

# Colby T. Ford, Ph.D.

*Computational Biologist,  
Cloud AI & Genomics Architect*

Phone: (828) 773-8009  
Academic Email: colby.ford@uncc.edu  
Consulting Email: colby@tuple.xyz  
Location: Charlotte, NC  
Website: colbyford.com

LinkedIn: colbyford  
GitHub: colbyford  
ORCID: 0000-0002-7859-3622  
Google Scholar: zetDB4EAAAAJ  
Google Presence: 11dXPqljsq  
NPPES NPI: 1255002408  
Microsoft MVP: 5005240

## Education

### *The University of North Carolina at Charlotte*

2018 Doctor of Philosophy in Bioinformatics and Computational Biology  
*Advisor: Daniel Janies, Ph.D.*  
*Wayland H. Cato Doctoral Fellow*

2015 Master of Science in Data Science and Business Analytics  
*Advisor: Mirsad Hadžikadić, Ph.D.*

2014 Bachelor of Arts in Applied Mathematics, Concentration in Statistics  
*Advisor: Mary Kim Harris, Ed.D.*  
*Coffey Scholar, C.L. Robbins Scholar*

## Industry Positions

### *Tuple (a Microsoft Partner)*

Founded a Microsoft and Databricks partner firm that provides data architecture and research consulting services based on the Azure cloud, focusing on cloud AI, life science, and genomics solutions. Co-wrote the Azure Data Scientist Microsoft Certified Solution Associate exam (DP-100) and the MS Learn Azure Data Scientist exam.

Created the azbasespace connector for Illumina BaseSpace to Azure and the ahab system for -omics in Kubernetes.

September 2019 - Present Founder & Principal Consultant

### *Amissa*

Co-founded Amissa, a digital health company that's using cloud-based AI and analytics with wearable devices to help in the fight against Alzheimer's disease. Designed and built the Sinuo platform, which applies state-of-the-art artificial intelligence techniques to sensory data, biomarkers, and additional patient metrics. Amissa's work has been funded by the NIH's National Institute for Aging and the NSF's I-Corps program.

March 2020 - Present Co-Founder & VP of Technology

### *Seagen*

Established a scalable cloud computing architecture on Azure to support the Translational Sciences group in the discovery of therapeutic targets for antibody development in oncology. This includes an automated Kubernetes-based workflow for scalable RNAseq, WES, and pharmacodynamics pipelines along with supporting data platform, data lake, and machine learning architecture. Collaborated with Research to design scalable, cloud-based protein structural analyses for antibody candidates against tumor targets. Created drug sensitivity machine learning models using gene expression data.

September 2021 - Present Mgr. Machine Learning, Bioinformatics  
March 2021 - September 2021 Sr. Scientist, Bioinformatics

*BlueGranite (a Microsoft Partner)*

Developed data platform and AI solutions using the Microsoft Azure cloud with services such as Azure Machine Learning, Cognitive Services, and Databricks (Apache Spark). Hosted many training workshops and gave conference presentations and demonstrations on Microsoft advanced analytics technologies. Started and led the genomics practice, building scalable solutions for bioinformatics analyses in Azure. (BlueGranite.com/Genomics)

May 2020 - October 2021	Principal Architect, Life Sciences
December 2020 - May 2020	Genomics Practice Lead
November 2018 - December 2020	Solution Architect, Artificial Intelligence
January 2017 - November 2018	Senior Data Scientist

*Edge Systems - The HydraFacial Company*

Lead the data science and cloud infrastructure team through research and development initiatives related to IoT medical device informatics. Responsible for innovation with product marketing and engineering to improve and enhance technology products and processes. Implemented computer vision capabilities on the HydraFacial skin analyzer device for detection and classification of dermatologic conditions.

July 2019 - March 2020    Director of Data Science

*Ember.AI*

Developed a custom artificial intelligence engine for building state-of-the-art natural language processing models for fraud and compliance-related use cases. Built a large-scale, massively parallel processing architecture using Amazon Web Services and Databricks (Apache Spark) for interactive model training.

August 2018 - March 2020    Chief Data Scientist

*AmerisourceBergen - Lash Group*

Developed machine learning experiments and data analysis workflows to aid in pharmaceutical client analytics. Created analysis pipelines for the discovery and understanding into patient drug adherence and rare disease drug access. Consulted in company data architecture including master data management and governance for future business strategy to grow value for client accounts.

February 2016 - February 2017    Senior Data Science Lead

*Mariner (a Microsoft Partner)*

Consulted in the development of machine learning experiments, including parametric and non-parametric models, statistical predictions, and data mining. Built Azure cloud-based solutions for data collection, processing, and storage using Microsoft Azure Technologies such as Data Science Virtual Machines, Machine Learning Studio, Azure SQL Database, and more. Designed and created interactive visualizations for both dashboarding and reporting using Microsoft Power BI, Tableau Desktop & Tableau Server.

October 2014 - February 2016    Data Scientist

## Academic Positions

### *The University of North Carolina at Charlotte*

Redesigned and taught *Cloud Computing for Data Analysis* (DSBA 6190) for the School of Data Science Master's program. Received a funding grant from Microsoft to modernize and rebuild the course content on the Azure cloud platform. Performing research in a broad range of fields in the bioinformatics and computational biology space including infectious diseases, epistasis, human genomics, and phylogenetics.

November 2021 - Present	Associate Faculty	School of Data Science
July 2021 - Present	University Business Partner	Research and Economic Development
January 2021 - May 2021	Public Health Researcher	Dept. of Public Health
January 2019 - July 2021	Bioinformatics Researcher	Dept. of Bioinformatics and Genomics
January 2019 - May 2021	Associate Faculty	School of Data Science
April 2018 - December 2018	Postdoctoral Researcher	Dept. of Bioinformatics and Genomics
May 2016 - April 2018	Research Assistant	Dept. of Bioinformatics and Genomics

### *Northeastern University*

Taught a summer session of machine learning to graduate students in the LEVEL analytics program.

May 2015 - August 2015   Guest Lecturer   Dept. of Business and Analytics

### *North Carolina New Schools*

Responsible for entire technology inventory: ordering, maintenance, management, etc. and maintained school website. Liaison between college (CCC&TI) and high school technology departments. Taught NCVPS mathematics courses, held additional teaching sessions in math and science and SAT & ACT preparation.

April 2014 - October 2014	Technology Facilitator	Caldwell Early College
April 2014 - October 2014	Mathematics Instructional Assistant	Caldwell Early College

## Publications

### Journal Articles

- J20. Nicholas J. Santistevan, **Colby T. Ford**, Cole S. Gilsdorf, and Yevgenya Grinblat. Behavioral and transcriptomic analyses of *mecp2* function in zebrafish. *In Progress.*, 2023
- J19. Daniel Kepple, **Colby T. Ford**, Jonathan Williams, Beka Abagero, Shaoyu Li, Jean Popovici, Delenasaw Yewhalaw, and Eugenia Lo. Comparative transcriptomics reveal differential gene expression in *Plasmodium vivax* geographical isolates and implications on erythrocyte invasion mechanisms. *bioRxiv*, 2023
- J18. **Colby T. Ford**, Shirish Yasa, Denis Jacob Machado, and Daniel A. White III, Richard Allen Janies. Predicting changes in neutralizing antibody activity for sars-cov-2 xbb.1.5 using in silico protein modeling. *Frontiers in Virology*, 3, 2023
- J17. **Colby T. Ford**, Cheikh Cambel Dieng, Anita Lerch, Dickson Doniou, Kovidh Vegesna, Daniel Janies, Liwang Cui, Linda Amoah, Yaw Afrane, and Eugenia Lo. Genetic variations of *Plasmodium falciparum* circumsporozoite protein and the impact on interactions with human immunoproteins and malaria vaccine efficacy. *Infection, Genetics and Evolution*, 2023
- J16. The Critical Assessment of Genome Interpretation Consortium. CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods, 2022
- J15. Cristian M. Galván-Villa, Francisco A. Solís-Marín, Karen Lopez, Janessa Cobb, Leopoldo Díaz-Pérez, Carlos R. Rezende-Ventura, **Colby T. Ford**, and Daniel A. Janies. Occurrence of the Indo-West Pacific starfish *Luidia magnifica* Fisher, 1906 (Echinodermata: Asteroidea) in the Mexican Pacific. *In Review. (Marine Biodiversity)*, 2022
- J14. Dorcas Bredu, George K. Ahadzi, Donu Dickson, Sherik-fa Anang, Alexander Asamoah, George Asumah, Nana Prepah, Dorcas Obiri Yeboah, **Colby T. Ford**, Eugenia Lo, Keziah Laurencia Malm, and Linda Eva Amoah. Nationwide surveillance of Pfhrp2 exon 2 diversity in *Plasmodium falciparum* circulating in symptomatic malaria patients living in Ghana. *American Journal of Tropical Medicine and Hygiene.*, 2022
- J13. **Colby T. Ford**, Denis Jacob Machado, and Daniel A Janies. Predictions of the SARS-CoV-2 Omicron Variant (B.1.1.529) Spike Protein Receptor-Binding Domain Structure and Neutralizing Antibody Interactions. *Frontiers in Virology*, 2022
- J12. Anthony Ford, Daniel Kepple, Jonathan Williams, Gabrielle Kolesar, **Colby T. Ford**, Abnet Abebe, Lemu Golassa, Daniel A. Janies, Delenasaw Yewhalaw, and Eugenia Lo. Gene Polymorphisms Among *Plasmodium vivax* Geographical Isolates and the Potential as New Biomarkers for Gametocyte Detection. *Frontiers in Cellular and Infection Microbiology*, 11, 2022
- J11. **Colby T. Ford**, Gezahegn Solomon Alemayehu, Kayla Blackburn, Karen Lopez, Cheikh Cambel Dieng, Lemu Golassa, Eugenia Lo, and Daniel Janies. Modeling *Plasmodium falciparum* Diagnostic Test Sensitivity Using Machine Learning With Histidine-Rich Protein 2 Variants. *Frontiers in Tropical Diseases*, 2:28, 2021
- J10. **Colby T. Ford**, Rachel Scott, Denis Jacob Machado, and Daniel Janies. Sequencing Data of North American SARS-CoV-2 Isolates Shows Widespread Complex Variants. *medRxiv*, 2021
- J9. Anthony Ford, Daniel Kepple, Beka Raya Abagero, Jordan Connors, Richard Pearson, Sarah Auburn, Sisay Getachew, **Colby T. Ford**, Karthigayan Gunalan, Louis H. Miller, Daniel A. Janies, Julian C. Rayner, Guiyun Yan, Delenasaw Yewhalaw, and Eugenia Lo. Whole genome sequencing of *Plasmodium vivax* isolates reveals frequent sequence and structural polymorphisms in erythrocyte binding genes. *PLOS Neglected Tropical Diseases*, 14(10):1–27, 10 2020
- J8. **Colby T. Ford**, Gabriel Lopez Zenarosa, Kevin B. Smith, David C. Brown, John Williams, and Daniel Janies. Genetic Capitalism and Stabilizing Selection of Antimicrobial Resistance Genotypes in *Escherichia coli*. *Cladistics*, 2020

- J7. **Colby T. Ford**, Jia Wen, Daniel Janies, and Xinghua Shi. A Parallelized Strategy for Epistasis Analysis Based on Empirical Bayesian Elastic Net Models. *Bioinformatics*, 03 2020. btaa216
- J6. **Colby T. Ford** and Daniel Janies. Ensemble machine learning modeling for the prediction of artemisinin resistance in malaria. *F1000Research*, 9(62), 2020
- J5. Adriano de Bernardi Schneider, **Colby T. Ford**, Reilly Hostager, John Williams, Michael Cioce, Ümit V. Catalyürek, Joel O Wertheim, and Daniel Janies. StrainHub: A phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*, 08 2019
- J4. Wyatt T. Clark, Laura Kasak, Constantina Bakolitsa, Zhiqiang Hu, Gaia Andreoletti, Giulia Babbi, Yana Bromberg, Rita Casadio, Roland Dunbrack, Lukas Folkman, **Colby T. Ford**, David Jones, Panagiotis Katsonis, Kunal Kundu, Olivier Lichtarge, Pier L. Martelli, Sean D. Mooney, Conor Nodzak, Lipika R. Pal, Predrag Radivojac, Castrense Savojardo, Xinghua Shi, Yaoqi Zhou, Aneeta Uppal, Qifang Xu, Yizhou Yin, Vikas Pejaver, Meng Wang, Liping Wei, John Moulton, Guoying Karen Yu, Steven E. Brenner, and Jonathan H. LeBowitz. Assessment of predicted enzymatic activity of  $\alpha$ -N-acetylglucosaminidase variants of unknown significance for CAGI 2016. *Human Mutation*, 40(9):1519-1529, 2019
- J3. **Colby T. Ford**, Aneeta Uppal, Conor M. Nodzak, and Xinghua Shi. Prediction of the effect of naturally occurring missense mutations on cellular N-acetyl-glucosaminidase enzymatic activity. *bioRxiv*, 2019
- J2. **Colby T. Ford**. An integrated phylogeographic analysis of the Bantu migration. *ProQuest Dissertations and Theses*, page 120, 2018
- J1. Daniel Janies, **Colby Ford**, Lambodhar Damodaran, and Zachary Faigen. Spread of Middle East Respiratory Coronavirus: Genetic versus epidemiological data. *Online Journal of Public Health Informatics*, 9(1), 2017

### Conference Papers and Presentations

- C8. Adriano de Bernardi Schneider, **Colby T. Ford**, Jakob McBroome<sup>1</sup>, Jennifer Martin, Daniel Janies, Yatish Turakhia, and Russel Corbett-Detig. Understanding the spread of SARS-CoV-2 clusters through an integrated pipeline using UShER, Cluster Tracker and StrainHub. *Genetics Society of America - Population, Evolutionary, and Quantitative Genetics Conference*, 2022
- C7. Cheikh Cambel Dieng, **Colby T. Ford**, Dickson Doniou, Jennifer Huynh, Lerch Anita, Daniel Janies, Linda Amoah, Yaw Afrane, and Eugenia Lo. Population structure and selection-mediated changes in *Plasmodium falciparum* by next-generation sequencing. *Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2020
- C6. **Colby T. Ford**. Visualizing transmission networks of pathogens using phylogenetic data with StrainHub. *I Meeting of Systematics, Biogeography, and Evolution - Symposium of Virology in the SARS-CoV-2 era*, 2020
- C5. Daniel Janies, **Colby T. Ford**, Kevin Smith, Gabriel Lopez Zenarosa, and John Williams. Evolution of gain and loss of antimicrobial resistance genes in *Escherichia coli*. *XXXVIII Annual Meeting of the Willi Hennig Society*, 2019
- C4. Kevin Smith, **Colby T. Ford**, Gabriel Lopez Zenarosa, John Williams, and Daniel Janies. Phylogenetic analysis of the genetic variation of multi-drug resistant *Escherichia coli*. *National Council on Undergraduate Research*, 2019
- C3. Jia Wen, **Colby T. Ford**, Daniel Janies, and Xinghua Shi. New strategies toward scaling up epistasis analysis on large-scale genomic datasets. *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2018
- C2. **Colby T. Ford** and Andy Lathrop. Predictive modeling of vegetation density using R and a cloud data platform. *Analytics for Social Good, University of Cincinnati*, 2017
- C1. **Colby T. Ford**, Ming Xue, Peter M. Whiteley, Ward Wheeler, Daniel A. Janies, and Xinghua Shi. Visualizing linguistic disparity of Uto-Aztecan languages and Bantu languages. *Society for Anthropological Sciences Annual Meeting*, 2016

## Software and Coding

- S3. **Colby T. Ford.** *msgen* - R functions for interfacing with the Microsoft Genomics service in Azure, January 2021
- S2. Adriano de Bernardi Schneider, **Colby T. Ford**, Reilly Hostager, John Williams, Michael Cioce, Ümit V. Catalyürek, Joel O Wertheim, and Daniel Janies. StrainHub: A phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*, 08 2019
- S1. **Colby T. Ford**, Jia Wen, Daniel Janies, and Xinghua Shi. A Parallelized Strategy for Epistasis Analysis Based on Empirical Bayesian Elastic Net Models. *Bioinformatics*, 03 2020. btaa216

## Books and Chapters

- B2. **Colby T. Ford.** *Genomics in the Azure Cloud*. O'Reilly Media, Dec 2022
- E2. **Colby T. Ford** and Larry Baker. *Building a Genomics Data Lake in Azure*. BlueGranite, Jan 2021. URL: <https://www.bluegranite.com/genomics-data-lake-ebook>
- C1. Cheikh Cambel Dieng, **Colby T. Ford**, Jennifer Huynh, Linda E. Amoah, Yaw A. Afrane, Daniel A. Janies, and Eugenia Lo. Progress in *Plasmodium* genomics and current challenges in malaria control. In *Current Topics and Emerging Issues in Malaria Elimination*. IntechOpen, London, 2021. In Progress. ISBNs: 978-1-83968-484-5, 978-1-83968-483-8, DOI: 10.5772/intechopen.96530
- E1. **Colby T. Ford.** *Sparkitecture - A collection of "cookbook-style" scripts for simplifying data engineering and machine learning in Apache Spark*. GitBook, October 2019
- B1. **Colby Ford.** *Caesura - late-intermediate piano technique book*. Lulu, March 2010. ISBN: 978-0-557-36832-7

## Other Professional Contributions

### Technical Articles:

- TA21. **Colby T. Ford.** What does ChatGPT think about using the Azure cloud for genomics?, January 2023. Medium
- TA20. **Colby T. Ford.** 5 Quick Things You Didn't Know About the Microsoft Cloud for Healthcare, November 2022. Medium
- TA19. **Colby T. Ford.** How to Deploy Shiny Apps in Azure and use your Domain Name, October 2022. Medium
- TA18. **Colby T. Ford.** Using Microsoft Dev Box and Azure Compute Galleries for Bioinformatics, August 2022. Medium
- TA17. **Colby T. Ford.** A Deep Dive Look Into Global Monkeypox Transmissions - from January to July 2022., July 2022. Medium
- TA16. **Colby T. Ford.** Deploying Azure Databricks with a Virtual Network and Network Security Group, July 2022. Medium
- TA15. **Colby T. Ford.** Quick Note: What's up with the new SARS-CoV-2 BA.2.75 and BA.5 variants?, July 2022. Medium
- TA14. **Colby T. Ford.** Am I Hallucinating or Can AI Now Design Cancer-Curing Antibodies?, June 2022. Medium
- TA13. **Colby T. Ford.** Scaling Genomics in the Cloud with Microsoft Azure, May 2022. Medium
- TA12. **Colby T. Ford.** How to predict many protein structures with AlphaFold2 at-scale in Azure Machine Learning, January 2022. Medium
- TA11. **Colby T. Ford.** Predicted Protein Interactions of the SARS-CoV-2 B.1.1.529 Variant with Neutralizing Antibodies, November 2021. Medium

TA10. **Colby T. Ford**. Protein Structure Prediction of the new B.1.1.529 SARS-CoV-2 Spike Variant with AlphaFold2, November 2021. Medium

TA9. **Colby T. Ford**. How Do We Fix Peer Review? (Because It's Broken...), October 2021. Medium

TA8. **Colby T. Ford**. I Asked GPT-3 to Write Some Chocolate Chip Cookie Recipes...and I Baked Them), June 2021. Medium

TA7. **Colby T. Ford**. What is a "variant" as it relates to COVID-19? (for non-geneticists), February 2021. Medium

TA6. **Colby T. Ford**. Automated Alignment and Variant Calling in Azure using the Microsoft Genomics service and the *msgen* R package, January 2021. Medium

TA5. **Colby T. Ford**. Better Evaluate COVID-19 Test Performance using Bayes Factors, December 2020. Medium

TA4. **Colby T. Ford**. Assessment of retail out-of-stock conditions using statistical inference. Technical report, Mariner, 2016

TA3. **Colby T. Ford** and Wayne Snyder. Revenue protection using machine learning for utilities management. Technical report, Mariner, 2015

TA2. **Colby T. Ford**. The allure of machine learning, now within reach in Microsoft Azure. Technical report, Mariner, 2015

TA1. **Colby T. Ford**. Demand forecasting using machine learning to reduce working capital. Technical report, Mariner, 2015

### Professional Blog Posts:

B30. **Colby T. Ford**. Case Study: Scaling Rare Disease Research in the Azure Cloud, 2022. *Tuple Case Study*

B29. **Colby T. Ford**. Using Azure CycleCloud and Batch for Scalable HPC Workloads, 2021. *BlueGranite Technical Blog*

B28. **Colby T. Ford**. Microsoft AI: Look, Listen, Innovate! Hackfest - BlueGranite's Smart City Solution, 2021. *BlueGranite Technical Blog*

B27. **Colby T. Ford**. Query Millions of Genomic Variants in Minutes with Azure Synapse, 2021. *BlueGranite Technical Blog*

B26. **Colby T. Ford**. Review: How 3 Life Science Organizations Harness the Power of the Cloud, 2021. *BlueGranite Technical Blog*

B25. **Colby T. Ford**. Introducing BlueGranite's Azure Data Factory Connector for Illumina BaseSpace, 2021. *BlueGranite Technical Blog*

B24. **Colby T. Ford**. Event Recap: Shape your Future with Azure Data and Analytics, 2020. *BlueGranite Technical Blog*

B23. **Colby T. Ford**. Reading Bioinformatics and Genomics Files in Power BI, 2020. *BlueGranite Technical Blog*

B22. Jon Gore, Thomas J. Weinandy, and **Colby T. Ford**. Take Back Your Data to Gain a Competitive Advantage, 2020. *BlueGranite Technical Blog*

B21. **Colby T. Ford**. A BlueGranite Blog Post Written (Mostly) by AI, 2020. *BlueGranite Technical Blog*

B20. **Colby T. Ford**. Give Your Genomics Pipeline a \*Glow\* Up in Azure Databricks, 2020. *BlueGranite Technical Blog*

B19. **Colby T. Ford**. SIR Modeling on Azure: COVID-19 Hospital Impact Model for Epidemics, 2020. *BlueGranite Technical Blog*

- B18. **Colby T. Ford.** Let's Talk About COVID-19, 2020. *BlueGranite Technical Blog*
- B17. **Colby T. Ford.** Comparing Azure Machine Learning Service and Azure Databricks, 2020. *BlueGranite Technical Blog*
- B16. **Colby T. Ford.** Recap of rstudio::conf(2020) for Data Science and Machine Learning, 2020. *BlueGranite Technical Blog*
- B15. **Colby T. Ford.** Scaling your Genomics Pipeline in the Cloud with Azure Databricks, 2019. *BlueGranite Technical Blog*
- B14. **Colby T. Ford.** Migrating & Scaling Machine Learning Models to Azure Databricks for Cloud-Powered AI, 2019. *BlueGranite Technical Blog*
- B13. **Colby T. Ford.** Introducing the Databricks Unified Analytics Platform for Genomics, 2018. *BlueGranite Technical Blog*
- B12. **Colby T. Ford.** Recap: Spark+AI Summit 2018, 2018. *BlueGranite Technical Blog*
- B11. **Colby T. Ford.** Cognitive Services Showcase: API Search Tools, 2018. *BlueGranite Technical Blog*
- B10. **Colby T. Ford.** Let Azure do the Heavy Lifting on Your AI Workload, 2018. *BlueGranite Technical Blog*
- B9. **Colby T. Ford.** Recap of rstudio::conf 2018, 2018. *BlueGranite Technical Blog*
- B8. **Colby T. Ford.** Microsoft Azure & Databricks = Cloud-Scale Spark Power, 2017. *BlueGranite Technical Blog*
- B7. **Colby T. Ford.** Maximize Your Customer Retention by Predicting Customer Churn, 2017. *BlueGranite Technical Blog*
- B6. **Colby T. Ford.** Become the Maestro of your Genomics Workflow with Bioconductor and Microsoft R Server, 2017. *BlueGranite Technical Blog*
- B5. **Colby T. Ford.** Publishing Predictive Web Services with Microsoft R Server, 2017. *BlueGranite Technical Blog*
- B4. **Colby T. Ford.** Data Visualization for Bioinformatics with R in Power BI, 2017. *BlueGranite Technical Blog*
- B3. **Colby T. Ford.** Webinar Recap: Distributed Computing & R Server, 2017. *BlueGranite Technical Blog*
- B2. **Colby T. Ford.** SAS Enterprise Guide vs. Microsoft Azure Machine Learning, 2017. *BlueGranite Technical Blog*
- B1. **Colby T. Ford.** ImpoRting and ExpoRting: Getting Data Into and Out of R, 2017. *BlueGranite Technical Blog*



## Conferences, Training, Press, and Speaking Engagements

Nov. 2014	Microsoft Roadmap Event	<i>Speaker</i>	Charlotte, NC
Sep. 2015	Microsoft Cortana Analytics Conference	<i>Attendee</i>	Seattle, WA
Sep. 2016	Advanced Pharma Analytics	<i>Attendee</i>	Newark, NJ
Feb. 2017	Analytics for Social Good - U. Cincinnati	<i>Speaker</i>	Cincinnati, OH
Mar. 2017	Society for Applied Anthropology Conference	<i>Speaker</i>	Santa Fe, NM
Mar. 2017	BlueGranite Distributed Computing Webinar	<i>Speaker</i>	Online
Oct. 2017	BlueGranite Retail Webinar	<i>Speaker</i>	Online
Nov. 2017	Nissan Analytics Expo	<i>Speaker</i>	Nashville, TN
Jan. 2018	rstudio::conf(2018)	<i>Attendee</i>	San Diego, CA
Jun. 2018	Spark + AI Summit 2018	<i>Attendee</i>	San Francisco, CA
Sep. 2018	Big Data Ignite	<i>Speaker</i>	Grand Rapids, MI
Sep. 2018	Microsoft Ignite	<i>Attendee</i>	Orlando, FL
Jan. 2019	rstudio::conf(2019)	<i>Attendee</i>	Austin, TX
Mar. 2019	Microsoft Azure AI Hackfest	<i>Attendee</i>	New York, NY
Apr. 2019	Azure Databricks Training Event	<i>Speaker</i>	Chicago, IL
May 2019	CECHS Commencement	<i>Speaker</i>	Lenoir, NC
Jun. 2019	BlueGranite Azure Databricks Retail Webinar	<i>Speaker</i>	Online
Jun. 2019	Azure Databricks Training Event	<i>Speaker</i>	Charlotte, NC
Jun. 2019	Azure Databricks Training Event	<i>Speaker</i>	Detroit, MI
Jul. 2019	Databricks Retail Webinar	<i>Speaker</i>	Online
Oct. 2019	UNCC CCI Ph.D. Open House Panel	<i>Panelist</i>	Charlotte, NC
Nov. 2019	UNCC Data Science Initiative Git Workshop	<i>Speaker</i>	Charlotte, NC
Jan. 2020	PyData Charlotte	<i>Speaker</i>	Charlotte, NC
Jan. 2020	rstudio::conf(2020)	<i>Attendee</i>	San Francisco, CA
Apr. 2020	WBTV COVID-19 Interview	<i>Interviewee</i>	Online
May 2020	CBS News COVID-19 Interview	<i>Interviewee</i>	Online
May 2020	UCSD HIV Dynamics Conference	<i>Speaker</i>	Online
Jun. 2020	Spark + AI Summit 2020	<i>Speaker/Panelist</i>	Online
Jul. 2020	SBE Symposium: Virology in the SARS-CoV-2 era	<i>Speaker</i>	Online
Aug. 2020	BlueGranite Scaling Genomic Analyses in Azure Webinar	<i>Speaker</i>	Online
Sep. 2020	TEDxUNCCCharlotte	<i>Speaker</i>	Charlotte, NC
Oct. 2020	BlueGranite Loan Risk Analysis Webinar	<i>Speaker</i>	Online
Jan. 2021	BlueGranite AI-in-a-Day Training Event	<i>Speaker</i>	Online
Jan. 2021	rstudio::global(2021)	<i>Attendee</i>	Online
Jan. 2021	Lenoir NewsTopic Interview	<i>Interviewee</i>	Online
Mar. 2021	Microsoft AI: Look, Listen, Innovate! Solution Hackathon	<i>Participant</i>	Online
May 2021	Data + AI Summit 2021	<i>Attendee</i>	Online
Jul. 2021	Microsoft Ignite 2021	<i>Attendee</i>	Online
Jul. 2021	Alzheimer's Association International Conference 2021	<i>Attendee</i>	Denver, CO
Oct. 2021	Microsoft Research Summit	<i>Attendee</i>	Online
Nov. 2021	Microsoft Ignite	<i>Attendee</i>	Online
Dec. 2021	The Economist Interview on Omicron Modeling	<i>Interviewee</i>	Online
Jan. 2022	WIRED Interview on Omicron Modeling	<i>Interviewee</i>	Online
Feb. 2022	United Nations ITU AI for Good Programme	<i>Speaker</i>	Online
May 2022	Manning Twitch Stream: Genomics in Azure	<i>Speaker</i>	Online
Jul. 2022	Microsoft Inspire	<i>Attendee</i>	Online
Jul. 2022	rstudio::conf(2022)	<i>Attendee</i>	Washington DC
Jul. 2022	Alzheimer's Association International Conference 2022	<i>Panelist</i>	San Diego, CA
Oct. 2022	UNCC Seminar: Getting Started with Git	<i>Instructor</i>	Charlotte, NC
Nov. 2022	Microsoft Cloud for Healthcare Event (TTT)	<i>Participant</i>	Seattle, WA
Jan. 2023	NC HIMSS Event	<i>Participant</i>	Charlotte, NC
Apr. 2023	Microsoft MVP Global Summit	<i>Awardee</i>	Redmond, WA
Jun. 2023	Data + AI Summit 2023	<i>Attendee</i>	San Francisco, CA
Jul. 2023	Alzheimer's Association International Conference 2023	<i>Presenter</i>	Amsterdam, NL
Sep. 2023	posit::conf(2023)	<i>Attendee</i>	Chicago, IL

## Skills

Languages:	Python • R • SQL • Javascript • SAS • Visual Basic
High Performance Computing:	Spark (Databricks) • K8s • CUDA • Hadoop • MPI
Cloud Computing:	Microsoft Azure • AWS • Docker • Terraform
Visualization:	Tableau • Power BI • Shiny • D3.js
Markup/Web:	Markdown • $\text{\LaTeX}$ • HTML5 • CSS3
Bioinformatics:	Phylogenetics • Haplotyping • Protein Modeling • Epistasis

## Professional Memberships

2013-Present	The Society for Industrial and Applied Mathematics
2014-Present	The American Statistical Association
2015-2017	UNCC Data Science Initiative Advisory Board
2015-2016	Northeastern University LEVEL Advisory Board
2016-2017	The Society for Anthropological Sciences
2018-2019	American Association for the Advancement of Science
2020-Present	International Society for Computational Biology
2021-Present	International Society for Infectious Diseases
2021-Present	International Society to Advance Alzheimer's Research and Treatment
2022-Present	American Cancer Society - Associate Board of Advisors

## Student Advising

Jan. 2019 - May 2019	Jainmary Jose	Cloud Computing Teaching Assistant
Aug. 2019 - Dec. 2019	Anjali Khushalani	Cloud Computing Teaching Assistant
Jan. 2020 - May 2020	Kovidh Vegesna	UNCC Data Science Practicum Intern
May 2020 - Sep. 2020	Somesh Kale	Amissa Developer Intern
May 2020 - Sep. 2020	Heet Detroja	Amissa Developer Intern
May 2020 - Sep. 2020	Chaitanya Darade	Amissa Developer Intern
Sep. 2020 - May 2021	Adesoji Ademiluyi	Bioinformatics Research Assistant
Sep. 2020 - Oct. 2021	Eh So Wah	Amissa Developer Intern
Jan. 2022 - May 2022	Paige Oldiges	Cloud Computing Teaching Assistant
Jun. 2022 - Sep. 2022	Carter DeMordaunt	Tuple Bioinformatics Intern
Jan. 2023 - May 2023	Utweij Sai Nalluri	Cloud Computing Teaching Assistant

## Awards

Apr. 2022	40 under 40	Charlotte Business Journal
Apr. 2022	Impact Award	Seagen, Inc.
Mar. 2023	Alumnus of the Year	UNC Charlotte
Apr. 2023	Most Valuable Professional (MVP)	Microsoft

## Certifications

### Technical Certifications

May 2018	Databricks Certified Developer - Apache Spark 2.x for Python	Certification #: 53873
Nov. 2019	Microsoft Azure Data Scientist Associate (DP-100)	Exam Writer
Sep. 2020	Microsoft Data Science Partnership Program	Custom Core Certification
Oct. 2020	Microsoft Azure AI Engineer Associate (AI-100)	Certification #: H532-5207
Oct. 2020	Microsoft Azure AI Fundamentals (AI-900)	Certification #: H533-2414
Nov. 2020	Databricks Partner Program - Developer Champion	Certification #: 917207
Feb. 2021	Microsoft Certified Trainer	Certification #: 17875868

### Research Certifications

Dec. 2019	CITI Program - Biomedical Researcher (IRB)	Record ID: 34535340
Dec. 2019	CITI Program - Data or Specimens Only Research (IRB)	Record ID: 34853732
Nov. 2020	CITI Program - Animal Biosafety	Record ID: 39539397
Nov. 2020	CITI Program - Laboratory Personnel BSS	Record ID: 39539395
Nov. 2020	CITI Program - USDA Permits	Record ID: 39539400
Nov. 2020	CITI Program - Shipping and Transport of Regulated Biological Materials	Record ID: 39539399
Nov. 2020	CITI Program - Principal Investigators BSS	Record ID: 39539394
Nov. 2020	CITI Program - Institutional Biosafety Committee (IBC) Members BSS	Record ID: 39539396
Jan. 2021	CITI Program - Social & Behavioral Research (IRB)	Record ID: 40316365
Jun. 2021	CITI Program - Good Clinical Practices (GCP)	Record ID: 43096988
Feb. 2023	CITI Program - Biomedical Sciences: Responsible Conduct of Research	Record ID: 54594415
Feb. 2023	CITI Program - Biomedical Investigators and Research Study Team	Record ID: 54601364

## Journal Peer Review Activity

Frontiers in Pediatrics	Pediatric Infectious Diseases
Frontiers in Chemistry	Chemical Biology
Frontiers in Immunology	Vaccines and Molecular Therapeutics
Frontiers in Physiology	Red Blood Cell Physiology
Frontiers in Medicine	Infectious Diseases - Surveillance, Prevention and Treatment
Science Advances	Biomedicine and Life Sciences

## Public Sequence Contributions

NCBI SRA	BioProject: PRJNA783000	<i>P. falciparum</i> CSP/ MSP-1 amplicon sequences	Ghana	90 samples
NCBI SRA	BioProject: PRJNA784582	<i>P. vivax</i> transcriptomes	Ethiopia	10 samples

## Funding

Dec. 2023 - Nov. 2024	NIH   NIA SBIR Phase I	Examining Biometric Measures During Perimenopause And Menopause To Predict Sleep Interruptions And Hot Flashes	\$299,971	Pending	Co-I
Aug. 2023 - Aug. 2024	NIH   NIA SBIR Phase II	Empowering Advanced Alzheimer's Disease and Dementia Research Through Remote Patient Monitoring and Cloud-Based Wearable Devices	\$2,500,000	Pending	Co-I
Aug. 2023 - Aug. 2025	US DoD   Army	Entomological Surveillance Dashboard, Harnessing Geographic Information Systems to Modernize Military Entomology	\$150,000	Funded	Site PI
Apr. 2022 - Aug. 2023	NC SBIR Match	A Multifaceted Digital Health Application to Advance Alzheimer's Disease Patient Monitoring, Safety, Caregiving, and Research	\$75,000	Funded	Co-I
Aug. 2021 - Aug. 2023	NIH   NIA SBIR Phase I	A Multifaceted Digital Health Application to Advance Alzheimer's Disease Patient Monitoring, Safety, Caregiving, and Research	\$495,787	Funded 1R43AG072981-01A1	Co-I
Apr. 2021 - Sep. 2022	NSF Innovation Corps	Investigating the Utility of Digital Biomarkers with Wearable Devices and Artificial Intelligence in Alzheimer's Disease and other Neurological Disorders	\$50,000	Funded 2127407	PI/TL
May 2020 - Dec. 2020	UNCC SDS Seed Grant	Analytical Modeling of IoMT data to predict onset and progression of Alzheimer's	\$20,000	Funded	Co-PI
Jan. 2019 - May 2019	Microsoft	Azure Funding for Cloud Computing Course	\$30,000	Funded	PI
Oct. 2014 - Present	Microsoft	Various Business Incentive Funds (BIF), Partner Investment Engine (PIE) Funds, GoFast Funding, and End Customer Investment Funds (ECIF).	>\$750,000	Funded	PI