

Nikhil Haas

nikhil.haas@gmail.com | 206-851-3416

EDUCATION

M.S. HEALTH INFORMATICS

UNIVERSITY OF SAN FRANCISCO
Dec. 2015

B.S. BIOMEDICAL ENG.

BOSTON UNIVERSITY
COLLEGE OF ENGINEERING
May 2009

LINKS

Github:// [nsh87](#), [nhaas-twist](#)LinkedIn:// [nikhilhaas](#)Website:// [nikhilhaas.com](#), [ishcra.com](#)

OPEN SRC. PROJECTS

ReceptorMarker: R package to analyze antibody/TCR sequences and markers**regressors**: Package for stats and plots with fitted scikit-learn regression models**EHRcorral**: Probabilistically generate master patient index – Python package**MIMIC**: Provision VM with ICU data**shinyVM**: Quickly create Shiny apps

SKILLS / KNOWLEDGE

Programming Languages & Markup:

Python • R • SQL • JavaScript
MATLAB • Markdown • C++ • HTML

Scientific Programming:

NumPy/SciPy • scikit-learn • MATLAB
Cython • ggplot2 • pandas • matplotlib

Frameworks, Tools, & Other:

AWS: EC2/RDS/S3 • Git • Unix • ETL
Postgres • Mongo • Drill • Django
Ansible • Celery • RabbitMQ • Travis CI
RESTful APIs • Vagrant • Fabric • D3.js
GraphQL • Flask • Docker • Samtools
BWA • FASTQ/A • NGS (Short Read)
pytest (unit testing) • testthat (R)

Stats & ML:

(Un+)Supervised Machine Learning
Clustering • Regression • Classification
Dimension Reduction • Correlation
Algorithm Implementation

EXPERIENCE

TWIST BIOSCIENCE | BIOINFORMATICS ENG. | SNR. DATA SCIENTIST

2016 - Present | San Francisco, CA

- Create and maintain production computational packages: sequence-based toolkits, annotator, custom NGS variant caller, uniform TE probe designer.
- Sped up codon optimizer 9x via Cython to allow its use on ecommerce platform.
- Created first of kind distributed biosecurity screening system for ordered DNA.
- Perform numerical and statistical analyses as needed to support software improvement, prototyping, or other teams needing assistance analyzing data.
- Blueprinting de novo assembler for pooled variant libraries using GATB (C++). Exposed to Rust-based sequence designer. Quantitative unit tests developed.
- Spin up APIs, create databases, utilize IaaS/SaaS, and generate infrastructure needed to support pipeline automation. Make design and implementation decisions to enable meeting performance, deployment, or other constraints.

STANFORD UNIV. SCHOOL OF MEDICINE | BIOENGINEER

2015 – Present | Stanford, CA

- Designed and built [receptormarker.com](#) for sequence analysis, clustering, similarity detection, and visualization of single-cell immunological data.
- Co-authored in R several JS and HTML-based visualizations, binary clustering methods, and callable dimension-reduction techniques for identifying T-cell responses and antibody binding region DNA sequences from biopsied patients.

UNIVERSITY OF SAN FRANCISCO | ADJUNCT PROFESSOR

2016 – 2017 | San Francisco, CA

- Pitched and taught Introduction to Exploratory Data Analysis to introduce M.S. students to data manipulation in Python, exploratory visualizations and correlation metrics, unsupervised learning and outlier / feature detection.
- Co-taught M.S. course in Semantic Organization of Health Information and Data Standards: SQL/NoSQL, data semantics, parsing, extraction, ontologies.

ACADEMIC WORK | M.S. HEALTH INFORMATICS STUDENT

2014 - 2015 | San Francisco, CA

- Focused on machine learning, statistical methods, and C.S. in biotech and healthcare:
- Studied the mathematical and statistical bases for machine learning algorithms; coded several (e.g. HMM) algorithms using their mathematical formulas.
 - Performed in-depth examinations of [PCA](#), regression and [shrinkage methods](#).
 - Compared [feature selection](#) methods for classification of heart failure.

BIO-RAD LABORATORIES | BUSINESS SYSTEMS ANALYST

2011 – 2014 | Hercules, CA

- Facilitated projects between IT, Product, and Global & Digital Marketing.
- Led functional design of major- and minor-release backend projects for website.
- Utilized internal search and AdWords APIs to create Django dashboard for analyzing customer intent and desired products, used by various product teams.

AMPED SYSTEMS, INC. | COFOUNDER

2010 – 2012 | Seattle, WA

- Created course material lookup and textbook price comparison website that integrated 535 college APIs; monetized through affiliate programs.

SELECT COURSES

Statistical Computing for
Biomedical Data Analytics //
Bioinformatics //
Linear Regression Analysis //
Intro to Machine Learning //
Clinical Decision Support and
Health Data Analytics //
NoSQL & Relational Databases //
Computation for Analytics //
Probability & Statistics //
Physiology and Cell Biology //
Thermodynamics //
Signals Processing //

GRANTS & SCHOLARSHIPS

2015	Computational & Systems Immunology	Gates Fnd. Grant	\$50K
2015	NumFOCUS Diversity Scholarship	SciPy2015 Conf.	\$2K
2014-2015	Trustee & Innovation Scholarship	University of SF	\$3K

RESEARCH & PUBLICATIONS

HARVARD-MIT HEALTH SCIENCES & TECHNOLOGY

2008 – 2009 | Cambridge, MA

Developed an approach to assemble cell-laden, vascularized 3D tissue constructs that mimicked microvasculature in the [Khademhosseini Lab](#).

- [1] Y. Du, M. Ghodousi, H. Qi, N. Haas, W. Xiao, and A. Khademhosseini, "Sequential assembly of cell-laden hydrogel constructs to engineer vascular-like microchannels," *Biotechnology and Bioengineering*, 2011.

STANFORD UNIVERSITY SCHOOL OF MEDICINE

2015 – 2017 | Stanford, CA

Aided in analysis and figure generation using self-built, open source analytics web platform for T-cell repertoire analysis developed with funding from the Gates Foundation through the [Mark M. Davis Lab](#).

- [1] J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. L. Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, and M. M. Davis, "Identifying specificity groups in the T cell receptor repertoire," *Nature*, 2017.