

Kevin L. Keys

SENIOR BIOINFORMATICS ENGINEER

AncestryDNA, 153 Townsend St #800, San Francisco, CA 94107

☎ (+1) 520-276-9017 | ✉ kkeys@ancestry.com | 🏠 www.kevinkeys.net | 📷 kkeys | 🌐 kkeys | 📞 0000-0003-4886-6666 | 🐦
@thetrueblueklk | 🎓 Kevin L. Keys

Education

University of California, Los Angeles

Los Angeles, CA

PH.D BIOMATHEMATICS

September 2011 - September 2016

- Dissertation title: Projection algorithms for large scale optimization and genomic data analysis
- Advisor: Kenneth Lange

University of California, Los Angeles

Los Angeles, CA

M.S. BIOMATHEMATICS

September 2011 - December 2012

The University of Arizona

Tucson, AZ

B.S. MATHEMATICS AND B.A. LINGUISTICS

August 2006 - May 2010

Skills

Computing

R, Bioconductor, Julia, MATLAB, Python, BASH, LaTeX, Git, SGE, AWS (EC2,S3,ECR,Batch,VPC,CF), Docker, PostgreSQL, Terraform, Jupyter, RMarkdown, RShiny, Jenkins, HarnessNG, SLURM

Bioinformatics

PLINK, ADMIXTURE, vcftools, bcftools, samtools, RFMix, EAGLE, GCTA, BOLT-LMM, PrediXcan, limma, GATK, RSEM, STAR, bbtools, FastQC, MultiQC, Cromwell, Nextflow, Snakemake, CellRanger, bcl2fastq, SPAdes, Seurat, scanpy, GSEA

Analysis

statistical genetics, RNA-Seq, scRNA-Seq, differential expression analysis, single cell expression analysis, multiomic analysis, genetic admixture, ancestry estimation, genetic risk scores, human genetic diversity, linear modeling, predictive modeling, machine learning, penalized regression, parallel computing, scientific computing, scientific software engineering

Languages

English, Spanish, Catalan, Portuguese

Experience

AncestryDNA

San Francisco, CA

SENIOR BIOINFORMATICS ENGINEER

Oct 2023 - current

- Improved scalability, efficiency, and reproducibility of genetic ancestry estimation
- Optimized scalability and cost-efficiency of cloud research computing infrastructure for population genetic analyses

AmplifyBio

South San Francisco, CA

SENIOR BIOINFORMATICS SCIENTIST

Feb 2023 - Oct 2023

- Designed and implemented cloud-enabled bioinformatics pipelines for critical company-wide sequencing data processing and analysis
- Maintained scalable cloud research computing infrastructure for cell therapy R&D needs
- Executed statistical analyses of high-dimensional single-cell data for cancer immunotherapy assay development

Ambys Medicines

SENIOR SCIENTIST I, COMPUTATIONAL BIOLOGY

South San Francisco, CA

July 2022 - Nov 2022

- Wrote and deployed pipelines for analysis of multi-omic data to genomically profile human livers for cell therapy applications
- Performed statistical analyses assays of liver function and regeneration for gene therapy development
- Structured and curated heterogeneous genomic data sources into a centralized company-wide dashboard for rapid therapeutic development

Ambys Medicines

SCIENTIST II, COMPUTATIONAL BIOLOGY

South San Francisco, CA

June 2020 - July 2022

University of California, San Francisco

POSTDOCTORAL SCHOLAR

San Francisco, CA

September 2016 - June 2020

- Studied the ability of transcriptome prediction models to predict across populations
- Supervised 9+ trainees in projects covering genetic epidemiology and statistical genetics
- Provided computational support for whole genome analysis, editorial support for manuscripts

University of California, Los Angeles

GRADUATE STUDENT

Los Angeles, CA

September 2012 - August 2016

- Worked on parallel computing frameworks for high dimensional statistical analysis with ADMM
- Developed proximal algorithms for high-dimensional mathematical (non)convex optimization
- Produced open-source software for hard thresholding algorithms for penalized regression in GWAS

Stanford University

VISITING GRADUATE RESEARCHER

Stanford, CA

September 2014 - June 2015

- Developed mathematical optimization code in Julia
- Translated signals processing algorithms into software suitable for penalized regression in GWAS

Freie Universität Berlin

SUMMER INTERN

Berlin, Germany

June 2014 - August 2014

- Participated in UCLA IPAM / FUB Konrad Zuse Zentrum Graduate Research in Industrial Projects for Students
- Studied feasibility of using SAP HANA computing platform to integratively analyze proteomic, transcriptomic, and imaging data for assisting in cancer diagnoses
- Used R and Bioconductor to analyze microarray data from NCBI GEO using LASSO regression
- Used HANA Predictive Analytics Library to train and test a support vector machine for binary classification of cancer status

University of Oxford

SUMMER INTERN

Oxford, United Kingdom

June 2011 - August 2011

- Participated in Oxford Computational Biology Summer School at Oxford Department of Statistics
- Used hidden Markov random field theory to study metabolic network evolution
- Implemented Markov chain simulator in R to model gain/loss of reactions
- Managed KEGG metabolic network XML data with Perl

Universitat Pompeu Fabra

FULBRIGHT STUDENT RESEARCHER

Barcelona, Spain

September 2010 - June 2011

- Studied fine-grained signals of molecular evolution in human metabolome
- Used BioPerl to query Ensembl gene ontologies and KEGG networks
- Analyzed molecular evolution in aligned primate and rodent orthologous genes with PAML software
- Provided computational and data management support in BASH, R, and Perl

The University of Arizona

STUDENT RESEARCH ASSISTANT

Tucson, AZ

June 2007 - May 2010

- Modeled population dynamics and coevolution of languages and genes in eastern Indonesia with genetic, linguistic, and cultural data
- Performed PCR, restriction enzyme assays, and linguistic phylogenetic construction with Swadesh word lists
- Studied hidden Markov models for haplotype imputation

Fellowships & Funding

2020-2025 Career Development Award (K01 HL153100) , NHLBI	San Francisco, CA
2018-2020 Data Science Fellow , BIDS-UCSF BCHSI	Berkeley, CA
2017-2020 Postdoc Supplement (R01 HL135156) , NIH	San Francisco, CA
2015-2016 Genomic Analysis Training Program (GATP) Trainee (T32 HG002536) , UCLA	Los Angeles, CA
2015-2016 Dissertation Year Fellowship , UCLA (declined)	Los Angeles, CA
2012-2016 Graduate Research Fellowship , National Science Foundation	Los Angeles, CA
2011-2012 Graduate Opportunity Fellowship , UCLA	Los Angeles, CA
2010-2011 Student Research Fellow , Fulbright	Barcelona, Spain
2009 Biomedical Research Abroad: Vistas Open! (BRAVO!) Fellow , UA	Barcelona, Spain
2008-2010 Maximizing Access to Research Careers (MARC-USTAR) Trainee (T34 GM008718) , UA	Tucson, AZ
2006-2010 National Hispanic Scholar Award , UA	Tucson, AZ

Service

Reviewer	<i>Bioinformatics, Diabetic Medicine, Journal of Medical Genetics, AJHG, PLOS Genetics, PLOS ONE, Pharmacogenomics, Journal of Statistical Computation and Simulation, HGG Advances, iScience, JACI</i>
Volunteer	SACNAS National Conference (presentation judge, abstract reviewer, social media team) (2016-)
Membership	SACNAS (2007-), ASHG (2017-), IGES (2017-2020), SACNAS at UCSF (2016-2020), SACNAS at UCLA (2012-2016) NSF Review Panel (2021), NSF GRFP Review Panel (2020-), SACNAS Grants and Development Subcommittee (2021-),
Committees	SACNAS Conference Planning Committee (2020-), SACNAS Membership Committee (2020-2022), SACNAS Development Committee (2021-), ImageXD 2019, JuliaCon 2017, UCLA Biological Sciences Council (2014)

Mentees

2017-2020 María G. Contreras , (undergraduate; now postbac researcher at UCSF)	UCSF
2017-2020 Pagé C. Goddard , (postbac; now graduate student at Stanford Genetics)	UCSF
2018-2020 Benjamin B. Chu , (PhD student; currently doctoral candidate at UCLA Biomathematics)	UCSF/UCLA
2017-2019 Annie Li , (high school; now undergraduate at the University of Michigan)	UCSF
2017-2019 Oona Risse-Adams , (high school; now undergraduate at the University of California, Los Angeles)	UCSF
2017-2018 Andrew M. Zeiger , (postbac; now medical student at Thomas Jefferson University)	UCSF
2017-2018 Peter Huang , (undergraduate; Johns Hopkins University Class of 2020)	UCSF
2017-2018 Joaquín Magaña , (postbac; now IT consultant in Oakland, CA)	UCSF

Oral Presentations

Aug 2020	Gene expression prediction models across human populations , Universidade Federal Rio Grande do Sul Statistics Seminar	Porto Alegre, Brasil
Feb 2020	Generalizability of gene expression prediction models across human populations , UC Irvine Statistics Seminar	Irvine, CA
Jan 2020	Generalizability of gene expression prediction models across human populations , CU Anschutz Biostatistics Seminar	Denver, CO
Oct 2019	Cross-population generalizability and statistical power of predictive models for gene expression , ASHG 2019	Houston, TX
Jul 2019	Cross-population portability of prediction models for gene expression , Universidad de La Laguna Jornada de Genética	Tenerife, Spain
Apr 2019	On the cross-population portability of gene expression prediction models , UCSF BCHSI Seminar	San Francisco, CA
Nov 2018	Genetics, Ancestry, and Disease: Pediatric Asthma in Puerto Rico , ALAB 2018	Mayagüez, PR
Oct 2018	Transethnic portability of gene expression models , UCSF SABRE Asthma Seminar	San Francisco, CA
Oct 2018	Transethnic portability of gene expression models , ASHG Annual Conference	San Diego, CA
Oct 2018	Transethnic portability of gene expression models , SACNAS National Conference	San Antonio, TX
Jan 2018	The transethnic portability of predicted gene expression from transcriptome repositories , UTHSC Genetics, Genomics, and Informatics Seminar	Memphis, TN
Oct 2017	Whole genome sequencing of bronchodilator response in minority children with asthma , SACNAS National Conference	Salt Lake City, UT
May 2016	Iterative hard thresholding for model selection in genome-wide association studies , UCLA GATP Symposium	Los Angeles, CA
Aug 2015	Model selection for genome-wide association studies , JSM 2015	Seattle, WA
Jul 2015	Model selection for genome-wide association studies , ISMP 2015	Pittsburgh, PA
Mar 2015	Model selection for genome-wide association studies , IPAM LAT2015	Los Angeles, CA
Oct 2014	MM Proximal Distance Algorithms , SACNAS National Conference	Los Angeles, CA
Oct 2011	Network paradigm for molecular evolutionary analysis of metabolic pathways , SACNAS National Conference	San Jose, CA
Feb 2010	A survey of the human skin pigmentation pathway for signature of selection , UA BECUR Conference	Tucson, AZ

Publications

- [1] Esther Herrera-Luís et al. “Epigenome-wide association study of lung function in Latino children and youth with asthma”. *Clinical Epigenetics* 14.9 (2022), 1–15. DOI: <https://doi.org/10.1186/s13148-022-01227-5>.
- [2] Linda Kachuri et al. “Gene expression in African Americans and Latinos reveals ancestry-specific patterns of genetic architecture”. *bioRxiv* (2022). DOI: <https://doi.org/10.1101/2021.08.19.456901>.
- [3] Elizabeth Wilson et al. “Human Hepatocytes Expanded in a Novel Rat Bioreactor Maintain Full Functionality In Vitro and In Vivo”. In: *Molecular Therapy*. Vol. 30. 4. Cell Press, 50 Hampshire St, Floor 5, Cambridge, MA 02139 USA. 2022, 184–184.
- [4] Javier Pérez-García et al. “Epigenome-wide association study of bronchodilator response in African Americans”. *European Respiratory Journal* 58.PA2389 (2021). DOI: 10.1183/13993003.congress-2021.PA2389.
- [5] Benjamin B. Chu et al. “Multivariate GWAS: Generalized Linear Models, Prior Weights, and Double Sparsity”. *GigaScience* 9.6 (June 2020), g1aa044. DOI: 10.1093/gigascience/g1aa044.
- [6] Pagé C. Goddard et al. “Integrative genomic analysis in African American children with asthma finds three novel loci associated with lung function”. *Genetic Epidemiology* (2020). DOI: <https://doi.org/10.1002/gepi.22365>.
- [7] Ester Herrera-Luís et al. “Genome-wide association study reveals a novel locus for asthma with severe exacerbations in diverse populations”. *Pediatric Allergy and Immunology* (2020). DOI: <https://doi.org/10.1111/pai.13337>.
- [8] Yale Jiang et al. “A genome-wide study of DNA methylation in white blood cells and asthma in Latino children and youth”. *Epigenetics* (2020), 1–9. DOI: <https://doi.org/10.1080/15592294.2020.1809872>.

- [9] **Kevin L. Keys**, Angel C.Y. Mak, et al. “On the cross-population portability of gene expression prediction models”. *PLOS Genetics* 16 (Aug. 2020), e1008927. DOI: [10.1371/journal.pgen.1008927](https://doi.org/10.1371/journal.pgen.1008927).
- [10] Eunice Y. Lee et al. “Whole Genome Sequencing Identifies Novel Functional Loci Associated with Lung Function in Puerto Rican Youth”. *American Journal of Respiratory and Critical Care Medicine* 202.7 (2020). DOI: <https://doi.org/10.1164/rccm.202002-03510C>.
- [11] Joaquín Magaña et al. “An epistatic interaction between pre-natal smoke exposure and socioeconomic status has a significant impact on bronchodilator drug response in African American youth with asthma”. *BioData Mining* (2020), 1–20. DOI: <https://doi.org/10.1186/s13040-020-00218-7>.
- [12] Angel C.Y. Mak et al. “Novel KITLG/SCF regulatory variants are associated with lung function in African American children with asthma”. *Genetics* 215.3 (2020), 869–886. DOI: <https://doi.org/10.1534/genetics.120.303231>.
- [13] Corbin Quick et al. “Sequencing and Imputation in GWAS: Cost-Effective Strategies to Increase Power and Genomic Coverage Across Diverse Populations”. *Genetic Epidemiology* 44.6 (June 2020), 537–549. DOI: [10.1002/gepi.22326](https://doi.org/10.1002/gepi.22326).
- [14] **Kevin L. Keys**, Hua Zhou, and Kenneth Lange. “Proximal Distance Algorithms: Theory and Examples”. *Journal of Machine Learning Research* 20.66 (2019), 1–38.
- [15] Eunice Y. Lee et al. “Ambient Air Pollution, Asthma Drug Response and Telomere Length in African American Youth”. *Journal of Allergy and Clinical Immunology* (2019). DOI: <https://doi.org/10.1016/j.jaci.2019.06.009>.
- [16] Hua Zhou et al. “OpenMendel: a cooperative programming project for statistical genetics”. *Human Genetics* (2019), 1–11. DOI: <https://doi.org/10.1007/s00439-019-02001-z>.
- [17] Alfonso Landeros et al. “BioSimulator.jl: Stochastic simulation in Julia”. *Computer Methods and Programs in Biomedicine* 167 (2018), 23–35. DOI: [http://doi.org/10.1016/j.cmpb.2018.09.009](https://doi.org/10.1016/j.cmpb.2018.09.009).
- [18] Angel C.Y. Mak et al. “Whole genome sequencing of pharmacogenetic drug response in racially diverse children with asthma”. *American Journal of Respiratory and Critical Care Medicine* 197.12 (2018), 1552–1564. DOI: [http://doi.org/10.1164/rccm.201712-25290C](https://doi.org/10.1164/rccm.201712-25290C).
- [19] Ludovica Montanucci et al. “Influence of pathway topology and functional class on the molecular evolution of human metabolic genes”. *PLOS ONE* 13.12 (2018), e0208782. DOI: [http://doi.org/10.1371/journal.pone.0208782](https://doi.org/10.1371/journal.pone.0208782).
- [20] Andrew M. Zeiger et al. “Genetic Determinants of Telomere Length in African American Youth”. *Scientific Reports* 8.1 (2018), 13625. DOI: [http://doi.org/10.1038/s41598-018-31238-3](https://doi.org/10.1038/s41598-018-31238-3).
- [21] **Kevin L. Keys**, Gary K. Chen, and Kenneth Lange. “Iterative hard thresholding for model selection in genome-wide association studies”. *Genetic Epidemiology* 41.8 (2017), 756–768. DOI: [http://doi.org/10.1002/gepi.22068](https://doi.org/10.1002/gepi.22068).
- [22] **Kevin Lawrence Keys**. “Projection algorithms for large scale optimization and genomic data analysis”. PhD thesis. University of California, Los Angeles, 2016.
- [23] Kenneth Lange and **Kevin L. Keys**. “The proximal distance algorithm”. In: *Proceedings of the International Conference of Mathematics*. Vol. 4. 2014, 95–116. DOI: [arXiv:1507.07598](https://arxiv.org/abs/1507.07598).
- [24] Giovanni M. Dall’Olio et al. “Ten simple rules for getting help from online scientific communities”. *PLoS Computational Biology* 7.9 (2011), e1002202. DOI: [http://doi.org/10.1371/journal.pcbi.1002202](https://doi.org/10.1371/journal.pcbi.1002202).