

DEVON KOHLER

Curriculum vitae

Boston, MA 02130 • kohler.d@northeastern.edu • (508) 680-4450 • www.devonkohler.com

RESEARCH SUMMARY

Computational biologist and machine learning researcher developing AI methods for large-scale biological data. My work focuses on integrating proteomics and multi-omics data with causal modeling frameworks to estimate intervention effects and support decision-making in drug development. Experienced in building scalable statistical workflows (MSstats ecosystem) widely adopted in academia and industry and collaborating with pharmaceutical partners on real-world datasets.

EDUCATION

Northeastern University, Boston, MA

Khoury College of Computer Sciences, Advisor: Olga Vitek

2026

PhD in Computer Science

Northeastern University, Boston, MA

Khoury College of Computer Sciences, GPA 4.0/4.0

2020

Masters of Science in Data Science

Union College, Schenectady, NY

Bachelor of Arts in Economics, Minor in Political Science

2013

RESEARCH EXPERIENCE

Northeastern University, Boston, MA

Apr 2020 – Jun 2026

PhD Candidate

- Developed causal AI frameworks to predict drug toxicity by integrating proteomics and transcriptomics; designed preprocessing pipelines, constructed biologically informed gene networks, and developed custom Bayesian causal models to estimate interventional effects from observational data
- Developed novel machine learning and statistical methods, including a custom Isolation Forest implementation to quantify peak quality in DIA proteomics data, improving differential analysis accuracy while controlling false discovery rates (FDR)
- Led development of the MSstats software ecosystem, providing scalable statistical workflows for quantitative proteomics; widely adopted across academia and pharmaceutical research for differential analysis and experimental design
- Built and deployed MSstatsShiny, a graphical user interface hosted on AWS EC2, enabling non-programmers in the proteomics community to perform reproducible statistical analyses without coding expertise
- Collaborated with pharma and biotech partners (Genentech, Pfizer, AstraZeneca, Talus Bio) to design computational solutions for real-world challenges; led biweekly results-driven meetings, resulting in multiple peer-reviewed publications.

PUBLICATIONS

370 total citations, h-index 5, i10-index 5 (Google Scholar)

First Author & Co-First Author Publications

1. **Kohler, D.**, Staniak, M., Yu, F., Nesvizhskii, A., & Vitek, O. (2024). An MSstats workflow for detecting differentially abundant proteins in large-scale data-independent acquisition mass spectrometry experiments with FragPipe processing. *Nature Protocols*. 19, 2915–2938.
2. Malinovska, L.*, Cappelletti, V.*, **Kohler, D.***, Piazza, I., Tsai, T. H., Pepelnjak, M., Stalder, P., Dörig, C., Sesterhenn, F., Elsässer, F., Kralickova, L., Beaton, N., Reiter, L., de Souza, N., Vitek, O., & Picotti, P. (2023). Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications. *Nature Protocols*. 18 (3), 659-682.
3. **Kohler, D.**, Staniak, M., Tsai, T. H., Huang, T., Shulman, N., Bernhardt, O., MacLean, B., Nesvizhskii, A., Reiter, L., Sabido, E., Choi, M., & Vitek, O. (2023). MSstats Version 4.0: Statistical Analyses of

Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. *Journal of Proteome Research*. 22(5), 1466-1482

4. **Kohler, D.***, Tsai, T. H.*, Verschueren, E., Huang, T., Hinkle, T., Phu, L., Choi, M., & Vitek O. (2023). MSstatsPTM: Statistical Relative Quantification of Posttranslational Modifications in Bottom-Up Mass Spectrometry-Based Proteomics. *Molecular & Cellular Proteomics*. 22 (1), 100477.
5. **Kohler, D.**, Kaza, M., Pasi, C., Huang, T., Staniak, M., Mohandas, D., Sabido, E., Choi, M., & Vitek, O. (2023). MSstatsShiny: A GUI for Versatile, Scalable, and Reproducible Statistical Analyses of Quantitative Proteomic Experiments. *Journal of Proteome Research*. 22 (2), 551-556.

Co-Author Publications

1. Staniak, M., Huang, T., Figueroa-Navedo, A., **Kohler, D.**, Choi, M., Hinkle, T., Kleinheinz, T., Blake, R., Rose, C., Xu, Y., Jean Beltran, P., Xue, L., Bogdan, M., Vitek, O. (2025). Relative quantification of proteins and post-translational modifications in proteomic experiments with shared peptides: a weight-based approach. *Bioinformatics*. 41 (3), btaf046.

Submitted Manuscripts

1. **Kohler, D.**, Dogu, E., Bhattacharya, M., Karayel, O., Magana, M., Wu, A., Anania, V. G., & Vitek, O. (2026). Accounting for longitudinal peak quality metrics with MSstats+ enhances differential analysis in proteomic experiments with data-independent acquisition. bioRxiv preprint. Manuscript under revision at *Nature Communications*.
2. Tewari, V.*, **Kohler, D.***, Huh, S., Sachs, K., Gyori, B. M., & Vitek, O. (2026). Simulated mass spectrometry-based protein abundances enable effective experimental design. Manuscript under revision at *Journal of Proteome Research*.
3. Szvetecz, S., **Kohler, D.**, Federspiel, J., Field, D., Jean-Beltran, P., Seward, R., Suh, H., Xue, L., Vitek, O. (2026). MSstatsResponse: Semi-parametric statistical model enhances detection of drug-protein interactions in chemoproteomics experiments. bioRxiv preprint. Manuscript under review at *Molecular & Cellular Proteomics*.

*Co-first authors

INDUSTRY COLLABORATIONS

Genentech, South San Francisco, CA 2020-2026

- Collaborated with Genentech's proteomics group to develop statistical methods for PTM analysis, resulting in the MSstatsPTM software package, an accompanying peer-reviewed publication, and internal adoption of the methods at Genentech; collaboration led directly to a computational proteomics internship where I personally integrated MSstatsPTM into their production workflows
- A subsequent collaboration on integrating longitudinal quality metrics into DIA proteomics analysis, partly driven by Genentech scientists attending my ASMS 2023 talk, resulted in the MSstats+ framework, a manuscript currently under revision at *Nature Communications*, and a follow-up oral presentation at ASMS 2025; methods are in active internal use at Genentech

Pfizer, Boston, MA 2022-2026

- Collaborated on adapting and extending MSstatsPTM, MSstatsShiny, and MSstatsBig to integrate with Pfizer's internal proteomics workflows
- Provided statistical consulting on a thermal protein profiling project, contributing to analytical workflows that supported a peer-reviewed publication
- Collaborated on a chemoproteomics project resulting in a submitted manuscript at *Molecular & Cellular Proteomics* (second author) and the MSstatsResponse software package, which Pfizer uses internally

AstraZeneca, Cambridge, UK 2024-2026

- Developed causal inference methods to model drug-induced toxicity in liver cells, integrating multi-omics data and prior biological knowledge; resulted in the Causomic framework for interventional estimation

- Initiated this collaboration after Talus Bio founder Lindsay Pino attended my talk at ASMS Asilomar; recognizing a shared interest in drug response estimation in proteomics, we established an ongoing collaboration focused on causal inference methods and correct experimental design integration, leveraging Talus's unique chemoproteomics datasets

SOFTWARE & AI SYSTEMS

Lead developer and maintainer of the **MSstats ecosystem**, a suite of statistical software for quantitative mass spectrometry proteomics widely adopted across academic and pharmaceutical research laboratories.

Software in the MSstats ecosystem exceeded 70,000+ downloads in 2025.

MSstats (Bioconductor)

Core statistical framework for differential abundance analysis in label-free quantitative proteomics.

Lead Developer & Maintainer (2020–Present)

Bioconductor ranking: **Top 12%**

Downloads (2025): **21,654**

MSstatsConvert (Bioconductor)

Data import infrastructure enabling standardized analysis across mass spectrometry processing platforms.

Lead Developer & Maintainer (2021–Present)

Bioconductor ranking: **Top 13%**

Downloads (2025): **18,456**

MSstatsTMT (Bioconductor)

Statistical modeling framework for tandem mass tag (TMT) labeled quantitative proteomics experiments.

Lead Developer & Maintainer (2021–Present)

Bioconductor ranking: **Top 17%**

Downloads (2025): **10,977**

MSstatsPTM (Bioconductor)

Statistical framework for relative quantification of post-translational modifications in bottom-up proteomics.

Creator & Lead Developer

Bioconductor ranking: **Top 22%**

Downloads (2025): **8,197**

MSstatsLiP (Bioconductor)

Statistical framework for limited proteolysis (LiP-MS) experiments measuring protein structural changes.

Creator & Lead Developer

Bioconductor ranking: **Top 42%**

Downloads (2025): **4,391**

MSstatsShiny (Bioconductor)

R-Shiny graphical user interface integrated with MSstats family of R packages.

Co-Creator & Lead Developer

Bioconductor ranking: **Top 54%**

Downloads (2025): **4,113**

MSstatsBig (Bioconductor)

Computational methods for scalable analysis of large, out-of-memory proteomics datasets.

Co-Creator & Lead Developer

Bioconductor ranking: **Top 83%**

Downloads (2025): **3,135**

MSstats+ (Method Development)

Quality-aware statistical modeling framework incorporating anomaly detection and robust weighting for data-independent acquisition (DIA) proteomics experiments.

Creator & Lead Developer

Integrated within the MSstats ecosystem; manuscript under revision at Nature Communications.

Causomic (GitHub)

Bayesian causal network discovery framework integrating biological prior knowledge with probabilistic programming for multi-omics causal inference.

Creator & Lead Developer

Active development for proteomic perturbation modeling and interventional analysis.

Github Repositories

Personal: <https://github.com/devonjkohler/>

Laboratory: <https://github.com/Vitek-Lab/>

TEACHING & EDUCATIONAL LEADERSHIP

Invited instructor and organizer of international training programs in quantitative proteomics, statistical modeling, and computational biology. Taught over 20 training courses across five countries, reaching over 500 researchers from more than 25 countries worldwide

Invited Instructor

- Ubuntu Proteomics Summer School, Stellenbosch, South Africa, (2023-2026)
- FEBS European Summer School on Advanced Proteomics, Brixen, Italy, (2023-2024)
- UW Targeted Mass Spectrometry Course, Seattle, WA, (2024)
- Targeted proteomics: Experimental design and data analysis, Barcelona, Spain, (2023)
- EMERGE Webinar Series – ReSyn Biosciences, Online, (2023)
- North American Mass Spectrometry Summer School, Madison, WI, (2023)
- EMBL: Quantitative Proteomics – Strategies and Tools to Probe Biology, Heidelberg, Germany, (2023)
- ASMS Fall Workshop, Online, (2020)

Course Organizer & Lead Instructor

- ASMS Annual Meeting (Short Course). (2022-2025)
- May Institute on Computation and Statistics for Mass Spectrometry and Proteomics, Boston, MA. (2022, 2024-2026)
- US HUPO Annual Meeting (Short Course), Chicago, IL, (2023)

Teaching Assistant

- *Statistical Inference in Computer Science*. Northeastern University, Boston, MA, (Fall 2021)
- *Algorithms*. Northeastern University, Boston, MA. (Fall 2019)

CONFERENCE PRESENTATIONS

Oral Presentations

Kohler, D., Dogu, E., Bhattacharya, M., Karayel, O., Magana, M., Wu, A., Anania, V. G., & Vitek, O. (2025, Jun). Integration of longitudinal quality metrics enhances differential analysis in noisy large-scale mass spectrometry-based proteomics experiments. Oral presentation, ASMS 2025.

Kohler, D., Sachs, K., Zucker, J., Gyori, B. M., Pino, L., & Vitek, O. (2024, Oct). Causal inference enables the estimation of outcomes of interventions from observational mass spectrometry-based proteomics experiments. Oral presentation, HUPO 2024.

Kohler, D., Sachs, K., Tapley Hoyt, C., Gyori, B., Zucker, J., & Vitek, O. (2023, Nov). Computational causal inference methods and single-cell proteomics enable the estimation of interventions from purely observational experiments. Oral presentation, ASMS Asilomar Conference 2023.

Kohler, D., Sachs, K., Tapley Hoyt, C., Gyori, B., Zucker, J., & Vitek, O. (2023, Oct). Single-cell mass spectrometry-based proteomics coupled with suitable experimental design enables the application of causal inference methods to observational experiments. Oral presentation, iSCMS 2023.

Kohler, D., & Vitek, O. (2023, Jun). Bayesian statistical modeling reveals missing value mechanisms in label-free mass spectrometry-based proteomics experiments. Oral presentation, ASMS 2023.

Poster Presentations

Kohler, D., Wu, A., & Vitek, O. (2026, Jun). LLM-enabled automated data conversion for reproducible MSstats differential analysis. Poster presentation, ASMS 2026.

Kohler, D., Staniak, M., Raju, D., Szvetecz, S., Wu, A., & Vitek, O. (2024, Jun). Enhanced usability in the MSstats family of statistical analysis software for quantitative mass spectrometry-based proteomics. Poster presentation, ASMS 2024.

Kohler, D., Sachs, K., Tapley Hoyt, C., Gyori, B., Zucker, J., & Vitek, O. (2023, Mar). Single-cell mass spectrometry-based proteomics enables causal inference in observational studies. Poster presentation and lightning talk, US HUPO 2023.

Kohler, D., Kaza, M., Pasi, C., Huang, T., Staniak, M., Mohandas, D., Sabido, E., Choi, M., & Vitek, O. (2022, Jun). MSstatsShiny: A multipurpose UI for reproducible analysis of quantitative proteomic experiments. Poster presentation, ASMS 2022.

Kohler, D., Zucker, J., Tewari, V., Sachs, K., Ness, R., & Vitek, O. (2021, Oct). Explorations of causal probabilistic programming approaches for rule-based models of biological signaling pathways. Poster presentation, ProbProg 2021.

Kohler, D., Tsai, T. H., Huang, T., Verschueren, E., Hinkle, T., Choi, M., & Vitek, O. (2021, Mar). MSstatsPTM: An R software for detecting quantitative changes in post-translational modifications. Poster presentation and lightning talk, US HUPO 2021.

MENTORSHIP

- Mentor for 10+ Master's students, Northeastern University. **2021–Present**
- Mentor for 3 visiting scholars in quantitative proteomics and computational biology. **2021–Present**
- Mentor for 3 incoming PhD students in data science and computational biology, Northeastern University. **2022–Present**

PROFESSIONAL EXPERIENCE

Genentech, South San Francisco, CA

May 2021 - Aug 2021

Computational Proteomics Intern

- Integrated the MSstatsPTM statistical tool into internal proteomics pipelines using Docker, Python, and R, enabling collaborators to access advanced analyses through a graphical interface
- Analyzed proteomic experiments across multiple projects, applying statistical inference and delivering results via interactive visualization platforms
- Collaborated in a software development team using version control and issue tracking (Git, Jira) to coordinate features, conduct code reviews, and implement updates via feature branching.

The TJX Companies, Framingham, MA

Jan 2020 - Mar 2020

Data Science Graduate Coop

- Implemented a time-series ML model to forecast online sales for the Sierra brand; applied cross-validation, outlier detection, and seasonal adjustment, outperforming prior forecasting methods
- Designed a customer churn detection algorithm, modeling each customer's purchasing history as a cumulative distribution function (CDF) and flagging churn based on anomalous inter-purchase times

Cybba Inc. (formerly VeInteractive), Boston, MA

Nov 2015 – Oct 2019

Data Scientist

- Built a Python/SQL data pipeline to transform millions of daily web events into performance metrics used internally for optimization and externally to bill clients, ensuring accuracy and transparency

- Developed a boosted regression tree model for product performance prediction, using lasso regression and PCA to optimize predictors, validated via cross-validation
- Designed and executed A/B testing workflows, incorporating power analysis for sample size determination and statistical testing to assess significance of product impact

HONORS & RECOGNITION

- Invited Participant, Dagstuhl Seminar: Computational Proteomics, Schloss Dagstuhl, Leibniz Center for Informatics (2025)
- Departmental Nominee, Outstanding PhD Student Award (Experiential Learning), Northeastern University (2026)
- Barnett Institute Collaborative Award (\$20,000 research award), (2024)
- Khoury Research Apprenticeship (tuition scholarship), Northeastern University (Fall 2020)

SERVICE TO PROFESSION/LEADERSHIP

Active contributor to international proteomics community through peer review, initiatives, and mentorship.

Conference Panels

- Single-Cell Proteomics Workshop, ASMS 2025, Baltimore, MD. Panelist. (Jun 2025)
- The Promises and Perils of MS Machine Learning, ASMS Asilomar Conference, Monterey, CA. Panelist. (Oct 2023)
- Data Independent Acquisition: After the Acquisition, ASMS 2023, Houston, TX. Panelist. (Jun 2023)

Committees

- Human Glycoproteomics Initiative – 2nd Study. **2022–Present.**

Peer Reviewer

- *Nature Communications*. **2025**
- *Bioinformatics*. **2022–2025**
- *Molecular & Cellular Proteomics*. **2022–2026**
- *Journal of Proteome Research*. **2023–2025**
- *European Conference on Computational Biology (ECCB)*. **2024, 2026**

Invited Articles

- **Kohler, D.**, Staniak, M., & Vitek, O. (2022). Recent Developments in MSstats Ecosystem: A collection of statistical methods for general scalable quantitative analysis of proteomic experiments. *HUPO News*. <https://hupo.org/News/12904194>.