

Frank Austin Nothhaft

fnothaft@alumni.stanford.edu • 202.340.0466

Work Experience

Tempus Labs

Senior Director, Platform Services Engineering **November 2020–present**

- I lead the engineering and computational biology teams supporting the development of the Tempus Lens platform, a software-as-a-service solution for oncology research. I have led multiple product launches across a broad range of platform components, including a web application for rapid biostatistical (e.g., rwOS) and molecular (e.g., co-mutation landscape) analysis of large multimodal oncology datasets, a cloud-hosted data science environment for computational biology analyses, and multiple new de-identification implementations.
- As a Senior Director accountable for the commercial success of the platform, I established both a pre-sales computational biology team and customer success/documentation team to support our strategic collaborations. I work closely with scientific and IT stakeholders at our research collaborators and within Tempus' professional science services teams to drive adoption of Lens. Internally, I work with key functions (site reliability engineering, finance/accounting) to drive improvements to core KPIs, including a multi-million dollar reduction in cloud costs, 10× reduction in latency through key product endpoints, and improvement of data science platform uptime SLI.
- As a hiring manager across multiple functions, I have created a hiring process that allowed our teams to scale from 10 to 40 employees while maintaining a high talent bar in both engineering and science functions. This hiring process also allowed our team to improve both gender and URM-status diversity by 20%, at both the IC and managerial level.

Databricks

Technical Director, Health/Life Sciences **January 2019–November 2020**

Go-To-Market Lead, Genomics **January 2018–December 2018**

Genomics SME (Consultant) **February 2017–December 2017**

- I led the development of the Databricks Runtime for Genomics from initial conception through general availability, across a team of three engineers and two bioinformaticians dedicated to the product. We delivered a high performance implementation of the GATK4 best practices germline, somatic, and joint genotyping pipelines, delivered as a managed cloud service. We achieved a 3× latency improvement over the Edico DRAGEN platform with comparable cost and accuracy.
- I supported multiple key pharmaceutical partners who were part of the UK Biobank exome sequencing effort. I led the development of a strategic product partnership with the Regeneron Genetics Center around Project Glow, an Apache Spark based tool for managing large genomic variation datasets that provided a 10× improvement in performance for genome wide association studies at the million patient scale.
- In the latter half of my tenure, I also assumed responsibility for all go-to-market activities (marketing, sales, partnerships/business development, and field engineering) across our healthcare/life sciences customers. From 2018 through the end of the first half of 2020, we grew committed customers by 7× and annual recurring revenue by 12×. Our clients included eight of the top ten global pharmaceutical companies, four of the top five US payers, and three of the top ten US hospitals.

**Work
Experience
(continued)**

**UC Berkeley AMPLab
Research Assistant**

August 2013–December 2017

- I led development of the ADAM/Big Data Genomics tool suite which provided an open source, Apache Spark-based implementation of the “GATK Best Practices” pipeline. Our pipeline could process a 30× WGS in less than 1 hour, with comparable accuracy to the GATK, and grew to have more than 75 open source contributors.
- In collaboration with UCSC, I contributed to the Toil workflow runner and worked to compare the performance of Kalisto vs. STAR/RSEM RNA-seq quantification pipelines through a joint analysis of more than 20,000 RNA-seq samples, culminating in a publication in Nature Biotechnology.
- As the primary grantwriter for the AMPLab Genomics team, I led cross-institute proposal efforts for successful NIH BD2K (with UCSC, OHSU, Sage, and UCSF), Cancer Cloud Pilot (with UCSC and Broad Institute), and SBIR grants (with Curoverse), as well as a Chan Zuckerberg Institute grant.

Color Genomics

Bioinformatics Consultant

October 2013–December 2014

Developed a germline copy number variant calling pipeline that ran in a CAP/CLIA validated setting as part of a targeted NGS panel for hereditary cancer screening.

Broadcom, R&D Engineer, IC Design 2

March 2016–June 2016

Broadcom, Engineer, Staff 1—IC Design

April 2012–March 2016

Broadcom, Engineer—IC Design

June 2011–April 2012

I developed novel validation techniques for cellular, BlueTooth, WiFi, and 802.11ad chipsets. My techniques enabled modeling circuit dynamics, and hardware accelerated circuit simulation.

Education

University of California, Berkeley

Doctor of Philosophy, Computer Science. August 2013–December 2017. GPA: 3.81.

Masters of Science, Computer Science. August 2013–May 2015. GPA: 3.79.

Advisors: Dave Patterson and Anthony Joseph.

Stanford University

Bachelor of Science with Honors, Electrical Engineering. September 2007–June 2011.

Minor in Management Science & Engineering. GPA: 3.24.

**Selected
Publications**

Full publication list available at <https://scholar.google.com/citations?user=BYSy-9oAAAAJ>.

- Lisa Wu, David Bruns-Smith, **Frank Austin Nothhaft**, Qijing Huang, Sagar Karandikar, Johnny Le, Andrew Lin, Howard Mao, Brendan Sweeney, Krste Asanović, David A Patterson, Anthony D Joseph. “FPGA Accelerated INDEL Re-alignment in the Cloud.” In Proceedings of the International Symposium on High Performance Computer Architecture, February 2019 (HPCA '19).
- Alyssa Kramer Morrow, George Zhixuan He, **Frank Austin Nothhaft**, Justin Paschall, Nir Yosef, Anthony D. Joseph. “Mango: Distributed Visualization for Genomic Analysis.” In Cell Systems, December 2019. Originally posted as *BioRxiv:360842*, July 2018.

**Selected
Publications
(continued)**

- Michael D. Linderman, Davin Chia, Forrest Wallace, and **Frank Austin Nothhaft**, “DECA: ScalableXHMM exome copy-number variant calling with ADAM and Apache Spark.” In BMC Bioinformatics, December 2019. Originally posted as *BioRxiv:062497*, September 2017.
- John Vivian, Arjun Rao, **Frank Austin Nothhaft**, Christopher Ketchum, Joel Armstrong, Adam Novak, Jacob Pfeil, Jake Narkizian, Alden D. Deran, Audrey Musselman-Brown, Hannes Schmidt, Peter Amstutz, Brian Craft, Mary Goldman, Kate Rosenbloom, Melissa Cline, Brian O’Connor, Megan Hanna, Chet Birger, W. James Kent, David A. Patterson, Anthony D. Joseph, Jingchun Zhu, Sasha Zaranek, Gad Getz, David Haussler, and Benedict Paten. “Toil enables reproducible, open source, big biomedical data analyses.” In *Nature Biotechnology*, April 2017. Originally posted as *BioRxiv:062497*, July 7, 2016.
- **Frank Austin Nothhaft**, Matt Massie, Timothy Danford, Zhao Zhang, Uri Laserson, Carl Yeksigian, Jey Kottalam, Arun Ahuja, Jeff Hammerbacher, Michael Linderman, Michael J. Franklin, Anthony D Joseph, and David A. Patterson. “Rethinking data-intensive science using scalable analytics systems.” In *Proceedings of the International Conference on Management of Data*, May 2015 (SIGMOD ’15).
- Selected industrial white papers:
 - Introducing GlowGR: An industrial-scale, ultra-fast and sensitive method for genetic association studies. *Joint white paper with the Regeneron Genetics Center discussing the sensitivity of the GlowGR GWAS method*. June 2020.
 - Accurately Building Genomic Cohorts at Scale with Delta Lake and Spark SQL. *Analysis of the accuracy and scalability of the Databricks joint genotyping pipeline*. June 2019.
 - Building the Fastest DNaseq Pipeline at Scale. *Performance benchmarking of Databricks Genomics pipeline vs. Edico DRAGEN*. September 2018.