Benjamin B. **Chu**

(+1) 909-569-4530 | bbchu@stanford.edu | biona001.github.io | biona001

Research Interests

Math and Statistics	Knockoffs, Penalized regression methods, variance component models, quasi-copulas
Computation	Parallel and High-Performance Computing, Big Data Analysis, Package Development
Biology	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*

Work Experiences

Postdoctoral researcher   *Stanford, California*

STANFORD UNIVERSITY, DEPARTMENT OF BIOMEDICAL DATA SCIENCES   *09/2021 - present*

- Collaborated with **8+** independent research groups across **4+** universities
- Distributed **6+** open source software packages in **Julia, Python**, and **R**
- Led the design, implementation, testing, maintenance, and analysis workflows for **4+ bioinformatics pipelines**

Software developer   *Remote*

GOOGLE SUMMER OF CODE, NUMFOCUS (JULIA COHORT)   *Summer 2018*

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

Research assistant   *Taipei, Taiwan*

INSTITUTE OF BIOMEDICAL SCIENCES, ACADEMIA SINICA   *Summer 2014, 2015, 2016*

- Used computational approaches to simulate zinc metalloproteins for drug-protein and drug-environment interactions

Notable Softwares

RootCauseDiscovery	**Python** and **Julia** package implementing the root-cause discovery algorithm for **gene expression data**
GhostKnockoffGWAS	Knockoff-based analysis of **GWAS summary statistics** data
Knockoffs.jl	Implements a variety of knockoff-filters for **high-dimensional feature selection**
EasyLD.jl	**Julia** utilities for handling LD (linkage disequilibrium) matrices in hail's BlockMatrix format
QuasiCopula.jl	Defines a new class of **multivariate distribution** that supports non-Gaussian marginals
knockoffspy/knockoffsr	**Python** and **R** packages that provide a direct wrapper over Knockoffs.jl
MendelIHT.jl	Efficient ℓ_0 **penalized regression** tool for **model selection**, optimized for **GWAS** analysis
MendelImpute.jl	Genotype **imputation**, **phasing** (haplotyping), and **admixture estimation**
VCFTools.jl	**Julia** utilities for handling **VCF** (variant call format) files
Thyrosim.jl	**Mechanistic model** for thyroid hormone regulation
MendelKinship.jl	Computation of **theoretical/empirical kinship** and other identity coefficients based on **pedigree/genotype data**
SnpArrays.jl	**Julia** utilities for handling **binary PLINK** (.bed/.bim/.fam) files

For more software projects I'm involved in, visit: `https://github.com/biona001`

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 PRECEPTORSHOP *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

FOLLOWSHIPS AND SCHOLARSHIPS

2024	**NLM T15 Postdoc Training Grant in Biomedical Data Science**,	*Stanford*
2018	**Genomics Analysis Training Grant (T32)**,	*UCLA*

MISCELLANEOUS

2017	**Certified in lifeguarding, first aid, AED, and CPR.**,	*Amer. Red Cross*
2015	**International inter-univeristy cube relay**, 10th place	*Cal Cube Club*
2014	**UC Berkeley semester Go (Weiqi) Tounament**, 1st place	*UCB Go Club*
2013	**3 by 3 rubik's cube speedsolve (10.52 seconds)**, 144th place in U.S, 910th in world	*World Cube Assoc*

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

13 total, 7 led as first or co-first authors (denoted by **)

Root cause discovery via permutations and Cholesky decomposition *arXiv*

J LI, **BB CHU**, I SCHELLER, J GAGNEUR, M MAATHIUS *2024 Oct 16; arXiv:2410.12151*
- Software available at `https://github.com/Jinzhou-Li/RootCauseDiscovery`

Second-order group knockoffs with applications to GWAS *Bioinformatics*

BB CHU, J GU, Z CHEN, T MORRISON, E CANDÈS, Z HE, C SABATTI *2024 Sep 28;btae580*
- Software available at `https://github.com/biona001/Knockoffs.jl`

A blended genome and exome sequencing method captures genetic variation in an unbiased, high-quality, and cost-effective manner *bioRxiv*

TA BOLTZ**, **BB CHU****, C LIAO**, JM SEALOCK**, ..., BM NEALE, DP HOWRIGAN, AR MARTIN *2024.09.06.611689v1*
- Software available at `https://github.com/atgu/bge_analysis`

Beyond guilty by association at scale: searching for causal variants on the basis of genome-wide summary statistics *bioRxiv*

Z HE**, **BB CHU****, J YANG**, J GU**, Z CHEN, L LIU, T MORRISON, ..., C SABATTI, E CANDÈS *2024.02.28.582621v1*
- Software available at `https://github.com/biona001/GhostKnockoffGWAS`

Controlled Variable Selection from Summary Statistics Only? A Solution via GhostKnockoffs and Penalized Regression *arXiv*

Z CHEN, Z HE, **BB CHU**, J GU, T MORRISON, C SABATTI, E CANDÈS *2024 Feb 2;arXiv:2402.12724*

Multivariate Genome-wide Association Analysis by Iterative Hard Thresholding *Bioinformatics*

BB CHU, S KO, JJ ZHOU, A JENSEN, H ZHOU, J SINSHEIMER, K LANGE *2023 Apr 1;39(4):btad193*
- Software available at `https://github.com/OpenMendel/MendelIHT.jl`

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

S Ko, **BB Chu**, D Peterson, C Okenwa, JC Papp, DH Alexander, EM Sobel, H Zhou, K Lange

American Journal of Human Genetics

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

- Software available at `https://github.com/OpenMendel/OpenADMIXTURE.jl`

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

Frontiers in Endocrinology

2022 Jul 14;13:888429

- Software available at `https://github.com/biona001/Thyrosim.jl`

A Flexible Quasi-Copula Distribution for Statistical Modeling

SM Ji, **BB Chu**, J Sinsheimer, H Zhou, K Lange

arXiv

2022 May 6;arXiv:2205.03505

- Software available at `https://github.com/OpenMendel/QuasiCopula.jl`

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

BB Chu, EM Sobel, R Wasiolek, J Sinsheimer, H Zhou, K Lange

Bioinformatics

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

- Software available at `https://github.com/OpenMendel/MendelImpute.jl`

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

BB Chu, KL Keys, CA German, H Zhou, Jin J. Zhou, J Sinsheimer, K Lange

GigaScience

2020 Jun;9(6):giaa044.

- Software available at `https://github.com/OpenMendel/MendelIHT.jl`

OpenMendel: A Cooperative Programming Project for Statistical Genetics

H Zhou, J Sinsheimer, D Bates, **BB Chu**, CA German, S Ji, KL Keys, ..., K Lange

Human Genetics

2020 Jan;139:61-71

- Software page available at `https://github.com/OpenMendel`

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.