Jeremy R. Ash

Homepage: jrash.github.io/ | jeremyryanash@gmail.com | Github: jrash | Linkedin: jeremyryanash

Experience

Johnson & Johnson AI/ML Drug Discovery Data Science — *Philadelphia, PA*

Feb 2022 – Pres

May 2025 – Pres Principal Scientist
Jul 2024 – May 2025 Senior Scientist
Feb 2022 – Jul 2024 Scientist

- Currently leading the development of an active learning platform, collaborating with a team of computational and experimental scientists automating the DMTA (Design-Make-Test-Analyze) cycle for small molecules with plans to scale to large molecules.
- Contributed to development of an end-to-end MLOps platform for ADMET prediction that deploys "global" models in LiveDesign and provides real-time support for all small molecule therapeutic area (TA) projects (50+ projects). Models use novel transformer architectures.
- Developed an active learning methodology to optimize data production for ADMET models, resulting in >40% reduction in experimental data expenditure across multiple ADMET endpoints used by the entire small molecule portfolio.
- Developed virtual screening workflows that have discovered novel progressible series advancing to hit-to-lead in multiple TA projects. Developed workflows for the following models: DNA-Encoded Library (DEL)-based machine learning models, high content imaging-based deep learning models, and chemical structure-based graphormer models.
- Developed a novel model performance evaluation tool embedded in LiveDesign, supporting an initiative to improve model usage by TA projects. Project teams use the tool extensively to assess when and where to apply their models, substantially improving model usage.
- Managed contractors and IT to establish infrastructure for MLOps and data engineering.

SAS Institute JMP Stats R&D — *Cary, NC*

Aug 2018 – Feb 2022

Oct 2021 – Feb 2022 Research Statistician Developer
Jan 2020 – Oct 2021 Research Statistician
Aug 2018 – Dec 2019 Graduate Industrial Trainee

- Researched and prototyped statistical methods for JMP statistical software using R, Python, SAS, and JSL (JMP scripting language).
- Developed statistical methods and other features in JMP using C++.
- Supported test-driven development by writing and managing suites of unit tests covering JMP's statistical methods.
- Selected projects: wavelet-based functional data analysis and design of experiments for chemometrics, multivariate control chart platform for high-dimensional process control, method for avoiding extrapolation in JMP's prediction profilers for machine learning models.

North Carolina State University Departments of Statistics and Bioinformatics — *Raleigh, NC*

Graduate Research Assistant | Advisors: Jacqueline Hughes-Oliver and Denis Fourches — *Aug 2014 – Dec 2019*

- Developed statistical methods for evaluation of virtual screening models in drug discovery.
- Developed descriptors based on molecular dynamics simulations to improve QSAR machine learning models.
- Developed chemmodlab, an R package for building cheminformatics machine learning models.

Louisiana State University Department of Biology — *Baton Rouge, LA*

Research Associate | Advisor: Jeremy Brown — *Feb 2014 – Aug 2014*

- Collaborated with the lab of Kyle Gallivan (Florida State University) developing TreeScaper, software for analysis of phylogenetic networks.
- Wrote documentation on nonlinear dimensionality reduction and network analysis methodologies for the phylogenetics community.

Skills

Core Competencies	AI/ML for drug discovery, virtual screening, ADMET modeling, active learning, cross-functional collaboration
ML & Data Science	Python ML toolkit (PyTorch, scikit-learn, chemprop), large data engineering (pandas, polars), deep learning and transformer architectures, statistical modeling, uncertainty quantification, data visualization
Comp Chem	Developing and applying cheminformatics methods, RDKit, Schrodinger Suite, KNIME, Spotfire, QSAR, DNA-encoded libraries, molecular dynamics
Programming	Python, R, SQL, AWS cloud services, development in team environments using DevOps and MLOps (e.g., SageMaker), git version control, documentation (Confluence), project management (JIRA)

Education

North Carolina State University — *Raleigh, NC*
Bioinformatics, Ph.D. | Statistics, M.S. — *Aug 2014 - Dec 2019*

The University of Texas at Austin — *Austin, TX*
Biology, B.S. — *Aug 2010 - Aug 2013*

Selected Publications

Google Scholar: https://scholar.google.com/citations?user=Au7CUSAAAAAJ&hl=en&oi=ao
ORCiD: https://orcid.org/0000-0002-8041-8524

1) Van Rompaey D, Ray Chaudhuri S, Ahmad M, Cisar J, Van Den Bergh A, **Ash JR**, Wu Z, Bryan MC, Edwards JP, DesJarlais R, Wegner JK. *Toward Dose Prediction at Point of Design.* Journal of Medicinal Chemistry, 2024, 67 (24), 22282–22290.
2) **Ash JR**, Wognum C, Rodríguez-Pérez R, Aldeghi M, Cheng AC, Clevert D-A, Engkvist O, Fang C, Price DJ, Hughes-Oliver JM, Walters WP. *Practically significant method comparison protocols for machine learning in small molecule drug discovery.* ChemRxiv, 2024. (submitted to Journal of Chemical Information and Modeling).
3) Wognum C, **Ash JR**, Aldeghi M, Rodríguez-Pérez R, Fang C, Cheng AC, Price DJ, Clevert D-A, Engkvist O, Walters WP. *A call for an industry-led initiative to critically assess machine learning for real-world drug discovery.* Nature Machine Intelligence, 2024, 6 (10), 1120–1121.
4) **Ash JR**, Hughes-Oliver JM. *Confidence bands and hypothesis tests for hit enrichment curves.* Journal of Cheminformatics, 2022, 14(1), 50.
5) **Ash JR**, King C, Lancaster L, Lekivetz R, Morgan J, Saanchi Y. *Testing the prediction profiler with disallowed combinations—A statistical engineering case study.* Quality Engineering, 2022, 34(4), 507-521.
6) Odenkirk MT, Zin PPK, **Ash JR**, Reif DM, Fourches D, Baker ES. *Structural-based connectivity and omic phenotype evaluations (SCOPE): a cheminformatics toolbox for investigating lipidomic changes in complex systems.* Analyst, 2020, 145 (22), 7197-7209. (Cover Article)
7) **Ash JR**, Kuenemann MA, Rotroff D, Motsinger-Reif A, Fourches D. *Cheminformatics Approach to Exploring and Modeling Trait-Associated Metabolite Profiles.* Journal of Cheminformatics, 2019, 11 (1), 43.
8) **Ash JR**, Hughes-Oliver JM. *Chemmodlab: A Cheminformatics Modeling Laboratory R Package for Fitting and Assessing Machine Learning Models.* Journal of Cheminformatics, 2018, 10 (1).
9) **Ash JR**, Fourches D. *Characterizing the Chemical Space of ERK2 Kinase Inhibitors Using Descriptors Computed from Molecular Dynamics Trajectories.* Journal of Chemical Information and Modeling, 2017, 57 (6), 1286–1299.
10) Huang W, Zhou G, Marchand M, **Ash JR**, Morris D, Van Dooren P, Brown JM, Gallivan KA, Wilgenbusch JC. *TreeScaper: Visualizing and Extracting Phylogenetic Signal from Sets of Trees.* Molecular Biology and Evolution, 2016, 33 (12), 3314–3316.

Patents

1) **Ash JR**, Ceulemans H, Allen S. *Systems and methods for drug discovery using cellular imaging and predictive models.* Non-provisional patent application filed Jan 31, 2025 (not published yet).
2) Parker RJ, Barker CA, **Ash JR**, Gotwalt CM. *Analytic system for interactive graphical model selection based on wavelet coefficients.* US Patent 11,361,255 B2. 2021.
3) **Ash JR**, Gotwalt CM, Lancaster LC. *Analytic system with extrapolation control in interactive graphical prediction evaluation.* US Patent 10,963,804. 2021.

Awards

1) ACS Division of Chemical Information Scholarship of Excellence for poster presentation, *Molecular Modeling of Differential ERK1/2-Ligand Dynamic Interactions*, 2018.
2) ACS Division of Chemical Information Scholarship of Excellence for poster presentation, *Cheminformatics Approach to Exploring and Modeling Trait-Associated Metabolite Profiles*, 2017.
3) ACS Editor's Choice for excellence in publication, *Characterizing the Chemical Space of ERK2 Kinase Inhibitors Using Descriptors Computed from Molecular Dynamics Trajectories*, 2017.
4) Triangle Center of Evolutionary Medicine Fellowship, 2017.
5) NCSU Graduate School Fellowship, 2014.

Selected Talks

1) *Practically significant method comparison protocols for machine learning in small-molecule drug discovery*. Polaris Webinar, Virtual 2024.
2) *Evaluating performance of global ADMET models for estimating properties within a drug discovery project's chemical series*. ACS Conference, Denver, CO 2024.
3) *Data to decisions: Machine learning and data visualization tools optimize compound selection for solubility predictive models in real-time support of small molecule discovery*. ACS Conference, New Orleans, LA 2023.
4) *Analyzing Spectral Data: Preprocessing and Beyond*. JMP Discovery Summit, Virtual 2021.
5) *Methods for Helping Users Avoid Extrapolation when Making Predictions with Statistical and Machine Learning Models*. Joint Statistical Meetings and JMP Discovery Summit, Virtual 2020.
6) *Fault Detection and Diagnosis of the Tennessee Eastman Process using Multivariate Control Charts*. JMP Discovery Summit, Virtual 2020.
7) *Structure-Based Approach to Exploring and Modeling Trait-Associated Metabolite Profiles*. ACS Conference, Boston, MA 2018.
8) *Characterizing the Chemical Space of Kinase Inhibitors Using Molecular Descriptors Computed from MD Trajectories*. ACS Conference, Washington, DC 2017.

Service

2018-Pres **Reviewer**, Scientific journals (e.g. Journal of Cheminformatics)

2018 **Organizer**, First Annual NCSU Genomic Sciences and Biomath Research Symposium

2017-2018 **President**, NCSU Genomic Sciences Graduate Student Association

References

Available upon request