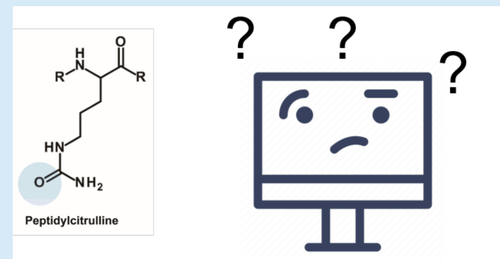
# Deep Learning Boosts Citrullination Identification in Mass Spectrometry-Based Proteomics

Rebecca Meelker Gonzalez [1,2], Wassim Gabriel [2,3], Sophia Laposchan [1,2], Erik Riedel [1,2], Mathias Wilhelm [2,3], Chien-Yun Lee [1,2]

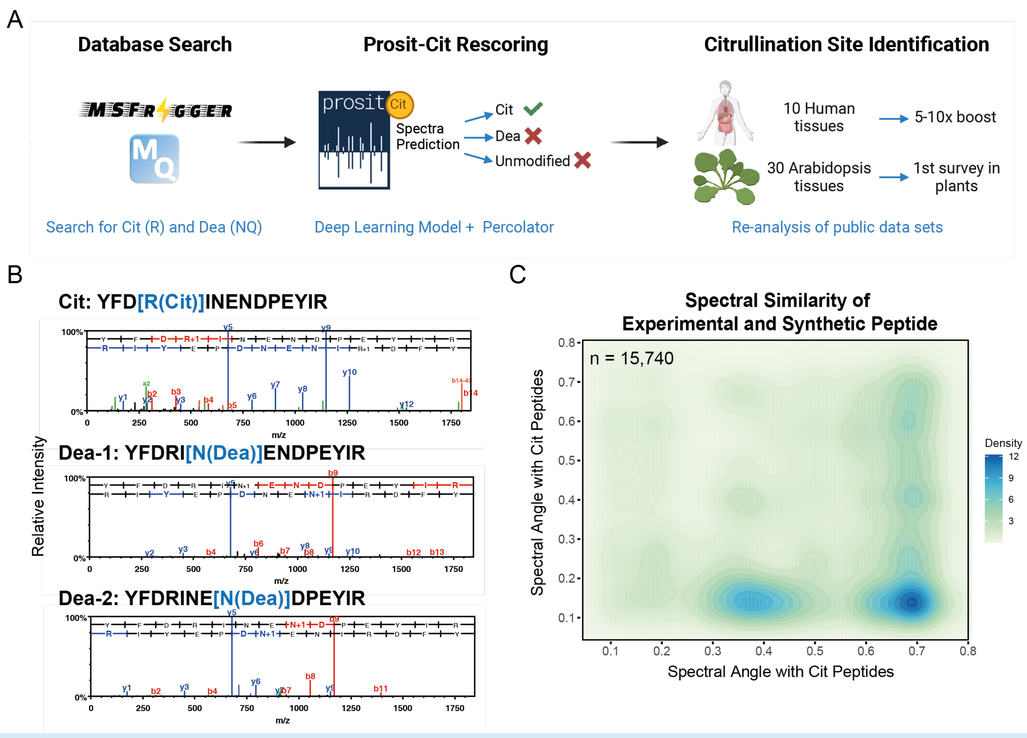
[1]Young Investigator Group: Mass Spectrometry in Systems Neurosciences, CLINSPECT-M consortium, Germany
[2]School of Life Sciences, Technical University of Munich (TUM), Germany
[3]Computational Mass Spectrometry, Technical University of Munich (TUM), Germany

- Citrullination is a largely unexplored protein PTM correlated to autoimmunity and inflammation.
- Direct identification of citrullination sites from deep proteomics profiling provides valuable insights but faces significant challenges (e.g. same mass as deamidation, $^{13}C$ isotope peaks).
- This study proposes a data analysis pipeline using deep learning model for high precision and throughput in large-scale proteomics data.



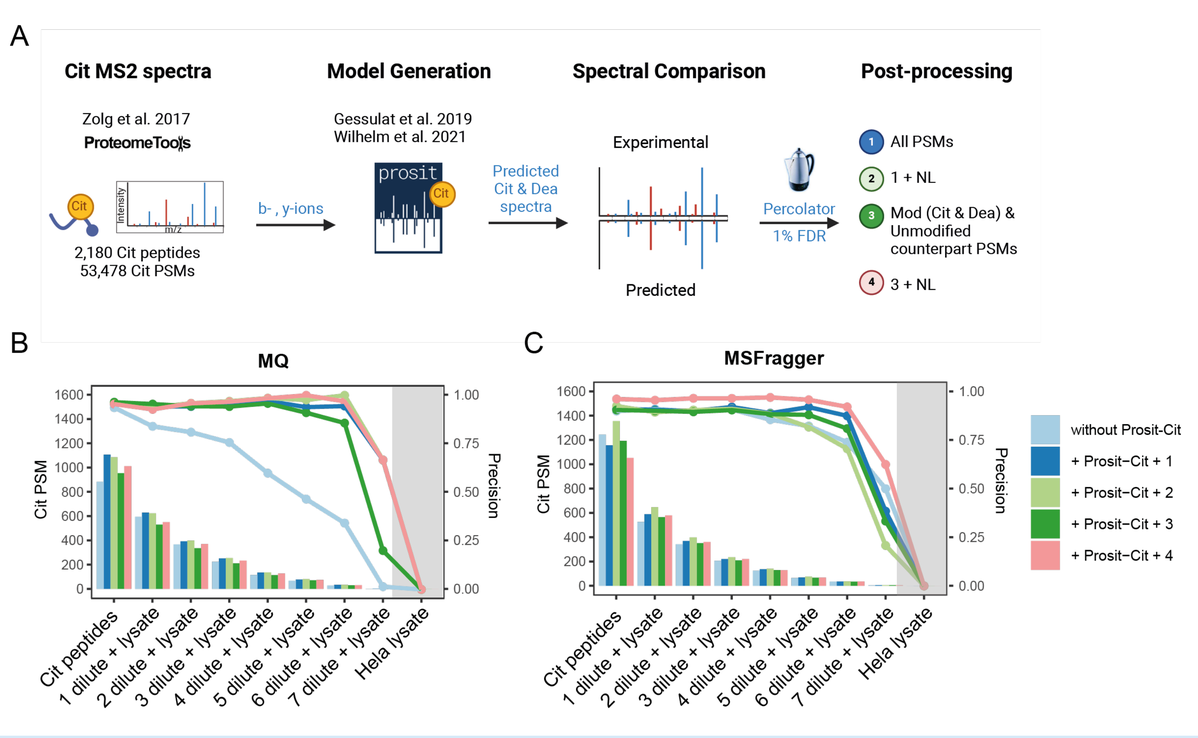

# RESULTS

## 1.Workflow and MS2 spectra similarity



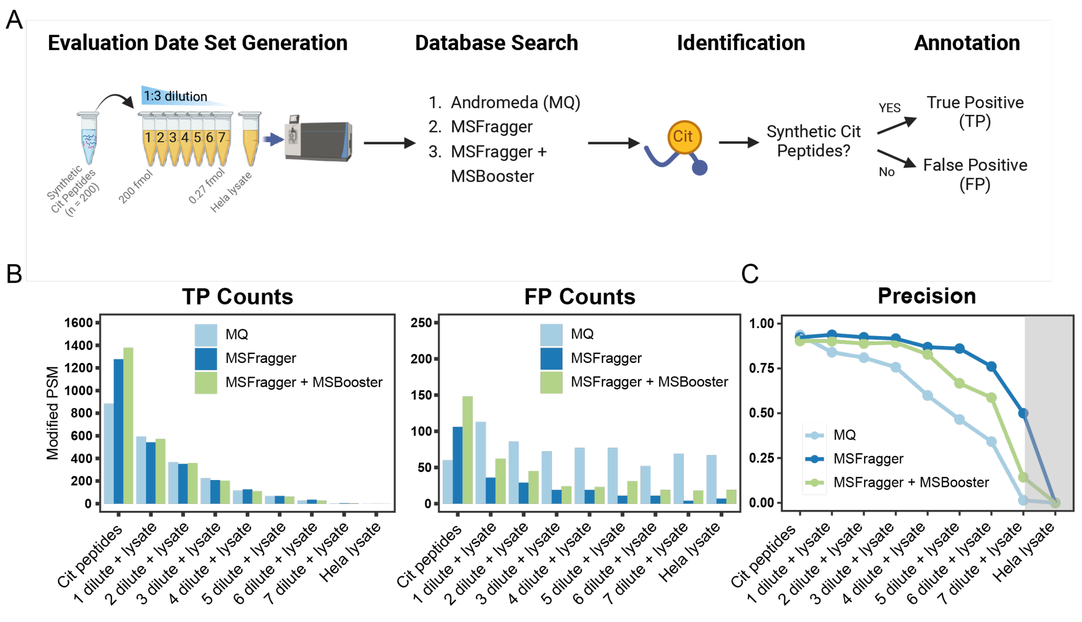

A. Workflow illustration. B. MS2 spectra of synthetic peptides containing citrullination or deamidation of gamma-adducin. C. Spectral contrast angle comparison of paired synthetic citrullinated and deamidated peptide spectra from Lee et al. Cit, citrullinated/citrullination; Dea, deamidated/deamidation.

## 3. Prosit-Cit enhances identification precision



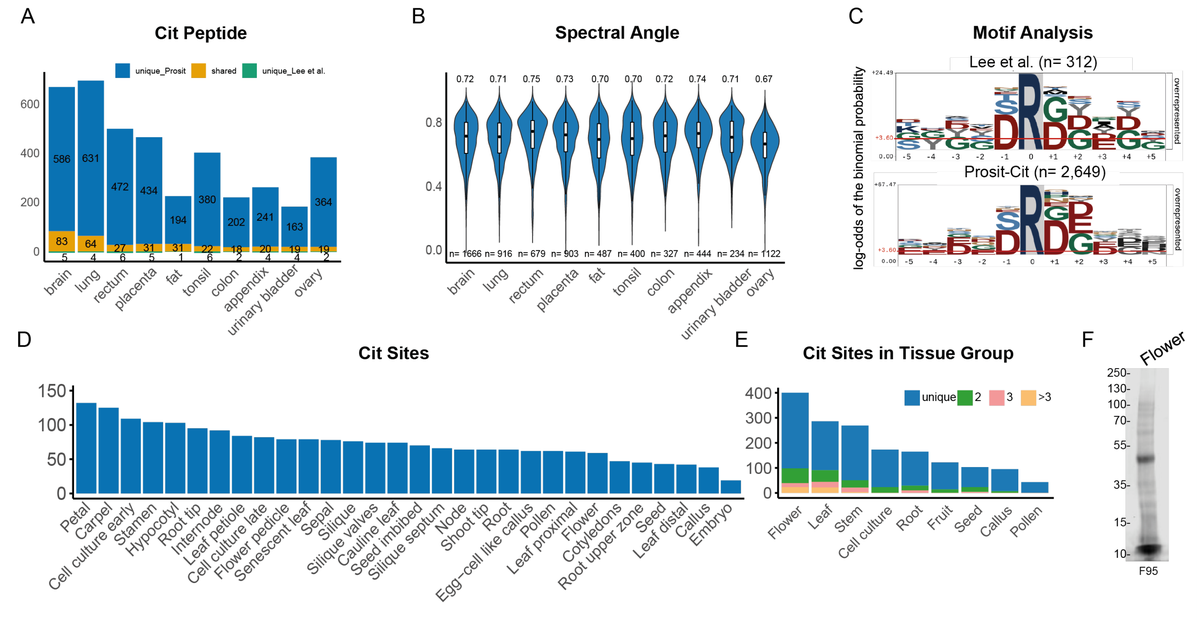

Prosit-Cit rescoring enhances the precision of citrullination identification. A. Workflow of Prosit-Cit rescoring and post-processing approaches. B-C. Citrullinated PSMs and precision of identification using different post-processing approaches following database searching with MaxQuant or MSFragger.

## 2. Comparing precision in MQ vs. MSFragger




A. Annotation of true positives and false positives at various ratios of synthetic citrullinated peptides spiked into HeLa tryptic digest. B. True and false positive counts of modified peptide-spectrum matches (PSMs). C. Precision of citrullination identification, calculated as TP/(TP+FP). MQ, MaxQuant. Cit, citrullinated.

## 4. Mining human and Arabidopsis tissue proteomes




A. Identified citrullination sites in 10 human tissue proteomes acquired by Wong et al. B. Spectral angle distributions in 10 human tissues. C. Motif analysis based on results from Lee et al. and Prosit-Cit. Numbers of citrullinated sequences used are shown. D. Identified citrullination sites in 30 Arabidopsis tissue proteomes acquired by Mergner et al. E. Identified citrullination sites in tissue groups. F. Western blot of flower tissue using the anti-citrullination antibody F95.

# CONCLUSION

- Combining MSFragger and Prosit rescoring significantly boosts identification and precision of citrullination from complex biological samples.
- We presents a precise, high-throughput pipeline for citrullination identification, marking the first survey of protein citrullination in plants and aiding understanding of protein citrullination.

The authors wish to acknowledge BMBF (161L0215) for financial support towards the project and attending this conference.

SPONSORED BY THE
Federal Ministry of Education and Research

https://mspec-sysns.de/ @msleemslab