

# False Discovery Rate Estimation in Spectral Deconvolution in Top-Down Proteomics

Ayesha Feroz<sup>1</sup>, Konstantin Nagornov<sup>3</sup>, Timo Sachsenberg<sup>1,2</sup>, Yury O.Tsybin<sup>3</sup>, Oliver Kohlbacher<sup>1,2,4</sup> and Kyowon Jeong<sup>1,2</sup>

<sup>1</sup>Applied Bioinformatics, Department for Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany, <sup>2</sup>Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany, <sup>3</sup>Spectroswiss, 1015 Lausanne, Switzerland, <sup>4</sup>Translational Bioinformatics, University Hospital Tübingen, Hoppe-Seyler-Str. 9, 72076 Tübingen, Germany

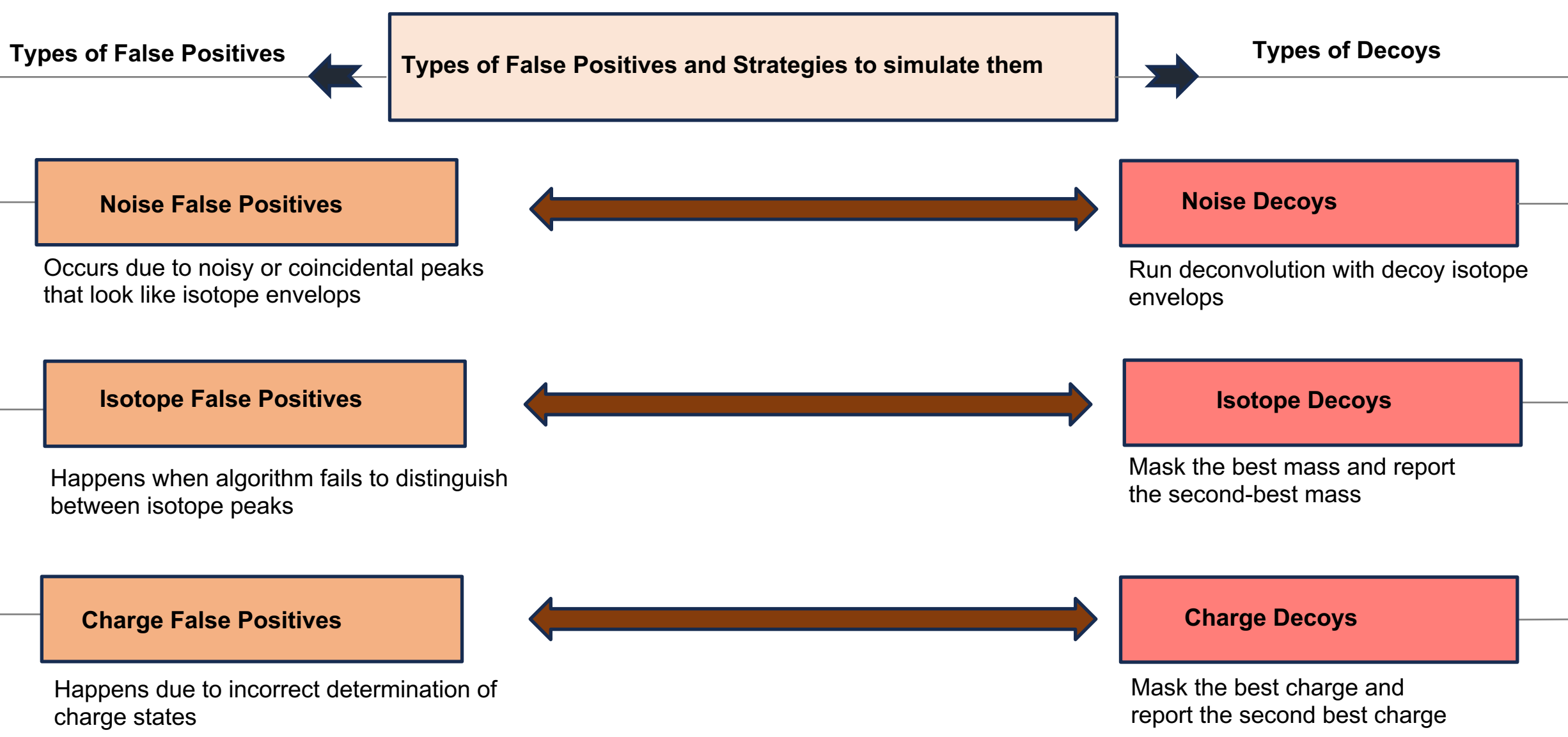
Spectroswiss

## Introduction

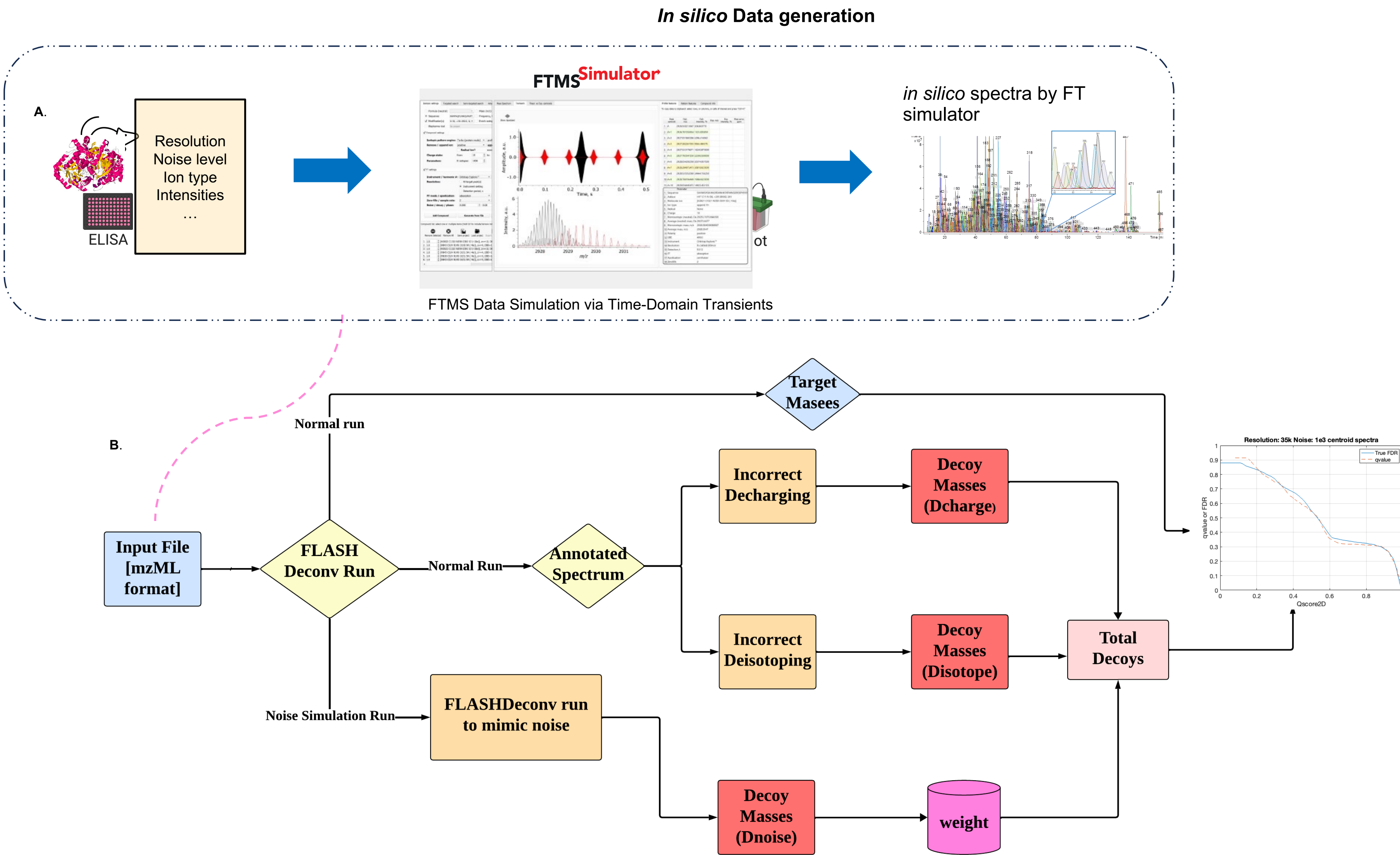
- The intricate ion structures of proteoforms present challenges in Top-Down Proteomics (TDP) analysis.
- Spectral deconvolution, crucial for simplifying TDP data, can lead to false positives if not accurately performed. Conventional methods may struggle to control this issue.
- To address it, we devised an **FDR estimation method** using decoy masses, simulating false positives. We extensively evaluated this method using both *in silico* and experimental spectra.
- Our FDR estimation method is now integrated into **FLASHViewer**, an open-source web app for TDP data visualization.

## FDR in deconvolution

FDR in deconvolution = # False deconvolved masses / # All deconvolved masses



## Method



## Results

### Evaluation on *in silico* MS Orbitrap dataset at different resolution (35k, 70k and 140k) and noise levels (1e<sup>3</sup>, 1e<sup>5</sup>)

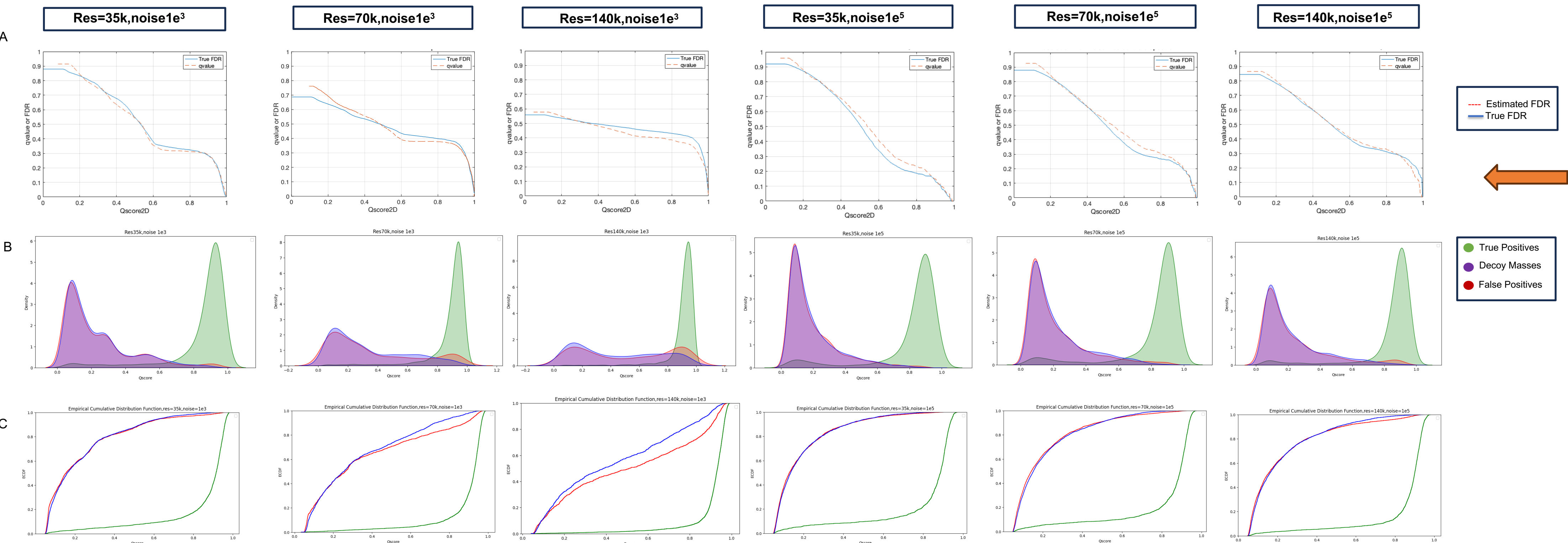
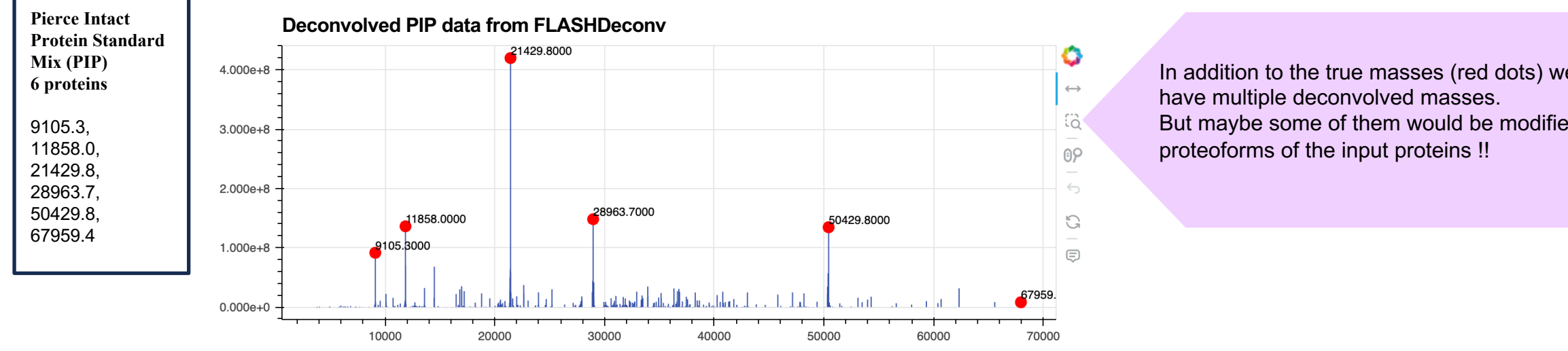


Table 1: Summary of In Silico and experimental datasets used for FDR estimation

Dataset type	Type of protein	instrument	Resolution level	Noise level	Ion types	Mass range of sequences
MS1 in silico	Human proteins	Orbitrap Fusion	35, 70, 140k@200m/z	1e3, 1e5	--	2-250KDa
MS1 in silico	Human proteins	TOF	30k@200m/z	1e3	--	2-250KDa
MS2 in silico	Carbonic anhydrase1	Orbitrap Fusion	70k@200m/z	1e3	b, y	28,870Da
MS experimental	Pierce Intact Protein Standard Mix	Q Exactive Orbitrap	--	--	--	9 - 70 kDa

### Evaluation on the experimental dataset



### Evaluation on *in silico* MS TOF dataset (30k resolution, noise1e3) and MS2 in silico datasets (70k resolution, noise level 1e3)



Figure 3: A. Plots of True Positives, False Positives, and Decoy masses of TOF MS datasets on 30k resolution and noise level 1e3 B. Plots of True positives, False Positives, and Decoy masses of MS2 datasets on 70k resolution and noise level 1e3

## Conclusion

- We have developed a novel **FDR estimation method** for deconvolution that minimizes proteoform-level biases by accounting for inherent false positives.
- Utilizing **Decoy Masses** our method effectively captures false positives, thereby enhancing the precision of FDR estimation.
- Our FDR estimation method enhances deconvolution, prevents false positive propagation, and provides a reliable framework for developing better scoring functions for proteoform identification.

## References

<sup>1</sup>Jeong, K., Kim, J., Gaikwad, M., Hidayah, S. N., Heikens, L., Schlitter, H., & Kohlbacher, O. (2020). FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell systems*, 10(2), 213–218.e6. <https://doi.org/10.1016/j.cels.2020.01.003>  
<sup>2</sup>Konstantin O. Nagornov, Anton N. Kozhinov, Natalia Gaslova, Laure Meunier, and Yury O. Tsybin *Journal of the American Society for Mass Spectrometry* **2022** 33 (7), 1113–1125 DOI: 10.1021/jasms.1c00336  
<sup>3</sup>Jeong, K., Kaulich, P. T., Jung, W., Kim, J., Tholey, A., & Kohlbacher, O. (2024). Precursor deconvolution error estimation: The missing puzzle piece in false discovery rate in top-down proteomics. *Proteomics*, 24, e2300068. <https://doi.org/10.1002/pmic.202300068>

FDR estimation is now in FLASHViewer!!  
Streamlit app for visualizing FLASHDeconv results

