

Siena Dumas Ang

siena.ang@gmail.com, siena@princeton.edu, (206) 914-8012

SUMMARY OF QUALIFICATIONS

- Strong background in pure Mathematics, Computer Science, and Computational Biology
- Experience in both academic research and industry research
- Works well independently and on team projects

Technical Skills:

- Experienced with probabilistic and mathematical modeling: Statistical Analysis, Computer Science and Applied Mathematics, Markov/Bayesian networks
- Language Proficiencies: Python, R, SQL, Java, git, Linux
- Bioinformatics: GWAS/EWAS, DNA methylation, meQTLs, biological sequence alignment, primer design, management of DNA synthesis and sequencing data, Hadoop/MapReduce

EDUCATION

Princeton University

Fall 2019-Present

Ph.D. Student in Quantitative and Computational Biology

Certificate in Statistics and Machine Learning

Advisor: Barbara Engelhardt, Computer Science

Graduate Fellowships and Awards:

Princeton Centennial Fellow

National Physical Science Consortium Fellow

Graduate Fellowship for STEM Diversity

Sponsored by the NSA

NSF Graduate Research Fellowship Honorable Mention 2019

Diversity & Leadership:

Princeton Insights Co-founder & Editor

QCB Scholars Planning Committee

Graduate Society of Women Engineers Executive Board

University of Washington, Cum Laude

Coursework Completed 2015

Bachelor of Science: Computer Science

Class of 2017

Bachelor of Science: Mathematics

Bachelor of Arts: Dance, Departmental Honors

Cumulative GPA: 3.82

Academy for Young Scholars, Robinson Center, 2010

Early entrance program at the University of Washington

Dance Program Scholarship Recipient, 2013

Annual Dean's List: 2011, 2012, 2013, 2014, 2015

Phi Beta Kappa, Alpha Chapter: National Academic Honors Society

RELATED EXPERIENCE

Research Assistant

Fall 2019 - Present

Lewis-Sigler Institute for Integrative Genomics

Princeton Computer Science – Princeton, NJ

Advisor: Barbara Engelhardt

- Expand statistical methods to identify genetic loci associated with sociological phenotypes on an under-served, racial minority and low socioeconomic status population.

- Develop methods to find and quantify differentially methylated regions (epigenetic markers), specifically related to early life stress & adversity.
- Utilize generalized linear models to perform GWAS and EWAS to associate genetic loci (SNPs) and epigenetic markers (DNA methylation) with phenotypes such as depression, bullying, neglect, and abuse.
- Analyze limitations of machine learning for complex data sets, in particular from the Fragile Families Challenge. This includes data imputation, handling missingness/sparsity, and applying influence functions to identify each individual's influence on prediction outcomes.
- Integrate a unique, complex, longitudinal dataset, the Fragile Families & Child Wellbeing Study, to examine genetic, epigenetic, development, and social influence to improve interventions for at-risk children.

Research Rotation

Winter/Spring 2020

Lewis-Sigler Institute for Integrative Genomics

Rotation Advisors: Olga Troyanskaya, Sarah Kocher

- Improve GWAS results to identify disease-gene associations using tissue-specific networks: Enhancing NetWAS for hypertension using various tissues and tissue combinations.
- Predict non-coding regulatory regions for social behavior in bees utilizing deep learning models for cis-regulatory codes trained on *Drosophila*.

Editor & Co-Founder

Fall 2019 - Present

Princeton Insights, insights.princeton.edu

- Edit reviews of scientific articles highlighting research at Princeton, written in the style of Nature's News & Views, to spotlight the most exciting and groundbreaking work at Princeton. Focus on making the reviews engaging, accessible, and scientifically accurate.
- Manage a graduate student organization with a staff of more than 15 post-doc, graduate student, and undergraduate writers.
- Collect recent scientific publications across all STEM fields at Princeton

Research Software Development Engineer 2

2018 – Summer 2020

Research Software Development Engineer

Summer 2015 – 2018

Microsoft Research – Redmond, WA

Project: DNA Storage

In partnership with University of Washington Molecular Information Systems Lab (MISL)

- Design and implementation of a pipeline for mathematical and statistical analysis of high-volume DNA synthesis and sequencing data from various sources to provide data visuals and metrics for biological experiments by measuring, quantifying, and identifying patterns in DNA errors.
- Experimental design in collaboration with the MISL team to explore the intersection of information technology and molecular-level manipulation using in-silico and wet lab experiments. For example, programmed the process of DNA random access by leveraging common biochemical techniques and utilizing existing alignment tools.
- Data handling of sequencing data to suggest optimizations of wet lab preparations and implement software processes, including encoding and decoding the library holding the current world record of data stored in DNA: 200MB.
- Management of a software group focused on integrating encoding, decoding, and analysis of DNA through an Azure cloud platform to create a WebApp to enable researchers and collaborators to utilize DNA Storage capabilities.

Artificial Intelligence Research Assistant

Fall 2014 - Spring 2015

University of Washington EE/CSE – Seattle, WA

Advisor/Research Scientist: Hannaneh Hajishirzi

- Natural language processing methods for solving word problems by analyzing each of the sentences in the problem statement to identify relevant variables and their values, and then mapping that information into an equation.
- Constructed a forest of trees that model equations for training data, complete with order of operations, entity matching, and incorporation of linguistic cues.
- Extracted units and converted them to the target unit type for word problem narratives.

Teaching Assistant

Fall 2014 - Spring 2015

University of Washington Computer Science Department – Seattle, WA

CSE 414/CSE 344, Introduction to Data Management (Database Systems)

Instructors: Alvin Cheung, Hal Perkins

- Topics of the course: introduction to database management systems, data models, query languages, language bindings, conceptual modeling, transactions, security, database tuning, data warehousing, parallelism, and web-data management.
- Weekly duties included teaching 106 students in 3 sections, planning meetings with full course staff, holding office hours, writing exams, and grading assignments.

Departmental Honors Dance Research

Spring 2014 - Spring 2015

University of Washington Dance Program – Seattle, WA

Advisor: Elizabeth Cooper, Divisional Dean of the Arts

- An investigation into the expression of emotion in dance, culminating in an evening length choreographic work and a formal academic research paper on findings.
- Presented independent research at the UW Undergraduate Research Symposium.

Software Engineering Intern-User Generated Content

Summer 2014

Groupon – Seattle, WA

- Creation of the internal User Generated Content Moderation Tool for monitoring tips and recommendations posted on the merchant pages.
- Full Website: programming in coffeescript, HTML/mustache, and CSS programming for a new Node.js application, as well as Java updates to the API and User Generated Content service clients.

Bioinformatics Research Assistant

Summer 2013

University of Washington Department of Medical Genetics – Seattle, WA

Hawkins Epigenomics Lab

- Mapped data to the human genome in a Unix environment, created scripts to find barcodes, alignment using bowtie2, data processing using Fastq and Sam files, data access in the UCSC Genome Browser
- R programming: differential peaks and table manipulation for RNA-Seq data

Category Insights Analyst Intern

Summer 2012

Clif Bar & Company – Emeryville, CA

- Comprehensive pricing analysis on mini bars: computed and evaluated price elasticity, compared to competitors, assessed how price compliance affects performance
- SKU assortment creation: worked on a team with field sales to create a new sku assortment for an account, evaluated a new sku rationalization
- Created an automated link between the Consumer Decision Trees and data
- Worked with Pivot tables and linear regressions on large Nielsen data sets

PUBLICATIONS

Probing the Physical Limits of Reliable DNA Data Retrieval

Nature Communications **11**, 2020.

Lee Organick, Yuan-Jyue Chen, [Siena Dumas Ang](#), Randolph Lopez, ..., Karin Strauss, Luis Ceze

Quantifying molecular bias in DNA data storage

Nature Communications **11**, 2020.

Yuan-Jyue Chen, C.N. Takahashi, L. Organick, C. Bee, [Siena Dumas Ang](#), ..., Karin Strauss

A comprehensive study of synthetic DNA preservation for DNA data storage

Submitted to ACS Nano, September 2020.

Lee Organick, Bichlien Nguyen, Rachel McAmis, Weida Chen, Xavier Kroll, [Siena Dumas Ang](#), ...

DNA Assembly for Nanopore Data Storage Readout

Nature Communications **10**, 2019.

Randolph Lopez, Yuan-Jyue Chen, [Siena Dumas Ang](#), Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, Georg Seelig, Karin Strauss, Luis Ceze

Random Access in Large-Scale DNA Data Storage

Nature Biotechnology **36**, 2018.

Lee Organick, [Siena Dumas Ang](#), Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, ..., Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze, and Karin Strauss

Clustering Billions of Reads for DNA Data Storage

Neural Information Processing Systems (NeurIPS), 2017.

Selected for a spotlight presentation (top 4.7% of submissions).

Cyrus Rashtchian, Konstantin Makarychev, Miklos Racz, [Siena Dumas Ang](#), Djordje Jevdjic, Sergey Yekhanin, Luis Ceze, and Karin Strauss

Parsing Algebraic Word Problems into Equations

Transactions of the Association for Computational Linguistics (ACL), 2016.

Rik Koncel-Kedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, [Siena Dumas Ang](#)

Scaling up DNA Data Storage

Synthetic Biology Conference (SEED), 2017.

Chris Takahashi, [Siena Dumas Ang](#), Lee Organick, Yuan-Jyue Chen, ...

End-to-end DNA Archival Storage System

American Chemical Society, 2017.

Lee Organick, [Siena Dumas Ang](#), Yuan-Jyue Chen, ...

PATENTS

Trace reconstruction from noisy polynucleotide sequencer reads

Patent filed, April 2017. Application: US20180211001A1, WO2017189469A1

Parikshit S. Gopalan, Sergey Yekhanin, [Siena Dumas Ang](#), Nebojsa Jojic, Miklos Racz, Karin Strauss, Luis Ceze

Polynucleotide sequencer tuned to artificial polynucleotides

Patent filed, March 2017. Application: US20180253528A1

Karin Strauss, [Siena Dumas Ang](#), Luis Ceze, Yuan-Jyue Chen, Hsing-Yeh Parker, Bichlien Nguyen, Robert Carlson

Primer design for retrieval of stored nucleotides

Patent filed, February 2017. Application: US20180223341A1, WO2018148099A1

Yuan-Jyue Chen, Karin Strauss, Luis Ceze, [Siena Dumas Ang](#), Sergey Yekhanin

Primer design for retrieval of stored polynucleotides

Patent filed, February 2017. Application: US20180223340A1, WO2018148085A1

Yuan-Jyue Chen, Luis Ceze, Sergey Yekhanin, Siena Dumas Ang, Karin Strauss

Efficient clustering of noisy polynucleotide sequence reads

Patent filed, September 2017. Application: WO2018063950A1

Luis Ceze, Sergey Yekhanin, Siena Dumas Ang, Karin Strauss, Cyrus Rashtchian, Ravindran Kannan, Konstantin Makarychev

Error correction for nucleotide data stores

Patent filed, January 2016. Application: US20170141793A1, WO2017083177A1

Karin Strauss, Siena Dumas Ang, Luis Ceze, James Bornholt

OTHER EXPERIENCE

Editor & Co-founder Princeton Insights	2019 – Present
QCB Scholars Planning Committee	2020 - Present
Assistant Instructor, Princeton Lewis Center of the Arts	Fall 2020 - Present
Graduate Society of Women Engineers Executive Board	Fall 2020 - Present
Princeton University Store Graduate Trustee	Fall 2020 – Present
Leader of Microsoft Research “Mavens” Book Club	2017 – 2019
Microsoft Grace Hopper Celebration Ambassador	Fall 2018
Principal Dancer & Aerialist, The Cabiri Aerial Company	2015 – 2019
Stage Manager, UW Dance Program	2014 – 2019
Dancer, Relay Dance Collective	2015 – 2018
Keynote Speaker & Judge, TeenTech WA	2017

References Available Upon Request