

Alex Rubinsteyn

alex.rubinsteyn@unc.edu | pirl.unc.edu | www.rubinsteyn.com

EDUCATION

Courant Institute, NYU

AUGUST 2009 — SEPTEMBER 2014

Ph.D. Computer Science

Thesis: *Runtime Compilation of Array-Oriented Python Programs*

Advisor: Dennis Shasha

Courant Institute, NYU

AUGUST 2007 — AUGUST 2009

M.Sc. Computer Science

University of Illinois Urbana-Champaign

AUGUST 2002 — MAY 2006

B.Sc. Individual Plan of Study in Neuroscience

Minor in Computer Science

WORK

University of North Carolina - Chapel Hill

Personalized Immunotherapy Research Lab

Assistant Professor

MARCH 2020 — Present

Started the Personalized Immunotherapy Research Lab (PIRL) with Dr. Ben Vincent. We pool resources and tightly coordinate research to craft machine learning driven personalized cancer immunotherapies.

Current research projects:

- Biological language models for generating antigen-specific T-cell receptor sequences.
- Designing and testing personalized TCR therapeutics (in vivo TCR-T and multi-TCR soluble T-cell engagers) which can be used with either endogenous or AI-generated T-cell receptor sequences (in collaboration with the Kuhlman Lab).
- Novel algorithms for uncovering large mutations (and the ORFs they generate) in cancer from long read DNA/RNA sequencing methods (Oxford Nanopore and Pacific Biosciences).
- Characterizing the surface peptide-MHC repertoire and T-cell epitope landscape of NUTM1 fusion driven carcinomas (rare deadly cancer with potential low hanging fruit for cure through personalized immunotherapy)
- Optimize therapeutic vaccine formulation for maximal T-cell responses and T-cell functionality (as a basis for future improved personalized cancer vaccine trials).

Previous:

- Improved genomics and machine learning tools used in upcoming PANDA-VAC personalized cancer vaccine trial.
- Collaborated with Jeremy Wang and the UNC clinical testing lab to start the UNC genomic pathogen surveillance lab.
- Created T-cell epitope directed peptide vaccine for SARS-CoV-2 which was immunogenic but, remarkably for SARS-CoV-2 vaccines, completely ineffective in mice.

Icahn School of Medicine at Mount Sinai School

Instructor & head of the OpenVax research group

JANUARY 2018 — FEBRUARY 2020

- Created new research group focused on open source software for mutation-specific tumor antigen discovery and helping run phase I personalized cancer vaccine clinical trials at Mount Sinai.

- In addition to two neoantigen vaccine clinical trials created as part of Hammer Lab, helped initiate third trial (NCT03359239) in bladder cancer, which combines our personalized vaccine with a checkpoint inhibitor.
- Responsible for all computational aspects of neoantigen vaccine trials at Mount Sinai, in addition to working with sequencing core and peptide manufacturer to identify and fix any problems that arise.
- Maintaining and improving immunogenomic software tools (e.g. improved filtering of somatic variants based on features determined from manual curation of previous patients).
- Other research projects:
 - evaluating concordance between different neoantigen prediction pipelines (low)
 - convolutional neural network Class II MHC binding model
 - harmonizing immune response data across vaccine trials to create ground truth datasets

Research scientist in the Hammer Lab

JANUARY 2014 — DECEMBER 2017

- Developed computational pipeline for personalized cancer vaccine trials at the Tisch Cancer Institute (NCT02721043, NCT03223103).
- Wrote portions of clinical trial protocol related to sequencing, bioinformatics, and neoantigen selection.
- Clinical trial operations between receipt of patient sequencing data up to ordering synthesis of personalized vaccine peptides, including detailed QC investigations of data anomalies and manual inspection of tumor mutational data.
- Created open source Python libraries for working with genomic mutations, predicting their protein-level effects, and prioritizing them for synthesis as part of peptide-based vaccine.
- Developed neural network models for predicting steps of antigen processing, such as binding of peptides to Class I MHC molecules (MHCflurry) and Class II MHC molecules.
- Helped build models of response to checkpoint blockade based on cohort sequencing data from clinical trials.
- Worked with Neon Therapeutics on a sponsored research agreement to predict antigen processing from mass spectrometry MHC ligandome data.

Capital K Partners

MAY 2011 — AUGUST 2012

Financial data analysis, trading algorithm design and system implementation

- Created data collection pipeline which gathers currency market order books from multiple venues, extracts multi-scale time-varying features, and stores these efficiently in HDF5 files.
- Implemented data analysis framework which distributes time-series feature extraction, ad-hoc queries, and model estimation across remote compute servers.
- Implemented trading algorithms in Python by interfacing with distributed trading system via ZeroMQ.

Google Inc.

SUMMER 2009

Software engineering intern

- Wrote MapReduce pipeline to compute term similarity statistics from co-occurrence data in medical corpora.
- Created framework for fusing heterogeneous health taxonomy data sources.
- Wrote prototype of query-rewriting health search using Python and Django.

BIR Training Center

Software Developer

MAY 2006 — JUNE 2007

- Developed student and teacher portals for grading, scheduling and attendance.
- Created development standards for all new software projects including use of version control and templating.
- Hired Senior Graphic Designer and Junior Web Developer.

TEACHING

Machine Learning for Computational Biology (BCB 726) FALL 2024.
Introduction to classical and deep learning methods in computational biology, culminating in applications of large language models to DNA and protein sequences.

Critical readings in biomedical statistics and machine learning (BCB 731) FALL 2024.
Graduate seminar at UNC exploring pitfalls of machine learning applied to high-dimensional biological and medical data.

Programming Languages (CS 2110) FALL 2011.
Teaching assistant with a weekly recitation for a Masters-level programming language theory/survey course at NYU.

GED / Basic Math OCTOBER 2009 — JUNE 2011.
Taught a weekly math class at the Fortune Society, which provides services and education to formerly incarcerated adults.

Tutoring FALL 2007 — FALL 2013.
Individual tutoring in calculus, undergraduate and Masters-level algorithms, and machine learning courses.

Teacher at cSplash 2008, 2009, 2014, 2019
cSplash is a day-long festival of math and science for high school students held at the Courant Institute.
Courses taught: *Automata and Language Models*, *Introduction to Machine Learning*, *This Machine Kills Cancer* and *Biology for Math Nerds*.

CONFERENCES AND WORKSHOPS

NUT Carcinoma Symposium MARCH 6TH, 2024
Gathered researchers and clinicians studying NUT carcinoma, a rare and lethal cancer without effective therapies, to foster collaboration and cross-pollination of research directions.

UNC Biotech Day APRIL 3RD, 2024
Organized one day conference featuring speakers from biotech companies in the Research Triangle Park to encourage industry/academic partnerships with students and faculty at UNC.

PyData Miami JANUARY 9TH-11TH, 2019
Lead organizer of a machine learning and data science conference in Miami which brought together researchers from industry and academia with practitioners and students.

PLDI: ARRAY JUNE 13TH, 2015
Started and co-chaired the ACM SIGPLAN International Workshop on Libraries, Languages, and Compilers for Array Programming, which is now hosted annually at the Programming Language Design and Implementation (PLDI) conference.

SELECT RESEARCH PUBLICATIONS

- **Conditional generation of real antigen-specific T cell receptor sequences**
Dhuvi Karthikeyan, Sarah Bennett, Amy Reynolds, Benjamin Vincent, Alex Rubinsteyn
Nature Machine Intelligence (2025)
- **PGV001, a multi-peptide personalized neoantigen vaccine platform: phase I study in patients with solid and hematologic malignancies in the adjuvant setting**
Mansi Saxena, Thomas Marron, Julia Kodysh, ..., Alexander Rubinsteyn, Nina Bhardwaj
Cancer Discovery (2025)
- **Multi-Task Training Increases Native Sequence Recovery of Antigen-Specific T-cell Receptor Sequences**
Dhuvi Karthikeyan, Alex Rubinsteyn
ICML Workshop on Efficient and Accessible Foundation Models (2024)
- **ACE configurator for ELISpot: optimizing combinatorial design of pooled ELISpot assays with an epitope similarity model**
Andy JS Lee, Dhuvi Karthikeyan, Misha Fini, Benjamin Vincent, Alex Rubinsteyn
Briefings in Bioinformatics (2024)
- **Landscape and selection of vaccine epitopes in SARS-CoV-2**

Christof C Smith, Kelly S Olsen, Kaylee M Gentry, Maria Sambade, Wolfgang Beck, Jason Garness, Sarah Entwistle, Caryn Willis, Steven Vensko, Allison Woods, Misha Fini, Brandon Carpenter, Eric Routh, Julia Kodysh, Timothy O'Donnell, Carsten Haber, Kirsten Heiss, Volker Stadler, Erik Garrison, Adam M Sandor, Jenny PY Ting, Jared Weiss, Krzysztof Krajewski, Oliver C Grant, Robert J Woods, Mark Heise, Benjamin G Vincent, Alex Rubinsteyn
Genome Medicine (2021)

- **MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing**
Timothy J. O'Donnell, Alex Rubinsteyn, Uri Laserson
Cell Systems (2020)
- **Bioinformatic methods for cancer neoantigen prediction**
Sebastian Boegel, John C. Castle, Julia Kodysh, Timothy O'Donnell, Alex Rubinsteyn
Book chapter in Progress in Molecular Biology and Translational Science (2019)
- **MHCflurry: Open-Source Class I MHC Binding Affinity Prediction**
Timothy O'Donnell, Alex Rubinsteyn, Maria Bonsack, Angelika B. Riemer, Uri Laserson, Jeff Hammerbacher
Cell Systems (2018)
- **Computational Pipeline for the PGV-001 Neoantigen Vaccine Trial**
Alex Rubinsteyn, Julia Kodysh, Nina Bhardwaj, Jeff Hammerbacher
Frontiers in Immunology (2017)
- **Predicting Peptide-MHC Binding Affinities With Imputed Training Data**
Alex Rubinsteyn, Timothy O'Donnell, Nandita Damaraju, Jeff Hammerbacher.
ICML Computational Biology Workshop (2016)
- **Using a Machine Learning Approach to Predict Outcomes after Radiosurgery for Cerebral Arteriovenous Malformations**
Eric Oermann, Alex Rubinsteyn, D. Ding, J. Mascitelli, R. Starke, J. Bederson, H. Kano, D. Lunsford, J. Sheehan, J. Hammerbacher, D. Kondziolka.
Scientific Reports (2016)
- **Mutation-Derived Tumor Antigens: Novel Targets in Cancer Immunotherapy**
John Finnigan, Alex Rubinsteyn, Jeff Hammerbacher, Nina Bhardwaj.
Oncology (December 2015)
- **Predicting Patient Outcomes After Radiosurgery For Arteriovenous Malformation**
Alex Rubinsteyn, Eric Oermann, Jeff Hammerbacher, D. Kondziolka.
NeurIPS Workshop on Machine Learning for Clinical Data Analysis, Healthcare and Genomics (2014)
- **Learning Random Forests on the GPU**
Yisheng Liao, Alex Rubinsteyn, Russell Power, Jinyang Li.
NeurIPS Workshop on Parallel and Large-scale Machine Learning (2013)
- **Parakeet: A Just-In-Time Parallel Accelerator for Python**
A. Rubinsteyn, N. Weinman, E. Hielscher, D. Shasha. HotPar 2012