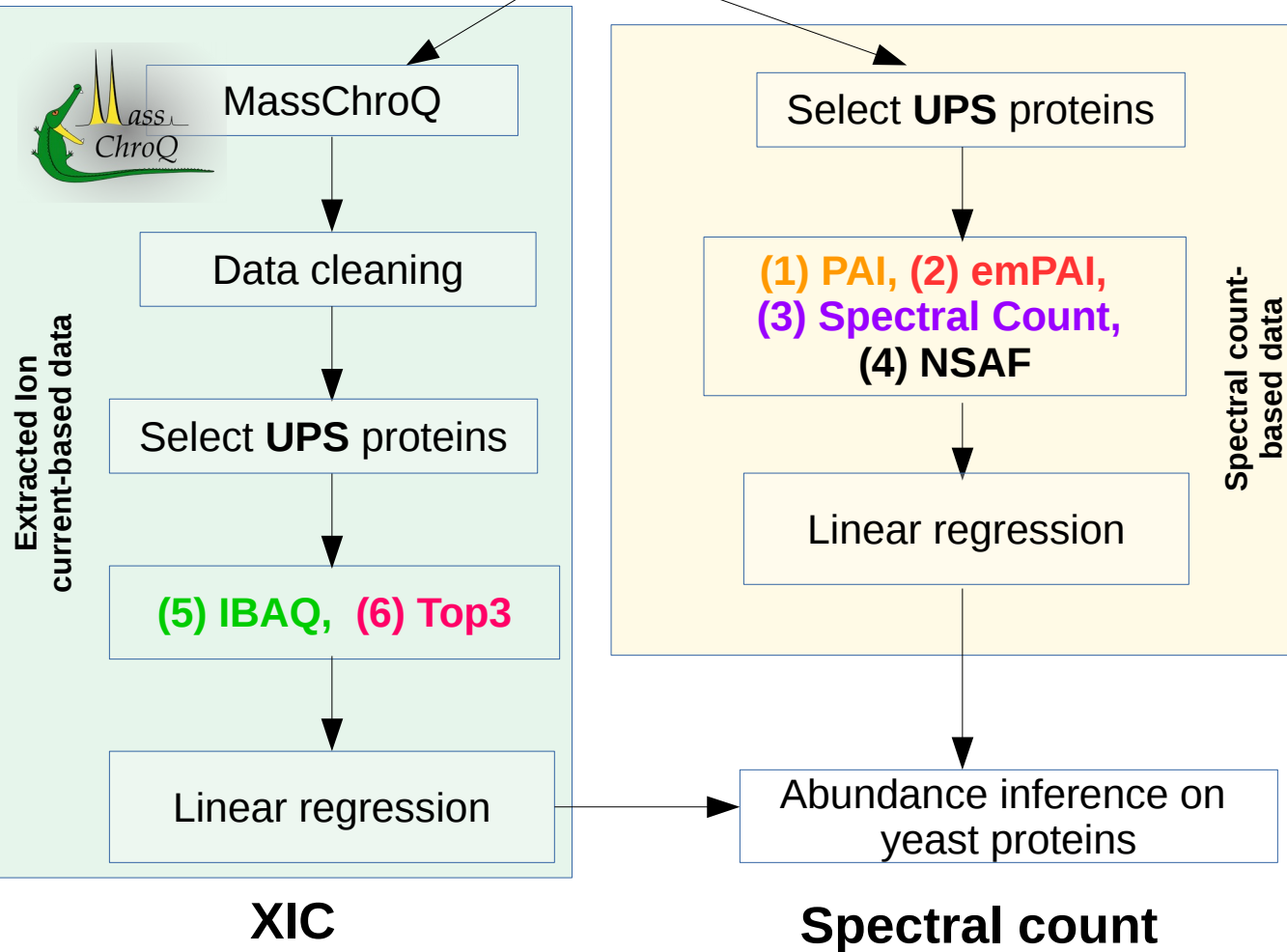
Spectral count

PAI: Observed peptides dived by the number of theoretically observable peptides

emPAI: 10^{PAI} -1

**Spectral Count (SC):** Number of MS2 spectra assigned to a protein

NSAF: $\dfrac{SC/L}{\sum (SC/L)}$

L= length of protein in aa

IBAQ: Sum of MS1 intensities of a protein divided by the number of theoretically observable peptides

Top3: Average of the 3 most intense MS1 intensities of a protein

Aiming to find the best protein abundance index

Variation of different abundance indexes

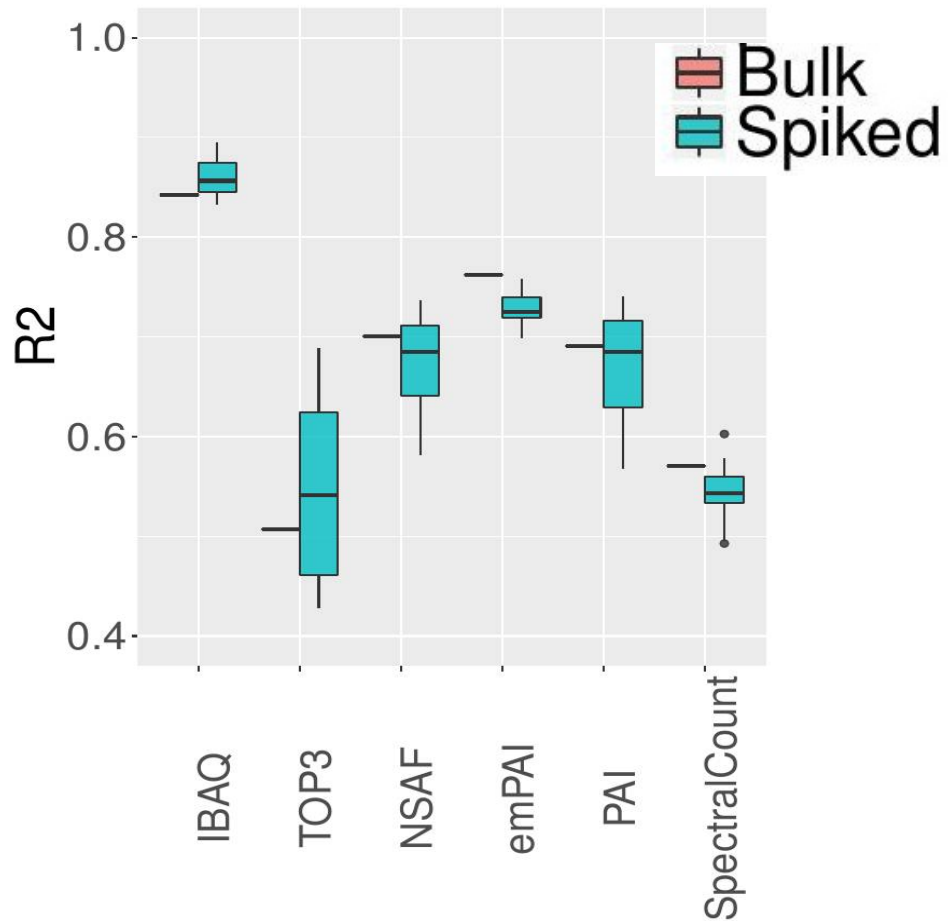Test conducted on the 31 UPS proteins
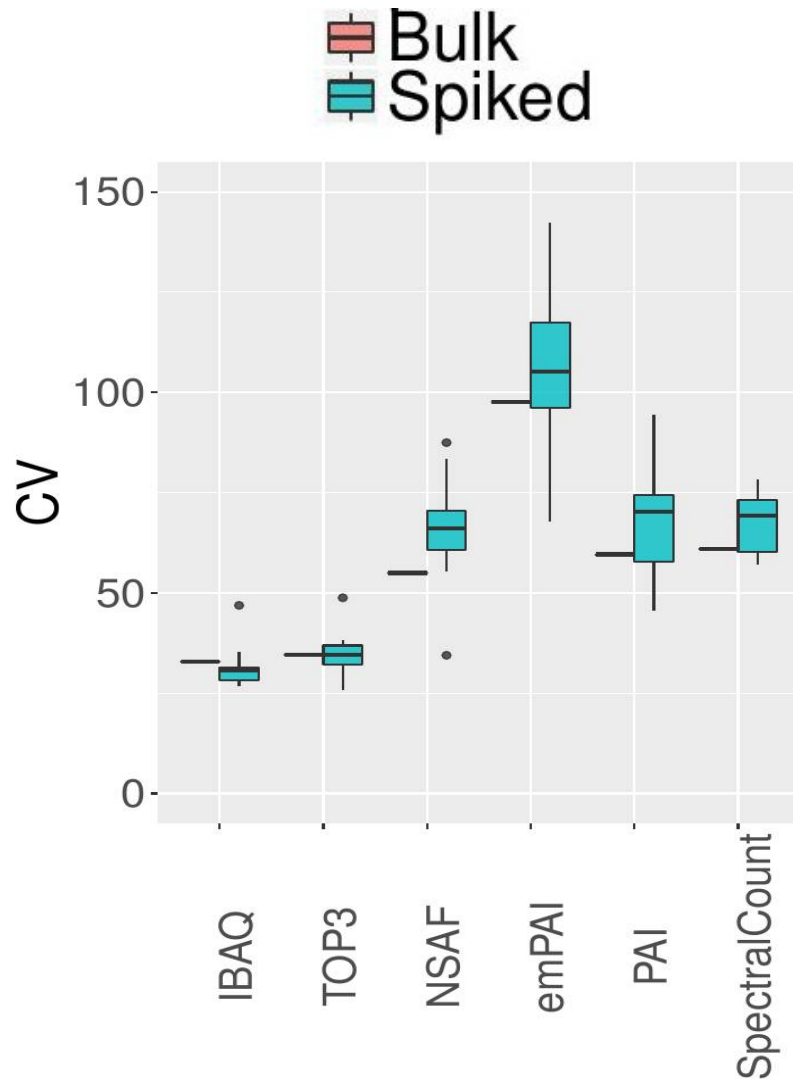
- Overall, the method **IBAQ** showed **the best coefficient of variation (CV)**.

- On every index **bulk** sampling showed a good coefficient of variation compared to spiked samples.
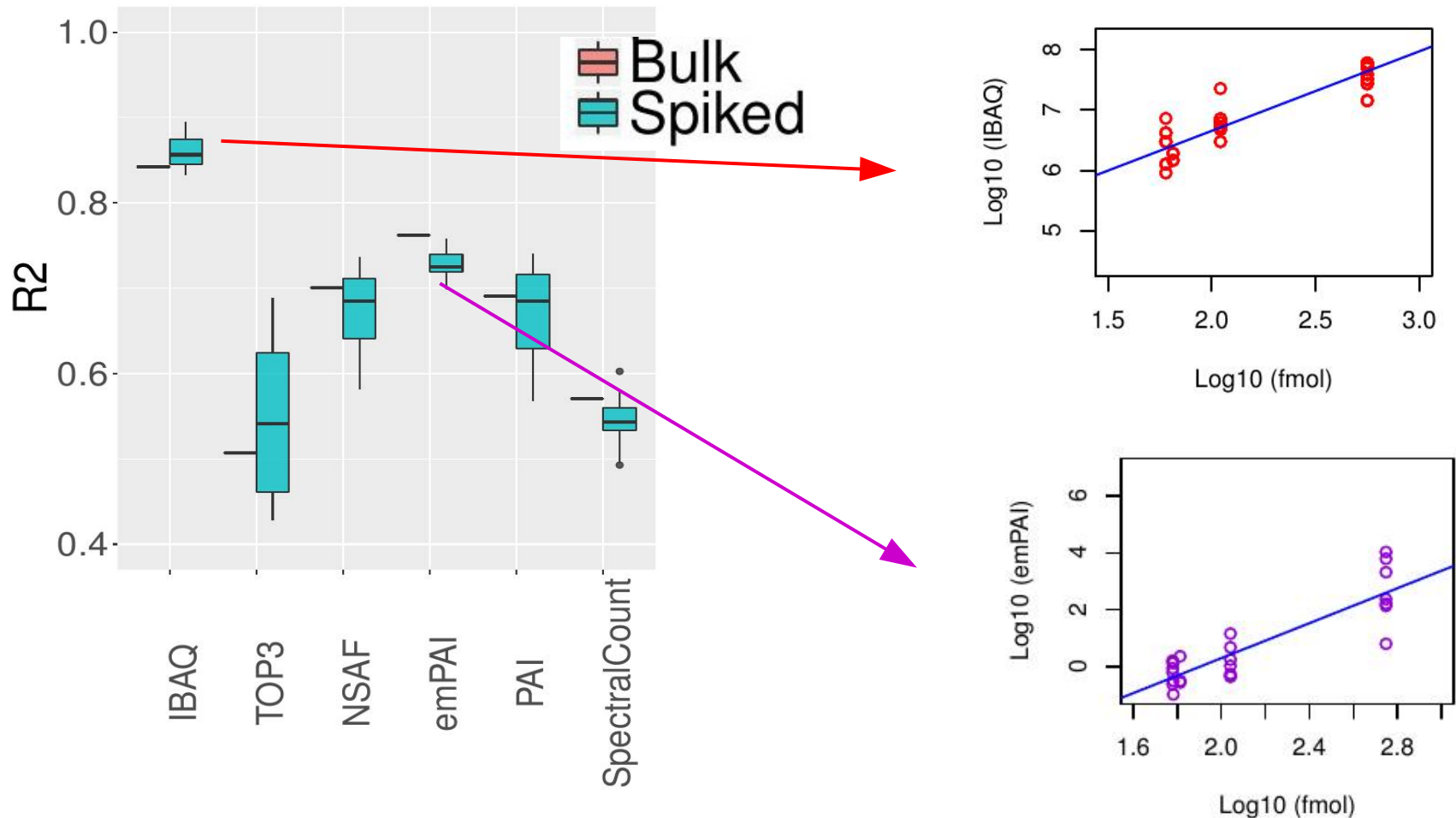
Coefficient of determination between the different abundance indexes

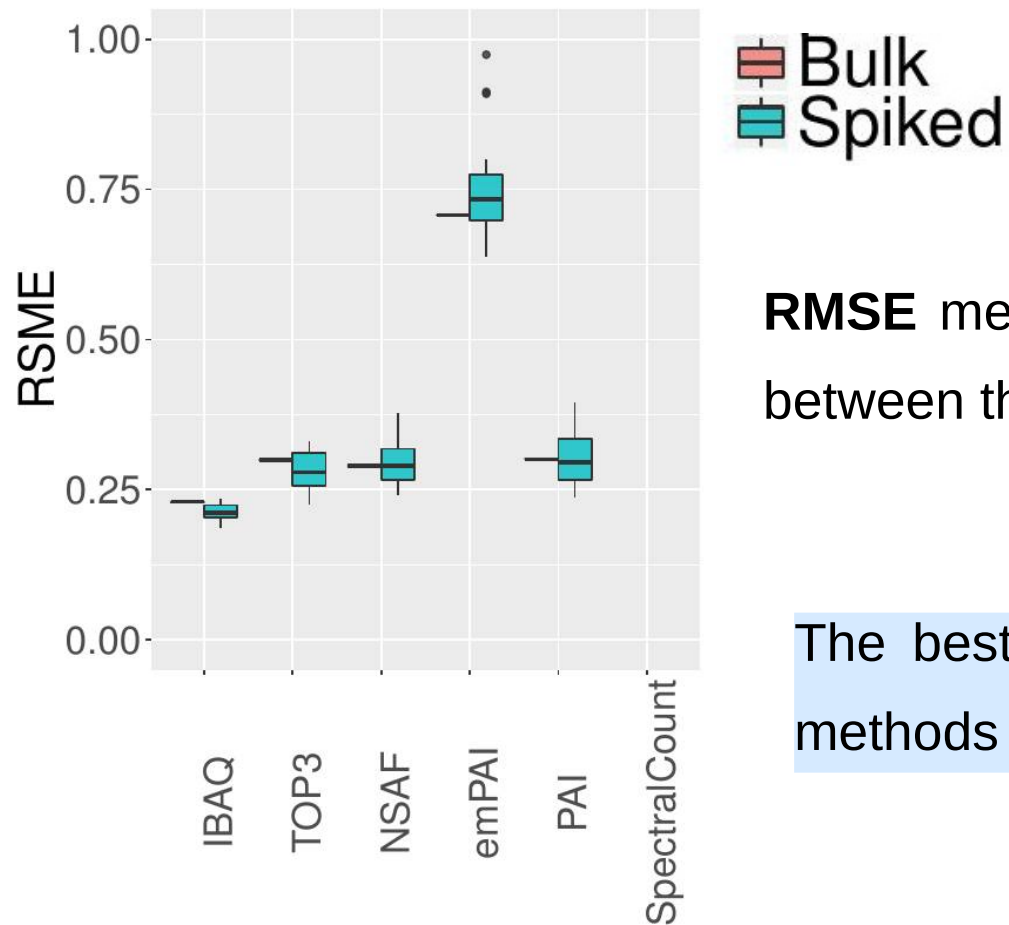Test conducted on the 31 UPS proteins

Coefficient of determination between the different abundance indexes

Test conducted on the 31 UPS proteins

- The method **IBAQ** showed the best linearity $R^2 = 0.89$

Goodness of the fit between the different abundance indexes

RMSE measures measures the quality of the fit between the actual data and the predicted model.

The best godness of fit were obtained by the methods **TOP3**, **NSAF** and **IBAQ**.

Test conducted on the 31 UPS proteins

Selection the abundance index to compute absolute quantification on yeast

Parameter	Indexes ordered by performance
CV median	Top3, **IBAQ**, **NSAF**, PAI, Spectral count, emPAI
R^2	**IBAQ**, emPAI, **NSAF**, PAI, Top3, Spectral count
RMSE	Top3, **IBAQ**, **NSAF**, PAI, emPAI, Spectral count

NSAF : 2739 proteins (ALL !)

IBAQ :1776 proteins

Top3 : 709 proteins

Selection the abundance index to compute absolute quantification on yeast

Parameter	Indexes ordered by performance
CV median	Top3, **IBAQ**, **NSAF**, PAI, Spectral count, emPAI
R^2	**IBAQ**, emPAI, **NSAF**, PAI, Top3, Spectral count
RMSE	Top3, **IBAQ**, **NSAF**, PAI, emPAI, Spectral count

NSAF : 2739 proteins (ALL !)

IBAQ :1776 proteins

Top3 : 709 proteins

IBAQ: Sum of MS1 intensities of a protein divided by the number of theoretically observable peptides

NSAF: $\dfrac{SC/L}{\sum(SC/L)}$

L= length of protein in aa

Validation of absolute values (in progress)

**Yeast proteins**
8 lab proteins
(known concentrations)

Sample digestion

Yeast + **8 lab proteins**
+ **UPS proteins**

Aiming to compare the estimated values against the theoretical values

Validation of absolute values (in progress)

Yeast proteins
8 lab proteins
(known concentrations)

Sample digestion

Yeast + **8 lab proteins**
+ **UPS proteins**

Yeast proteins

8 lab proteins
(known concentrations)

Sample digestion

Yeast + **8 lab proteins** + **UPS proteins**
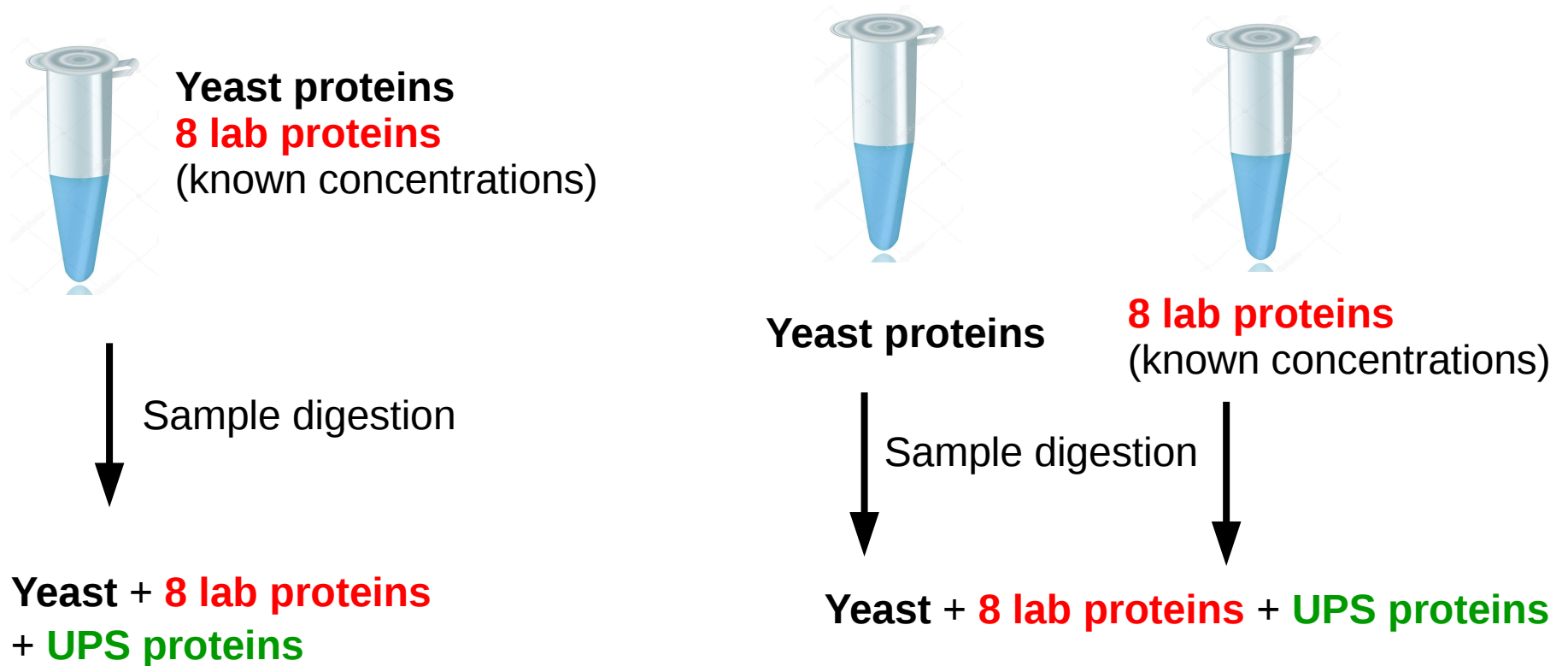
Aiming to compare the estimated values against the theoretical values

Conclusion

- Considering the performance of the 6 indexes we propose the use of **NSAF** and **IBAQ**

- The use of **bulk** for linear regression represent a solution to the high costs of UPS standards.

- The UPS standards are a good choice to perform absolute quantification suitable for modeling (4 orders of magnitude, 31/48 UPS proteins in a complex mixture)
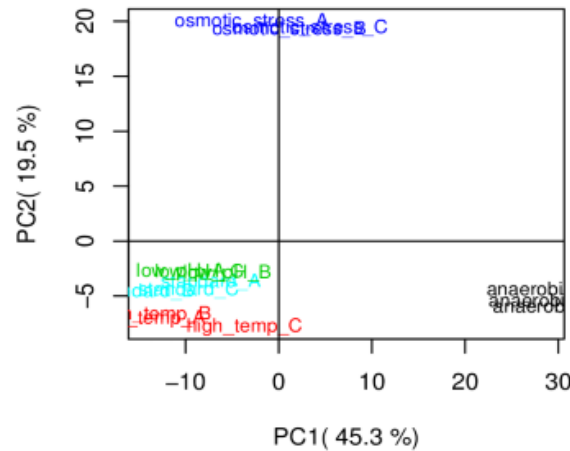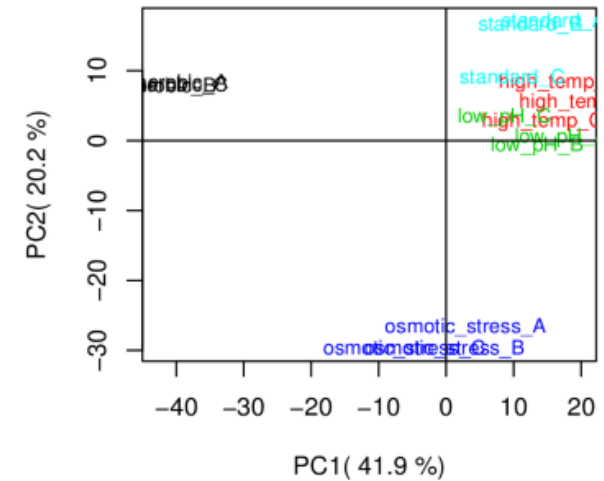
Thank you
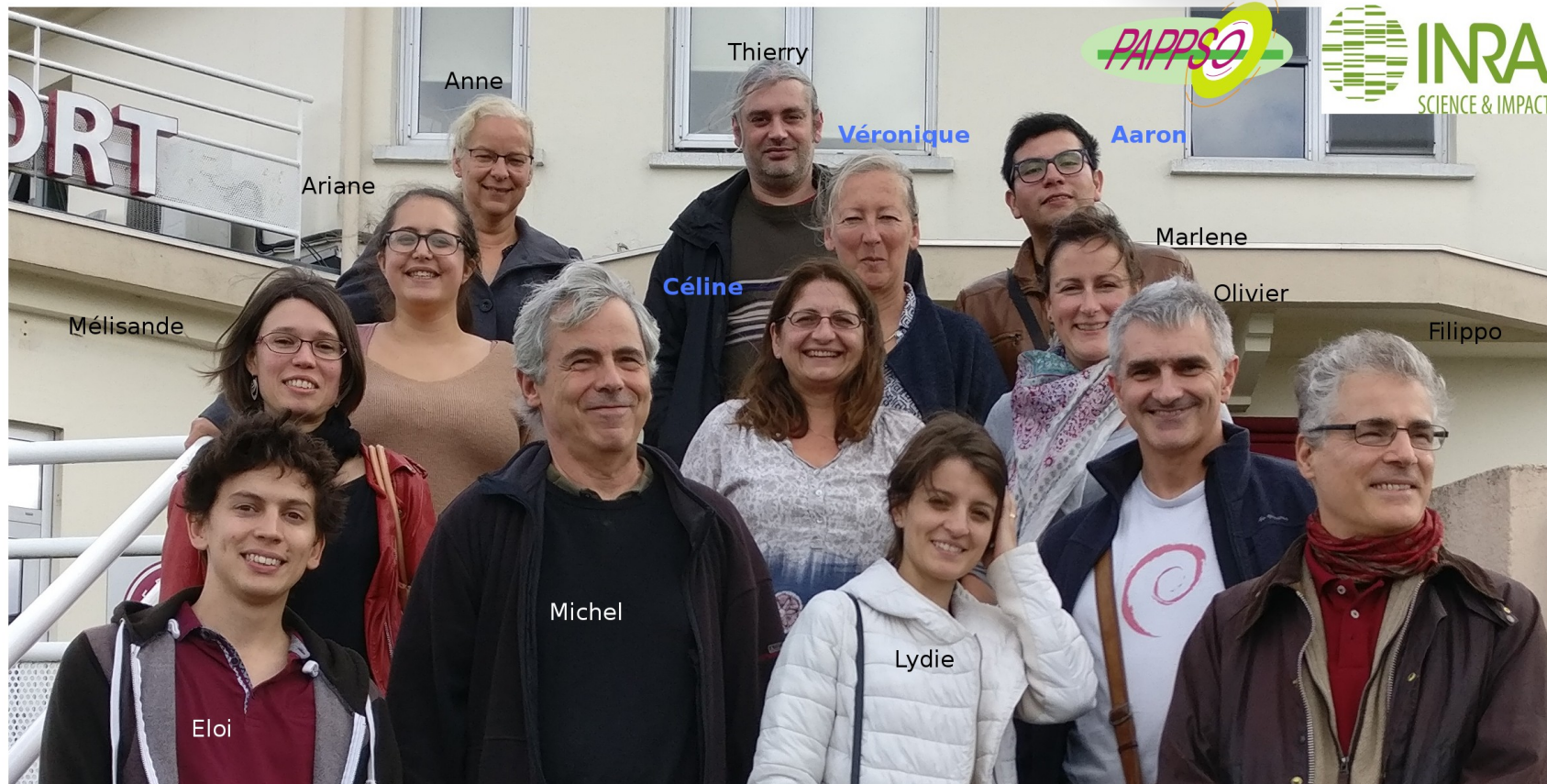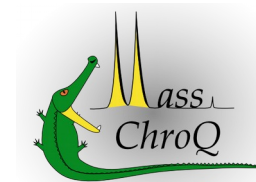
PCA of "absolute" and relative quantification on yeast extracts

RELATIVE QUANTIFICATION

Spectral Count: 2316 proteins
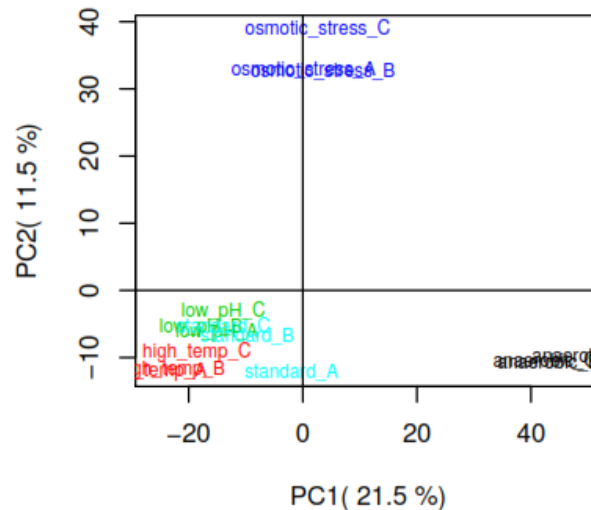**XIC: 1575 proteins**

PCA based on Spectral Counts

PCA based on XIC

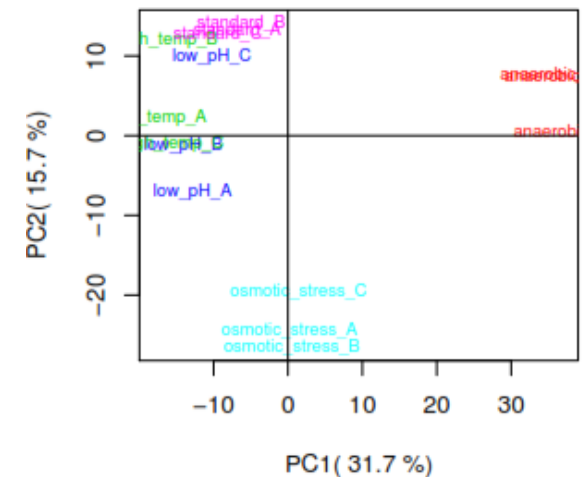**ABSOLUTE QUANTIFICATION**

NSAF : 2739 proteins
IBAQ : 1776 proteins

PCA based on NSAF -bulk

PCA based on IBAQ -bulk