Jeremy R. Ash

Bioinformatics Ph.D. · Statistics M.S.

Address: 1613 Spruce St. Apt. 4, Philadelphia, PA, 19103.

jeremyryanash@gmail.com | jrash.github.io/ | jrash | jeremyryanash | @jeremyr_ash | Google Scholar | ORCID iD | Publons

Education

North Carolina State University — *Raleigh, NC, USA*

Bioinformatics, Ph.D. | Statistics, M.S. — *Aug. 2014 - Dec. 2019*

The University of Texas at Austin — *Austin, TX, USA*

Biology, B.S. — *Aug. 2010 - Aug. 2013*

Skills

Research Areas	Machine learning, Statistics, Data science, Cheminformatics, Bioinformatics, Chemometrics
Programming	R, Python (pytorch, scikit-learn, etc.), C++, Bash, HPC, SQL, LaTeX, Schrodinger, Knime, Software development (CI/CD, Version control, Unit testing, Documentation), MLOps, Data Engineering
Computational	Machine learning, Statistical modeling (Frequentist, Bayesian, Multivariate, Non-Parametric), Simulation, Methods development, Probability theory, Numerical methods

Experience

AI/ML Drug Discovery Data Scientist

Spring House, PA

Johnson & Johnson Innovative Medicine — *Feb. 2022 - PRES*

- Member of the chemical property prediction team, developing an end-to-end MLOps platform for ADMET prediction. Platform utilizes in house transformer based NLP methods as well as other deep learning methods and serves ”global” models that are utilized across the therapeutics discovery pipeline for real time support of molecular design.
- Development of uncertainty quantification methods for property prediction and active learning methods for systematic data production to support ML models.
- Deep learning modeling methods for cell painting imaging data for chemical toxicity prediction and image based virtual screening.
- Improving tools for model evaluation that are relevant to business impact and are user friendly for chemists.

Research Statistician Developer

Cary, NC

SAS Institute, JMP Division, Stats R&D — *Oct. 2021 - Feb. 2022*

- Feature development in C++ for multivariate statistics platforms.
- Selected projects (with links): wavelet based functional data analysis and design of experiments for chemometrics, randomized SVD in PCA platform, improved utilization of BLAS in numerical routines.

Research Statistician

Cary, NC

SAS Institute, JMP Division, Stats R&D — *Jan. 2020 - Oct. 2021*

- Research and prototype methodologies in consideration for JMP, using R, Python, SAS and JSl (JMP scripting language).
- Evaluate methodologies using customer and simulation data. Present findings to development as they plan new features.
- Support test driven development by writing and managing suites of unit tests and confirming accuracy in daily software builds.
- Engage with customers through conference presentations, published research and beta feedback.
- Selected projects (with links): multivariate control chart platform for detecting outliers in high dimensional data, method for avoiding extrapolation in JMP’s prediction profilers for machine learning models.

Graduate Industrial Trainee

Cary, NC

SAS Institute, JMP Division, Stats R&D — *Aug. 2018 - Dec. 2019*

- Year round internship program. 20hrs per week with the JMP Stats R&D team on research projects.

Graduate Research Assistant

NCSU

Advisors: Jacqueline Hughes-Oliver and Denis Fourches

Aug. 2015 - Dec. 2019

- Statistical methods development for QSAR models – models predicting the bioactivity of small molecules given their chemical structure.
- Uncertainty quantification in virtual screening: confidence band and hypothesis tests methods for hit enrichment curves at extremely small testing fractions.
- Using information from molecular dynamics simulations to improve machine learning methods for drug activity prediction.
- Developing **chemmodlab**, an R package for building cheminformatics machine learning models.

Graduate Research Assistant

NCSU

Advisor: Jeff Thorne

Aug. 2014 – July 2015

- Group project with the labs of Scott Schmidler (Duke), and Jotun Hein (Oxford) hosted by The Statistical and Applied Mathematical Sciences Institute.
- Improved the biological realism of insertion and deletion in a Bayesian model for the joint estimation of phylogeny and protein structure alignment. Implemented in the software, **StatAlign**.

Research Associate

LSU

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 non-linear dimensionality reduction and network analysis methodologies for the phylogenetics community.

Undergraduate Research Assistant

UT Austin

Advisors: Harold Zakon, David Hillis, Daniel Johnston

Oct. 2012 – Dec. 2013

- Used phylogenetic methods to predict functional divergences of sodium channels during their evolutionary history.
- Studied the involvement of the prefrontal cortex in trace eyelid conditioning in mice using behavioral training, genetic engineering, and histology.

Selected Publications

In Preparation:

11) **Ash JR**, John C, Kolashetti A, Nelis P, Sanders M, Pavlovskii A, Ahmad M, Du Jardin M, Peat A, Gijsen H, Ceulemans H, Stokbroekx S, Varsakelis C. *Data to decisions: Machine learning and data visualization tools optimize compound selection for solubility predictive models in real-time support of small molecule discovery*. (in prep for J. Med. Chem.)

Published (Titles link to papers):

10) **Ash, J. R.**, Hughes-Oliver, J. M. *Confidence bands and hypothesis tests for hit enrichment curves.* Journal of Cheminformatics, 2022, 14(1), 50.

9) **Ash, J.**, 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.

8) Odenkirk, M. T.; Zin, P. P. K.; **Ash, J. R.**; Reif, D. M.; Fourches, D.; Baker, E. S. *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** *Methods Development for Quantitative Structure-Activity Relationships.* North Carolina State University, PhD Dissertation. 2019.

6) Fourches, D.; **Ash, J.** *4D-Quantitative Structure–Activity Relationship Modeling: Making a Comeback.* Expert Opin. Drug Discov. 2019, 14 (12), 1227–1235.

5) **Ash, J. R.**; Kuenemann, M. A.; Rotroff, D.; Motsinger-Reif, A.; Fourches, D. *Cheminformatics Approach to Exploring and Modeling Trait-Associated Metabolite Profiles.* J. Cheminform. 2019, 11 (1), 43.

4) Burnum-Johnson, K. E.; Zheng, X.; Dodds, J. N.; **Ash, J.**; Fourches, D.; Nicora, C. D.; Wendler, J. P.; Metz, T. O.; Waters, K. M.; Jansson, J. K.; Smith, R. D.; Baker, E. S. *Ion Mobility Spectrometry and the Omics: Distinguishing Isomers, Molecular Classes and Contaminant Ions in Complex Samples.* TrAC Trends Anal. Chem. 2019, 116, 292–299.

3) **Ash, J. R.**; Hughes-Oliver, J. M. Chemmodlab: A Cheminformatics Modeling Laboratory R Package for Fitting and Assessing Machine Learning Models. J. Cheminform. 2018, 10 (1).

2) **Ash, J.**; Fourches, D. *Characterizing the Chemical Space of ERK2 Kinase Inhibitors Using Descriptors Computed from Molecular Dynamics Trajectories.* J. Chem. Inf. Model. 2017, 57 (6), 1286–1299.

- Awarded ACS Editor's Choice.

1) Huang W.; Zhou G.; Marchand M.; **Ash J. R.**; Morris D.; Van Dooren P.; Brown J. M.; Gallivan K. A.; Wilgenbusch J. C. *TreeScaper: Visualizing and Extracting Phylogenetic Signal from Sets of Trees.* Mol. Biol. Evol. 2016, 33 (12), 3314–3316.

Patents

Parker, R. J.; Barker, C. A.; **Ash, J. R.**; Gotwalt, C. M. *Analytic system for interactive graphical model selection based on wavelet coefficients.* US Patent 11,361,255 B2.

Ash, J. R.; Gotwalt, C. M., Lancaster, L. C. *Analytic system with extrapolation control in interactive graphical prediction evaluation.* US Patent 10,963,804. 2021.

Selected Presentations

(Titles link to presentations)

American Chemical Society Conference
New Orleans, LA

Presentation — *Mar. 2023*

Ash JR, John C, Kolashetti A, Nelis P, Sanders M, Pavlovskii A, Ahmad M, Du Jardin M, Peat A, Gijsen H, Ceulemans H, Stokbroekx S, Varsakelis C. *Data to decisions: Machine learning and data visualization tools optimize compound selection for solubility predictive models in real-time support of small molecule discovery*.

JMP Discovery Summit America
Virtual

Presentation — *Sep. 2021*

Worley B, **Ash JR**. *Analyzing Spectral Data: Preprocessing and Beyond*.

JMP Discovery Summit America
Virtual

Presentation — *Oct. 2020*

Ash JR. *Fault Detection and Diagnosis of the Tennessee Eastman Process using Multivariate Control Charts*.

Joint Statistical Meetings
Virtual

Presentation — *Aug. 2020*

Ash JR, Lancaster L. *Methods for Helping Users Avoid Extrapolation when Making Predictions with Statistical and Machine Learning Models.*

Bioinformatics Ph. D. Defense
Raleigh, NC

Presentation — *June 2021*

Ash JR *Methods Development for Quantitative Structure-Activity Relationships*.

American Chemical Society Conference
Boston, MA

Presentation — *Aug. 2018*

Ash JR, Kuenemann MA, Rotroff D, Motsinger-Reif A, and Fourches D. *Structure-Based Approach to Exploring and Modeling Trait-Associated Metabolite Profiles.*

Poster

Ash JR, Hughes-Oliver JM and Fourches D. *Molecular Modeling of Differential ERK1/2-Ligand Dynamic Interactions*.

- ACS Division of Chemical Information Scholarship of Excellence.

American Chemical Society Conference
Washington, DC

Presentation — *Aug. 2017*

Ash JR and Fourches D. *Characterizing the Chemical Space of Kinase Inhibitors Using Molecular Descriptors Computed from MD Trajectories.*

Poster

Ash JR, Kuenemann MA, and Fourches D. *Cheminformatics Approach to Exploring and Modeling Trait-Associated Metabolic Profiles.*

- ACS Division of Chemical Information Scholarship of Excellence.

American Chemical Society Conference *Philadelphia, PA*

Poster *Aug. 2016*

Ash JR and Fourches D. *Leveraging GPU-Accelerated Molecular Dynamics Simulations to Compute and Analyze the 4D Chemical Descriptor Space of ERK2 Kinase Inhibitors.*

Graduate Coursework

Statistics Ph. D.

- Statistical Theory I & II
- Linear Models and Variance Components
- Computing for Statistical Research
- Experimental Statistics For Biological Sciences
- Real Analysis

Statistics Masters

- Introduction to Statistical Learning
- Applied Bayesian Analysis
- Linear Models and Regression

Bioinformatics Ph. D.

- Bioinformatics I & II
- Computational Methods for Molecular Biology
- Bioinformatics Consulting
- Molecular Genetics
- Functional Genomics

Awards

Oct 2019 International Conference on Statistical Distributions and Applications Travel Grant

Aug 2018 ACS Division of Chemical Information Scholarship of Excellence

Fall 2017 Triangle Center of Evolutionary Medicine Fellowship

Aug 2017 ACS Division of Chemical Information Scholarship of Excellence

Fall 2014 NCSU Graduate School Fellowship

Service

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

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

2018-Pres. **Reviewer,** Journal of Cheminformatics.

References

- Christopher Gotwalt, JMP Stats R&D Director (Christopher.Gotwalt@jmp.com)
- Jacqueline Hughes-Oliver, PhD Co-chair (hughesol@ncsu.edu)
- Jeffrey Thorne, PhD committee and StatAlign collaborator (thorne@ncsu.edu)