**Zachary Apell** <https://zapell.github.io>

Ann Arbor, MI 48104 [856.298.0064](TEL:856.298.0064)

[www.linkedin.com/in/zach-apell](http://www.linkedin.com/in/zach-apell) [zapell@umich.edu](mailto:zapell@umich.edu)

# Education

**MS,** Biostatistics

**University of Michigan School of Public Health**, Ann Arbor, MI

* **Genome Science Training Program (GSTP) Predoctoral Trainee** — NIH T32HG000040, National Human Genome Research Institute (2025 – Present)
* **Research Advisor**: Dr. Minji Kim, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI
* **Relevant Coursework**:
  + **Genetics and Bioinformatics**: Molecular, Cellular, and Organismal Genetics; Introductory Biology for Computational Scientists; Bioinformatics Concepts and Algorithms
  + **Statistical Methods:** Applied Bayesian Inference; Theory and Application of Longitudinal Analysis; Statistical Models and Numerical Methods in Human Genetics; Probability and Distribution Theory; Biostatistical Inference
  + **Computing and Data Analysis:** Statistical Computing
* **Student Representative** on Biostatistics Alumni Relations Committee
* **Peer Mentor Volunteer**, Department of Biostatistics

**BS,** Data Science

**University of Michigan**, Ann Arbor, MI

* **Relevant Coursework**: Machine Learning, Data Mining, Bayesian Data Analysis, Web Systems, Database Management Systems, Honors Multivariable Calculus, Computer Vision

# Technical Skills / Expertise

* **Programming Languages**: R, Python, Rcpp, C++, Cython, SQL, PySpark, SAS, Shell scripting
* **Bioinformatics & Genomics Tools:** bedtools, samtools, MEME Suite, Juicer, deepTools
* **Computing Platforms:** Great Lakes (SLURM High-Performance Computing (HPC)), Amazon Web Services (AWS), Amazon SageMaker
* **Development & Collaboration Tools:** Git, GitHub, Jupyter, RStudio
* **Data Processing & Visualization:** ggplot2, matplotlib, seaborn, tidyverse, pandas, NumPy, SciPy

RESEARCH AND PUBLICATIONS

**PUBLICATIONS**

* Gupta, N., **Apell, Z**., Kim, M. (2025) Chrom-Sig: de-noising 1-dimensional genomic profiles by signal processing methods (under review)

**Zachary Apell** [**zapell@umich.edu**](mailto:zapell@umich.edu) **Page Two**

**MINJI LAB**, Ann Arbor, MI 2024 – Present

* **Role of microsatellite repeat elements in B-Lymphoblastic Leukemia (B-ALL)**
  + Constructed method to normalize 1-d genomic data for an unlimited number of samples of different sequencing depth and background noise.
  + Evaluated 3 different subtypes of B-ALL to determine the differentiating effect of GGAA repeat elements to understand the functionality of an understudied but related B-ALL subtype.
* **Mouse Spermatid**
  + Adapted Juicer pipeline to run on specific HPC architecture for Hi-C data processing and analysis of early and late stage mouse spermatid cells.
  + Called and compared compartments, TADs, and loops for early and late stage spermatids as well as the combined spermatid Hi-C data to embryonic stem cells to understand the timing of chromatin architectural change across the development of spermatids.
* **Chrom-Sig**
  + Contributed to development of Chrom-Sig to de-noise 1-d genomic data such as ATAC-seq, ChIP-seq, and CUT&RUN assays.
  + Validated Chrom-Sig on GM12878 CTCF CUT&RUN data, increasing peak precision of CTCF binding sites from 39% (original data) to 74% and confirming biologically relevant signal is maintained.
  + Applied MEME Suite motif discovery to Chrom-Sig processed data, identifying CTCF motifs with a dramatically lower E-value (~10^450-fold more significant) compared to the original data.
  + Compared ChromHMM state distributions between original and Chrom-Sig processed data showing Chrom-Sig increased enhancer/promoter states by ~12% in K562 RNA Polymerase II ChIP-Seq experiments and in GM12878 ATAC-seq experiments Chrom-Sig reduced unwanted transcription/heterochromatin states from 29% to 13%, highlighting improved signal capture.
* **MIA-Sort**
  + Optimized MIA-sort, an algorithm to identify chromatin hubs by quickly extracting chromatin contacts involving any number of genomic loci.
  + Designed data structures and an indexing scheme to reduce the search space of MIA-sort, improving the computational efficiency of large data files like SPRITE.

**HE LAB**, Ann Arbor, MI 2019

**Statistical Researcher,** Advisor: Dr. Xuming He

* Implemented a Random Forests model to determine if a high school football player needed to undergo further concussion testing based on the registered impact in their helmet.
* Evaluated the predictiveness of daily impact density and other head impact metrics.

Professional Experience

**COOPER UNIVERSITY HEALTHCARE**, Camden, NJ 2024

**Population Health Intern**

* Engineered dataset including determinants of health and medical diagnoses to develop a model for risk stratification analysis of entire patient population resulting in a 66% increase in identification of high-risk patients.

**Zachary Apell** [**zapell@umich.edu**](mailto:zapell@umich.edu) **Page Three**

**ROCKET MORTGAGE**, Detroit, MI 2022 - 2023

**Data Scientist**

* Worked on dial prioritization ML models predicting likelihood of lead to close.
* Developed lifecycle data science product from raw state through deployment, using Amazon Athena, Sagemaker, and S3 for purchase and refinance retention models.
* Leveraged PySpark to build dataset joining legacy data lake with new AWS storage.
* Designed preprocessors and unit testing for categorical and numerical values used in lambda handler function for inferencing.
* Created hyperparameter tuning jobs, deciding best ML model parameters.
* Implemented Platt calibration for comparison of true probabilities between models due to highly imbalanced datasets.
* Improved top 3 decile customer conversion performance 75% compared to old standard.
* Increased net lead-to-allocate conversion 47% post-deployment compared to pre-deployment average.
* Contributed to prioritized engine model (PEM) pre-credit purchase test.
* Designed query recreating multifaceted business logic that identified control group with 95% accuracy.
* Built XGBoost model predicting loans most likely to reach pre-approval letter (PAL) status that converted at higher rate than control group.
* Improved folder conversion 31% using escalation logic overall, with top 2 deciles converted at 150% better rate.

**FREEDOM MORTGAGE**, Mt. Laurel, NJ 2020 - 2022

**Data Scientist**

* Designed SQL query connected to Tableau in real time for catalog of rule triggers allowing transparency for compliance, regulations, and operations business divisions.
* Leveraged experience with data governance, extract / transform / load (ETL), and marketing analytics business groups to create SQL queries linking data from marketing, servicing, originations, and customer sources for use in lead prioritization and customer satisfaction models.

**UNIVERSITY OF MICHIGAN ADVANCED RESEARCH COMPUTING CENTER**, Ann Arbor, MI 2019-2020

**Data Science Intern**

* Gained hands-on experience with HPC clusters, including job scheduling and resource allocation.
* Processed large-scale genomic data using ADAM, leveraging Spark to parallelize genomic data analysis across clusters.

**WACKER CHEMICAL CORPORATION**, Ann Arbor, MI 2019

**Data Science Intern**

* Led sales forecasting project.
* Decreased forecast deviation by 10%, experimenting with long short-term memory recurrent neural networks, support vector regression, and seasonal autoregressive integrated moving average (SARIMA) models to forecast sales with 79% accuracy and order entries with 87% accuracy using internal data and external economic factors.

**Zachary Apell** [**zapell@umich.edu**](mailto:zapell@umich.edu) **Page Four**

* Prepared dashboards for business teams using Tableau, visualizing customer and product trends and insights.
* Applied natural language processing (NLP) techniques using Twitter application programming interface (API) to calculate sentiment analysis and extract key phrases.

**THE BURR PROJECT**, Voorhees, NJ 2018

**Data Analyst Intern**

* Collected and cleaned data using pandas to build database of politicians and voting records, facilitating transparency between candidates and constituents.

**VOLUNTEER EXPERIENCE**

**STATCOM,** University of Michigan, Ann Arbor, MI 2024 - 2025

**Huron Valley Humane Society**

* Discovered trends in return to owner rates by seasonality and species to increase understanding of lost and found pet patterns.
* Contributed to geospatial heatmap application that allows users to view geographic locations of found animals.

**PROJECTS**

* Optimized R Implementation of MatrixNormal Distribution
* Developed and published [Fast Matrix Normal](https://github.com/ziyuliu1999/fastMN) to improve the computational efficiency of sampling and pdf calculations by an order of magnitude using Rcpp and properties of matrices.
* Live Betting Database
* Automated process using Python and Selenium remote web driver to scrape dynamic data on AWS EC2 instance to PostgreSQL database hosted on AWS RDS instance.
* Utilized (cron) jobs and AWS CloudWatch to schedule bash scripts and lambda functions to minimize total cost.
* Diagnosing Diabetes in Pima Indians
* Compared effectiveness of Bayesian Logisitic Regression using STAN and Logistic Regression on predicting diabetes in underrepresented Pima Indian community.

# Additional Training

* AWS Cloud Computing Certified