

EXPERIENCE

Team Lead Data Engineering at Ginkgo Bioworks 2022–Present

- Lead and managed team of data engineers, system administrators, statisticians, bioinformaticians, and scientists at the PhD level working within the AgBio unit of Ginkgo Bioworks.
- Mentored and coached team members in data science, bioinformatics, data engineering, and statistics.
- Key leadership role in successful merger of AgBio unit with Ginkgo, including all relevant R&D business applications and data-adjacent systems.

Team Lead Data Engineering at Bayer Crop Science 2018–2022

- Hired, managed, and developed team of 5+ Data Engineers, Systems Administrators, and Business Analysts working within the Biologics R&D unit of Bayer Crop Science enabling data capture, data integration, and operationalization of data analysis pipelines
- Developed and supervised implementation of data capture, integration, and analysis strategies to increase the value of genomics, metabolomics, transcriptomics, spectroscopic, phenotypic (*in vitro* and *in planta*), and fermentation/formulation process data for discovery and development
- Lead the development of multiple systems while coaching, mentoring, and developing developers and engineers
- Served as a key collaborator on multiple cross-function and cross-divisional projects, including leading the architecture of a life science collaboration using serverless architecture to provide machine-learning estimates of critical parameters from spectrographic measurements
- Established and developed network of internal and external contacts for technical implementation of Bayer program goals.

Debian Developer 2004–Present

- Maintained, managed configurations, and resolved issues in multiple packages written in R, perl, python, scheme, C++, and C.
- Resolved technical conflicts, developed technical standards, and provided leadership as the elected chair of the Technical Committee.
- Developer of Debbugs, a perl and SQL-based issue-tracker with ≥ 100 million entries with web, REST, and SOAP interfaces.
- Provided vendor-level support for complex systems integration issues on Debian GNU/Linux systems.

Research Scientist at UIUC 2015–2017

- Planning, design, organization, execution, and analysis of multiple complex epidemiological studies involving epigenomics, transcriptomics, and genomics of diseases of pregnancy and post-traumatic stress disorder.
- Published results in scientific publications and presented results orally at major scientific conferences.
- Wrote and completed grants, including budgeting, scientific direction, project management, and reporting.
- Mentored graduate students and collaborated with internal and external scientists.
- Performed literature review, training, and applied new techniques to maintain abreast of current scientific literature, principles of scientific research, and modern statistical methodology.
- Wrote software and designed relational databases using R, perl, C, SQL, make, and very large computational systems (Blue Waters)

Postdoctoral Researcher at USC 2013–2015

- Design, execution, and analysis of an epidemiological study to identify genomic variants associated with systemic lupus erythematosus using targeted deep sequencing.
- Wrote multiple pieces of software to reproducibly analyze and archive large datasets resulting from genomic sequencing.
- Coordinated with clinicians, molecular biologists, and biologists to produce analyses and major reports.

Postdoctoral Researcher at UCR 2010–2012

- Executed and analyzed an epidemiological study to identify genomic variants associated with systemic lupus erythematosus using prior information and array based approaches in a trio and cross sectional study of individuals from the Los Angeles and greater United States.
- Wrote and maintained multiple software components to reproducibly perform the analyses.

EDUCATION

Doctor of Philosophy (PhD) in Cell, Molecular and Developmental Biology UC Riverside
Bachelor of Science (BS) in Biology UC Riverside

SKILLS

Leadership and Mentoring

- Lead teams of PhD and MD scientists in multiple scientific and industrial programs
- Mentored graduate students and Outreachy and Google Summer of Code interns
- Former chair of Debian's Technical Committee
- Head developer behind <https://bugs.debian.org>

Bioinformatics, Genomics, and Epigenomics

- NGS and array-based Genomics and Epigenomics of complex human diseases using RNA-seq, targeted DNA sequencing, RRBS, Illumina bead arrays, and Affymetrix microarrays from sample collection to publication
- Reproducible, scalable bioinformatics analysis using make, nextflow, and cwl based workflows on cloud- and cluster-based systems on terabyte-scale datasets
- Alignment, annotation, and variant calling using existing and custom software, including GATK, bwa, STAR, and kallisto

- Using evolutionary genomics to identify causal human variants

Statistics

- Statistical modeling (regression, inference, prediction, and machine learning in very large (> 1TB) datasets) using R and python.
- Correcting & experimental design to overcome multiple testing, confounders, and batch effects (both Bayesian and frequentist)
- Reproducible research

Software Development

- Languages: python, R, perl, C, C++, python, groovy, sh (bash, POSIX, and zsh), make
- Collaborative Development: git, Jira, gitlab CI/CD, github actions, Aha!, continuous integration & deployment, automated testing
- Web, Mobile: Shiny, jQuery, JavaScript
- Databases: Postgresql (PL/SQL), SQLite, Mysql, NoSQL

Big Data

- Parallel and Cloud Computing (slurm, torque, AWS, OpenStack, Azure)
- Inter-process communication: MPI, OpenMP
- Filestorage: Gluster, CEFS, GPFS, Lustre
- Linux system administration

Applications and Daemons

- Web: apache, nginx, varnish (load balancing/caching), REST, SOAP, Tomcat
- Build Tools: GNU make, cmake
- Virtualization: libvirt, KVM, qemu, VMware, docker
- VCS: git, mercurial, subversion
- Mail: postfix, exim, sendmail, spamassassin
- Configuration Infrastructure: puppet, hiera, etckeeper, git
- Documentation: \LaTeX , confluence, emacs, Markdown, MediaWiki, ikiwiki, trac
- Monitoring: munin, nagios, icinga, prometheus
- Issue Tracking: Debbugs, Request Tracker, Trac, JIRA
- Office Software: Gnumeric, Libreoffice, \LaTeX , Word, Excel, Powerpoint

Networking

- Hardware, Linux routing and firewall experience, ferm, DHCP, openvpn, bonding, NAT, DNHS, SNMP, IPv4, and IPv6.

Operating systems

- GNU/Linux (Debian, Ubuntu, Red Hat)
- Windows
- MacOS

Communication

- Strong written communication skills as evidenced by publication record
- Strong verbal and presentation skills as evidenced by presentation, leadership, and teaching record

AUTHORED OPEN SOURCE SOFTWARE

- **Debbugs**: Bug tracking software for the Debian GNU/Linux distribution. <https://bugs.debian.org>
- **CairoHacks**: Bookmarks and Raster images for large PDF plots in R.
- **Function2Gene**: Gene selection tool based on literature mining which enables Bayesian approaches to significance testing.
- **Helical Wheel Projections**: Web-based tool to draw helical wheel protein projections.

PUBLICATIONS AND PRESENTATIONS

- 24 peer-reviewed publications cited over 3000 times: <https://dla2.us/pubs>
- Publication record in GWAS, transcriptomics, SLE, GBM, epigenetics, comparative evolution of mammals, and lipid membranes
- H index ≥ 20
- Multiple presentations on EWAS of PTSD, genetics of SLE, and Open Source: <https://dla2.us/pres>

FUNDING AND AWARDS

Grants

- 2017 R Consortium: **Adding Linux Binary Builders to R-Hub** Role: Co-PI
- 2015 Blue Waters Allocation Grant: **Making ancestral trees using Bayesian inference to identify disease-causing genetic variants** Role: Primary Investigator
- **Tracking placenta and uterine function using urinary extracellular vesicles** (R21 RFA-HD-16-037) Role: Key Personnel
- **NIAMS R01-AR045650-04 Genetics of Childhood Onset SLE** to Chaim O. Jacob. Role: Key Personnel

Scholarships and Fellowships

- 2001–2003: University of California, Riverside Doctoral Fellowship
- 1997–2001: Regents of the University of California Scholarship.