Joshua A. Klein, Ph.D.

Senior Scientist II, Bioinformatics @ Gritstone bio

EMAIL: joshua.adam.klein@gmail.com
PHONE: 845-891-9438
LOCATION: Boston, Massachusetts, USA
LINKEDIN: in/joshua-klein-a6155956/
GITHUB: mobiusklein

Education

2013 – 2019	Ph.D., Bioinformatics *Algorithms for Integrated Analysis of Glycomics and Glycoproteomics by LC-MS/MS*	Boston University
2008 – 2012	Bachelor of Arts, Computer Science *with honors* Bachelor of Arts, Neuroscience	Vassar College

Research & Work Experience

Scientist 2019 – 2022 Senior Scientist 2022-	> Integrated immunopeptide identifications from multiple mass spectrometry search algorithms to characterize a large cohort of human leukocyte antigen (HLA) class I and II datasets from internal and published sources. Used a collection of canonical and non-canonical proteomes for comparison of identification rates and tradeoffs. Developed large-scale spectrum clustering and archiving tools to track antigens shared across experiments and to facilitate mining of unassigned spectra for novel antigens. > Trained and benchmarked deep learning models for predicting whether a peptide would be presented by HLA class I and II receptors for oncology and infectious diseases. Oncology model further developed for use in a reproducible, good manufacturing practices (GMP) grade software pipeline for personalized cancer vaccine design. > Built target selection workflow to predict HLA-presented (shared) neoantigen and cancer-associated antigen candidates from RNAseq experiments for validation by mass spectrometry. > Contributed to the design of hybrid B-cell and T-cell vaccines for COVID-19 under clinical trial funded by NIAID and BARDA [2]. Developed tools for modeling population-scale antigen coverage to compare vaccine designs and to visualize the impact of mutations on vaccine component antigens. > Contributed to design and optimization of multi-strain HPV T-cell vaccines partnered with the Bill and Melinda Gates Foundation. Mined public HLA datasets to develop target population models to test our designs against. > Lead the development of a gene fusion calling and validation pipeline from RNAseq for use in a GMP setting combining open-source tools and internal development.	Gritstone bio
2019 –	> Created a local linear model for the prediction of glycopeptide LC-MS relative retention time. Presented as an effective validation measure against ammonium adduction and other pervasive monosaccharide substitution events [3]. An updated version linked with glycopeptide fragmentation modeling and glycan network smoothing was published in 2024 [1]. > Participate in the Proteomics Standards Initiative (PSI). Contributed to specification proposals for ProForma proteoforms notation, mzSpecLib spectral library formats, and mzIdentML. > Contribute to and maintain a variety of open-source projects for computational mass spectrometry [4,5].	Independent Work
2014 – 2018	> Built an analysis pipeline to interpret raw data from a wide range of complex liquid chromatography coupled mass spectrometry (LC-MS) experiments spanning multiple vendors and instrument classes for glycomics and glycoproteomics, including the design of per-study search spaces.	Boston University Joseph Zaia Luis Carvalho

- Devised scoring models for glycan and glycopeptide LC-MS and LC-MS/MS profiles for identifying compositions groups and structures.
- Implemented a collection of denoising, centroiding, deisotoping and charge state deconvolution, and precursor recalibration algorithms for complex mass spectra.

Selected Publications

[1] **Klein, J.**, Carvalho, L., & Zaia, J. (2024). Expanding N-glycopeptide identifications by modeling fragmentation, elution, and glycome connectivity. *Nature Communications*, *15*(1), 6168

[2] Palmer, C. D., Scallan, C. D., Kraemer Tardif, L. D., Kachura, M. A., Rappaport, A. R., Koralek, D. O., Uriel, A., Gitlin, L., **Klein, J.**, Davis, M. J., Venkatraman, H., Hart, M. G., Jaroslavsky, J. R., Kounlavouth, S., Marrali, M., Nganje, C. N., Bae, K., Yan, T., Leodones, K., … Ustianowski, A. (2023). GRT-R910: a self-amplifying mRNA SARS-CoV-2 vaccine boosts immunity for ≥6 months in previously-vaccinated older adults. Nature Communications, 14(1).

[3] **Klein, J.** & Zaia, J. (2020). Relative Retention Time Estimation Improves N-Glycopeptide Identifications by LC–MS/MS. Journal of Proteome Research, 19(5), 2113-2121.

[4] **Klein, J.**, & Zaia, J. (2019). glypy: An Open Source Glycoinformatics Library. Journal of Proteome Research.

[5] Levitsky, L. I., **Klein, J. A.**, Ivanov, M. V., & Gorshkov, M. V. (2019). Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. Journal of Proteome Research, 18(2), 709–714.

[6] **Klein, J.**, Carvalho, L., & Zaia, J. (2018). Application of Network Smoothing to Glycan LC-MS Profiling. Bioinformatics

[7] **Klein, J.** A., Meng, L., & Zaia, J. (2018). Deep sequencing of complex proteoglycans: a novel strategy for high coverage and site- specific identification of glycosaminoglycan-linked peptides. Molecular & Cellular Proteomics, 17

[8] Khatri, K., **Klein, J. A.**, Haserick, J. R. J. R., Leon, D. R. D. R., Costello, C. E. C. E., McComb, M. E. M. E., & Zaia, J. (2017). Microfluidic Capillary Electrophoresis-Mass Spectrometry for Analysis of Monosaccharides, Oligosaccharides, and Glycopeptides. Analytical Chemistry, 89(12).

[9] Khatri, K., **Klein, J. A.**, White, M. R., Grant, O. C., Lemarie, N., Woods, R. J., … Zaia, J. (2016). Integrated omics and computational glycobiology reveal structural basis for Influenza A virus glycan microheterogeneity and host interactions. Molecular & Cellular Proteomics: MCP, 13975(615).

Programming Languages and Environments

Advanced:	Python, C, Rust, JavaScript+WebDev
Intermediate:	Java, C#, C++, R, Lisp, Bash, Git, Nextflow, SQL
Basic:	Perl, Ruby

Computational Techniques and Tools

Task and Data Parallelism, Relational Database Design and Optimization, Web APIs, Semantic Graphs, Data Storage Optimization, AWS Batch, S3, Grid Engine, Docker, TensorFlow, PyTorch

Data Analysis Methods

Linear Models, Generalized Linear Models, Feature Engineering, Mixture Models, Probability Theory, Data Visualization, Deep Learning