



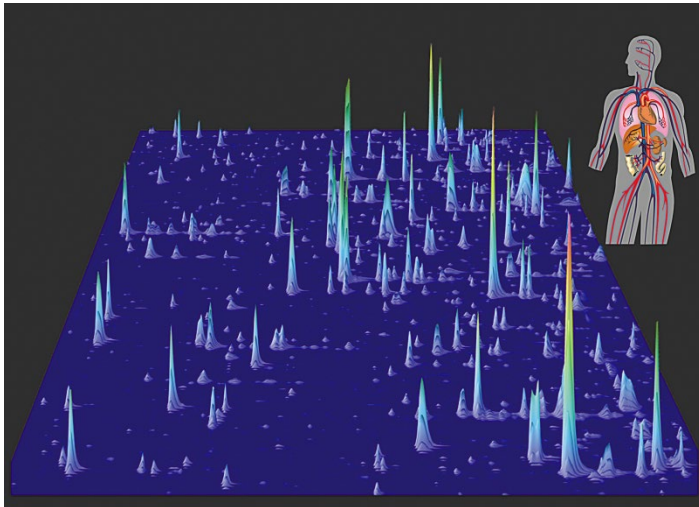
university of  
 groningen

faculty of science  
 and engineering

analytical biochemistry

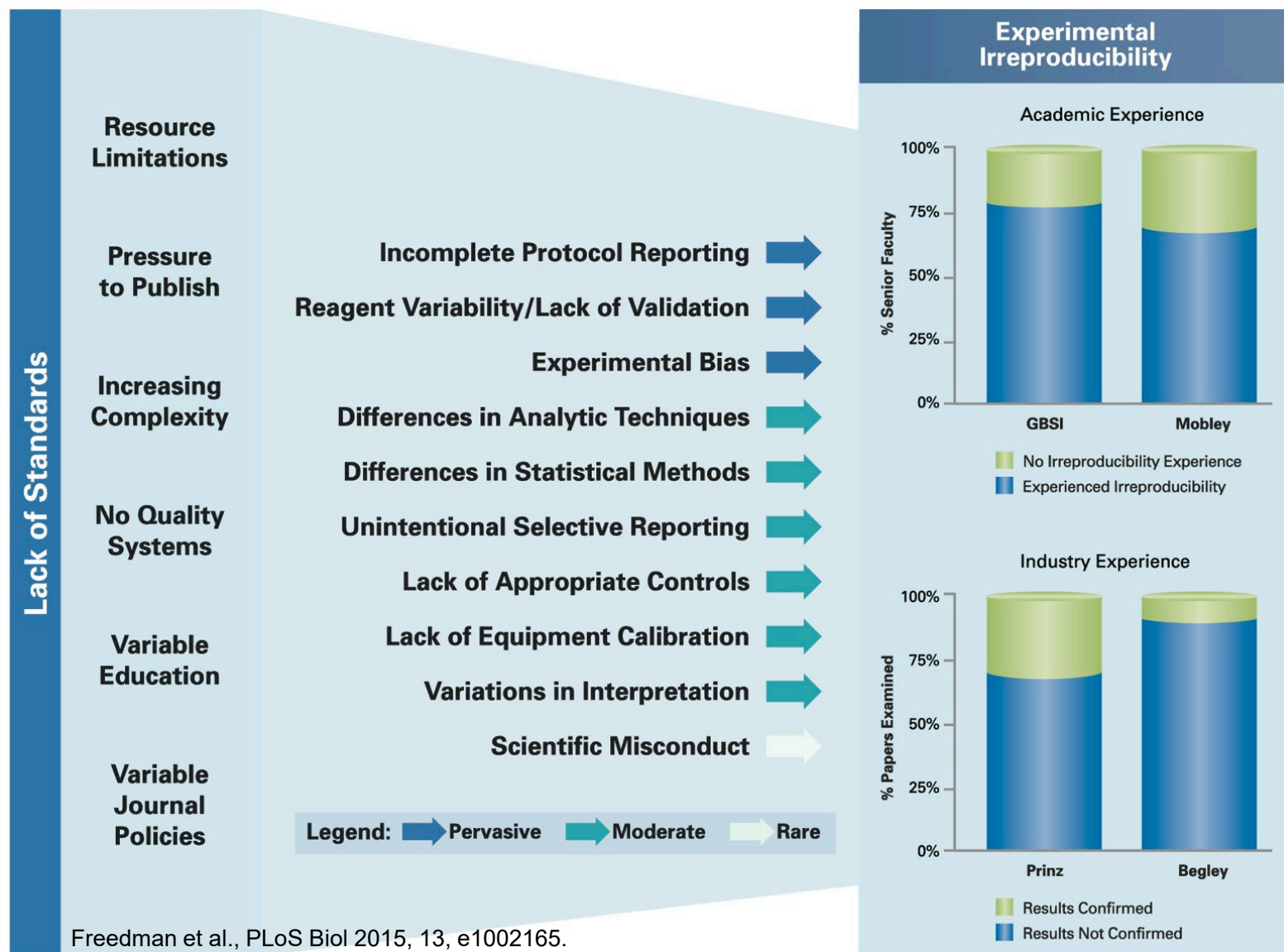
# Biomarker Best Practices: confounding factors, preanalytics and method validation

*Rainer Bischoff*  
*r.p.h.bischoff@rug.nl*



EU-COST - Basel, March 28, 2019

# Causes of irreproducibility



# Research antibodies

- Standards to validate antibodies used in basic or clinical research, and in bio-pharma
- Standards to ensure reproducibility of antibodies over time between manufacturers and lots
- Best practices, such as correct use of negative and positive controls in antibody-based tests

## RELIABLE BINDING REAGENTS FOR ALL

Making the sequences of all binding reagents freely available would give researchers and suppliers a universal reference system.

### REAGENT SEQUENCES IN DATABASE



### DISTRIBUTION METHODS

- In-house production**  
Researchers order gene sequences and make their own reagents in house.
- Commercial distribution**  
Companies stockpile commonly used reagents or generate reagents on demand.
- Non-profit distribution**  
Non-profits stockpile commonly used reagents or generate reagents on demand.

### STANDARDIZED BINDING REAGENTS



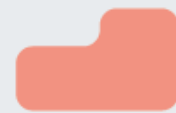
Recombinant monoclonal antibodies



Aptamers (nucleic-acid-based reagent)



Recombinant antibodies



Non-antibody binding reagents

# Biological specimen

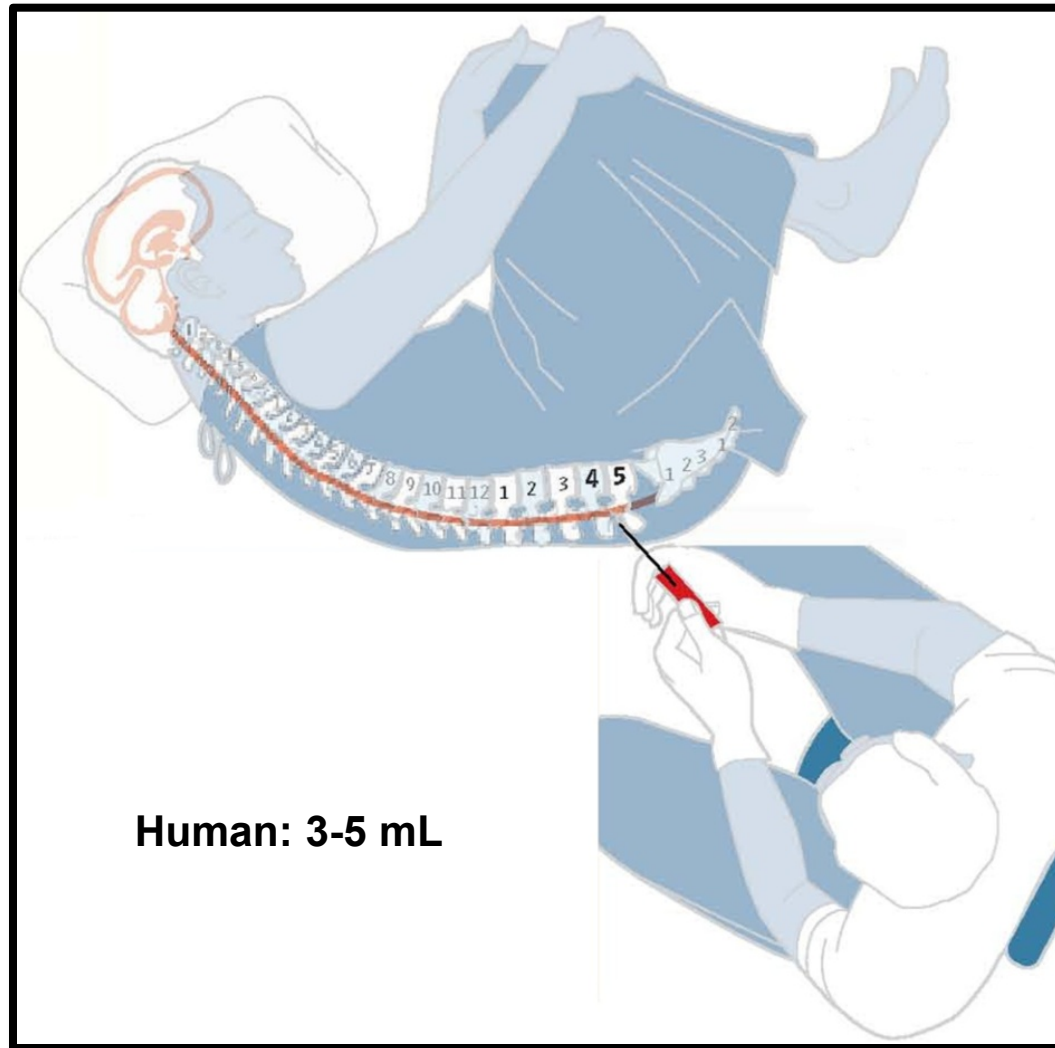
**Pre-analytical variables** affect outcomes unrelated to biology:  
Surgical procedures, storage, freeze-thaw cycles, analysis, etc.  
Gene expression, protein levels, post-translational protein modification, etc.

**Ideally, we would:**

Carefully follow a standard operating procedure (SOP)  
Link all specimens to the SOP used to collect them



# Example CSF: preanalytical factors





# Sample handling & stability

---

## Delayed storage after sampling

CSF was left at room temperature directly after sampling for 0, 30, 120 min

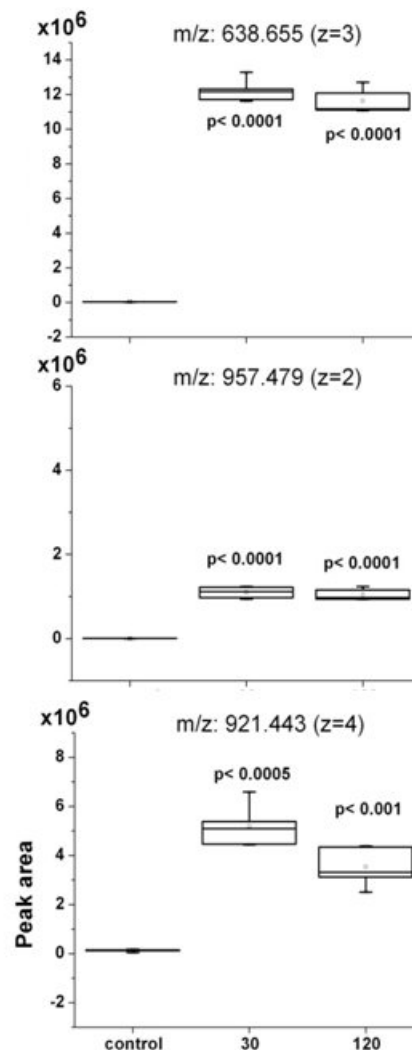
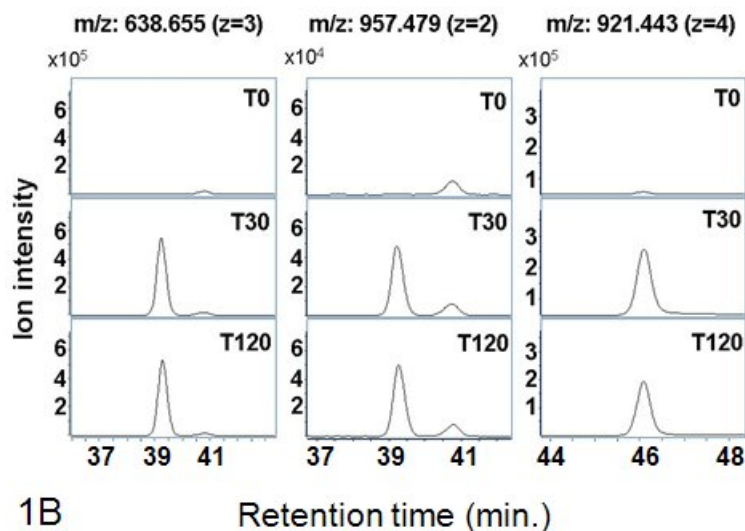
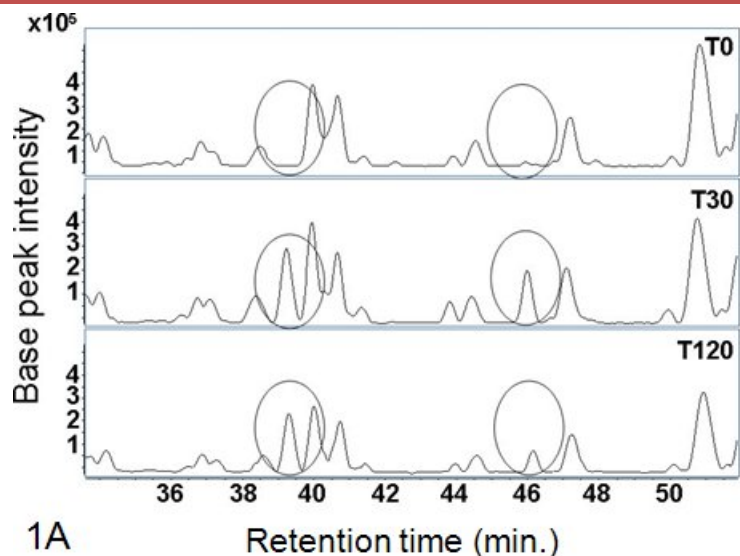
## Freeze/thaw cycles

Snap freeze in liq. N<sub>2</sub> and thaw on ice 1-10 times

## Autosampler stability

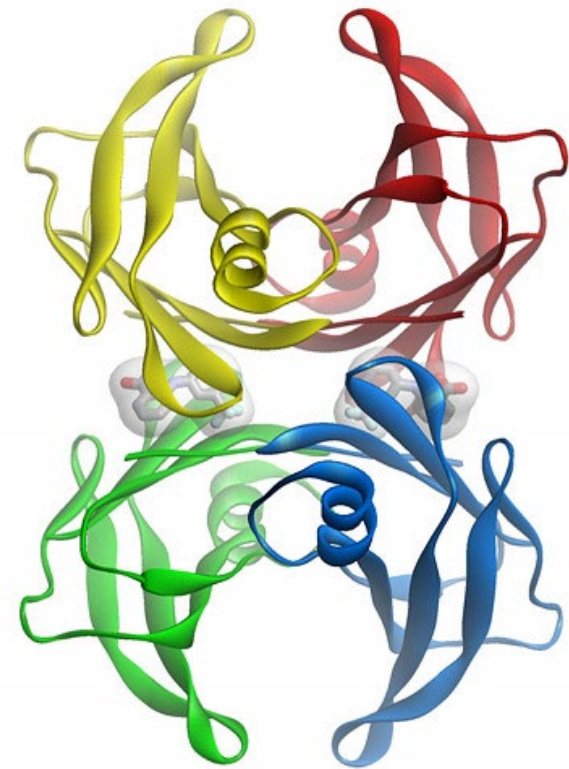
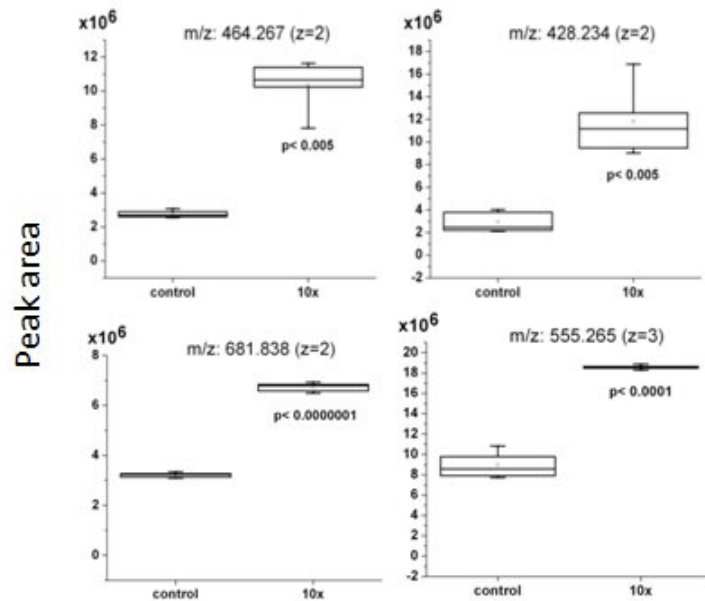
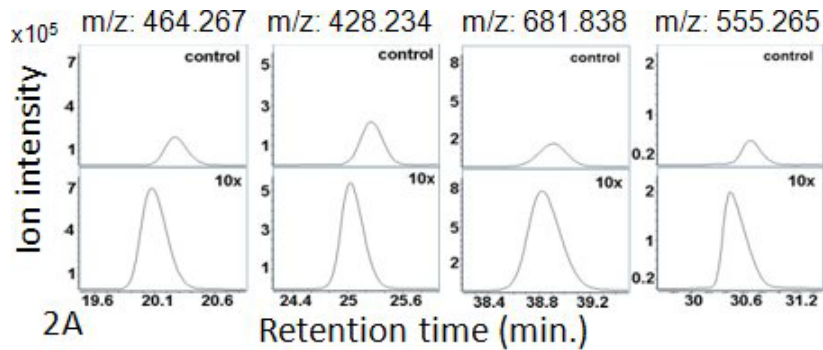
Leave in autosampler for extended time periods

# Delayed storage after sampling



1C Incubation at room temp. (min.)

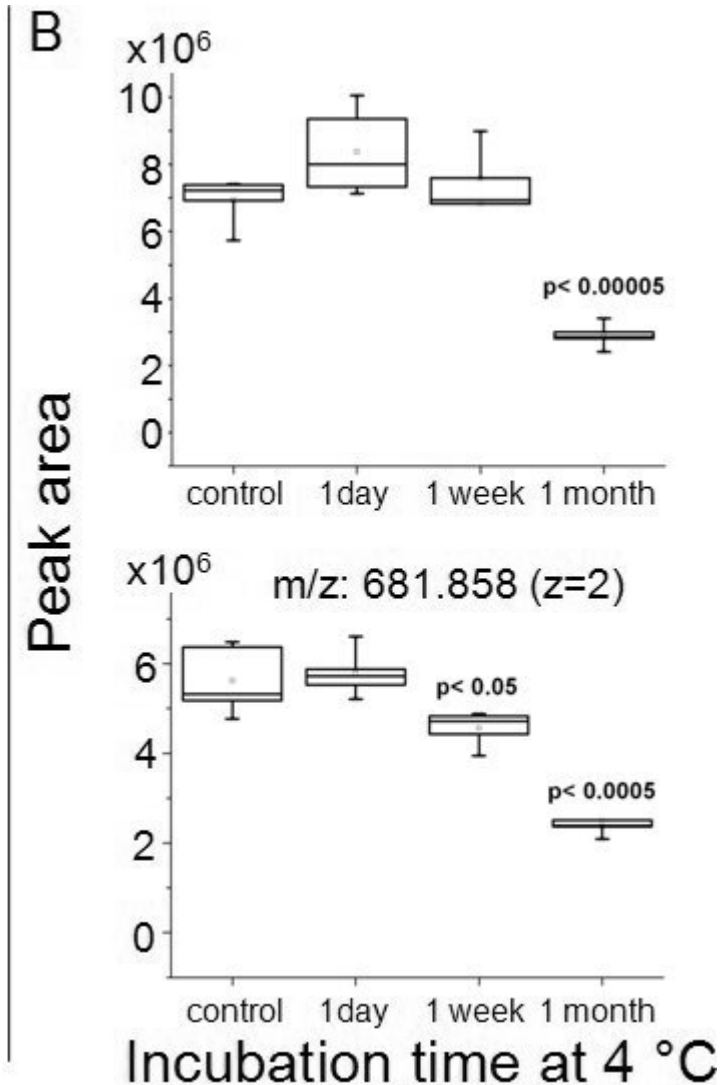
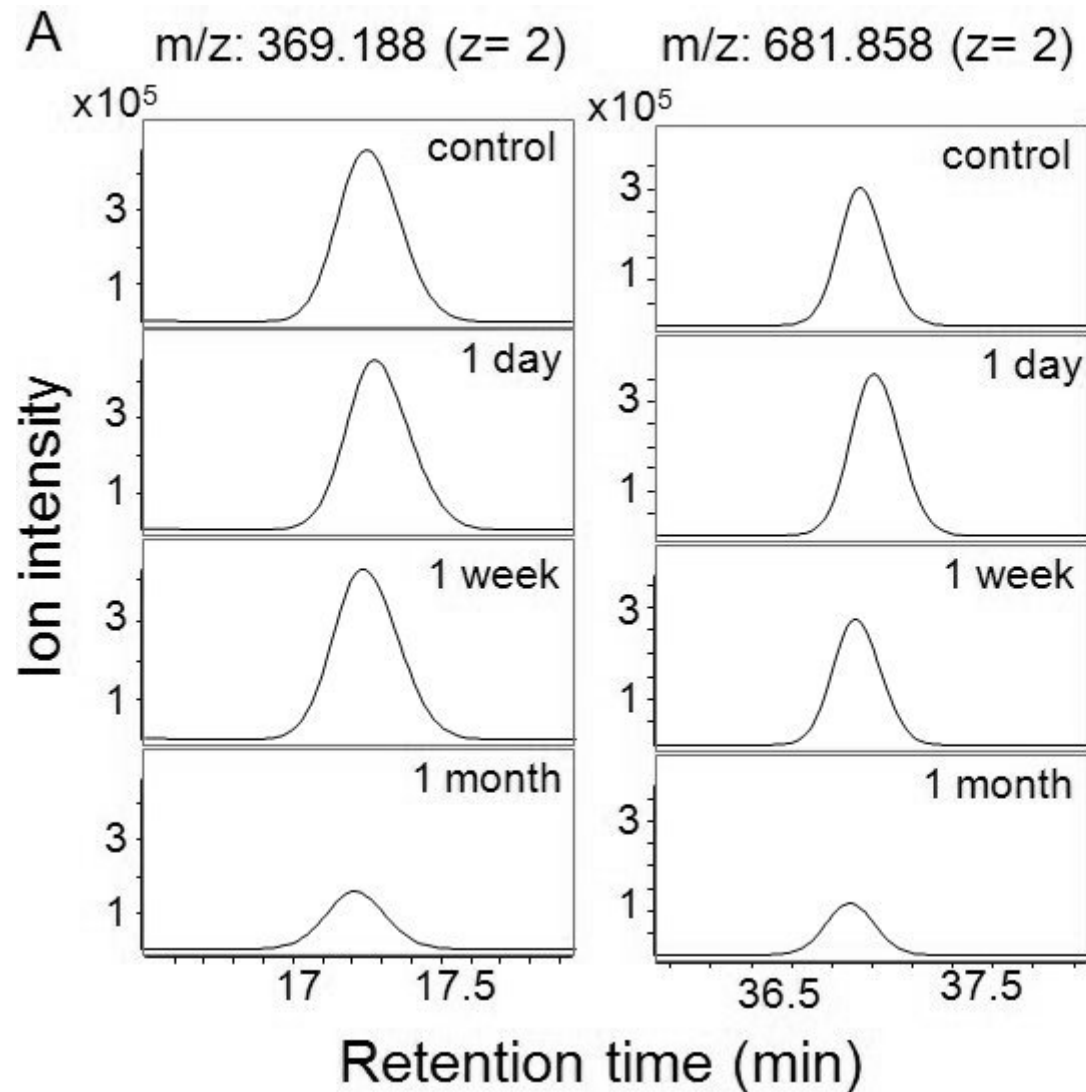
# Freeze-thaw cycles



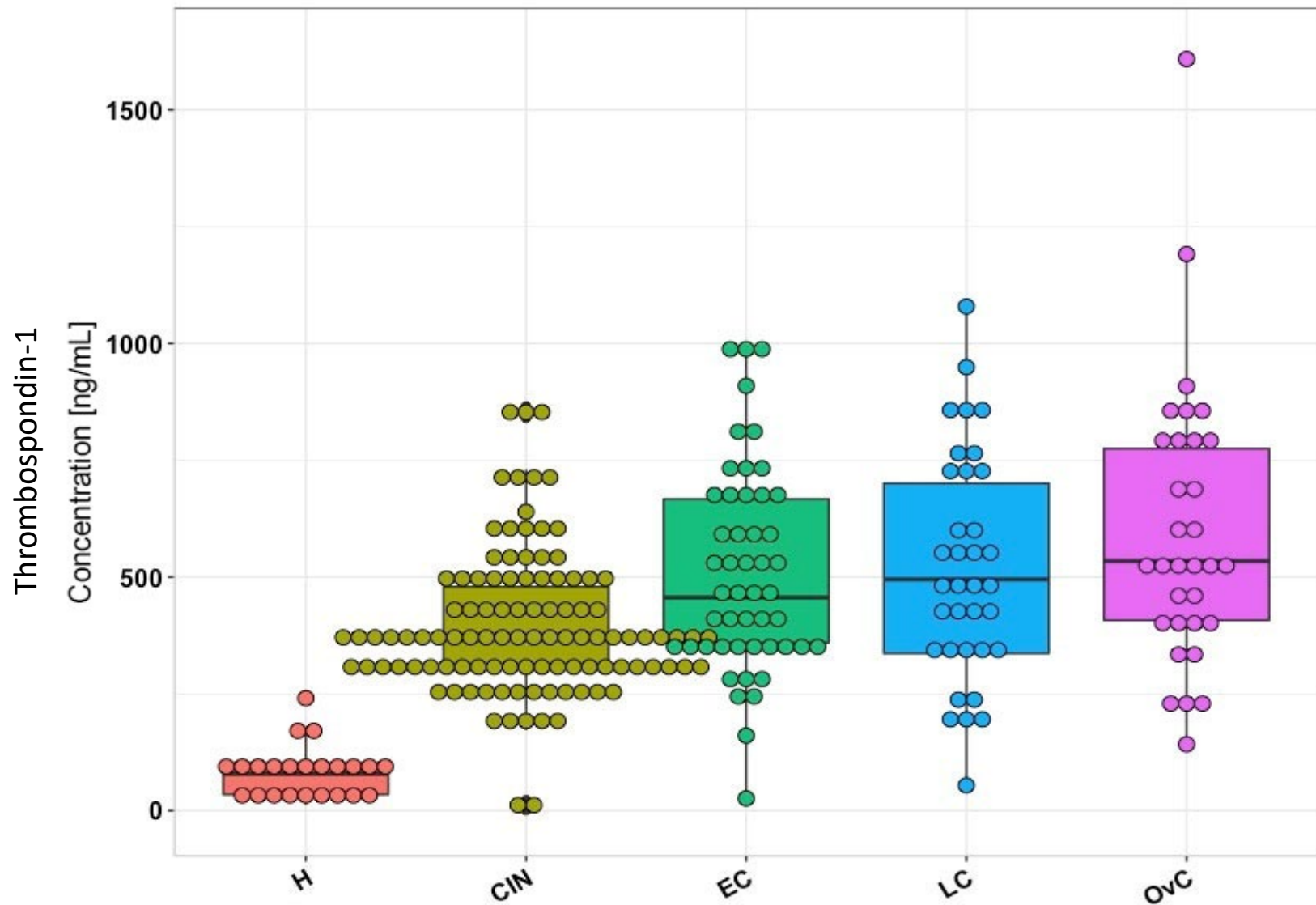
Transthyretin



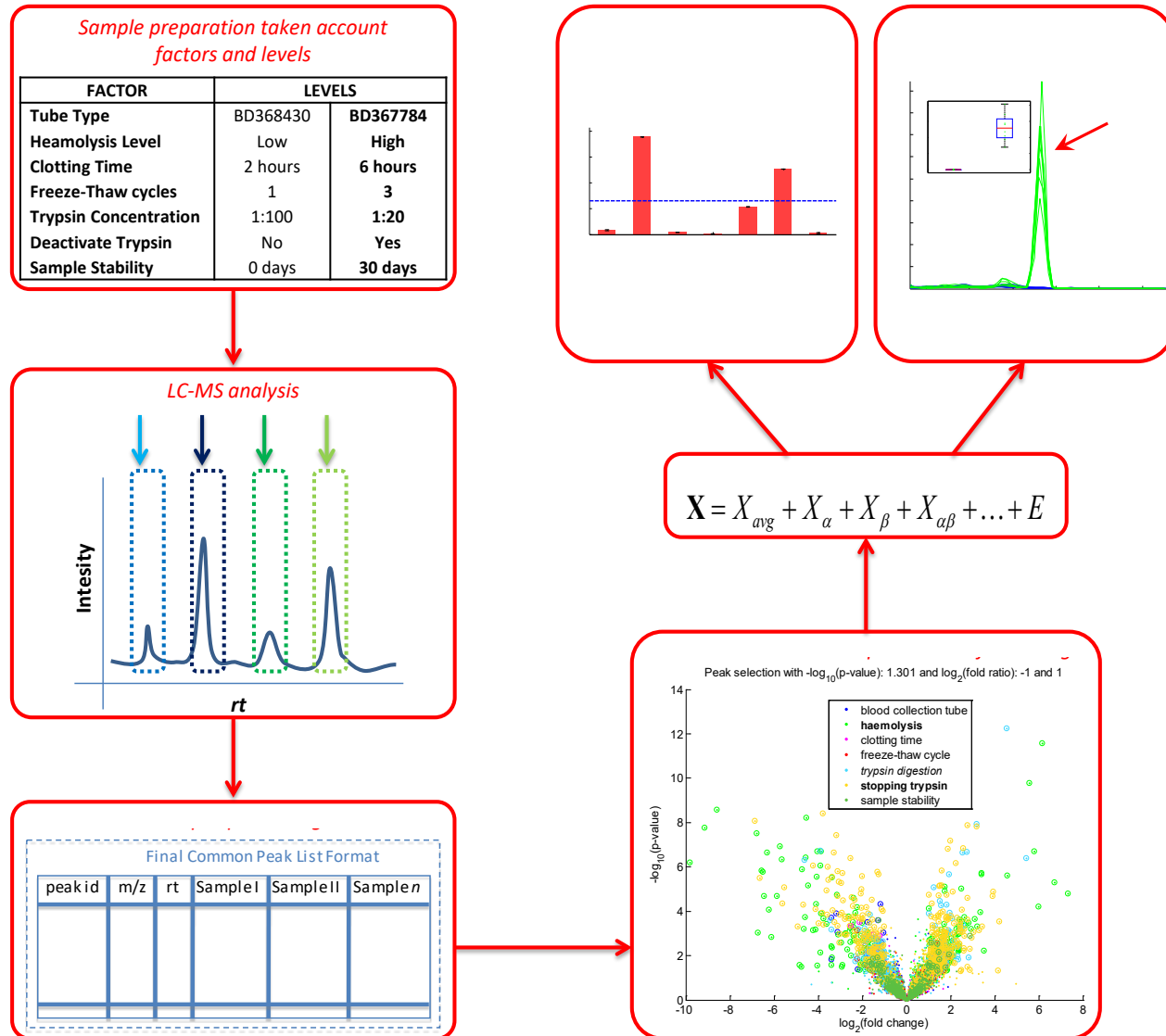
# Storage in autosampler at 4°C



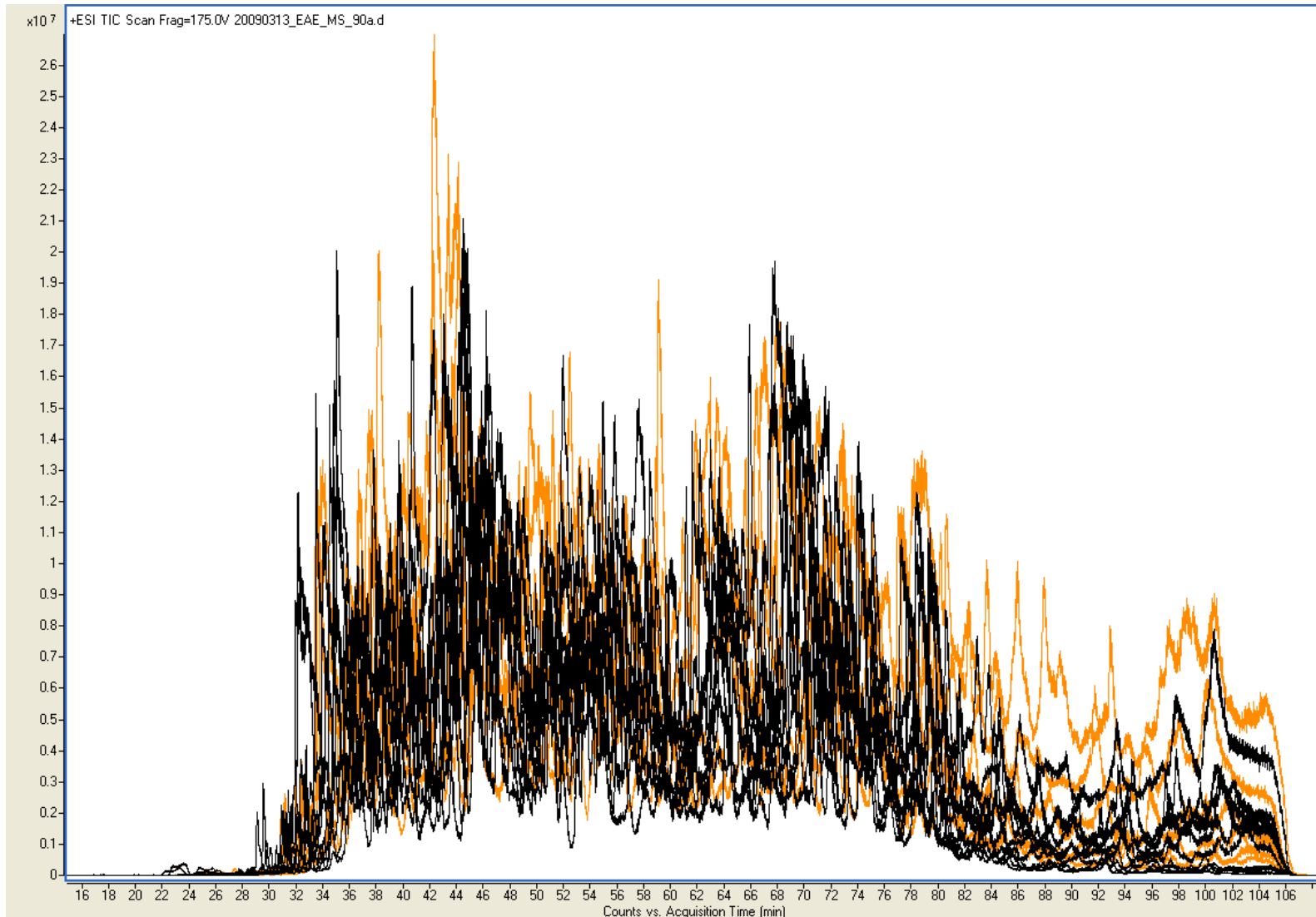
# Conditions of blood coagulation



# How to assess the relevance of preanalytical factors?



# Effect of data (pre)processing



# From raw data to results



1. Sample analysis/Data acquisition



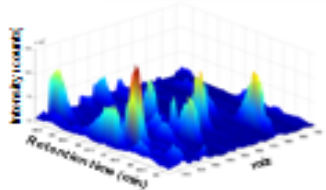
2. Data pre-processing  
(e.g. 2D-smoothing)



3. Geometrical peak picking with local  
baseline and peak volume calculation



4. 2D time alignment



5. Peak matching across multiple chromatograms  
(generation of aligned quantitative peak matrix)



6. Principal Component Analysis  
(PCA) of complete peak matrix



7. Nearest Shrunken Centroid (NSC)  
classification with double cross validation

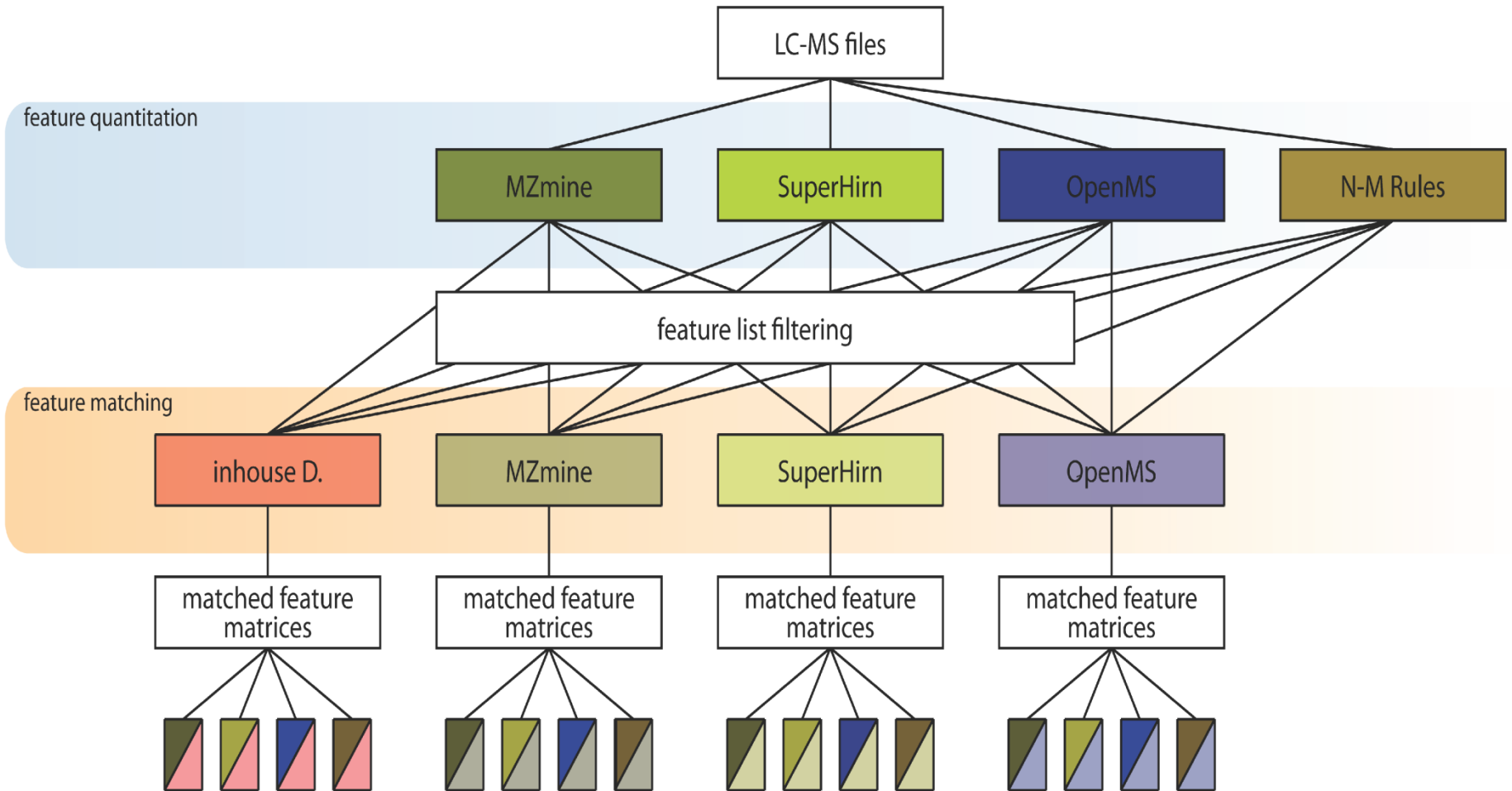


8. PCA of selected variables

Suits et al., Anal. Chem., 80: 3095-3104 (2008)  
Christin et al., Anal. Chem., 80: 7012-7021 (2008)  
Suits et al., Anal. Chem., 83: 7786-7794 (2011)  
Hoekman et al., Mol. Cell. Proteomics, 11 (6) (2012)  
Christin et al., Mol. Cell. Proteomics, 12, 263-276 (2013)



# Same data – different results



# A scoring system to assess the quality of data (pre)processing

Formula parameters  
 $p = 1, x = 1$

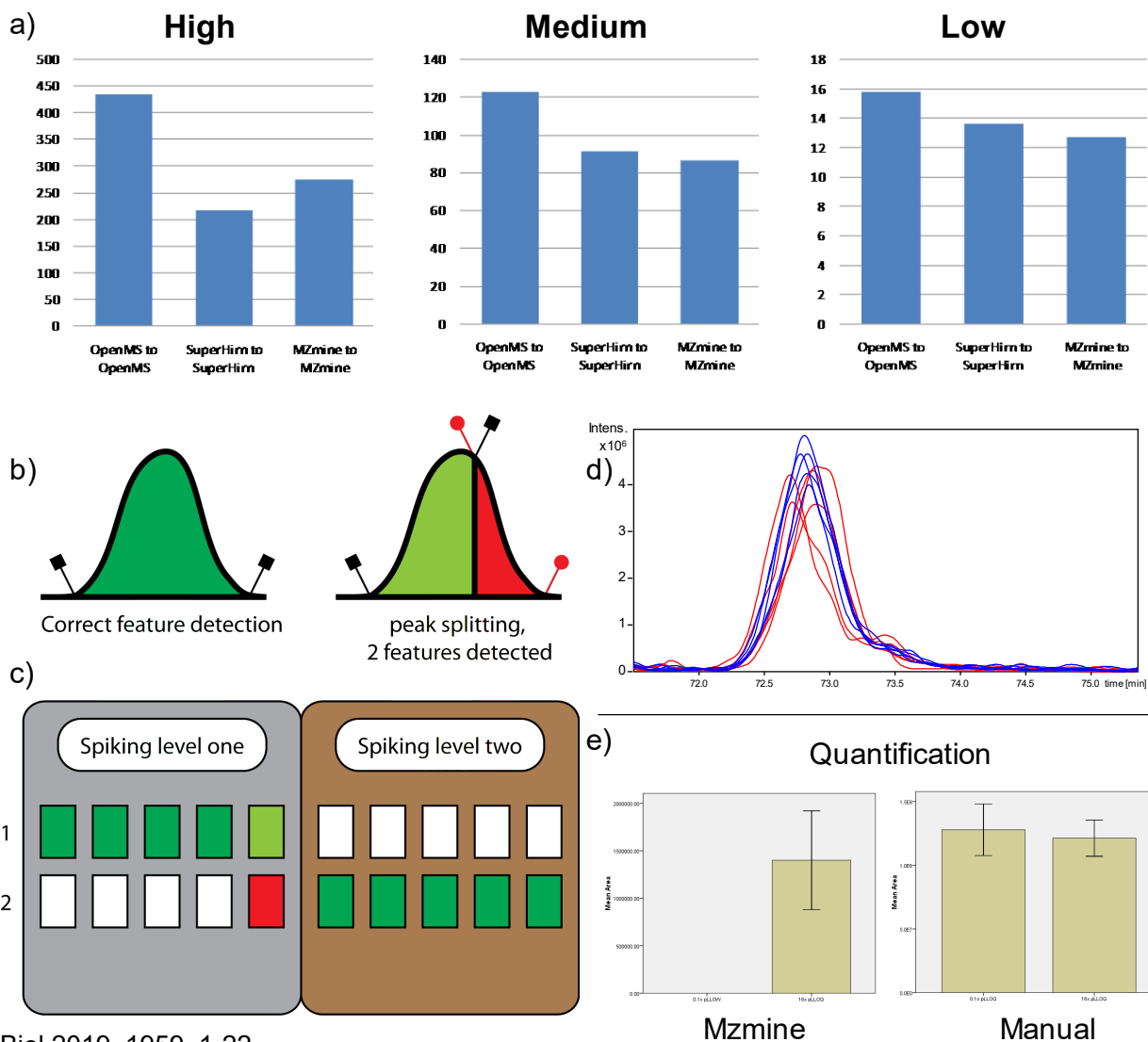
Formula parameters  
 $p = 5, x = 1$

Peaks sorted by t-value

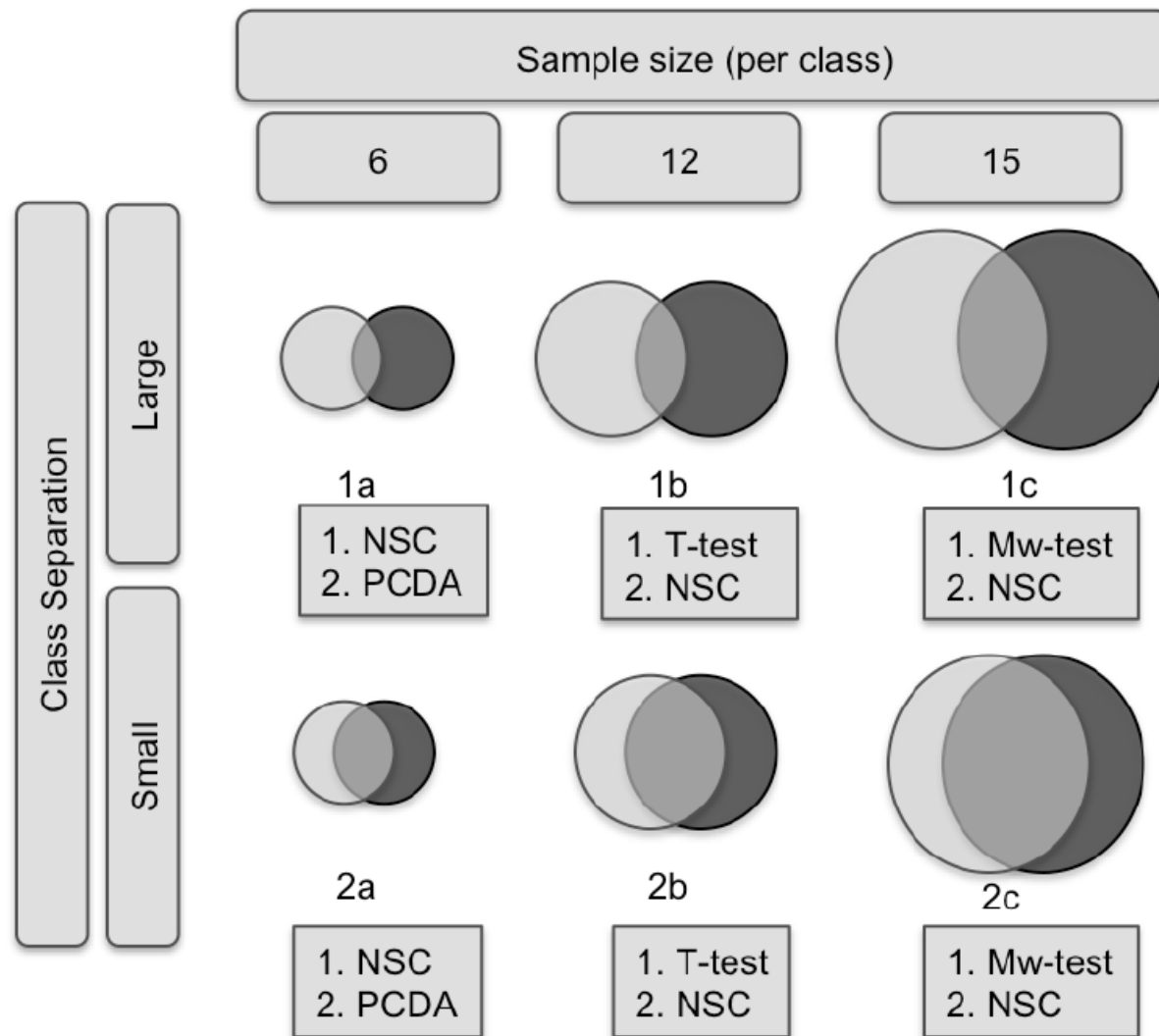
Score	Score	Score	Score
● 1	● 1	● 1	○ 0
● 1	● 1	● 1	● 0.5
○ 0	○ 0	● 1	● 0.5
○ 0	● 0.5	● 1	● 0.5
○ 0	○ 0	○ 0	● 0.5
○ 0	● 0.33	● 0.5	● 0.5
○ 0	○ 0	○ 0	○ 0
○ 0	● 0.25	○ 0	○ 0
+ + + +			
2	3.1	4.5	2.5

Score	Score	Score	Score
● 1	● 1	● 1	○ 0
● 1	● 1	● 1	● 0.83
○ 0	○ 0	● 1	● 0.83
○ 0	● 0.83	● 1	● 0.83
○ 0	○ 0	○ 0	● 0.83
○ 0	● 0.71	● 0.83	● 0.83
○ 0	○ 0	○ 0	○ 0
○ 0	● 0.63	○ 0	○ 0
+ + + +			
2	4.2	4.8	4.2

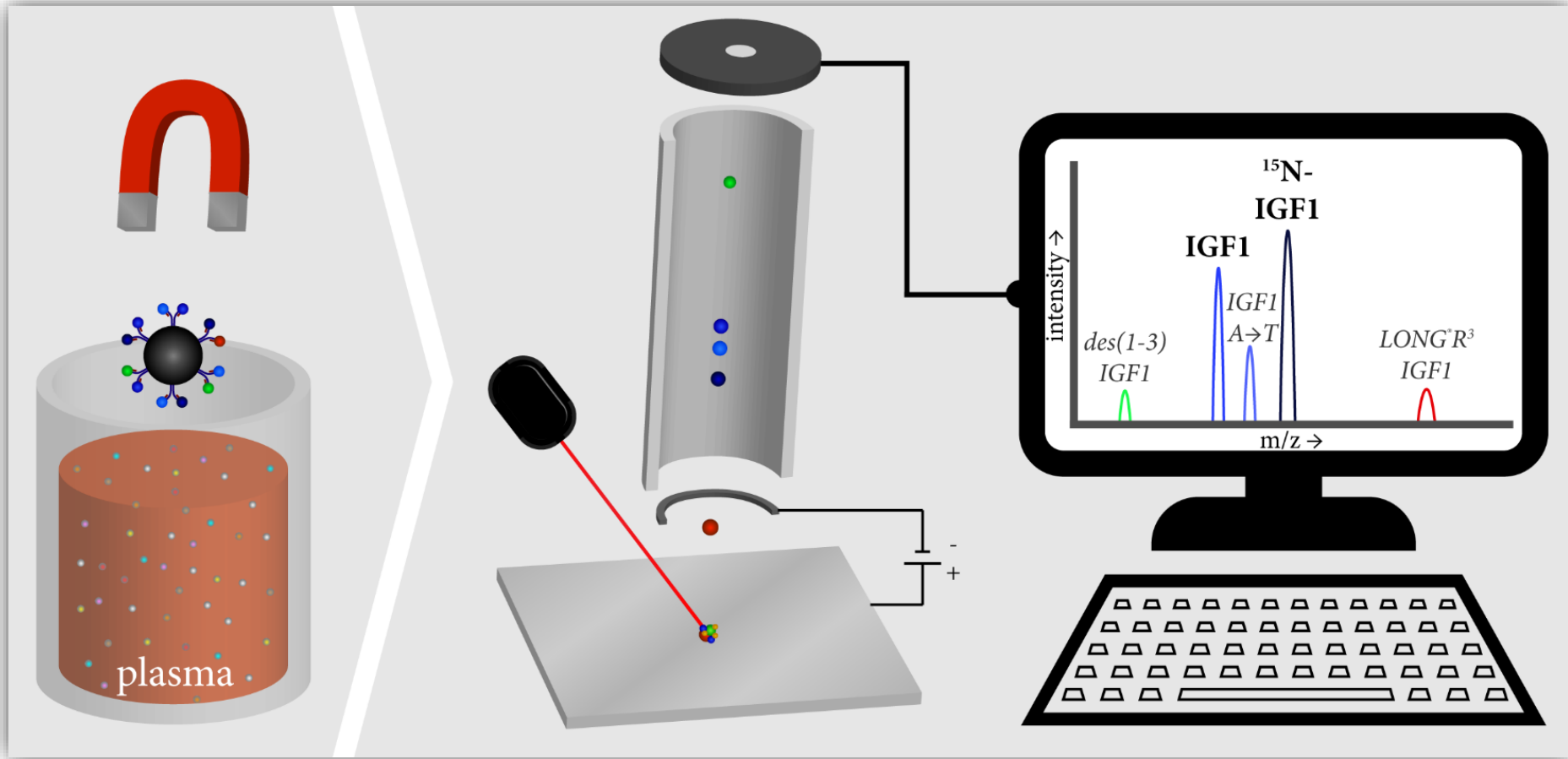
# Peak splitting: one possible source of errors



# Same processed data – different results



# Method validation: Quantification of IGF-1 by MALDI-MS



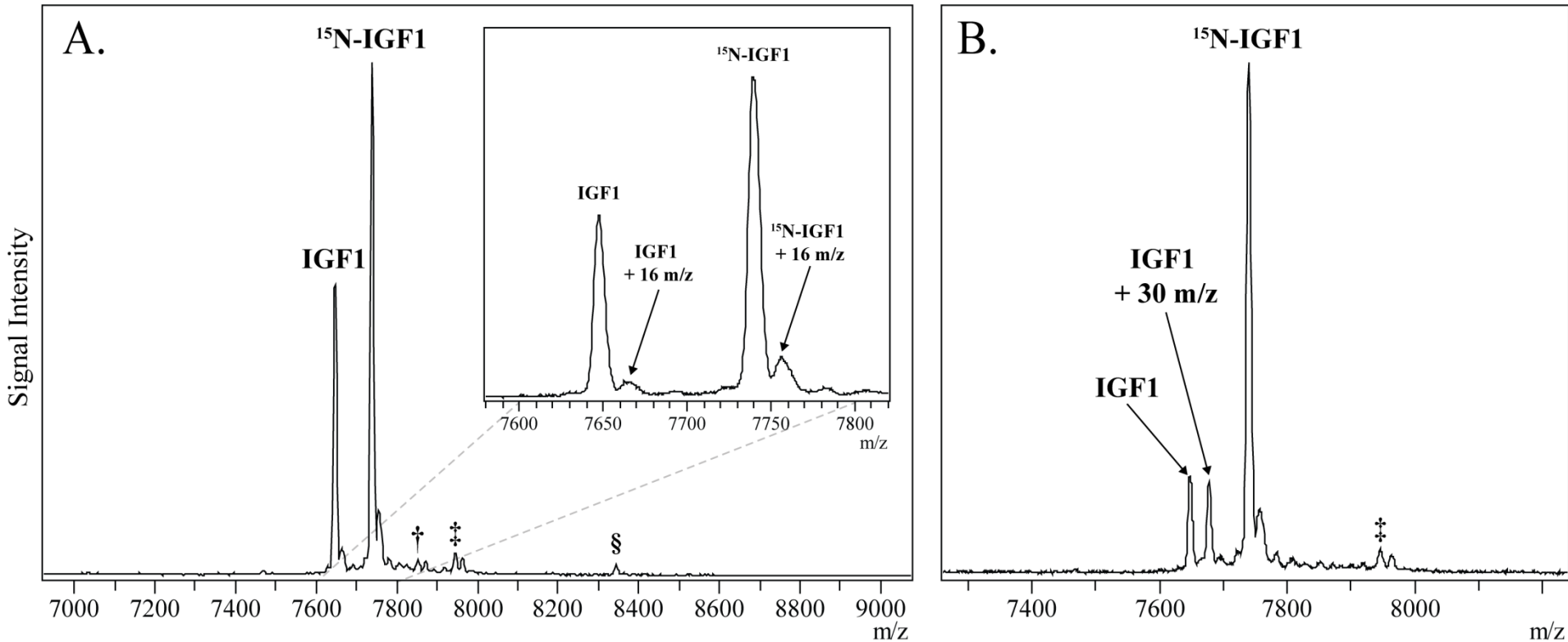


# Method development

---

- ›  $^{15}\text{N}$ -IGF1 as internal standard
  - reliable compensation for sources of variability
- › Rat plasma as calibration matrix
  - adequate mimicking of human plasma
- › SDS-based IGF1/IGFBP-complex dissociation
  - proper liberation of IGF1 from its binding proteins
- › Workflow automation
  - matching MALDI-TOF's high throughput capabilities
- › Identification of quality indicators
  - improving MALDI-TOF performance & data quality

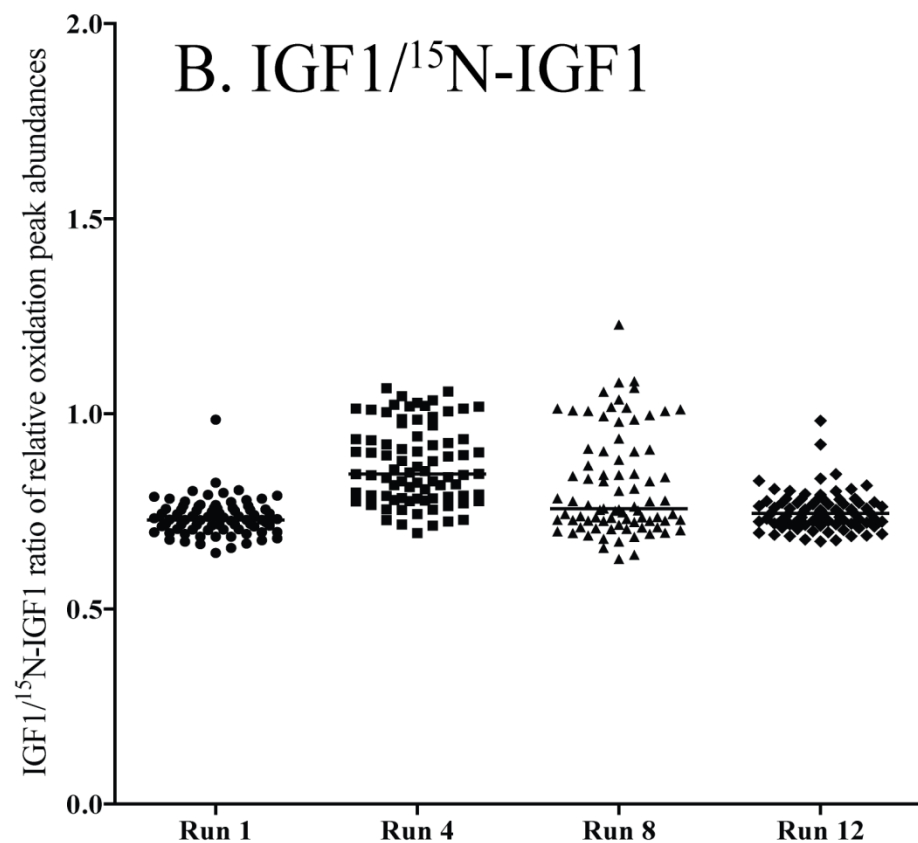
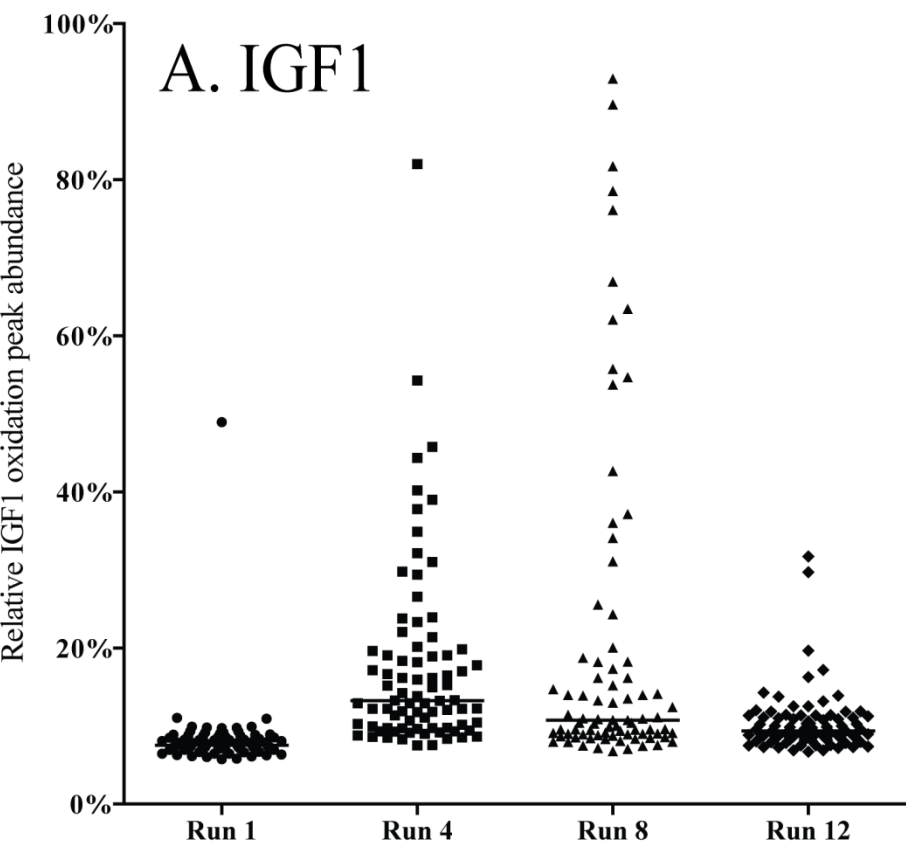
# Relevance of the internal standard



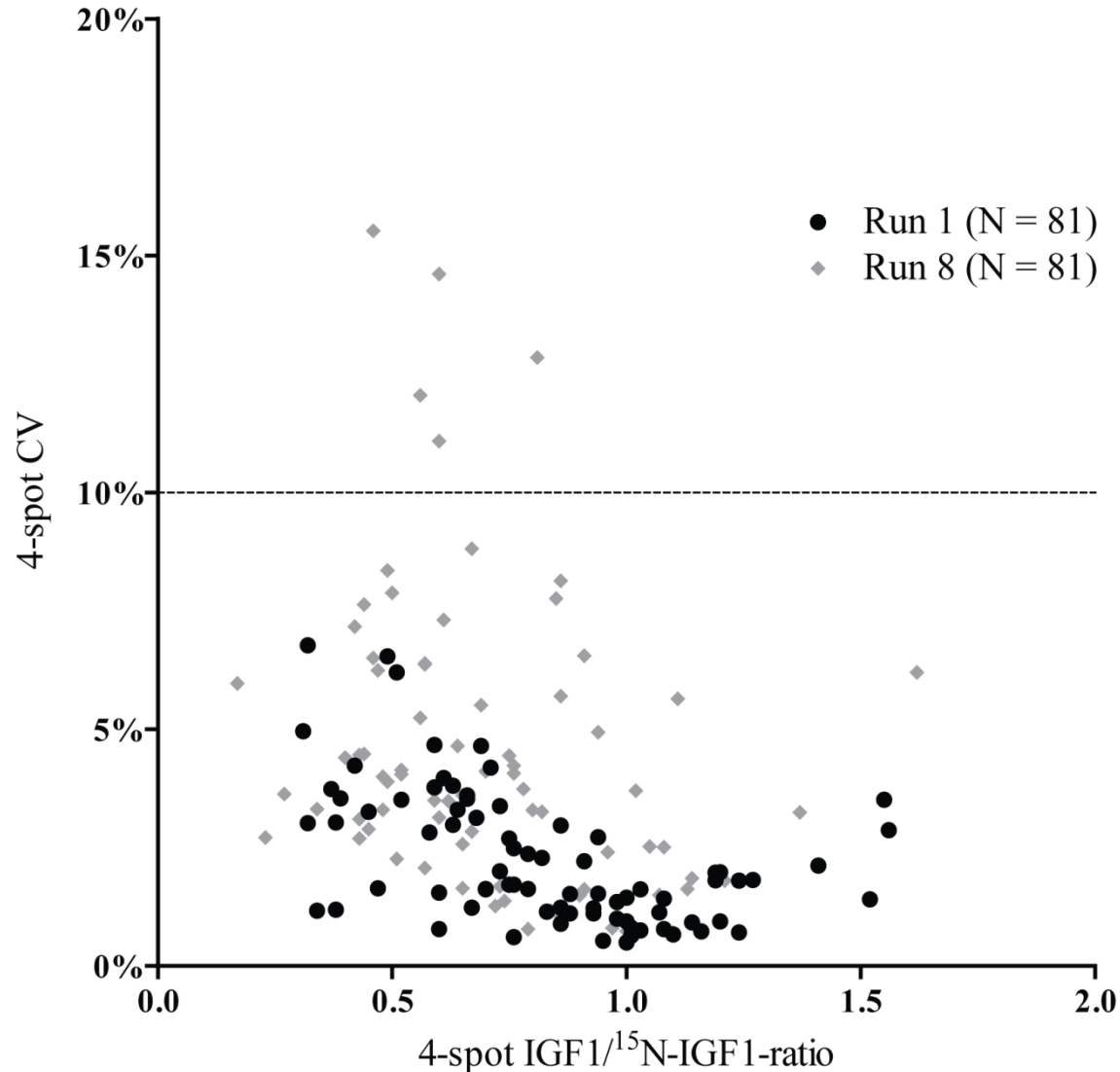
# Summary of validation data

		QC-low		QC-medium		QC-high	
		CV	Bias	CV	Bias	CV	Bias
Accuracy & precision (3 runs in 6-fold)	run 1	5%	1%	5%	8%	13%	4%
	run 2	6%	-4%	4%	-6%	15%	-10%
	run 3	10%	2%	4%	-2%	15%	6%
Bench-top stability (24h, in 3-fold)		14%	-9%	-	-	1%	9%
Freeze-thaw stability -20 °C (3 cycles, in 3-fold)		6%	-13%	-	-	4%	12%
MALDI-sample stability (7 days, in 6-fold)	day 0	5%	1%	5%	8%	13%	4%
	day 7	3%	10%	4%	10%	4%	12%
		20 ng/mL calibrant		QC-low		QC-medium	
		CV	Bias	CV	Bias	CV	Bias
IGFBP3 challenge test (in 5-fold)		5%	-7%	4%	-6%	1%	-2%
		+ 25 ng/mL		+ 100 ng/mL		+ 500 ng/mL	
		CV	Bias	CV	Bias	CV	Bias
Spike recovery (6 different plasma samples)		9%	4%	7%	11%	12%	4%

# Quality indicator: oxidation

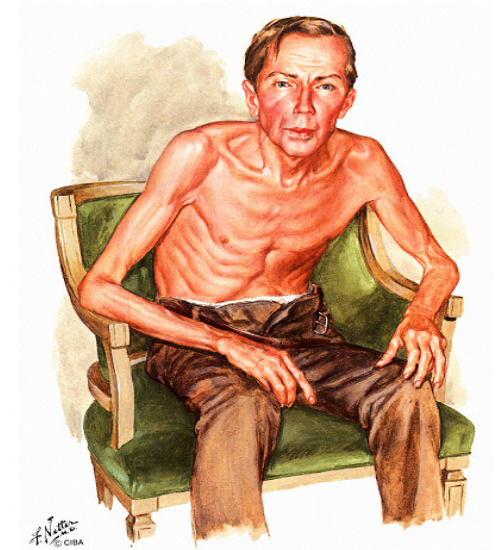
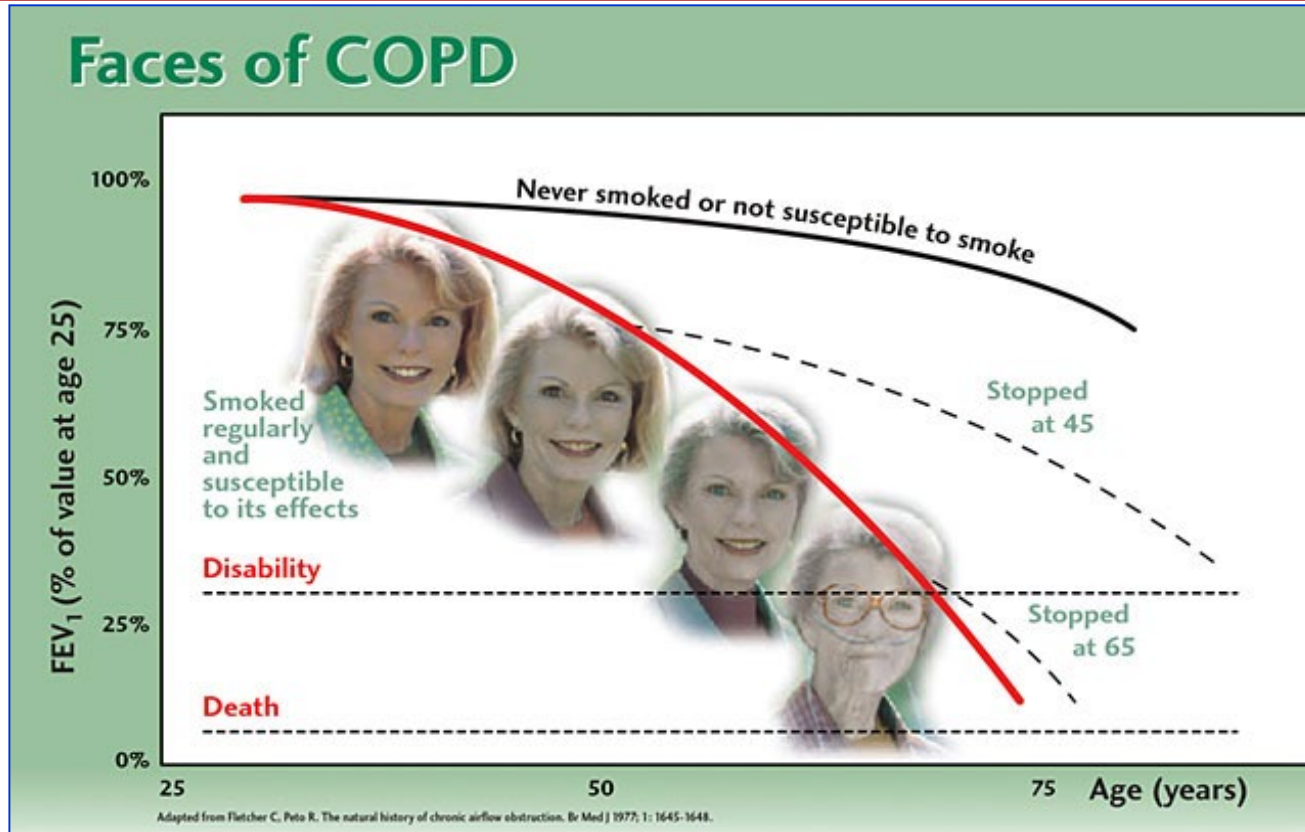


# Quality indicator: 4-spot variability



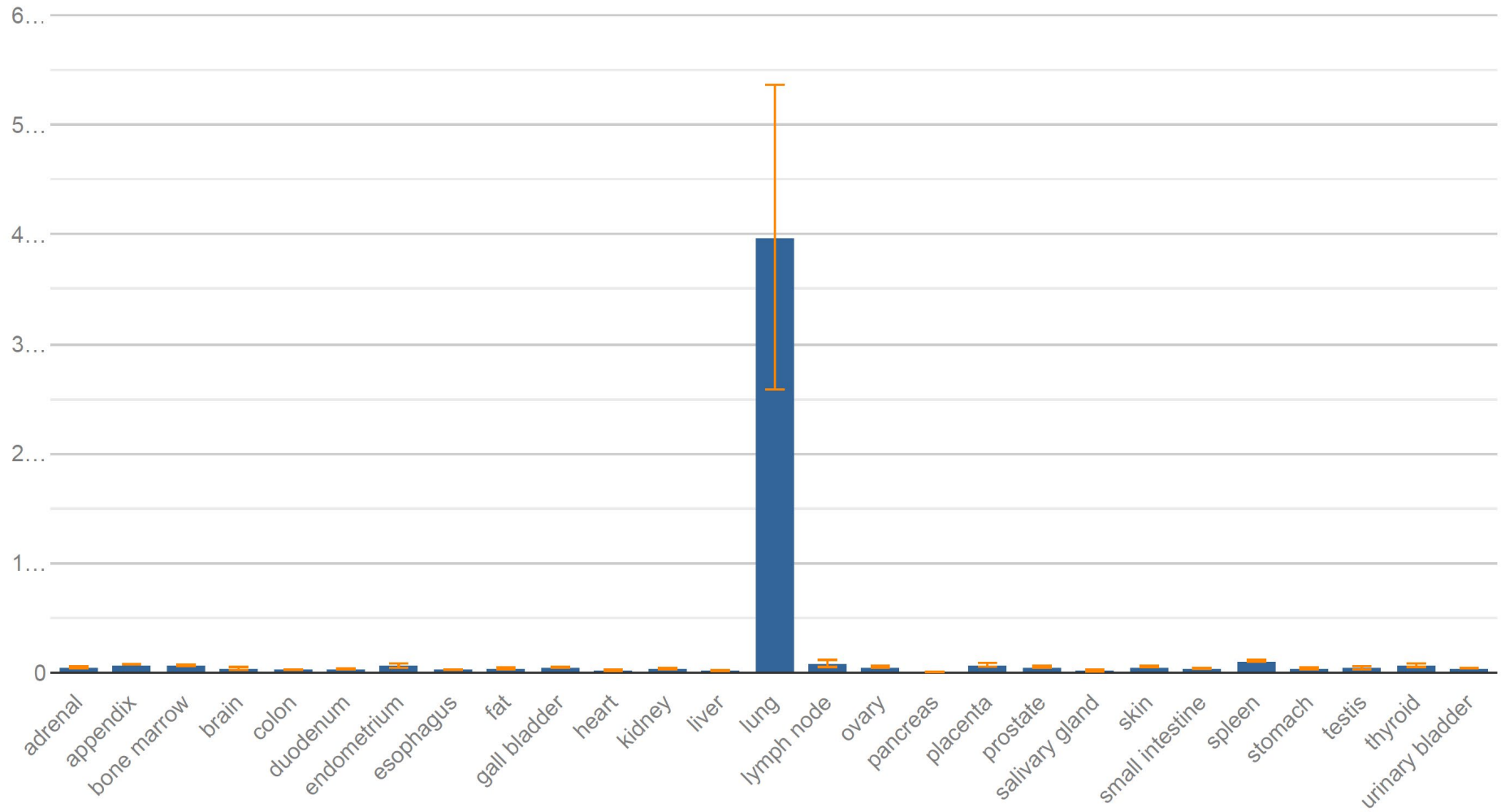


# Confounding factors: sRAGE and COPD

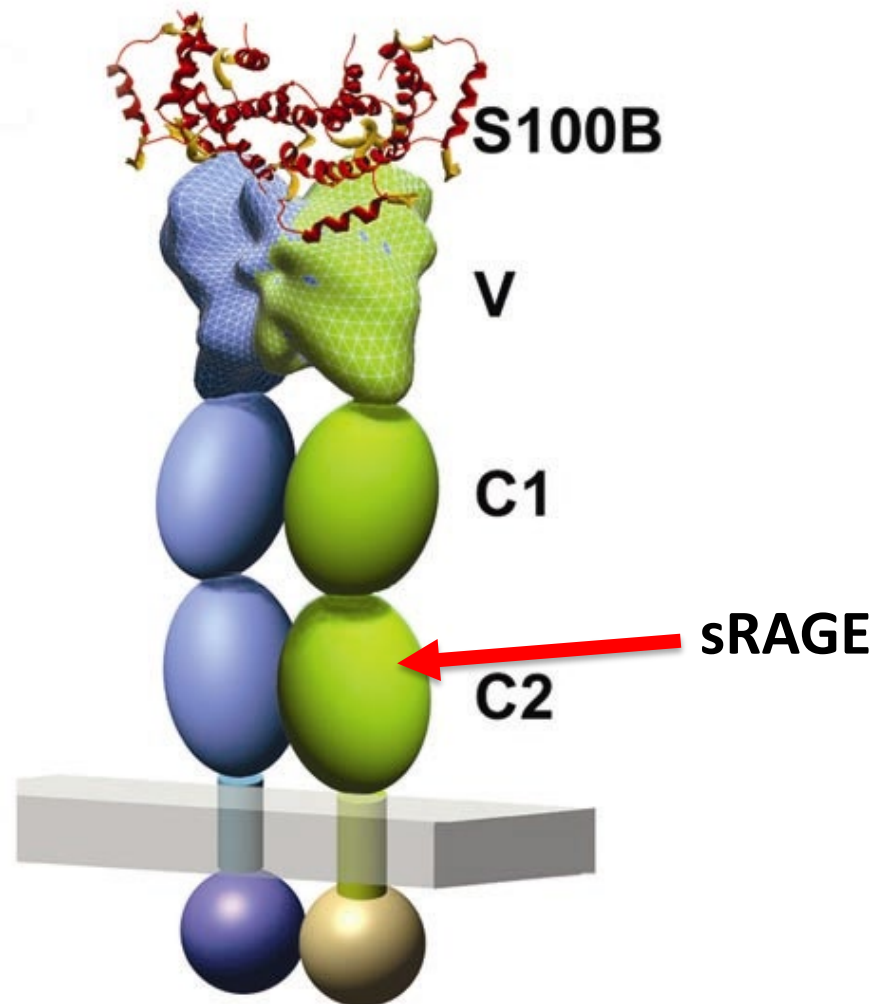


- 20% of the smokers develop COPD, more than 200 million people have COPD
- Progressive loss of lung function with a large impact on the quality of life
- Insufficient insight in the molecular mechanisms of COPD
- Limited therapeutic options

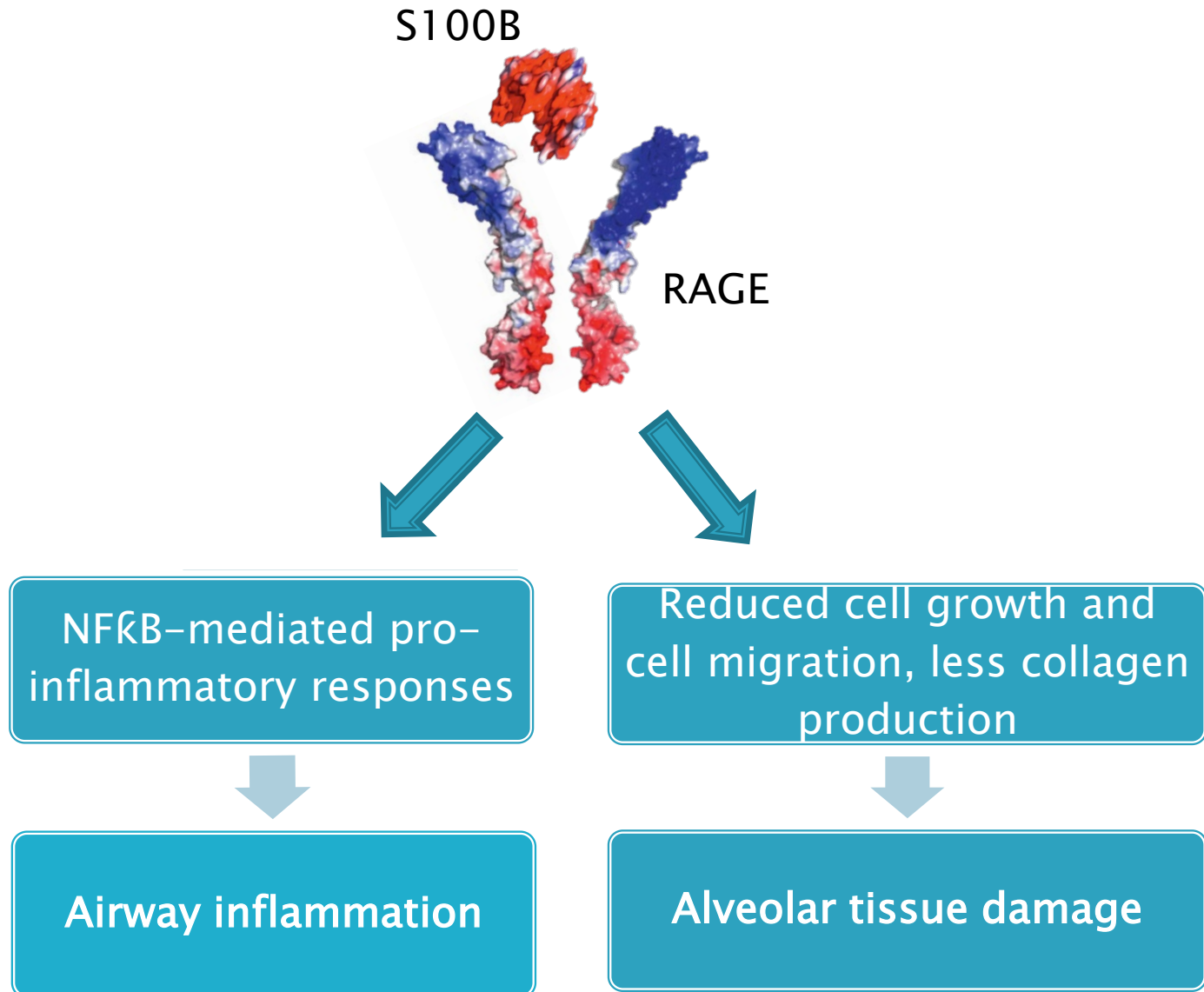
# Expression of the AGER Gene



# Receptor for Advanced Glycation Endproducts (RAGE)



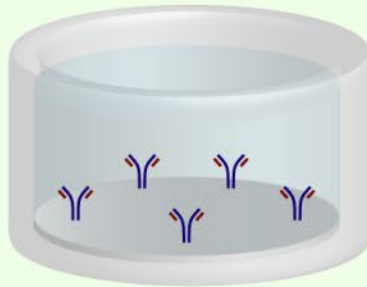
# RAGE effects



# sRAGE methodology

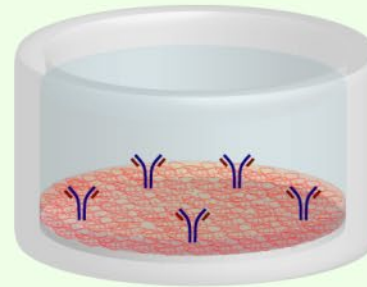
**1.**

**plate coating**



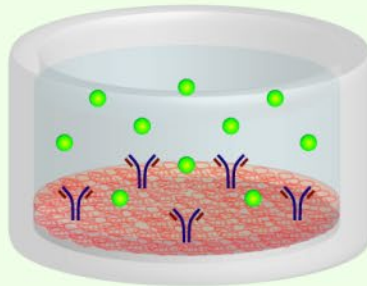
**2.**

**plate blocking**



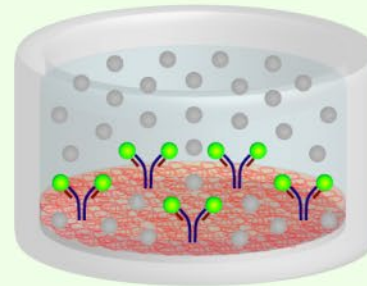
**4.**

**analyte collection**



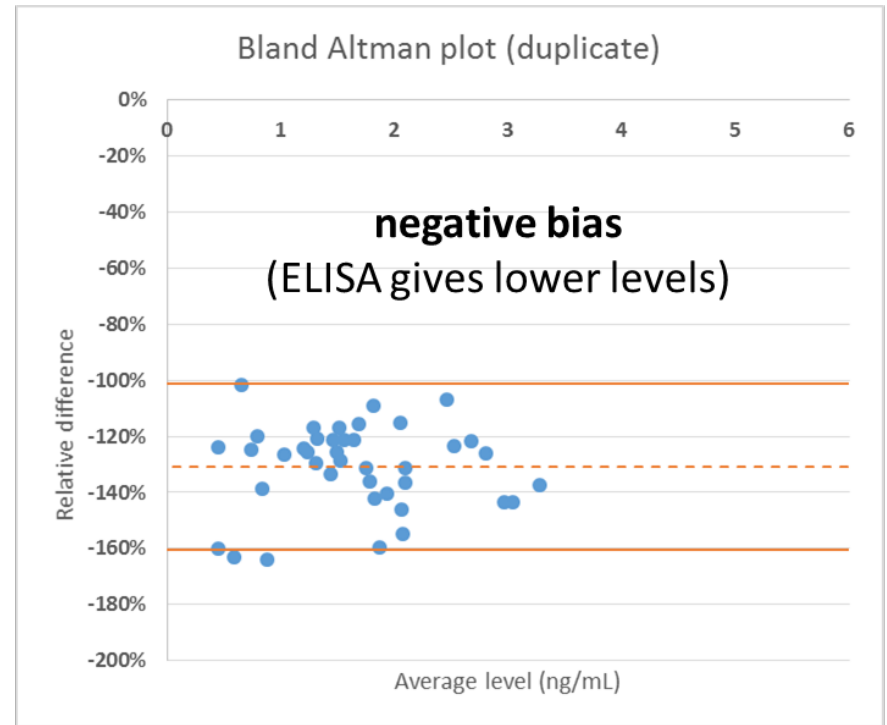
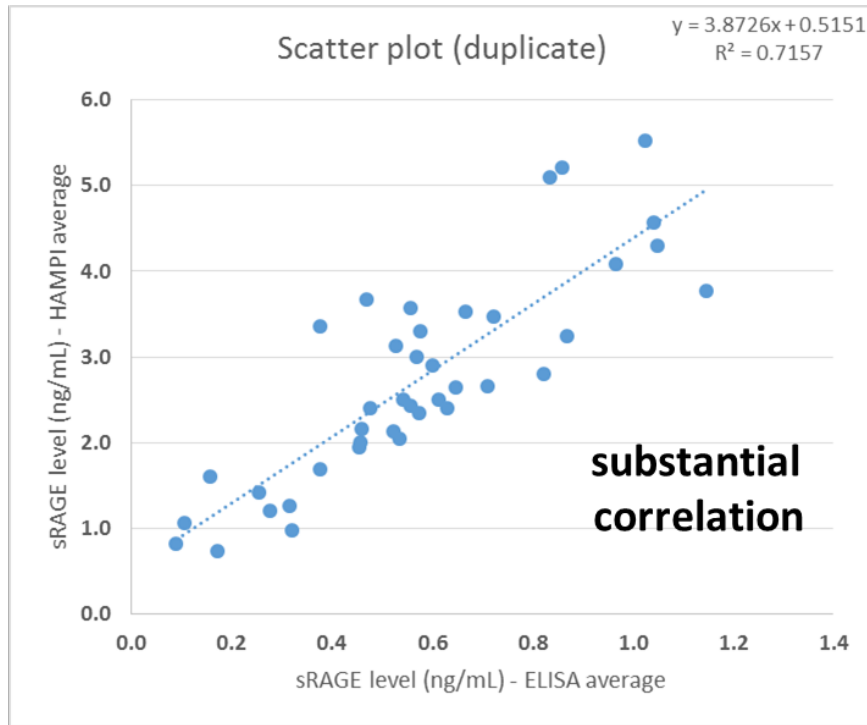
**3.**

**immunocapture**



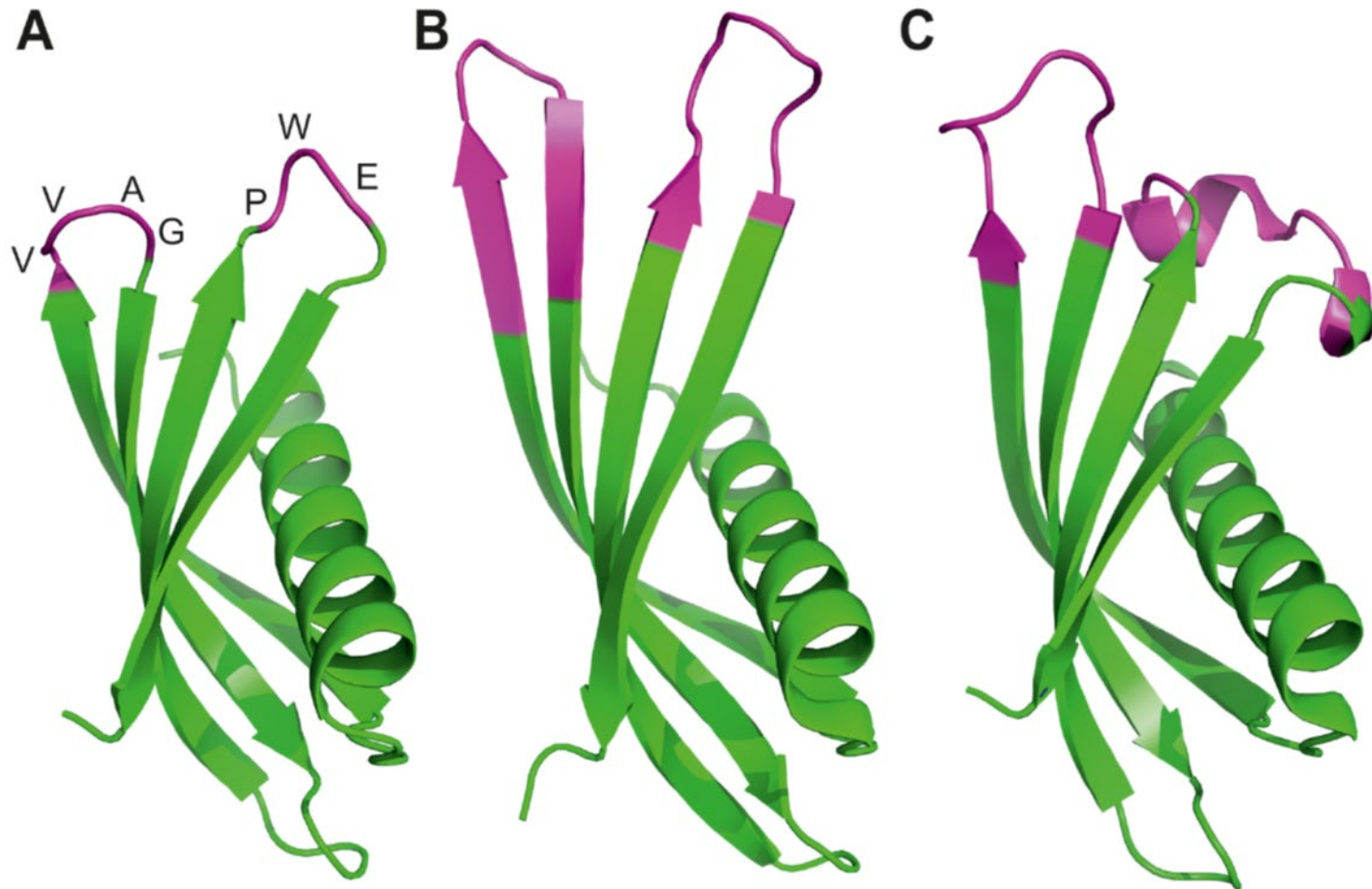


# Immunoaffinity LC-MS vs. ELISA



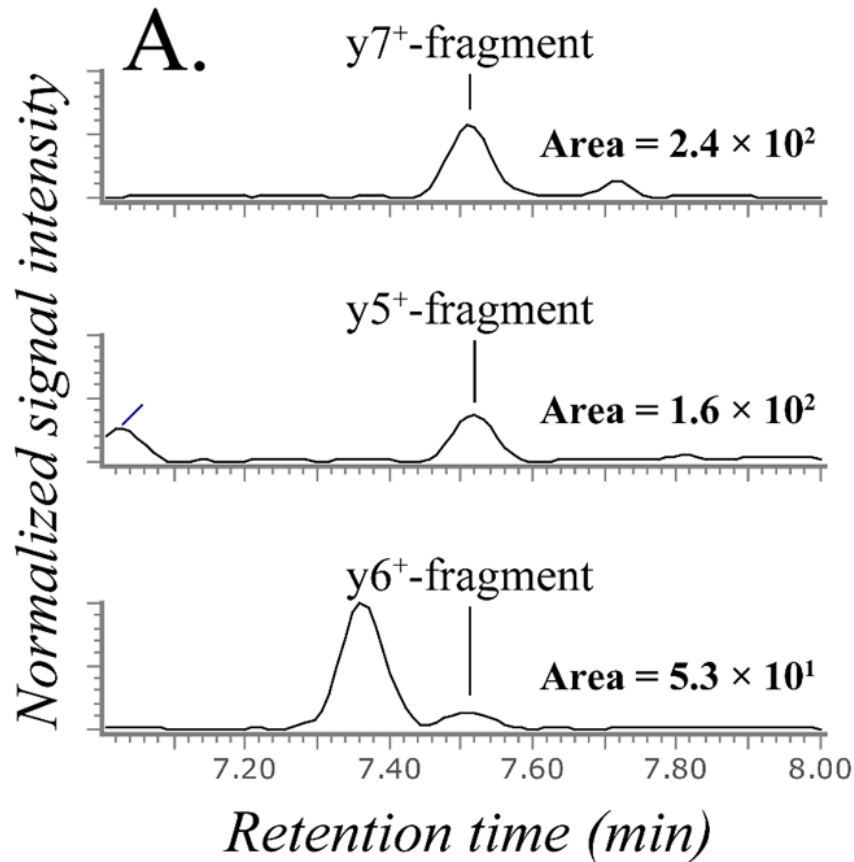
→ Lower levels by ELISA are due to an insufficient amount of primary antibody

# Alternative Affinity Agents: Affimers

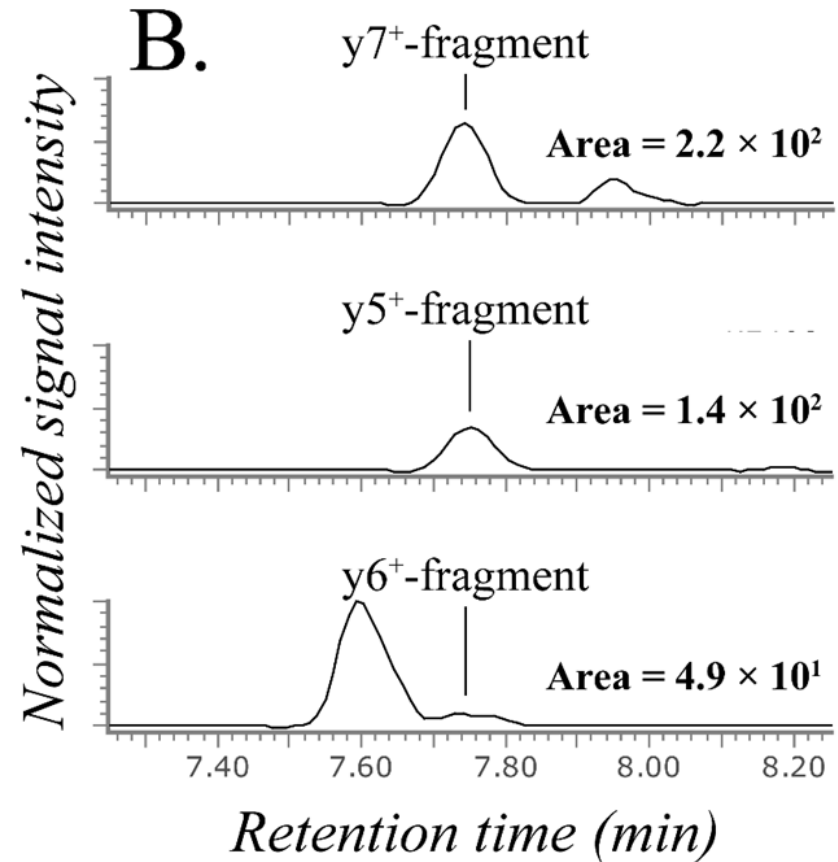


# Affimer-Based Enrichment vs. Antibody-Based Enrichment (1)

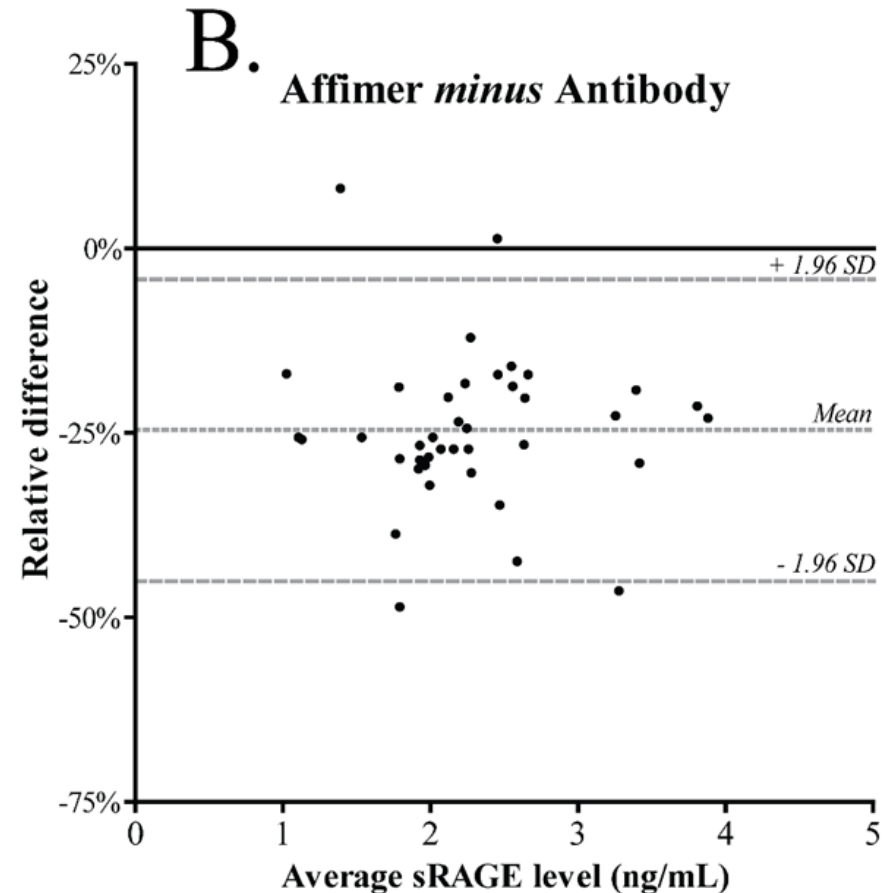
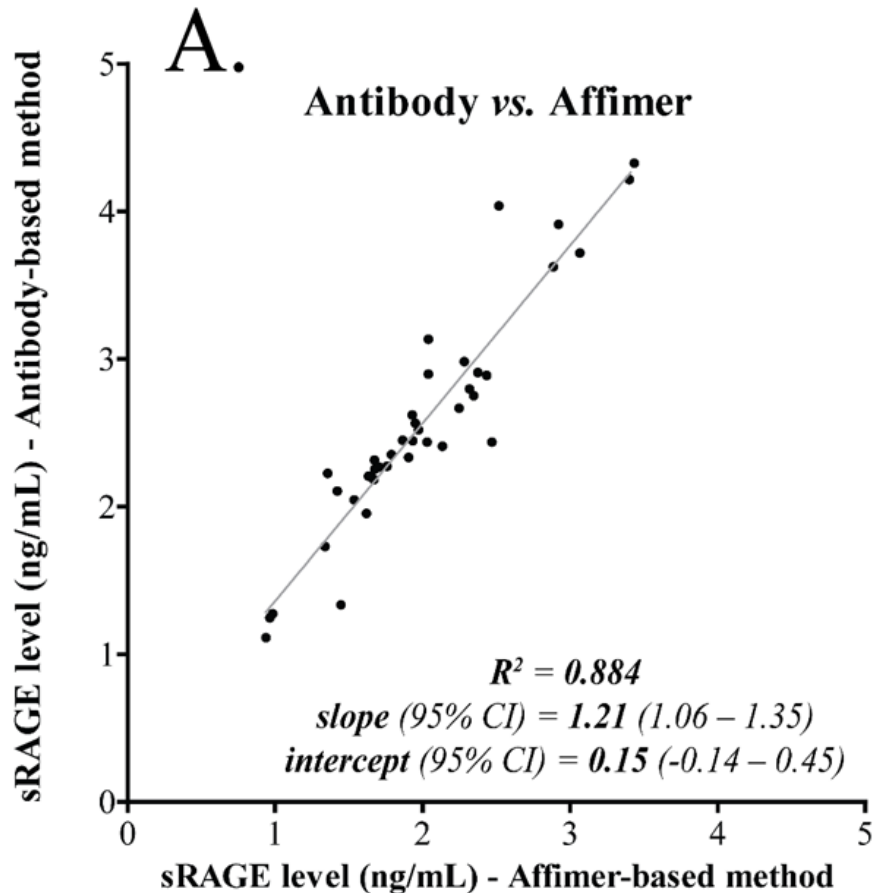
## Affimer



## Antibody



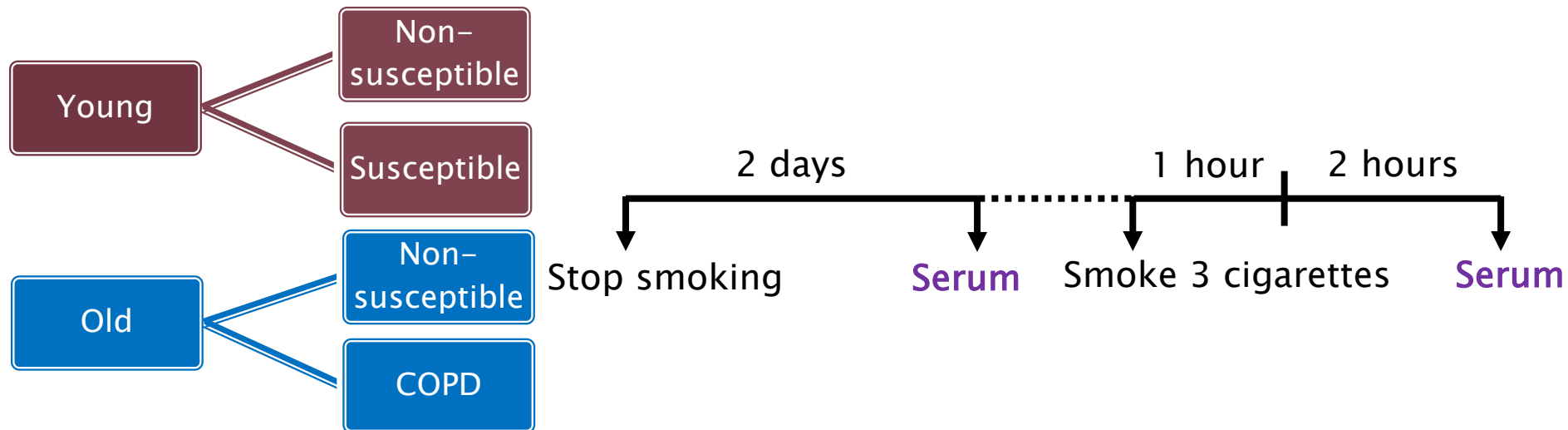
# Affimer-based Enrichment vs. Antibody-based Enrichment (2)



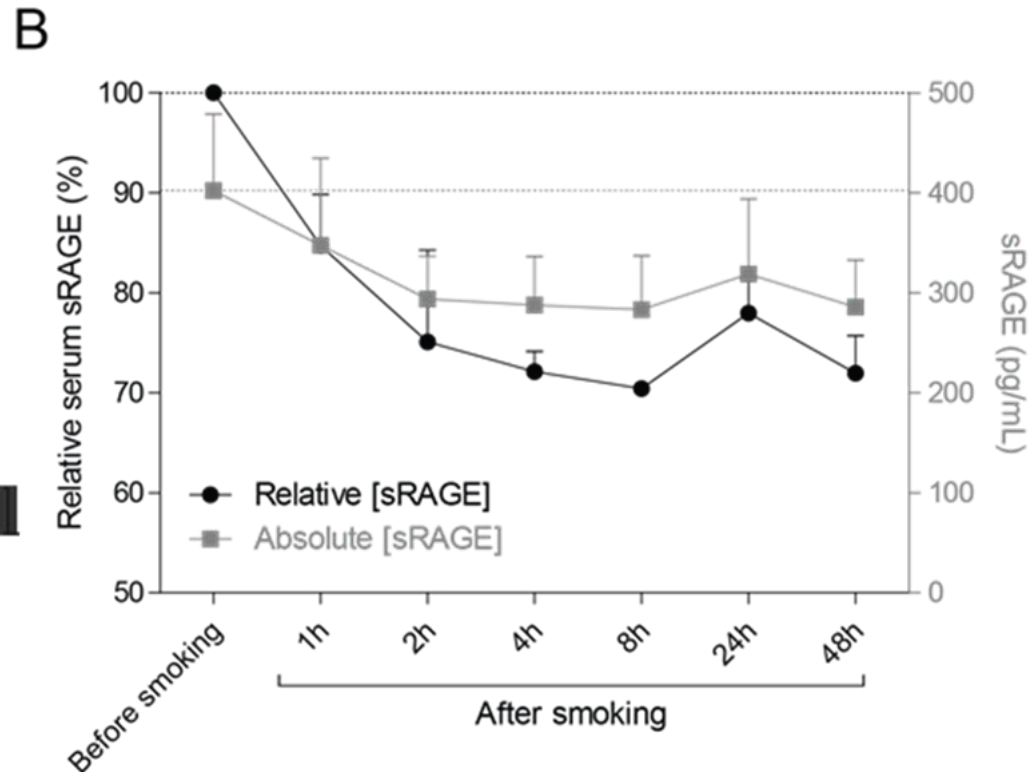
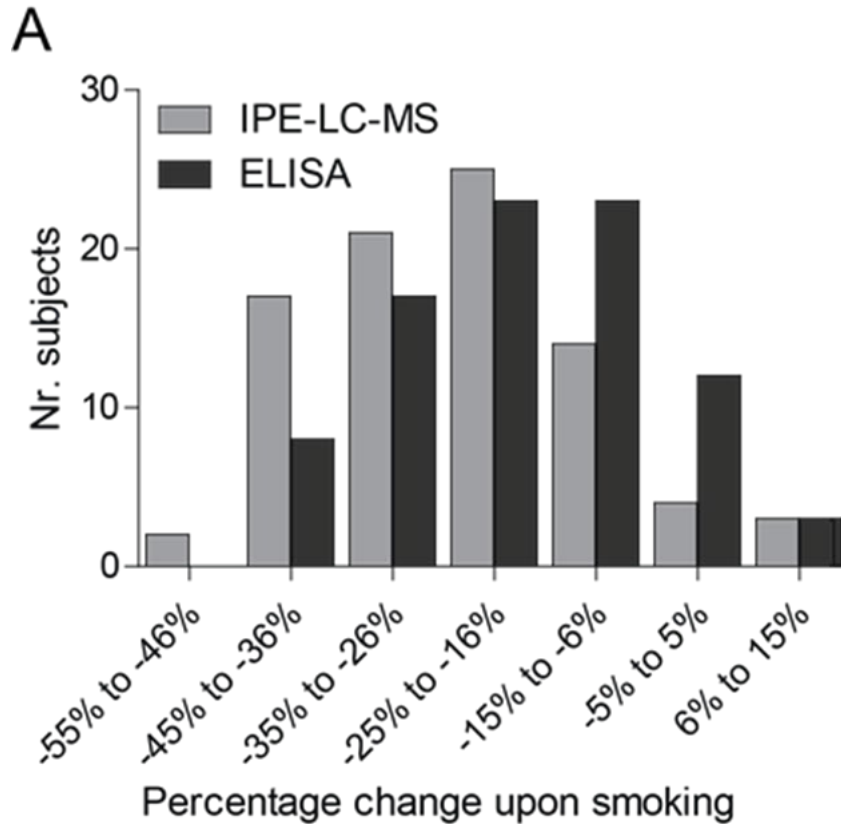
# sRAGE and acute smoking

## Serum samples

- ▶ COPD patients GOLD I–IV
- ▶ Healthy controls young (20–40 y) or old (40–80 y) that are either smokers or non-smokers.



# Confounding factor: acute smoke exposure



# Confounding factors: urinary desmosine

Independent Variables	Dependent Variable: InDES	
	$\beta$	<i>p</i> -value
<b>Age (year)</b>	0.350	2.75 E-12
<b>Gender (male/female)</b>	0.301	4.64 E-10
<b>BMI (kg/m<sup>2</sup>)</b>	0.121	0.012
<b>Current Smoking (vs. Never Smokers)</b>	0.436	7.97 E-9
<b>Former Smoking (vs. Never Smokers)</b>	0.280	1.59 E-4

The standardized coefficient ( $\beta$ ) and corresponding regression significance (*p*-value) are shown.

BMI=body mass index

# Summary & Conclusions

---

- Biomarker discovery and validation studies suffer from a lack in well-characterized biochemical reagents and a lack in consistent guidelines (such as are available for regulated, quantitative bioanalysis) as well as from a lack of compliance with such procedures/guidelines when available.
- Preanalytical factors may affect sample quality and ultimately the reliability of results.
- Confounding factors need to be taken into account when it comes to designing biomarker discovery and validation studies.
- Data (pre)processing as well as the ensuing statistical analysis may affect the final results considerably even when the (pre)processed data were identical.



# Acknowledgements



## ***Biomarker Development Center***

Alain van Gool (Radboud UMC)

Theo Luider (EMC)

Nick ten Hacken (UMCG)

## ***Preanalytics & Confounders***

Natalia Govorukhina

Therese Rosenling

Sara Ongay

## ***IGF-1 & sRAGE***

Frank Klont

Daan Pouwels (UMCG)

Marrit Hadderingh

Stefan Bakker (UMCG)

## ***Data (Pre)processing & Statistics***

Christin Christin

Berend Hoekman

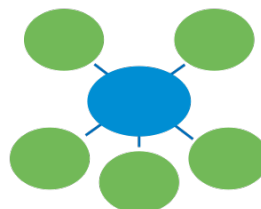
Frank Suits (IBM)

Peter Horvatovich

***University of Groningen***



Applied and  
Engineering Sciences



**Biomarker  
Development  
Center**



**umcg**

# The Analytical Biochemistry Research Group

