

# Benjamin B. Chu

655 Bradford Street, Apt 633, Redwood City, CA, 94063, USA

☎ (+1) 909-569-4530 | ✉ [bbchu@stanford.edu](mailto:bbchu@stanford.edu) | 🏠 <https://biona001.github.io/> | 💻 <https://github.com/biona001>

## Research Interests

<b>Math and Statistics</b>	Knockoffs, Penalized regression methods, Generalized linear models, Variance component models
<b>Computation</b>	Parallel and High-Performance Computing, Big Data Analysis, Package Development
<b>Biology</b>	Genome-Wide Association Studies, Phasing and imputation, admixture estimation, polygenic risk scores

## Education

### University of California, Los Angeles

Los Angeles, California

PH.D BIOMATHEMATICS

09/2016 - 08/2021

- Advisors: Kenneth Lange and Janet Sinsheimer
- Thesis: Scalable Algorithms for Genetic Association Studies, Genotype Imputation, and Ancestry Inference

### University of California, Berkeley

Berkeley, California

B.A., APPLIED MATHEMATICS

09/2012 - 05/2016

## Positions

### Postdoctoral Researcher, Stanford University

Stanford, California

DEPARTMENT OF BIOMEDICAL DATA SCIENCE (ADVISOR: CHIARA SABATTI)

09/2021 - present

## Experiences

### Google Summer of Code

NumFOCUS - Julia Cohort

STUDENT SOFTWARE DEVELOPER

Summer 2018

- An unique summer program where students work on self-proposed open source projects
- Added 3 additional features to IHT.jl to integrate it with the Open Mendel umbrella program

### Academia Sinica

Institute of Biomedical Sciences

RESEARCH ASSISTANT WITH PROF CARMAY LIM

Summer 2014, 2015, 2016

- Using computational approaches to model drug-protein and drug-environment interactions

## Publications (Google Scholars: <https://scholar.google.com/citations?user=jT7L8egAAAAJ>)

\*\* denotes co-first authorship

### Second-order group knockoffs with applications to GWAS

arXiv preprint

BB CHU, J GU, Z CHEN, T MORRISON, E CANDÈS, Z HE, C SABATTI

2023 Oct 23;arXiv:2310.15069

### Multivariate Genome-wide Association Analysis by Iterative Hard Thresholding

Bioinformatics

BB CHU, S KO, JJ ZHOU, A JENSEN, H ZHOU, JS SINSHEIMER, K LANGE

2023 Apr 1;39(4):btad193

### A Flexible Quasi-Copula Distribution for Statistical Modeling

arXiv preprint

SM JI, BB CHU, JS SINSHEIMER, H ZHOU, K LANGE

2022 May 6;arXiv:2205.03505

### Unsupervised Discovery of Ancestry Informative Markers and Genetic Admixture Proportions in Biobank-Scale Data Sets

American Journal of Human Genetics

S KO, BB CHU, D PETERSON, C OKENWA, JC PAPP, DH ALEXANDER, EM SOBEL, H ZHOU, K LANGE

2023 Feb 2;110(2):314-25

## Optimized Replacement T4 & T3 Dosing in Male & Female Hypothyroid Patients with Different BMIs using a Personalized Mechanistic Model of Thyroid Hormone Regulation Dynamics

M CRUZ-LOYA\*\*, **BB CHU**\*\*, J JONKLAAS, DF SCHNEIDER, J DiSTEFANO III (\* DENOTES CO-FIRST AUTHORSHIP)

*Frontiers in Endocrinology*

2022 Jul 14;13:888429

## A Fast Data-Driven Method for Genotype Imputation, Phasing, and Local Ancestry Inference: MendelImpute.jl

**BB CHU**, EM SOBEL, R WASIOLEK, JS SINSHEIMER, H ZHOU, K LANGE

*Bioinformatics*

2021 Dec 15;37(24):4756-63

## Iterative Hard Thresholding in GWAS: Generalized Linear Models, Prior Weights, and Double Sparsity

**BB CHU**, KL KEYS, CA GERMAN, H ZHOU, JIN J. ZHOU, JS SINSHEIMER, K LANGE

*GigaScience*

2020 Jun;9(6):giaa044.

## OpenMendel: A Cooperative Programming Project for Statistical Genetics

H ZHOU, JS SINSHEIMER, D BATES, **BB CHU**, CA GERMAN, S JI, KL KEYS, G MOSHER, J PAPP, EM SOBEL, J ZHAI, JJ ZHOU, K LANGE

*Human Genetics*

2020 Jan;139:61-71

## An Efficient Protocol for Computing the pKa of Zn-Bound Water

C GRAUFFEL, **BB CHU**, C LIM

*Physical Chemistry Chemical Physics*

2018 Nov;20(47):29637-47.

Papers in preparation:

## A Quasi-Copula model for Multivariate and Longitudinal GWAS via Automatic Differentiation

**BB CHU**, SM JI, H ZHOU, C SABATTI, K LANGE

*in preparation*

- Prototype software available at [https://github.com/OpenMendel/QuasiCopula.jl/tree/longitudinal\\_gwas](https://github.com/OpenMendel/QuasiCopula.jl/tree/longitudinal_gwas)

## In silico identification of putative causal genetic variants

Z HE\*\*, **BB CHU**\*\*, J YANG\*\*, J GU\*\*, Z CHEN, L LIU, T MORRISON, ME BELLOW, X QI, N HEJAZI, M MARTHUS, Y LE GUEN, H TANG, T HASTIE, I IONITA-LAZA, C SABATTI, E CANDÈS

*in preparation*

- Prototype software available at <https://github.com/biona001/GhostKnockoffGWAS>

## Controlled Variable Selection with Summary Statistics via Knockoffs

Z CHEN, Z HE, T MORRISON, J GU, **BB CHU**, C SABATTI, E CANDÈS

*in preparation*

## Public Talks

### False discovery rate control for polygenic risk prediction

CENTER FOR ADMIXED POPULATIONS AND HEALTH EQUITY (ONLINE)

*University of California, Los Angeles*

April 21, 2022

### Scalable algorithms for genetic association studies, genotype imputation, and ancestry inference

UNIVERSITY OF SOUTHERN CALIFORNIA, BIOSTATISTICS SEMINAR (ONLINE)

*University of Southern California*

January 28, 2021

### Scalable algorithms for GWAS, genotype imputation, and ancestry inference

UNIVERSITY OF MICHIGAN, BIOSTATISTICS SEMINAR (ONLINE)

*University of Michigan*

December 14, 2020

### Julia Meets Mendel: Algorithms and Software for Modern Genomic Data Analysis

AMERICAN SOCIETY OF HUMAN GENETICS VIRTUAL MEETING 2020

*Online*

October 27-30, 2020

### OpenMendel Programming Workshop

INAUGURAL LANGE SYMPOSIUM

*University of California, Los Angeles*

February 22, 2020

### A Multiple Regression Approach for GWAS and High Dimensional Inference

QUANTITATIVE AND COMPUTATIONAL BIOSCIENCES RESEARCH LUNCH

*University of California, Los Angeles*

February 8, 2020

### MendelIHT.jl: Generalized Linear Models for High Dimensional Genetics (GWAS) Data

JULIACON 2019, AVAILABLE AT: [WWW.YOUTUBE.COM/WATCH?V=UPIKAFShwFW](https://www.youtube.com/watch?v=UPIKAFShwFW)

*University of Maryland Baltimore*

July 25, 2019

# Teachings and Mentoring

## Inclusive Mentoring in Data Science

Stanford

POSTDOCTORAL MENTOR

2022

- Participated as a mentor to increase access to higher education and data science for underrepresented minorities

## Guest Lecturer, Biomath 203

UCLA

BIOMATH PRECEPTORSHIP

2020

- Taught 2 lectures at graduate probability course (Biomath 203) on random graph theory
- Wrote lecture notes, prepared and graded homework assignments, presented real-time simulation demos
- Materials available at: <https://github.com/biona001/teaching>

## Bruins-In-Genomics (B.I.G) summer research program

UCLA

GRADUATE STUDENT MENTOR

Summer 2018, 2019, 2020, 2021

- Mentors 2 undergraduates for 8 weeks each summer.
- Proposed student projects. Reviewed student code. Guided them through technical aspects of genetics research.

## Math 98 and 198 course

UC Berkeley

UNDERGRADUATE STUDENT INSTRUCTOR

2013 - 2016

- Taught 7 semesters of beginner/advanced Rubik's cube course (2 units), with about 15 students per semester.
- Average instructor rating 4.8/5.0

# Honors & Awards

## FELLOWSHIPS AND SCHOLARSHIPS

- 2019 **JuliaCon Financial Assistance for Travel,**
- 2019 **Doctoral Student Travel Grant,**
- 2018 **JuliaCon Financial Assistance for Travel,**
- 2018 **Genomics Analysis Training Grant (T32),**

Baltimore, MD  
Pisa, Italy  
London, UK  
NIH/UCLA

## MISCELLANEOUS

- 2017 **Certified in lifeguarding, first aid, AED, and CPR.,**
- 2015 **International inter-university cube relay, 10th place**
- 2014 **UC Berkeley semester Go (Weiqi) Tournament, 1st place**
- 2013 **3 by 3 rubik's cube speedsolve (10.52 seconds), 144th place in U.S, 910th in world**

Amer. Red Cross  
Cal Cube Club  
UCB Go Club  
WCA

# Notable Softwares

<b>Knockoffs.jl</b>	Fast and robust variable selection with Knockoffs
<b>GhostKnockoffGWAS.jl</b>	Prototype software for conducting knockoff-based analysis of GWAS summary statistics data
<b>EasyLD.jl</b>	Julia utilities for downloading and reading LD (linkage disequilibrium) matrices
<b>QuasiCopula.jl</b>	New class of distribution that captures correlation among non-Gaussian random variables efficiently
<b>knockoffspy</b>	A Python package that provides a direct wrapper over Knockoffs.jl
<b>knockoffsr</b>	An R package that provides a direct wrapper over Knockoffs.jl
<b>MendelIHT.jl</b>	Efficient $\ell_0$ penalized regression tool for model selection, optimized for gene-mapping (GWAS) analysis
<b>MendelImpute.jl</b>	Fast data-driven genotype imputation, phasing, and admixture estimation
<b>VCFTools.jl</b>	Julia utilities for handling VCF (variant call format) files
<b>Thyrosim.jl</b>	Individualized mechanistic Thyroid hormone simulator for optimizing LT4 and LT4+LT3 dosing
<b>fastPHASE.jl</b>	Julia wrapper for the famous fastPHASE genetics software. Allows the original program to run on binary PLINK data.
<b>MendelKinship.jl</b>	Computation of theoretical/empirical kinship and other identity coefficients based on pedigree/genotype data
<b>SnpArrays.jl</b>	Package for handling binary PLINK 1 formatted files