# Velocity DIA: Enhanced data acquisition and data evaluation strategies for Orbitrap research

# and

# Introducing WWA and Chimerys

Dr Joanna Kirkpatrick
Proteomics TSS

X-Omics Festival 17/4/2023

The world leader in serving science

# Why quantify proteins?

**More than just IDs - adding quantitative information makes the difference**

Confident characterization of differences in biological systems needs exact measurements for hypothesis testing of predictive theoretical models

Quantitative Proteomics

Translational Research

Biomarker Discovery

Single Cell Proteomics

**Accuracy**

determine correct protein abundance

**Precision**

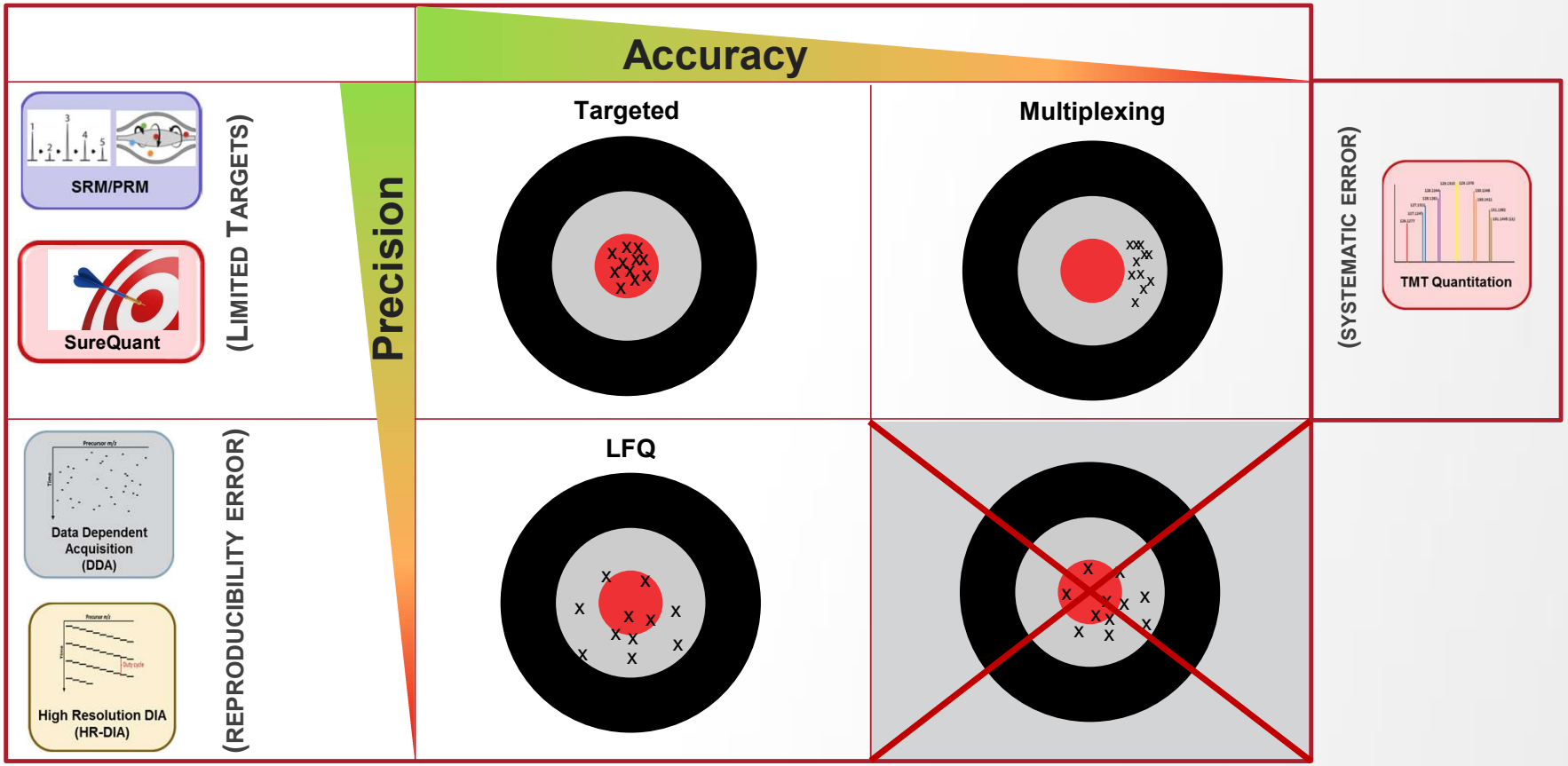
reduces number of measurements

**Throughput**

yields statistical significance

# Choices for proteomic quantitation

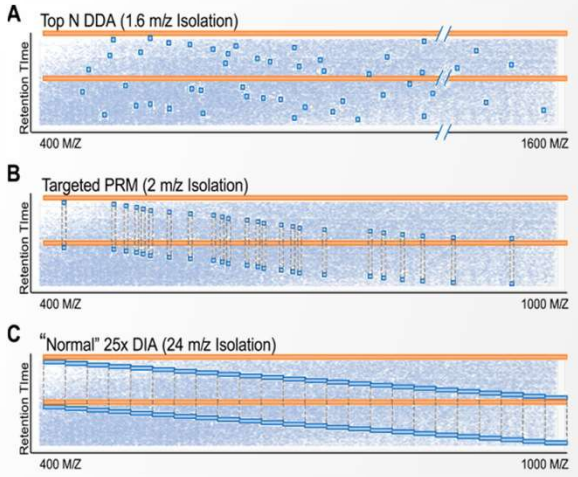
## Each method is a fit-for-purpose assay

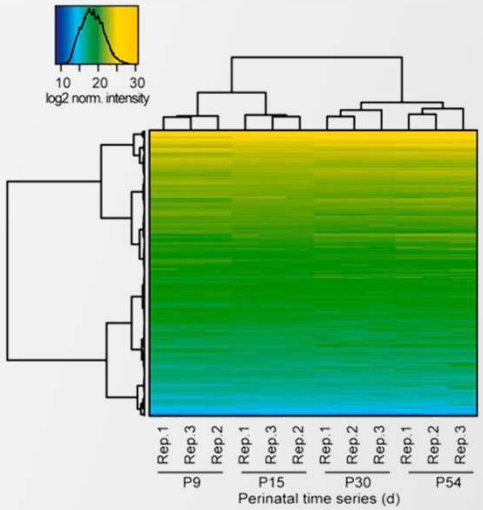


# High throughput data-independent acquisition platform

## Why use DIA for large scale proteomics?

- Large sample cohorts need to be analyzed in clinical studies
- Confident and reproducible identification and quantitation is needed
- Missing values need to be as minimal as possible
- High throughput is needed while maintaining accurate quantitation of as many proteins as possible



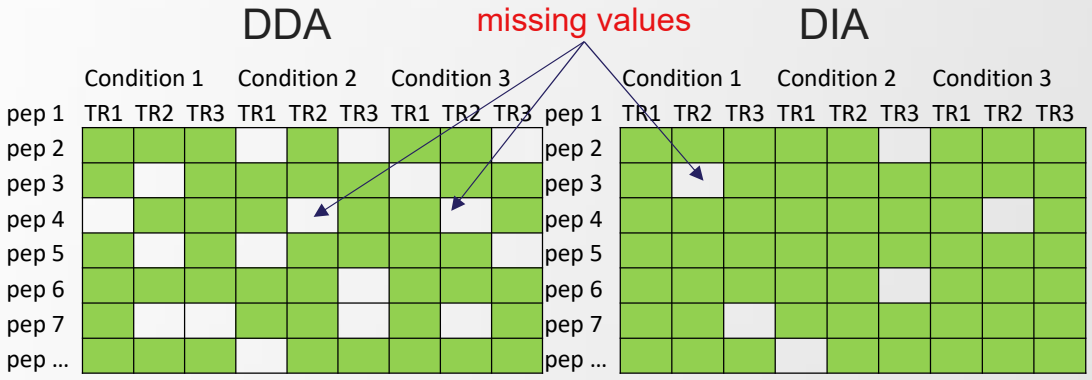




Figure adapted from Bruderer et al., Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results, MCP 2017, doi 10.1074/mcp.RA117.000314

# Principles of data-independent acquisition

## A fine balance between protein IDs, quantitation accuracy, and throughput

- In DIA mode, all ions isolated within a certain mass range are fragmented under equal conditions

- Important parameters include:
  - Isolation width
  - Mass range of isolation
  - Resolution in DIA scans
  - Window overlap
  - Collision energy

Determine speed, proteome coverage and quantitation accuracy

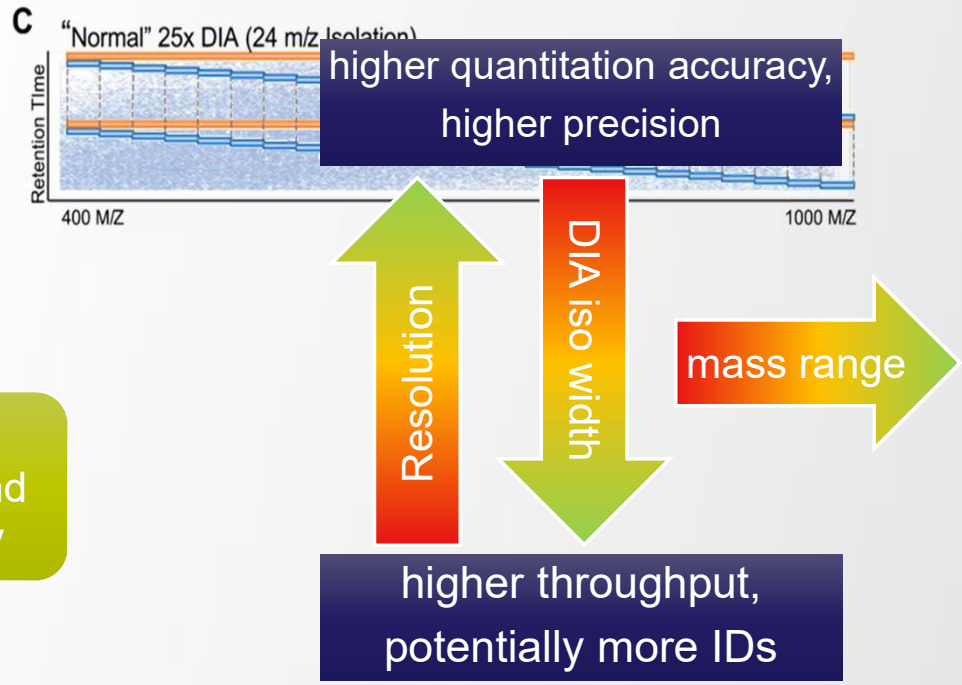


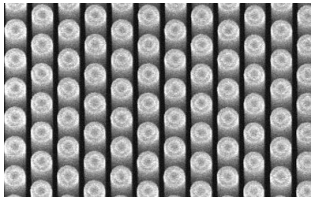

Figure adapted from Pino et al., Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries, MCP 2020 (https://doi.org/10.1074/mcp.P119.001913)

# High-throughput DIA workflow setup

## Workflow for high-throughput label-free quantitation and proteome

Data Independent Acquisition (DIA)

Thermo Scientific™ Vanquish™ NEO UHPLC



Thermo Scientific™ µPAC™ NEO 50cm

Thermo Scientific™ EASY-Spray™ Nano Source

Thermo Scientific™ Orbitrap Exploris™ 240 or 480 MS

Biognosys Spectronaut™

**Quantitation**

Accuracy | Precision

robust
reproducible
consistent

**Throughput**

(sample per day)

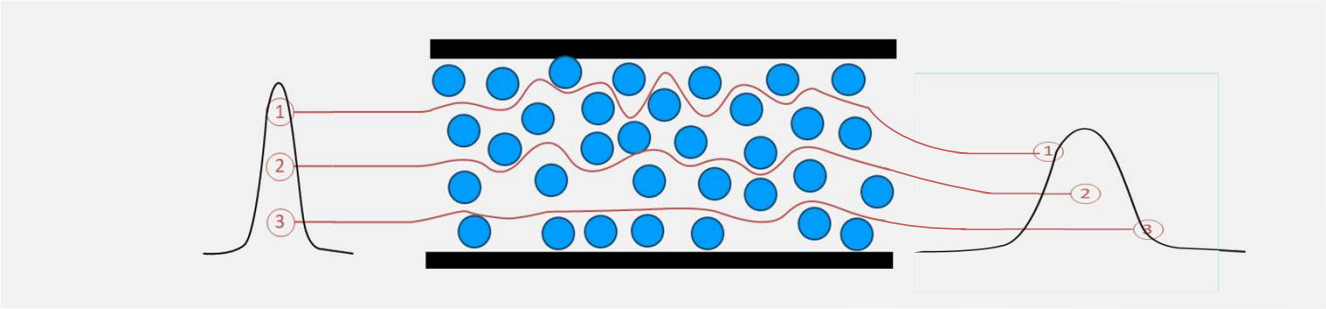
**Identifications**

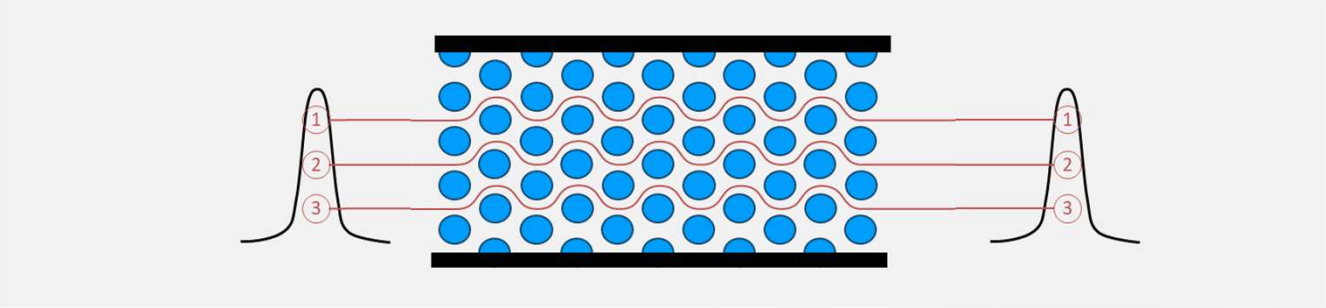
# peptides/proteins

Spectronaut is a registered trademark of Biognosys

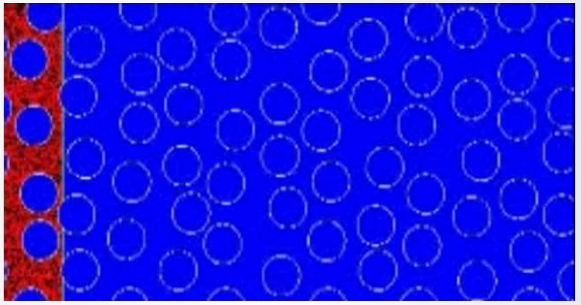
# Characteristic of the µPAC HPLC columns

## Perfect order in chromatography with micropillar array-based technology

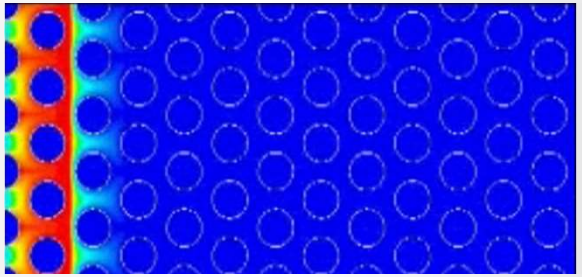



Disorder – Packed bed



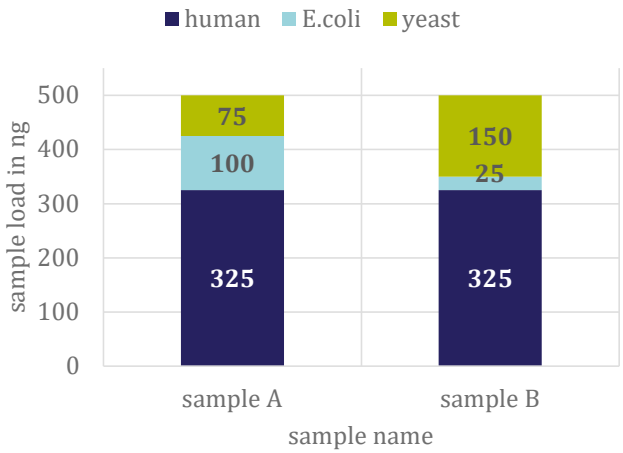
Order – Pillar Array



Reduced eddy-dispersion results in sharper peaks – higher intensity

# Experimental design

## Sample mixtures with different ratios

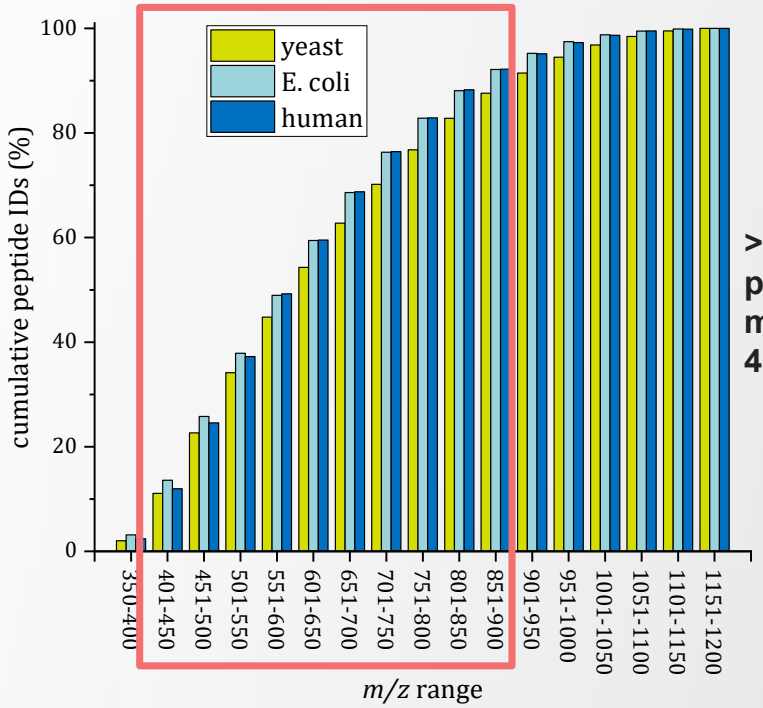



| species | ratio A:B |
|---|---|
| human | 1:1 |
| yeast | 1:0.5 |
| E.coli | 1:4 |

***Three-Proteome Mix***

- High human background levels
- 30 min gradients optimized DIA
- Spectronaut 16 directDIA™ processing (without library)

## How large does the mass range have to be?

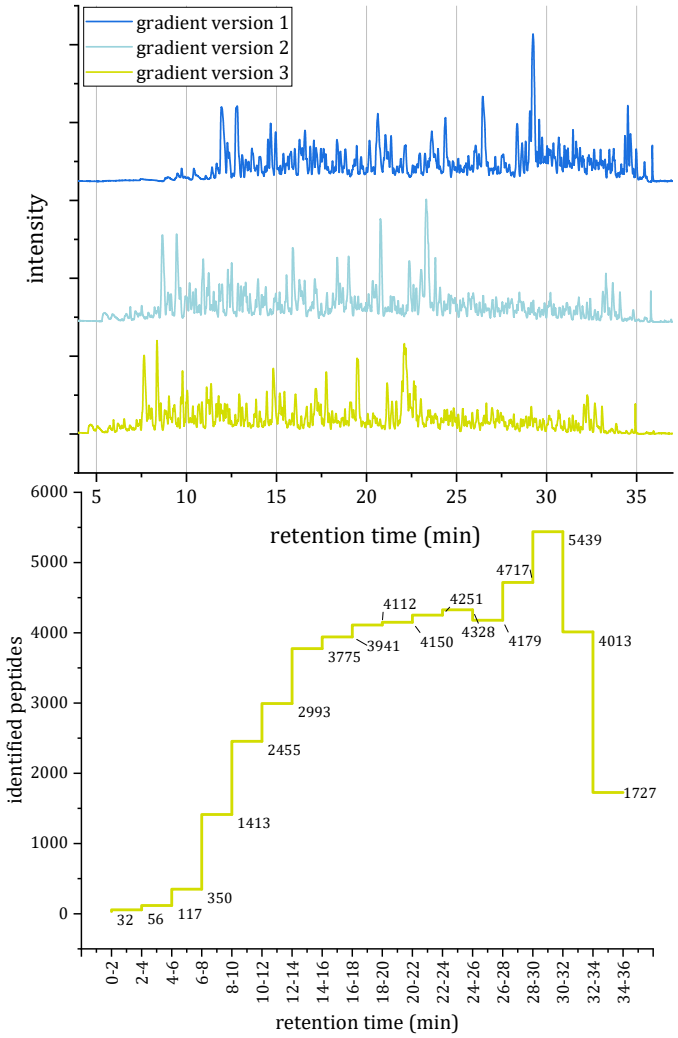



**>90% of identified peptides in the mass range 400-900 Th**

- Mass range and window size determine the duty cycle time
- 120 min DDA runs for all three species
  - mass range 350 – 1,200 Th

# Optimization of LC gradient and flow rate

## Optimal conditions for the evenly distributed elution and IDs should be determined




**LC method**

- Direct injection setup
- Flow a bit higher than usual
- Gradient optimized for µPAC Neo column

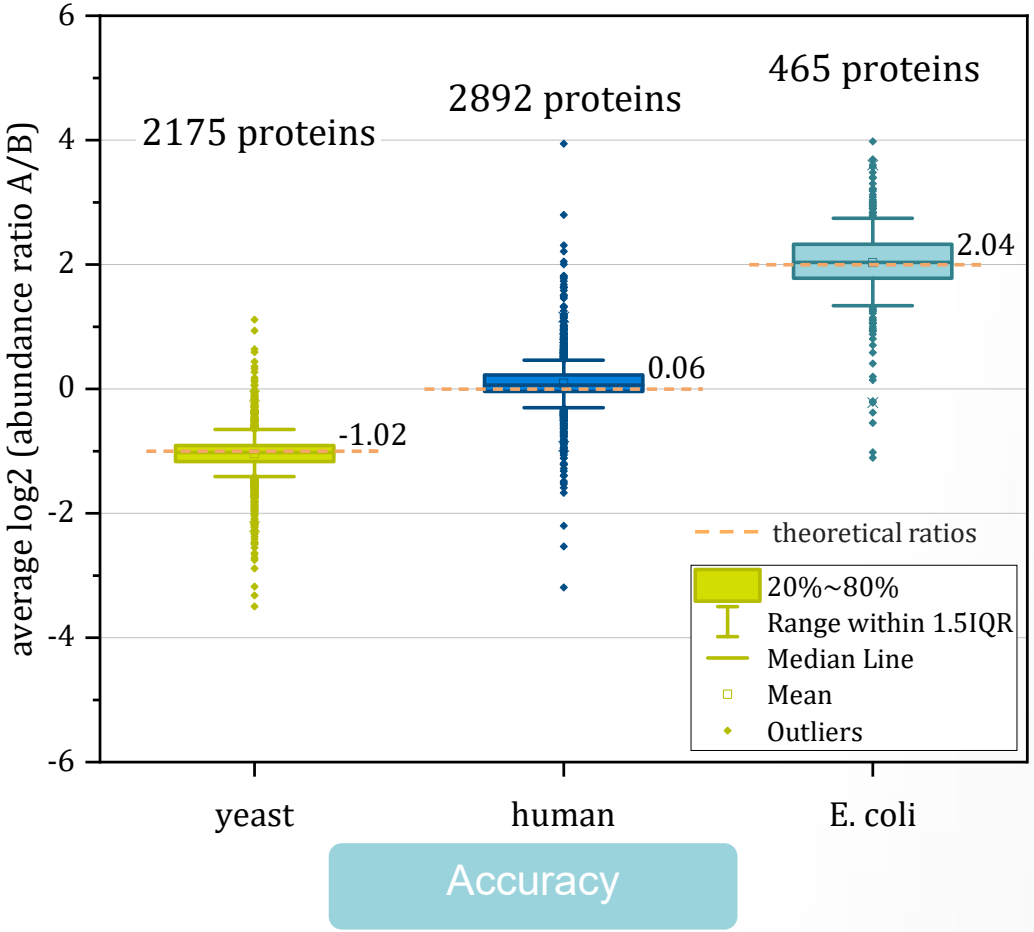
| No | Time | Duration [min] | Flow [µl/min] | %B | Volume [µl] | No. of Column Volumes |
|---|---|---|---|---|---|---|
| 1 | 0.000 | Run | | | | |
| 2 | 0.000 | 0.000 | 0.350 | 4.0 | 0.00 | 0.00 |
| 3 | 22.500 | 22.500 | 0.350 | 30.0 | 7.88 | 5.32 |
| 4 | 30.000 | 7.500 | 0.350 | 45.0 | 2.63 | 1.77 |
| 5 | 30.000 | Column Wash | | | | |
| 6 | 30.100 | 0.100 | 0.350 | 97.5 | 0.04 | 0.02 |
| 7 | 33.000 | 2.900 | 0.350 | 97.5 | 1.02 | 0.69 |
| 8 | 33.100 | 0.100 | 0.350 | 4.0 | 0.04 | 0.02 |
| 9 | 39.000 | 5.900 | 0.350 | 4.0 | 2.07 | 1.40 |
| 10 | 39.000 | Stop Run | | | | |
| 11 | 39.000 | Column Equilibration | | | | |

**MS DIA method:**

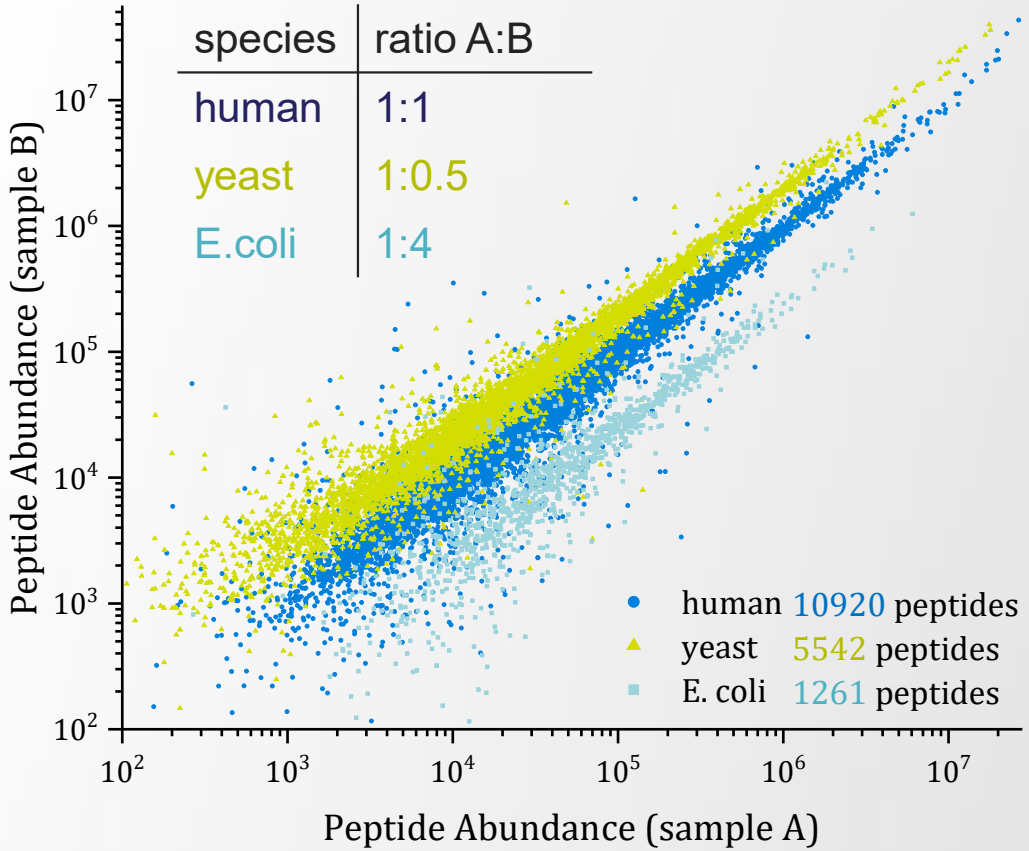
- MS1 resolution: 60k
- MS1 AGC target: 300%
- MS2 (DIA) resolution: 15k
- MS1 mass range (*m/z*): 400-900
- Isolation width: 12 Th
- DIA scan range (*m/z*): 145 - 1,450
- MS2 AGC target: 800%
- HCD NCE: 30

# Quantitation accuracy in three-proteome mixtures

## LFQ benchmark: a measure for quantitation accuracy with Spectronaut 17

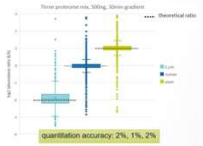



Accuracy



# Optimized HR-DIA on Orbitrap Exploris 240/480 MS is the right combination for your lab

## Vanquish Neo UHPLC | µPAC Neo column | Orbitrap Exploris 240/480 MS

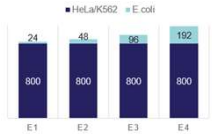



**Quantitation**
Accuracy |Precision

**Robustness**
reproducible performance

robust
reproducible
consistent

**Throughput**
samples per day




**Identifications**
# peptides/proteins

| | IDs |
|---|---|
| Peptides | > 94,000 |
| Proteins | > 12,500 |
| Protein Groups | > 9,500 |

# Summary of Velocity DIA workflow

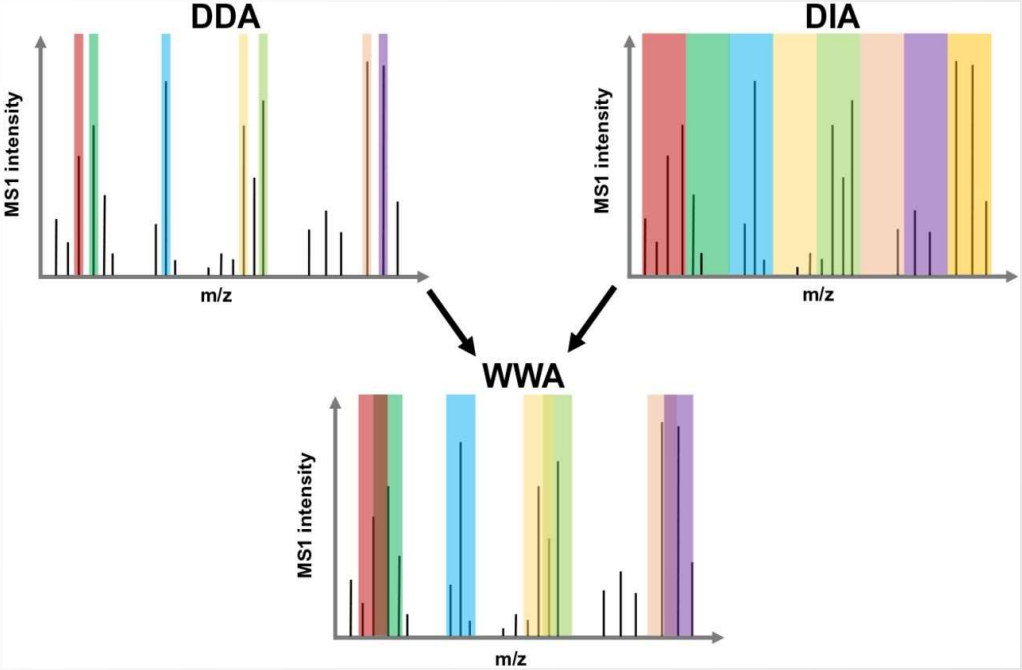
**Orbitrap Exploris 240/480 provides excellent quantitation and robust proteome coverage**

- The combination of Vanquish Neo, µPAC Neo column and Orbitrap Exploris 240/480 MS represents an ideal workflow for high-throughput label-free proteome ID and quan by DIA
- A short (30 min) gradient was developed, providing high throughput
- directDIA provides deep proteome coverage with > 99% data completeness
- DIA data is **high quality**, leading to excellent **quantitation** results as demonstrated with mixed proteome experiments, where median CVs of quantified peptides below 7 % and high ratio accuracy were obtained

# Introducing CHIMERYS and Wide-Window Acquisition (WWA)

- Limited input sample amounts - results in low proteome coverage. Even employing DIA, with low signal, many of the MS1 windows selected may contain few to no precursors of sufficient intensity to generate good quality MS2.
  - Solution: Increase injection times to allow more ions to accumulate for MS2
  - Problem: lower duty cycle
- Increase resolution in Orbitrap at no cost to cycle time = overkill for DDA on single precursor
  - Solution: Increase precursor isolation width in DDA mode
  - Problem: Generates Chimeric MS2 spectra (but with high resolution)
  - Benefit: Only regions of MS1 which are populated get selected for MS2

  RESULT:
- Generation of high-resolution MS2 spectra containing complex ion populations that can be resolved, making the most of the long injection times
  - Problem: How to search this data?
  - Solution: CHIMERYS software developed by MSAID (Munich, Germany) and included with Proteome Discoverer 3.0 is employed, which can resolve >10 precursors per MS2 spectrum.




WWA serves as a hybrid between DDA and DIA with a specific precursor being isolated for fragmentation (as with DDA), while the wide isolation windows allow for co-fragmentation and analysis of untargeted neighbouring precursors (as with DIA).

*Figure adapted from* Truong *et al. BioRxiv (2022) https://doi.org/10.1101/2022.10.18.512791*

# CHIMERYS: rethinking the identification of MS2 spectra

- Integrated into Proteome Discoverer 3.0 software.

CHIMERYS powered by INFERYS

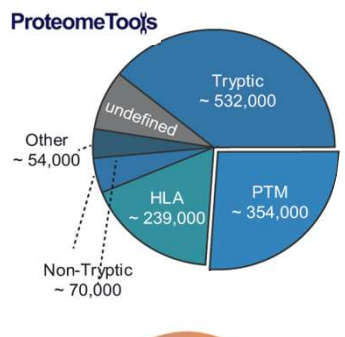
Deep learning pipeline

HQ spectra

Data selection

GPU-based RNN/CNN training

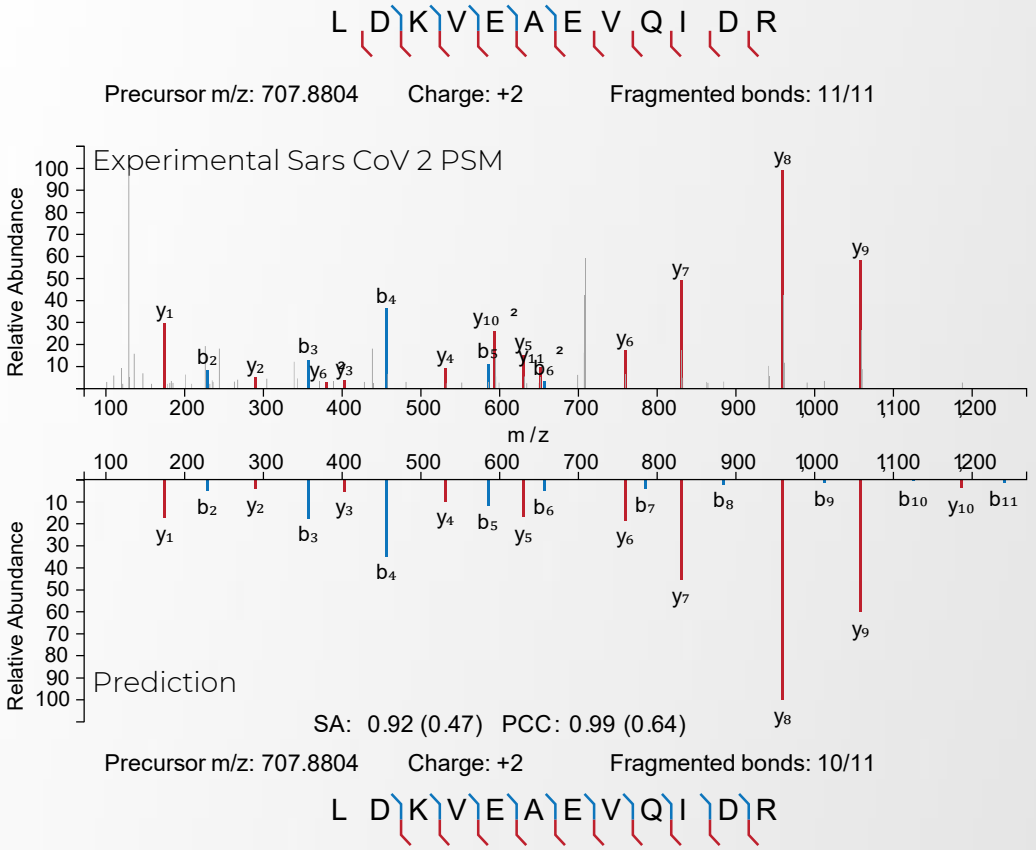
Prediction models




INFERYS

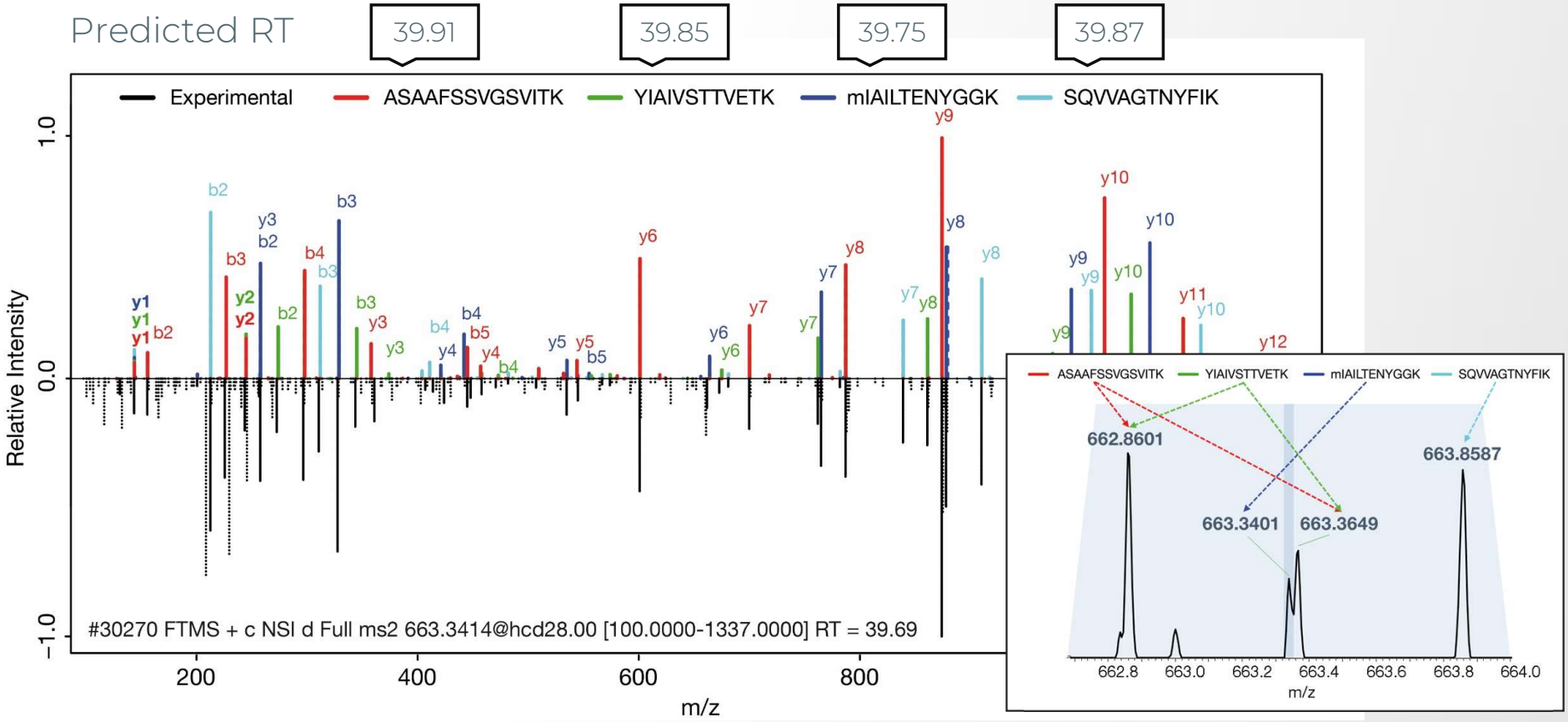
Advances in deep learning combined with high quality training data = INFERYS deep learning framework.

## Accurate prediction of spectra with the help of deep learning

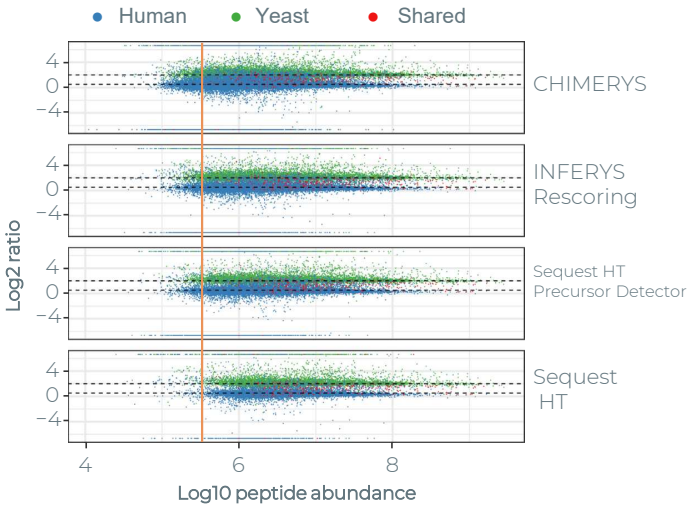


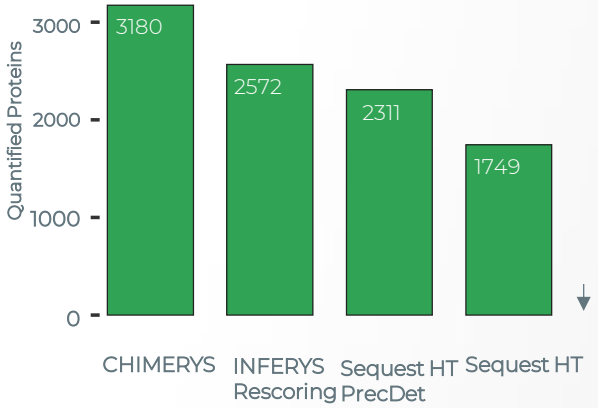
# Chimeric spectra in proteomic data: a novel, AI-based approach

Intensity-based deconvolution of chimeric spectra based on accurate predictions.

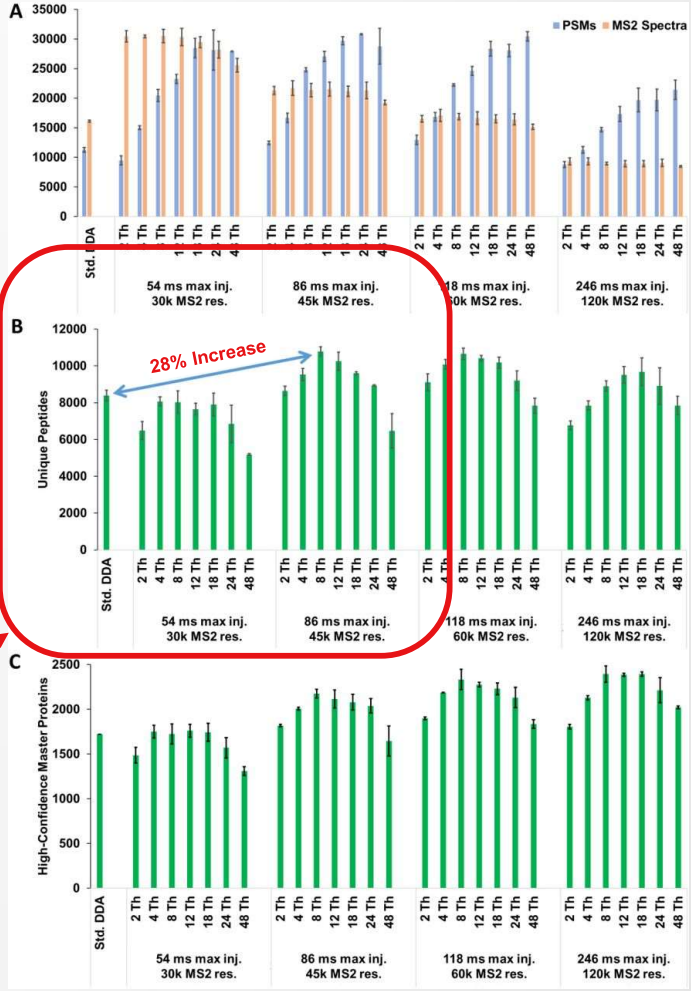


# CHIMERYS ’ sensitivity enables the identification of low-abundant peptides




1h DDA HeLa/Yeast 250ng/125ng lysate analyzed in three replicates on an Orbitrap Exploris 480 mass spectrometer




**Parameter optimization experiments for 40-min gradients using 0.2 ng aliquots of HeLa digest.**

- A) Number of collected MS2 spectra and PSMs as a function of maximum injection time/resolution and MS2 isolation window.
- (B) Number of unique peptides identified as a function of MS acquisition settings.
- (C) Number of high-confidence master proteins (1% FDR). All identifications are based solely on MS2 identification (no MBR).

*BioRxiv (2022) https://doi.org/10.1101/2022.10.18.512791*

# Acknowledgements

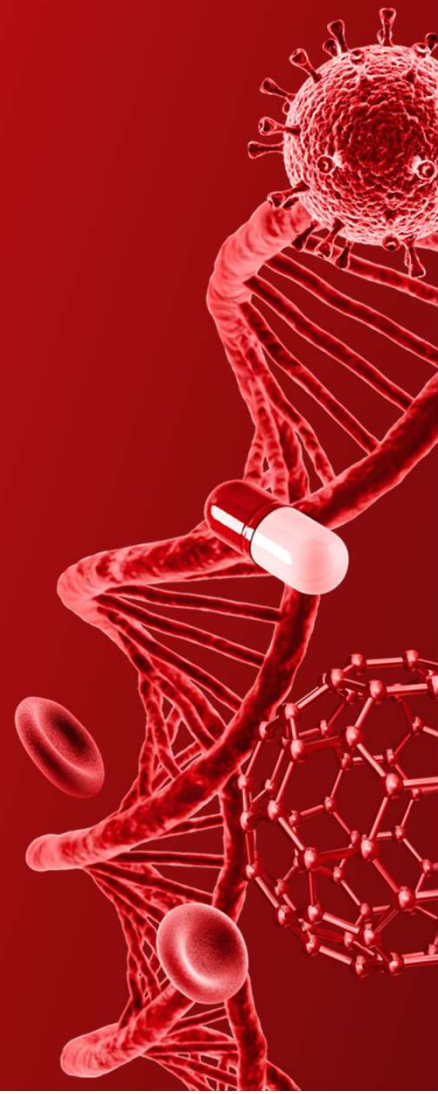

**Thermo Fisher Scientific**

- Frank Berg
- Kai Fritzemeier
- Waqas Nasir
- Pedro Navarro
- Jordi Bertran Vicente
- Katja Tham
- Tamara Vrublevskaya
- Shijun Li
- Terry Zhang
- Bernard Delanghe
- Christoph Henrich
- Alexander Harder
- Tabiwang Arrey
- Daniel Hermanson
- Yang Liu
- Amirmansoor Hakimi
- Yu Zhou
- Mick Greer
- Aaron Robitaille
- Daniel Lopez-Ferrer
- David Bergen
- Romain Huguet
- Vlad Zabrouskov
- Andreas Huhmer
- Mark Sanders
- Sergei Snovida
- Ryan Bomgarden
- Rosa Viner

**IMP Vienna:** Karl Mechtler

**Brigham Young University**: Ryan Kelly


# Thank you



**For Research Use Only. Not for use in diagnostic procedures.**