

Colby T. Ford, Ph.D.

*Computational Biomathematician,
Cloud AI Architect, & Data Scientist*

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

LinkedIn: colbyford
GitHub: colbyford
ORCID: 0000-0002-7859-3622
Google Scholar: zetDB4EAAAAJ
Google Presence: 11d xpqljsq

Education

The University of North Carolina at Charlotte

- 2018 Doctor of Philosophy in Bioinformatics and Computational Biology
Advisor: Daniel Janies, Ph.D.
Dissertation: An Integrated Phylogeographic Analysis of the Bantu Migration
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, Minor in Psychology
Advisor: Mary Kim Harris, Ed.D.
Coffey Scholar, C.L. Robbins Scholar

Industry Positions

Amissa

Amissa is a digital health company that's using cloud-based AI and analytics with wearable devices to help in the fight against Alzheimer's disease. Responsible for the Sinuo platform, which applies state-of-the-art artificial intelligence techniques to sensory data, health data, and additional patient metrics.

March 2020 - Present Co-Founder & VP of Technology

BlueGranite

Developed data platform and AI solutions using the Microsoft Azure cloud with services such as Azure Machine Learning Service, Cognitive Services, and Databricks (Apache Spark). Managed client engagements including requirements gathering, project planning, and budgeting. Hosted many training workshops and gave conference presentations and demonstrations on Microsoft advanced analytics technologies.

November 2018 - Present AI Solution Architect
January 2017 - November 2018 Senior Data Scientist

Edge Systems - The HydraFacial Company

Lead data science 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 skin 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 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

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*

Teaching *Cloud Computing for Data Analysis* (DSBA 6190) for the Data Science Initiative 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, and human phylogenetics.

| | | |
|----------------------------|---------------------------|--------------------------------------|
| January 2019 - Present | Associate Faculty | School of Data Science |
| January 2019 - Present | Bioinformatics Researcher | Dept. of Bioinformatics and Genomics |
| April 2018 - December 2018 | Postdoctoral Researcher | Dept. of Bioinformatics and Genomics |
| May 2016 - April 2018 | Graduate 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 LEVEL

North Carolina New Schools

Responsible for entire technology inventory: ordering, maintenance, management, etc. and maintained school website. Liaison between college & 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

- J11. Gabriel Lopez Zenarosa, **Colby T. Ford**, David Brown, Kevin Smith, and Daniel Janies. Extraction of gene associations in colistin-resistant superbugs in *Escherichia coli*. *mBio*, 2020. In Progress
- J10. Colby T Ford, Gezahegn Alemayehu, Kayla Blackburn, Karen Lopez, Cheikh Cambel Dieng, Eugenia Lo, Lemu Golassa, and Daniel Janies. Modeling *Plasmodium falciparum* Diagnostic Test Sensitivity using Machine Learning with Histidine-Rich Protein 2 Variants. *medRxiv*, 2020. In Review (BMC Malaria)
- J9. Anthony Ford, Daniel Kepple, Beka Raya, Richard Pearson, Sarah Auburn, **Colby T. Ford**, Karthigayan Gunalan, Louis H. Miller, Daniel A. Janies, Julian C. Rayner, Delenasaw Yewhalaw, Guiyun Yan, and Eugenia Lo. Whole genome sequencing of *Plasmodium vivax* isolates reveals frequent sequence and structural polymorphisms in erythrocyte binding genes. *PLoS Tropical Diseases*, 2020. In Press
- 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 α -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

- 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**. Sparkitecture - A collection of "cookbook-style" scripts for simplifying data engineering and machine learning in Apache Spark., October 2019
- 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

Other Professional Contributions

{B: Blog, TA: Technical Article}

- 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*
- 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 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 |

Conferences, Training, 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 |
| 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 | TEDx UNCCCharlotte | <i>Speaker</i> | Charlotte, NC |
| Oct. 2020 | BlueGranite Loan Risk Analysis Webinar | <i>Speaker</i> | Online |

Advising

| | | |
|-----------------------|-------------------|---------------------------------------|
| Jan. 2020 - May 2020 | Kovidh Vegesna | DSBA MS Practicum Internship |
| May 2020 - Sep. 2020 | Somesh Kale | Amissa Developer Internship |
| May 2020 - Sep. 2020 | Heet Detroja | Amissa Developer Internship |
| May 2020 - Sep. 2020 | Chaitanya Darade | Amissa Developer Internship |
| Sep. 2020 - Dec. 2020 | Adesoji Ademiluyi | Bioinformatics Research Assistantship |

Certifications

| | | |
|-----------|--|----------------------|
| May 2018 | Databricks Certified Developer - Apache Spark 2.x for Python | Certificate #: 53873 |
| Dec. 2019 | CITI Program - Biomedical Researcher (IRB) | Record ID: 34535340 |

Skills

| | |
|------------------------|--|
| Languages: | R • Python • SQL • SAS • Visual Basic • Javascript |
| Distributed Computing: | Spark (Databricks) • Hadoop • MPI • SNOW |
| Cloud Computing: | Microsoft Azure • AWS • Docker/Kubernetes |
| Visualization: | Tableau • Power BI • Shiny • D3.js |
| Markup/Web: | Markdown • \LaTeX • HTML5 • CSS3 |

Funding

| | | | | | |
|--------------------------|---------------------|--|-------------|---------|-------|
| Nov. 2020 - Oct. 2022 | NIH RADx-RAD | Exhaled Breath Collection and VOC Profiling: Screening Mechanisms for SARS-CoV-2 | \$1,628,016 | Pending | Co-I |
| Jan. 2021 - Jan 2022 | NGA BIG-R BAA | Comprehensive Data Mining and Modeling across the Sociocultural Spectrum for COVID-19 Pandemic Recovery | \$596,458 | Pending | Co-I |
| Jan. 2021 - Dec. 2021 | NIH SBIR | A Multifaceted Digital Health Application to Advance Alzheimer's Disease Patient Monitoring, Safety, Caregiving, and Research | \$500,000 | Pending | Co-I |
| Jul. 2020 - Jun. 2022 | Google.org | A Novel Deep Learning Modeling - Transfer Learning (DL-TL) Paradigm to Accurately Characterize the COVID-19 Pandemic Across Scales. | \$321,000 | Pending | Co-PI |
| Apr. 2020- May 2020 | Microsoft | COVID-19 Research on Azure | \$35,000 | Funded | PI |
| 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). | >\$400,000 | Funded | PI |

Last updated: September 15, 2020
www.colbyford.com