

David A. Knowles

I am a primarily computational geneticist developing machine learning, statistical genetics, and functional genomics methods to map the causal molecular pathways from genetic variation to disease risk, with a particular focus on alternative splicing and neurodegeneration.

<https://daklab.github.io/>

Positions held

2019-present	CORE FACULTY MEMBER New York Genome Center
2025-present	ASSOCIATE PROFESSOR (Computer Science)
2019-2024	ASSISTANT PROFESSOR (Computer Science) AFFILIATE MEMBER (Data Science Institute) INTERDISCIPLINARY APPOINTEE (Systems Biology) Columbia University
2014-2018	POSTDOCTORAL RESEARCHER (Genetics, Pathology) Stanford University Co-advisors: Prof. Jonathan Pritchard, Prof. Sylvia Plevritis
2012-2014	POSTDOCTORAL RESEARCHER (Computer Science) Stanford University Advisor: Prof. Daphne Koller
2008-2012	PHD Candidate, Roger Needham Scholar, Wolfson College, University of Cambridge Machine Learning Group, Cambridge University Engineering Department
2006	Summer Undergraduate Research Fellow California Institute of Technology

Education

2008-2012	PHD Engineering (Machine Learning) University of Cambridge Thesis: <i>Bayesian non-parametric models and inference for sparse and hierarchical latent structure</i> Advisor: Prof. Zoubin Ghahramani
2007-2008	MSc Bioinformatics and Systems Biology - Distinction Imperial College London Thesis: <i>Statistical tools for ultra-deep pyrosequencing of fast evolving viruses.</i> Advisor: Prof. Susan Holmes, Statistics Department, Stanford University.
2003-2007	MENG Engineering - Distinction University of Cambridge Thesis: <i>A non-parametric extension to Independent Components Analysis.</i> Advisor: Prof. Zoubin Ghahramani. BA Natural Sciences (Physics) - First Class

Honours & Awards

2017	Stanford Cancer Systems Biology Symposium — Poster Award
2014	The International Society for Bayesian Statistics Travel Award for best invited Bayesian paper
2014	The International Society for Bayesian Statistics Dennis V. Lindley Prize for innovative research in Bayesian Statistics
2007	Charles Lamb University prize for first place in Information Engineering Sir Joseph Larmor Silver Plate and three other college prizes (Cargill, Cunningham and College)
2005	Wright Prize for ranking 5/600 in Natural Sciences Earle Year Prize for top 4 students across all subjects at St. John's College Hollinshead-Howles Prize for top Natural Scientist at St. John's College BP Prize for Advanced Physics
2004	Gaskell Year Prize for ranking 9/600 in Natural Sciences Hollinshead-Howles Prize for Part IA
2003	Top 50 nationally in Royal Society of Chemistry Olympiad
2001	Top 50 nationally in Mathematics Olympiad

Fellowships

2008-2012	Roger Needham Scholar, <i>Wolfson College, University of Cambridge</i> , funded by Microsoft Research
2006	Summer Undergraduate Research Fellow, <i>California Institute of Technology</i>

Industry Positions

2025-ongoing	Scientific co-founder NeoSplice Therapeutics
2013-2014	Contract research scientist Ventana Roche Medical Systems <i>Deep learning/computer vision based prognosis of early stage breast cancer.</i>
2009-2012	Contract Software Engineer Microsoft Research Cambridge <i>Extending the probabilistic programming language Infer.NET.</i>
2009	Research Intern Microsoft Research Cambridge <i>Non-conjugate Variational Message Passing.</i> Supervisor: Thomas P. Minka.
2005	Equity Research Intern UBS Investment Bank, London
2004	Software Engineering Intern Data Connection Limited, London

Patents

2025	PCT/US2025/015271: Methods For Drug Target Deconvolution. With Rahul Satija (NYU/NYGC).
2024	PCT/US2024/033813: A deep learning framework to predict on-target and off-target activity of

CRISPR guide RNAs. With Neville Sanjana (NYU/NYGC).

2015 EP3155592B1: Predicting breast cancer recurrence directly from image features computed from digitized immunohistopathology tissue slides. With Ventana Roche.

Teaching

2019-ongoing Machine learning for functional genomics

Course #	Semester	Enrollment	Class score		Instructor score	
			Mean	Median	Mean	Median
COMS4995	Fall 2019	39	4.15	5.0	4.50	5.0
COMS4995	Spring 2021	63	4.35	5.0	4.46	5.0
COMS4762	Spring 2022	71	4.04	4.0	4.24	5.0
COMS4762	Fall 2022	72	4.33	5.0	4.50	5.0
COMS4762	Fall 2023	90	4.32	4.0	4.53	5.0
COMS4762	Fall 2024	115	4.0	4.0	4.0	5.0
COMS4762	Fall 2025	94	4.45	5.0	4.42	5.0

2022-ongoing BMEN4500 Functional Genomics Methods & Applications guest lecturer

2019 Columbia University Functional Genetics Boot Camp

2019 Weill Cornell Medical College Computational Genomics

2017 CBO 244: Lecture Series in Cancer Systems Biology

2013-2015 Statistical, Mathematical, and Computational Consulting (SMACC) hosted by Stanford ICME

2013 Stanford Statistics guest lecture on Bayesian nonparametrics

2009-2011 Statistics advisor, Cambridge University Statistics Clinic

2009-2011 Supervisor, Cambridge University Engineering Department, *Digital Signal Processing*

2009-2011 Demonstrator, Cambridge University Engineering Department, *C++ programming*

2007-2009 Private Mathematics tutor, Camtutors

Mentoring

UNDERGRADUATES

2025-ongoing Louis Ke, Columbia University

2025-ongoing Nicole Cui, Columbia University

2024-2025 Molly Durawa, Barnard College

Summer 2024 Sharvani Vadlamani, Columbia University

Summer 2024 Rohan Nidumbur, University of Maryland (co-mentored by Gao Wang, Columbia)

2024-ongoing Clara Yu, Caltech

Summer 2023 Fionna Feerick, Stevens Institute of Technology (NYGC/MacMillan diversity fellow)

2022-2024 Gilead Turok, Columbia University

2022-2024 Eric Kim, Columbia University

2022-2023 Ekene Ezeunala, Columbia University

2022-2023 Siddhant Sanghi, Columbia University (now a PhD student at UC Davis with Gerald Quon)

2022-2025 Gillian Bartels-Quansah (Science Path Scholar), Barnard College

2022-2024 Sophia Sowinski, Barnard College

Summer 2022 Mariya Semchuk (S Jay Levy diversity fellow), City College of New York

2021-2024 Stella Park, Columbia University

2021-2022 Daniel Um, Columbia University (now a Machine Learning Analyst at Bracebridge Capital)

2021-2022 Claire Johnson, Barnard (now a PhD student at Harvard Medical School with Eliezer Van Allen)

2019-2023 Collin Wang, Columbia University (now at Jane Street Capital)

2019-2020 Kevin Wang, Columbia University
 2019-2021 Udai Nagpal, Columbia University (now at Goldman Sachs)

MASTERS STUDENTS

2025-ongoing Eddie Kong, Columbia University
 2025-ongoing Vera Mazeeva, Columbia University
 2025-ongoing Nicholas Keung, Columbia University
 2024-ongoing Smriti Vaidyanathan, Columbia University
 2024-2025 Shree Raghavendra, Columbia University
 2023-ongoing Julia Lewandowski, Columbia University
 2023-2025 Zaikang Lin, Columbia University
 2022-2024 Xingpei Zhang, Columbia University
 2022-2024 Ting Chen, Columbia University
 2022-2023 Joshua Segovia (DREAM Bridge student), Columbia University
 2022-2023 Shouvik Mani, Columbia University with Elham Azizi (now a PhD student at Stanford University)
 2021-2022 Hyun Dong Lee, Columbia University
 2021-2022 Chloé Terwagne, Université Libre de Bruxelles (now a PhD student with Greg Findlay at The Francis Crick Institute, London).
 2020-2021 Nasrine Metic, École polytechnique fédérale de Lausanne (now a PhD student with Mirjana Efre-mova at Barts Cancer Institute, London).
 2019-2021 Stephen Malina, Columbia University (now Machine Learning Team Lead at Dyno Therapeutics).

PHD STUDENTS

2024-ongoing Arghamitra Talukder, Computer Science, Columbia University (co-mentored by Itsik Pe'er)
 2023-ongoing Daniel Meyer, Computer Science, Columbia University
 2022-ongoing Sei Chang, Computer Science, Columbia University
 2022-ongoing Anjali Das, Computer Science, Columbia University
 2021-2026 Jiayu Su, Systems Biology, Columbia University (co-mentored by Raúl Rabadán)
 2020-ongoing Shashaank Narayanan, Computer Science, Columbia University (co-mentored by David Sulzer)
 2021-2025 Karin Isaev, Systems Biology, Columbia University. Now at Pioneer Labs.
 2019-2025 Andrew Stirn, Computer Science, Columbia University. Now at Synthesize Bio.

POSTDOCTORAL SCHOLARS

2024-ongoing Aaron Zweig, NYGC/Columbia (co-mentored by Elham Azizi, Columbia)
 2024-ongoing Alex Tokolyi, NYGC
 2023-ongoing Isabella Grabski, NYGC (co-mentored by Rahul Satija, NYU)
 2023-ongoing Tatsuhiko Naito, NYGC (co-mentored by Towfique Raj, MSSM)
 2022-ongoing Aline Réal, NYGC
 2021-ongoing Scott Adamson, NYGC (co-mentored by Tuuli Lappalainen)
 2019-2025 Megan Schertzer, NYGC/Columbia. Now a Senior Staff Scientist at UVA.
 2019-2024 Brielin Brown, DSI fellow, NYGC/Columbia (co-mentored by Tuuli Lappalainen). Now Assistant Professor at the University of Pennsylvania in the Department of Biostatistics, Epidemiology and Informatics.

PREDOCTORAL RESEARCH SCIENTISTS

2024-2025 Stella Park, NYGC
 2022-2024 Claire Harbison, NYGC (joint with Tuuli Lappalainen)

2021-2024 Jui-Shan (Teresa) Lin, NYGC
 2021-2022 Peter Halmos, NYGC (now a PhD student with Ben Raphael at Princeton).

POSTDOCTORAL RESEARCH SCIENTISTS

2023-ongoing Saikat Banerjee, PhD, NYGC
 2020-ongoing Chirag Lakhani, PhD, NYGC
 2019-2021 Laura Pereira, PhD, NYGC (now Team Lead at Spark Therapeutics).

VISITING SCIENTISTS

2023-2024 Giacomo Cavalca, PhD student at University of Bologna. Co-mentored with Dr Gao Wang, Columbia Neurology.
 Summer 2023 Aditya Sriram, PhD student at University of Pittsburgh.

Service

TO THE DEPARTMENT

2023, 2024 Co-organized CS PhD visit
 2022-2024 Distributed Research Apprenticeships for Master's (DREAM) MS-to-PhD Bridge program mentor
 2022 Organized junior faculty session at annual CS faculty retreat
 2021-ongoing MS advisor for ML track

TO THE SCHOOL OF ENGINEERING

2021 Reviewed for Columbia Precision Medicine Initiative Pilot Grants Program
 2021-ongoing Reviewed for Columbia Engineering SIRS/STAR Funding

TO THE UNIVERSITY

2023-ongoing Advisor for Computational Biology BA.
 2021-2023 Created the new Computational Biology BA degree (w/ Itsik Pe'er): developed curriculum, partnered with Biological Sciences, and advocated to the University.
 2021 DSI molecular tumor board representative
 2021-ongoing Reviewing for Data Science Institute MS applicants
 2020 DSI capstone project advisor (with Johnson & Johnson)

TO NYGC

2025-ongoing AI Taskforce Co-chair
 2022 CSHL - NYGC faculty retreat planning committee
 2022-ongoing Statistical and population genetics working group co-chair
 2022-ongoing NYGC DEI summer internship program mentor & speaker
 2021-ongoing Faculty representative on academic DEI committee
 2021 Faculty recruitment committee co-chair
 2020-ongoing Weekly NYGC-wide lab meeting organizer
 2020-ongoing Annual postdoc & student award external judge organizer
 2021 AI Applications to Disease Research/Fundamentals of AI Workshop

TO THE PROFESSION

2026	Co-chair of MLCSB COSI track at the ISMB annual meeting.
2023-ongoing	Co-chair of the annual Machine Learning in Computational Biology (MLCB)
2020-2021	Machine Learning in Computational and Systems Biology (MLCSB) Communities of Special Interest (COSI) board member, Organization: International Society for Computational Biology (ISCB)
2020-2022	Co-organizer of MLCB.

PHD COMMITTEES

2025-2026	Jacob Blindenbach, Computer Science, Columbia University
2025-2026	Marliette Matos, Albert Einstein College of Medicine
2025-ongoing	Annie Yoolim Choi, Biomedical Informatics, Columbia University
2025-ongoing	Feiyue Yang, Systems Biology, Columbia University
2024-2025	Philippe Chlenski, Computer Science, Columbia University
2024-2025	Yining Liu, Computer Science, Columbia University
2024-ongoing	Charlotte Rochereau, Systems Biology, Columbia University
2024-ongoing	Christina Floristean, Computer Science, Columbia University
2024-2026	Justin Hong, Computer Science, Columbia University
2024-ongoing	Fion Shiau, Systems Biology, Columbia University
2023-ongoing	Yige Zhao, Systems Biology, Columbia University
2024-2025	Achille Nazaret, Computer Science, Columbia University
2022-2024	Sydney Hart, New York University
2022-2024	Sitara Persad, Computer Science, Columbia University
2022-2023	Erica Wolin, Biological Sciences, Columbia University
2021-ongoing	Izaak Coleman, Systems Biology, Columbia University
2022-ongoing	Mert Ketenci, Computer Science, Columbia University
2021-2024	Guojie Zhong, Systems Biology, Columbia University
2021-2022	Jonah Einson, Systems Biology, Columbia University
2021-2024	Raiyan Rashid Khan, Computer Science, Columbia University
2020-2022	Daniel Li, Computer Science, Columbia University
2019-2022	Akash Sookdeo, New York University
2019-2022	Neville Dusaj, Weill Cornell Medical Campus
2020-2021	Jie Yuan, Computer Science, Columbia University
2020-2021	Antonio Khalil Moretti, Computer Science, Columbia University
2020	External Examiner for Clint Lombard, Electrical and Electronic Engineering, Stellenbosch University, South Africa.
2020	Hooshmand Shokri Razaghi, Computer Science, Columbia University
2019-2020	Hanna Levitin, Systems Biology, Columbia University
2019	Gabriel Loaiza, Statistics, Columbia University

POSTDOCTORAL FELLOWSHIP & TRANSITION AWARD COMMITTEES

2026-ongoing	Sangho Yoon, Landau Lab, New York Genome Center
2026-ongoing	Yuexuan Xu, Mayeux & Vardarajan labs, Columbia University
2025-ongoing	Upendra Chalise, University of Minnesota (Career Mentoring and training in Omics and Data for Early-stage investigators mentee)
2025-ongoing	Zhaoning Wang, Ren lab, Columbia (K99)
2025-ongoing	Amanda Wilson, Landweber lab, Columbia (F32)
2024-ongoing	Iraj Eshghi, Imielinski lab, New York Genome Center (BWF CASI)
2024-ongoing	Mariela Cortes Lopez, Landau lab, New York Genome Center (K99)
2021-2023	John Morris, Lappalainen lab, New York Genome Center (K99)

2022-2024 Adam Widman, Landau lab, New York Genome Center
 2020-2022 Tim Stuart, Satija lab, New York Genome Center (K99)

Reviewing

2023-ongoing Poster judge for 1st NYC RNA Symposium at Rockefeller University
 2017 Advances in Neural Information Processing Systems (NeurIPS) Best Reviewer Award

JOURNALS

Nature, Science, Cell, Nature Genetics, Nature Communications, Nature Bioengineering, Nature Machine Intelligence, eLife, PLOS Genetics, PLOS Computational Biology, Bioinformatics, BMC Bioinformatics, npj Genomic Medicine, Transactions on Machine Learning Research, Science Advances, Genome Research, Genome Biology, Biological Psychiatry, Journal of Mathematical Biology, Journal of Machine Learning Research, Journal of Computational and Graphical Statistics, Journal of the Royal Statistical Society: Series B, IEEE Transactions on Pattern Analysis and Machine Intelligence, Annals of Applied Statistics. Machine Learning, Journal of the American Statistical Association, Statistics and Computing (Springer), Bayesian Analysis.

CONFERENCES

Advances in Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), International Conference on Artificial Intelligence and Statistics (AISTATS), Pacific Biocomputing Symposium (PBS), International Joint Conferences on Artificial Intelligence (IJCAI), International Conference on Learning Representations (ICLR), Intelligent Systems for Molecular Biology.

WORKSHOPS

PC member for NeurIPS Workshop on Computational Biology, PC member for NeurIPS Workshop on Optimization in Machine Learning

GRANT REVIEWING

2025 NIH New Innovator Award Program (DP2)
 2025 NIH Genomics, Computational Biology and Technology study section
 2025 Oak Ridge Associated Universities Ralph E Powe Junior Faculty Enhancement Awards
 2022 NIH:NHGRI Special Emphasis Panel for Supporting Talented Early Career Researchers in Genomics
 2022 Ad-hoc reviewer for NSF Directorate for Biological Sciences/Molecular and Cellular Biosciences Division/Systems and Synthetic Biology Team
 2021 NIH/NIA special emphasis panel for PAR-19-269 "Cognitive Systems Analysis of Alzheimer's Disease Genetic and Phenotypic Data".
 2021 NIH Director's Early Independence Awards (DP5) Subject Matter Expert.
 2020 US-Israel Binational Science Foundation reviewer.

TENURE AND PROMOTION REVIEWING

2023 Technical University of Denmark
 2022 Institut Pasteur
 2022 Rockefeller University

PROFESSIONAL AFFILIATIONS

Organizer for Machine Learning in Computational Biology (MLCB) 2019-2022, MLCB Co-chair 2023-ongoing, ISMB MLCB COSI member, Early Stage Investigators in Cancer Systems Biology Steering Committee, International Society of Bayesian Analysis (ISBA) Chair of the Continuing Education Committee.

Selected Talks

2026	Johns Hopkins University Data Science & AI Institute, invited speaker.
2026	Norwegian Centre for Molecular Biosciences and Medicine, Oslo, Norway. Invited speaker.
2026	Gene Regulation (CSHL), invited speaker.
2025	Medical Research Council Laboratory of Medical Sciences, London, UK. Invited speaker.
2025	UCLA Computer Genomics Summer Institute invited tutorial
2025	Yale RNA AI meeting invited speaker
2025	Alzheimer's disease Gordon Research Conference invited speaker
2025	Systems Biology session, Probabilistic Genomics (CSHL), co-chair and speaker
2025	Stony Brook Single Cell Genomics invited speaker
2024	Pacific Biosciences PRISM Boston invited speaker
2024	UCLA Computational Genomics Summer Institute
2024	Generative AI in Life Science, Copenhagen, invited speaker
2024	Brown/Lifespan Center for Clinical Cancer Informatics and Data Science
2024	Memorial Sloan Kettering Cancer Center csBio Seminar, invited speaker
2024	NYU Institute for Systems Genetics Nano Seminar, invited speaker
2023	MIT Bioinformatics Seminar Series
2023	Engineering in Medicine Symposium, Columbia University
2023	Quantitative Biology Seminar Series, Cold Spring Harbor Laboratories
2023	Genetics and Genomic Sciences External Speaker Seminar series, Mount Sinai School of Medicine
2022	Computational Genomics Summer Institute, UCLA
2022	Simons Institute <i>From Algorithms to Discovery</i> Conference
2021	Machine Learning in Computational Biology seminar series, University of Washington
2020	Mount Sinai Medical School Neurogenomics seminar series
2020	Einstein Developmental and Molecular Biology seminar series
2020	New York Stem Cell Foundation Invited Speaker
2020	NIH Single Cell Genomics Hackathon Invited Speaker
2019	Washington University Department of Genetics invited speaker
2019	NYC Human Genetics meeting invited speaker
2019	UCLA Computational Genomics Summer Institute
2018	UCLA Computational Genomics Winter Institute
2017	American Society of Human Genetics Annual Meeting
2017	Statistical and Computational Challenges in Large Scale Molecular Biology at the Banff International Research Station
2015	Highlights from Bayesian Analysis (Joint Statistical Meeting session), invited speaker
2015	Statistical and Computational Challenges In Bridging Functional Genomics, Epigenomics, Molecular QTLs, and Disease Genetics at the Banff International Research Station
2015	The Biology of Genomes meeting at Cold Spring Harbor Laboratory
2014	NeurIPS Variational Inference Workshop invited speaker
2013	RECOMB/ISCB Conference on Regulatory & Systems Genomics
2012	Stanford Biostatistics Workshop
2012	Collegio Carlo Alberto Statistics Seminar
2011	Cambridge Statistics Initiative One Day Special Meeting

2011 27th Conference on Uncertainty in Artificial Intelligence (UAI)
 2011 NeurIPS Workshop on Predictive Models in Personalized Medicine
 2011 Eighth Workshop on Bayesian Nonparametrics; Veracruz, Mexico
 2009 NeurIPS Workshop on Computational Biology

Grants awarded

2026-2027 **Splicing in Space and Time: Computational Modeling of Isoform Regulation in Neurodevelopmental Disorders.** Funding body: Columbia DSI. Role: MPI. Lead PI: Sanja Vickovic (Columbia/NYGC).
 2025-2027 **Cross-Species Analysis of Microglial States in Heterogeneous Alzheimer's Disease Progression**
 Funding body: Alzheimer's Association. Role: Co-I. Lead PI: Towfique Raj (MSSM).
 2025-2029 **FunGen-xQTL: Unraveling the genetic basis of molecular functions in Alzheimer's Disease**
 Award number: R01AG086467. Funding body: National Institute of Aging. Role: Co-I. Lead PI: Gao Wang (Columbia).
 2025-2026 **Disentangling Tissue and Ome Shared and Private Effects of Exercise** Award number: R03AG102031.
 Funding body: National Institute of Aging. Role: PI.
 2024-2025 **The genomic and synaptic basis of language disorder and learned sound association.** SIRS Blavatnik fund. Role: lead PI. Co-I: David Sulzer (Psychiatry, Neurology, Pharmacology).
 2023-2026 **Inferring cell-type specific gene regulatory and disease causal networks underlying cancer progression and drug resistance.** Funding body: MacMillan Center for the Study of the Non-Coding Cancer Genome. Role: Lead principal investigator (PI). Co-I: Elham Azizi, Biomedical Engineering, Columbia University.
 2024-2025 **A deep learning based approach for phenome-wide association studies.** Columbia Precision Medicine Pilot Award. Role: Co-I. Lead PI: Gamze Gürsoy (Columbia).
 2022-2025 **CAREER: Unifying short and long read RNA-seq analysis of alternative splicing using network flow models.** Award number: DBI2146398. Funding body: National Science Foundation Biological Informatics program. Role: PI.
 2022-2027 **Delineating the network effects of mental disorder-associated variants using convex optimization methods.** Award number: R01MH130879. Funding body: National Institute of Mental Health. Role: Lead PI. Co-PI: Niamh Mullins, Mount Sinai School of Medicine.
 2022-2024 **A CRISPR/Cas13 approach for identifying individual transcript isoform function in cancer.** Award number: R21CA272345. Funding body: National Cancer Institute. Role: Lead PI. Co-I: Neville Sanjana, New York University.
 2021-2022 **Identifying genetic and transcriptomic drivers of Parkinson's disease progression.** Award number: U01NS120256. Funding body: National Institute of Neurological Disorders and Stroke. Role: Co-Investigator.
 2021-2023 **Resolving complex alternative splicing of psychiatric disease genes using single-cell approaches.** R01MH125579. Funding body: National Institute of Mental Health. Role: Co-I. Multi-PIs: Gang Fang (MSSM), Kristen Brennand (Yale).
 2021-2023 **Fine-mapping psychiatric disease variants that affect post-transcriptional gene regulation.** Award number: R56MH127844. Funding body: National Institute of Mental Health. Role: Lead PI. With Kristen Brennand (Co-I, Yale) and William Fairbrother (Co-I, Brown).
 2020-2025 **Learning the Regulatory Code of Alzheimer's Disease Genomes.** Award number: U01AG068880. Funding body: National Institute on Aging. Role: Multiple PI. With Towfique Raj, Mount Sinai School of Medicine.

GRANTS AWARDED TO TRAINEES

2025 **Geometric modeling of perturbed gene regulation dynamics across time and space.** Damon Runyon Quantitative Biology Fellowship Award. Trainee: Aaron Zweig.

2024 **A probabilistic framework for deconvolving causal mechanisms of cancer therapeutics with genetic perturbation screens.** Damon Runyon Quantitative Biology Fellowship Award. Trainee: Isabella Grabski.

2024 **The genomic and synaptic basis of language disorder and learned sound association.** NIH/NIDCD. Award number: F31DC022183. Trainee: Shashaank Narayanan.

2022 **Perturbation-response approaches to determining the regulatory networks underlying human complex traits.** NIH/NHGRI. Award number: K99-HG012373. Trainee: Brielin Brown (postdoc).

2021 **Measuring allele and isoform-specific RBP binding to improve predictive models of RNA splicing.** NIH/NIGMS. Award number: F32GM142213. Trainee: Megan Schertzer (postdoc).

NSF Graduate Research Fellowships awarded to Udai Nagpal, Sei Chang, and Shashaank Narayanan. CSGrad4US fellowships awarded to Daniel Meyer and Arghamitra Talukder.

Publications

I have published in two fields, machine learning and genetics. For both, the first author(s) contributed most to the work, and the senior authors appear last. In machine learning, proceedings for the top conferences (i.e., NeurIPS, ICML, AISTATS) are as prestigious as any journal, whereas in genetics conference abstracts are not considered formal publications.

Total citations: **13594**. *h*-index: **40** (retrieved from Google scholar, 2/23/2026).

Google Scholar profile: <https://scholar.google.com/citations?user=N9QGV6YAAAAAJ>.

ORCID: <https://orcid.org/0000-0002-7408-146X>

Lab members are highlighted in **bold**.

Highlighted publications

- [1] **Jiayu Su**, Yiming Qu, **Megan Schertzer**, Haochen Yang, Jiahao Jiang, Tenzin Lhakhang, Theodore M Nelson, **Stella Park**, Qiliang Lai, Xi Fu, Seung-Won Choi, **David A Knowles**[†], and Raul Rabadan[†]. A computational framework for mapping isoform landscape and regulatory mechanisms from spatial transcriptomics data. *Nature Biotechnology* (2026). [†]Co-corresponding. **Summary:** SPLISOSM (SpatiaL ISOform Statistical Modeling) performs kernel-based nonparametric tests of isoform usage in spatial transcriptomics. SPLISOSM identifies over 1,000 spatially variable isoforms in the mouse brain, primarily in synaptic signaling pathways, and uncovers regulation by specific RNA binding proteins. Analysis of human glioblastoma demonstrates pervasive isoform diversity in antigen presentation and adhesion genes associated with specific microenvironmental conditions.
- [2] **Brielin C Brown**, John A Morris, Tuuli Lappalainen, and **David A Knowles**. Large-scale causal discovery using interventional data sheds light on gene network structure in K562 cells. *Nature Communications* (2025). **Summary:** we introduce inverse sparse regression (inspre), which extends instrumental variable analysis to infer causal networks that can include loops. inspre is a directed analogue of the graphical lasso, deconvolving direct effects from those mediated through the gene network. Applied to genome-wide Perturb-seq from K562, inspre detects a sparse network with pervasive mediation.
- [3] **Anjali Das**, Chirag Lakhani, Chloé Terwagne, Jui-Shan T Lin, Tatsuhiko Naito, Towfique Raj, and **David A Knowles**. Leveraging functional annotations to map rare variants associated with Alzheimer’s disease with gruyere. *American Journal of Human Genetics* (2025). **Summary:** gruyere is a rare variant association testing framework that links noncoding variants to genes and, unlike existing methods, learns genome-wide variant annotation impor-

tance to improve power. We used gruyere to analyse 21k whole genomes from the Alzheimer’s Disease Sequencing Project, finding 13 disease-associated genes missed by other methods, including the nuclear pore protein NUP93.

- [4] **Megan D Schertzer, Andrew Stirn, Keren Isaev, Laura Pereira, Anjali Das, Claire Harbison, Stella H Park**, Hans-Hermann Wessels, Neville E Sanjana, and **David A Knowles**. Cas13d-mediated isoform-specific RNA knockdown with a unified computational and experimental toolbox. *Nature Communications* (2025). **Summary:** to empower functional studies of alternative isoforms, we tested a novel strategy for isoform-specific knockdown: targeting unique exon-exon junctions in mature RNA with CRISPR/Cas13d. We characterise the determinants of effective knockdown and validate using long-read RNA-seq.
- [5] **Brielin C Brown†, Collin Wang**, Silva Kasela, François Aguet, Daniel C Nachun, Kent D Taylor, Russell P Tracy, Peter Durda, Yongmei Liu, W Craig Johnson, David Van Den Berg, Namrata Gupta, Stacy Gabriel, Joshua D Smith, Robert Gerzsten, Clary Clish, Quenna Wong, George Papanicolaou, Thomas W Blackwell, Jerome I Rotter, Stephen S Rich, R Graham Barr, Kristin G Ardlie, **David A Knowles†**, and Tuuli Lappalainen†. Multiset correlation and factor analysis enables exploration of multi-omics data. *Cell Genomics* (2023). †Co-corresponding. **Summary:** We developed MCFA, a scalable, interpretable extension of PCA to multiomics data. MCFA disentangles shared vs dataset-private factors of variation. Applied to the TOPMED MESA cohort MCFA discovers interpretable factors associated with metabolism and immune response.
- [6] Hans-Hermann Wessels*, **Andrew Stirn***, Alejandro Méndez-Mancilla, **Eric J Kim**, Sydney K Hart, **David A Knowles†**, and Neville E Sanjana†. Prediction of on-target and off-target activity of CRISPR–Cas13d guide RNAs using deep learning. *Nature Biotechnology* (2023). *Equal contribution. †Co-corresponding. **Summary:** we developed a state-of-the-art deep learning method for predicting CRISPR/Cas13d RNA-targeting guide efficacy. We provide both code and an easy to use website tiger.nygenome.org that has been widely adopted by the community.
- [7] Yang I. Li*, **David A. Knowles***, Jack Humphrey, Alvaro N. Barbeira, Scott P. Dickinson, Hae Kyung Im, and Jonathan K. Pritchard. Annotation-free quantification of RNA splicing using LeafCutter. *Nature Genetics* (2017). *Equal contribution. **Summary:** LeafCutter performs scalable, flexible quantification of alternative splicing from short-read RNA-seq. LeafCutter has been widely used for splicing quantitative trait locus mapping, including by standard-setting consortia such as GTEx and PsychENCODE.

Under submission

- [8] **Jiayu Su**, Jun Hou Fung, Haoyu Wang, Dian Yang, **David A Knowles†**, and Raul Rabadan†. On the consistent and scalable detection of spatial patterns. *arXiv* (2026). †Co-corresponding.
- [9] **Saikat Banerjee**, Shane O’Connell, Sarah M C Colbert, Niamh Mullins, and **David A Knowles**. Convex approaches to isolate the shared and distinct genetic structures of subphenotypes in heterogeneous complex traits. *medRxiv & under review at Bioinformatics* (2025).
- [10] Lindsay A Becker, Sofia A Quinodoz, Troy J Comi, Ofer Kimchi, **David A Knowles†**, and Clifford P Brangwynne†. Genome-wide mapping of mesoscale neuronal RNA organization and condensation. *bioRxiv* (2025). †Co-corresponding.
- [11] Winston H Dredge, B P Kailash, **Aline Réal**, Benjamin Z Muller, Beomjin Jang, Ting Fu, Hyomin Seo, Hong-Hee Won, Michael S Breen, Xinshu Xiao, **David A Knowles**, Jack Humphrey, and Towfique Raj. Meta-analysis of genetic regulation of RNA editing in the human brain identifies new genes underlying neurological disease. *medRxiv & under review at Cell* (2025).

- [12] **Isabella N Grabski**, Junsuk Lee, John Blair, Carol Dalgarno, Isabella Mascio, Alexandra Bradu, **David A Knowles**[†], and Rahul Satija[†]. Mapping transcriptional responses to cellular perturbation dictionaries with RNA fingerprinting. *bioRxiv & in revision at Cell* (2025). [†]Co-corresponding.
- [13] **Keren Isaev** and **David A Knowles**. Aging-associated alternative splicing programs conserved between human and mouse tissues. *bioRxiv & under review at Nature Aging* (2025).
- [14] **Chirag M Lakhani**, Jui-Shan T Lin, Anjali Das, Tatsuhiko Naito, Towfique Raj, and **David A Knowles**. Integration of deep learning annotations with functional genomics improves identification of causal Alzheimer’s disease variants. *medRxiv* (2025).
- [15] **Chirag M. Lakhani**, Giacomo Cavalca, Anjing Liu, Rohan Nidumbur, Ru Feng, Towfique Raj, Phillip De Jager, The Alzheimer’s Disease Functional Genomics Consortium, Gao Wang[†], and **David A. Knowles**[†]. Machine learning-based prediction of cell-type resolved brain eQTLs enhances discovery of variants explaining Alzheimer’s disease heritability. *medRxiv & under review at Nature Genetics* (2025). [†]Co-corresponding.
- [16] Chia-Yi Lee, Ashvin Ravi, Tzu-Chieh Huang, Briana Wyman, Ariana Chriss, Jack Humphrey, **David A Knowles**, Chris Cotsapas, Laura Huckins, Towfique Raj, and Kristen Brennand. Regulatory landscape of Alzheimer’s disease variants in human microglia. *medRxiv* (2025).
- [17] **Zaikang Lin**, Sei Chang, Aaron Zweig, Minseo Kang, Elham Azizi, and **David A Knowles**. Interpretable neural ODEs for gene regulatory network discovery under perturbations. *arXiv [cs.LG]* (2025).
- [18] Lingjun Meng, Alice Yu, Huanwei Huang, Jodie Meng, Yongjin Yoo, Katie Schaukowitch, Justyna A Janas, and **David A Knowles**. Oncogenic mutations disrupt lineage differentiation via epigenetic perturbations. *bioRxiv* (2025).
- [19] **Tatsuhiko Naito**, Kosei Hirata, Beomjin Jang, **Chirag M Lakhani**, Alice Buonfiglioli, Wan-Ping Lee, Otto Valladares, Li-San Wang, Yukinori Okada, Hong-Hee Won, Eirini P Papatrou, Samuele G Marro, **David A Knowles**, and Towfique Raj. Mosaic chromosomal alterations in blood are associated with an increased risk of Alzheimer’s disease. *medRxiv & under review at Nature Genetics* (2025).
- [20] **Aline Réal**, Kailash Bp, Winston H Cuddleston, Benjamin Z Muller, Beomjin Jang, **Alex Tokolyi**, Hong-Hee Won, Jack Humphrey, Towfique Raj[†], and **David A Knowles**[†]. Mapping genetic effects on splicing in ten thousand post-mortem brain samples reveals novel mediators of neurological disease risk. *medRxiv & under review at Cell* (2025). [†]Co-corresponding.
- [21] **Megan D Schertzer**, Stella H Park, Jiayu Su, Gloria M Sheynkman, and **David A Knowles**. Perplexity as a metric for isoform diversity in the human transcriptome. *bioRxiv & under review at Genome Biology* (2025).
- [22] **Alex Tokolyi**, Saikat Banerjee, and **David A Knowles**. Uncovering splicing-mediated disease associations in the UK Biobank and All of Us using a TWAS with Bayesian joint fine-mapping. *medRxiv & under review at Nature Genetics* (2025).
- [23] **Smriti Vaidyanathan**, **Keren Isaev**, Aaron Zweig, and **David A Knowles**. Robust integration of sparse single-cell alternative splicing and gene expression data with SpliceVI. *bioRxiv* (2025).
- [24] **Aaron Zweig**, **Zaikang Lin**, Elham Azizi, and **David A. Knowles**. Towards identifiability of interventional stochastic differential equations. *arXiv [cs.LG]* (2025).
- [25] **Aaron Zweig**, Mingxuan Zhang, Elham Azizi, and **David A Knowles**. Energy guided geometric flow matching. *arXiv* (2025).

Journal articles (genetics)

- [26] Michael B Fernando, Yu Fan, Yanchun Zhang, **Alex Tokolyi**, Aleta N Murphy, Sarah Kam-mourh, P J Michael Deans, Sadaf Ghorbani, Ryan Onatzevitch, Adriana Pero, Christopher Padilla, Sarah Williams, Erin K Flaherty, Iya A Prytkova, Lei Cao, **David A Knowles**, Gang Fang, Paul A Slesinger, and Kristen J Brennand. Phenotypic complexities of rare heterozygous neurexin-1 deletions. *Nature* (2025).
- [27] Beomjin Jang, Kailash Bp, **Alex Tokolyi**, Winston H Cuddleston, Ashvin Ravi, Sang-Hyuk Jung, **Tatsuhiko Naito**, Beomsu Kim, Min Seo Kim, Minyoung Cho, Mi-So Park, Mikaela Rosen, Joel Blanchard, Jack Humphrey, **David A Knowles**, Hong-Hee Won, and Towfique Raj. SingleBrain: A meta-analysis of single-nucleus eQTLs linking genetic risk to brain disorders. *Nature Genetics* (2025).
- [28] Alan M Moses, Jason E Stajich, Audrey P Gasch, and **David A Knowles**. Inferring fungal cis-regulatory networks from genome sequences. *Genetics* (2025).
- [29] Markus Terrey, Georgii Krivoshein, **Scott I Adamson**, Elena Arystarkhova, Laura Anderson, John Szvec, Shelby McKee, Holly Jones, Sara Perkins, Vijay Selvam, Pierre-Alexandre Picc, Dweet Chhaya, Ari Dehn, Aamir Zuberi, Stephen A Murray, Natalia S Morsci, Kathleen J Sweadner, **David A Knowles**, Else A Tolner, Arn M J M van den Maagdenberg, and Cathleen M Lutz. Alternating hemiplegia of childhood associated mutations in *Atp1a3* reveal diverse neurological alterations in mice. *Neurobiology of disease* (2025).
- [30] Shane O’Connell, **Brielin C Brown**, Dara M Cannon, Pilib Ó Broin, **David A Knowles**, Nadine Parker, Dag Alnæs, Lars T Westlye, **Saikat Banerjee**, Leila Nabulsi, Emma Corley, Ole A Andreassen, and Niamh Mullins. Deriving Mendelian Randomization-based causal networks of brain imaging phenotypes and bipolar disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2024).
- [31] Cameron Y Park*, **Shouvik Mani***, Nicolas Beltran-Velez, Katie Maurer, Teddy Huang, Shuqiang Li, Satyen Gohil, Kenneth J Livak, **David A Knowles**, Catherine J Wu, and Elham Azizi. A Bayesian framework for inferring dynamic intercellular interactions from time-series single-cell data. *Genome research* (2024). *Equal contribution.
- [32] Mariela Cortés-López, Paulina Chamely, Allegra G Hawkins, Robert F Stanley, Ariel D Swett, Saravanan Ganesan, Tarek H Mouhieddine, Xiaoguang Dai, Lloyd Kluegel, Celine Chen, Kiran Batta, Nili Furer, Rahul S Vedula, John Beaulaurier, Alexander W Drong, Scott Hickey, Neville Dusaj, Gavriel Mullokandov, Adam M Stasiw, **Jiayu Su**, Ronan Chaligné, Sissel Juul, Eoghan Harrington, **David A Knowles**, Catherine J Potenski, Daniel H Wiseman, Amos Tanay, Liran Shlush, Robert C Lindsley, Irene M Ghobrial, Justin Taylor, Omar Abdel-Wahab, Federico Gaiti, and Dan A Landau. Single-cell multi-omics defines the cell-type-specific impact of splicing aberrations in human hematopoietic clonal outgrowths. *Cell stem cell* (2023).
- [33] **Shashaank Narayanan**, Mahalakshmi Somayaji, Mattia Miotto, Eugene V Mosharov, Emily A Makowicz, **David A Knowles**, Giancarlo Ruocco, and David L Sulzer. Computational models of dopamine release measured by fast scan cyclic voltammetry in vivo. *PNAS nexus* (2023).
- [34] **Jiayu Su**, Jean-Baptiste Reynier, Xi Fu, Guojie Zhong, Jiahao Jiang, Rydberg Supo Escalante, Yiping Wang, Luis Aparicio, Benjamin Izar, **David A Knowles**, and Raul Rabadan. Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. *Genome biology* (2023).
- [35] Julia J Wattacheril, Srilakshmi Raj, **David A Knowles**, and John M Greally. Using epigenomics to understand cellular responses to environmental influences in diseases. *PLoS genetics* (2023).

- [36] Rockwell J Weiner, **Chirag Lakhani**, **David A Knowles**, and Gamze Gürsoy. LDmat: efficiently queryable compression of linkage disequilibrium matrices. *Bioinformatics* (2023).
- [37] Jack Humphrey, Sanan Venkatesh, Rahat Hasan, Jake T Herb, Katia de Paiva Lopes, Fahri Küçükali, Marta Byrska-Bishop, Uday S Evani, Giuseppe Narzisi, Delphine Fagegaltier, NYGC ALS Consortium, Kristel Slegers, Hemali Phatnani, **David A Knowles**, Pietro Fratta, and Towfique Raj. Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nature neuroscience* (2022).
- [38] **Stephen Malina**, Daniel Cizin, and **David A. Knowles**. Deep Mendelian randomization: Investigating the causal knowledge of genomic deep learning models. *PLOS Computational Biology* (2022).
- [39] Brunilda Balliu, Ivan Carcamo-Orive, Michael J Gloudemans, Daniel C Nachun, Matthew G Durrant, Steven Gazal, Chong Y Park, **David A Knowles**, Martin Wabitsch, Thomas Quertermous, Joshua W Knowles, and Stephen B Montgomery. An integrated approach to identify environmental modulators of genetic risk factors for complex traits. *American Journal of Human Genetics* (2021).
- [40] **Brielin C Brown** and **David A Knowles**. Welch-weighted Egger regression reduces false positives due to correlated pleiotropy in Mendelian randomization. *American Journal of Human Genetics* (2021).
- [41] Kévin Contrepois, Si Wu, Kegan J Moneghetti, Daniel Hornburg, Sara Ahadi, Ming-Shian Tsai, Ahmed A Metwally, Eric Wei, Brittany Lee-McMullen, Jeniffer V Quijada, Songjie Chen, Jeffrey W Christle, Mathew Ellenberger, Brunilda Balliu, Shalina Taylor, Matthew G Durrant, **David A Knowles**, Hani Choudhry, Melanie Ashland, Amir Bahmani, Brooke Enslen, Myriam Amsallem, Yukari Kobayashi, Monika Avina, Dalia Perelman, Sophia Miryam Schüssler-Fiorenza Rose, Wenyu Zhou, Euan A Ashley, Stephen B Montgomery, Hassan Chaib, Francois Haddad, and Michael P Snyder. Molecular Choreography of Acute Exercise. *Cell* (2020).
- [42] Andrew J Gentles, Angela Bik-Yu Hui, Weiguo Feng, Armon Azizi, Ramesh V. Nair, **David A. Knowles**, Alice Yu, Youngtae Jeong, Alborz Bejnood, Erna Forgó, Sushama Varma, Yue Xu, Amanda Kuong, Viswam S. Nair, Rob West, Matt van de Rijn, Chuong D. Hoang, Maximilian Diehn, and Sylvia K. Plevritis. Clinically-relevant cell type cross-talk identified from a human lung tumor microenvironment interactome. *Genome Biology* (2020).
- [43] Kevin Hadi, Xiaotong Yao, Julie M. Behr, Aditya Deshpande, Charalampos Xanthopoulos, Joel Rosiene, Madison Darmofal, Huasong Tian, Joseph DeRose, Rick Mortensen, Emily M. Adney, Zoran Gajic, Kenneth Eng, Jeremiah A. Wala, Kazimierz O. Wrzeszczyński, Kanika Arora, Minita Shah, Anne-Katrin Emde, Vanessa Felice, Mayu O. Frank, Robert B. Darnell, Mahmoud Ghandi, Franklin Huang, John Maciejowski, Titia De Lange, Jeremy Setton, Nadeem Riaz, Jorge S. Reis-Filho, Simon Powell, **David A. Knowles**, Ed Reznik, Bud Mishra, Rameen Beroukhi, Michael Zody, Nicolas Robine, Kenji M. Oman, Carissa A. Sanchez, Mary K. Kuhner, Lucian P. Smith, Patricia C. Galipeau, Thomas G. Paulson, Brian J. Reid, Xiaohong Li, David Wilkes, Andrea Sboner, Juan Miguel Mosquera, Olivier Elemento, and Marcin Imielinski. Novel patterns of complex structural variation revealed across thousands of cancer genome graphs. *Cell* (2020).
- [44] Brunilda Balliu, Matthew Durrant, Olivia de Goede, Nathan Abell, Xin Li, Boxiang Liu, Michael J Gloudemans, Naomi L Cook, Kevin S Smith, **David A Knowles**, Mauro Pala, Francesco Cucca, David Schlessinger, Siddhartha Jaiswal, Chiara Sabatti, Lars Lind, Erik Ingelsson, and Stephen B Montgomery. Genetic regulation of gene expression and splicing during a 10-year period of human aging. *Genome Biology* (2019).

- [45] Diego Calderon, Michelle L.T. Nguyen, Anja Mezger, Arwa Kathiria, Vinh Nguyen, Ninnia Lescano, Beijing Wu, John Trombetta, Jessica V. Ribado, **David A. Knowles**, Ziyue Gao, Audrey V. Parent, Trevor D. Burt, Mark S. Anderson, Lindsey A. Criswell, William J. Greenleaf, Alexander Marson, and Jonathan K. Pritchard. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics* (2019).
- [46] Kishore Jaganathan, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F McRae, Siavash Fazel Darbandi, **David Knowles**, Yang I Li, Jack A Kosmicki, Juan Arbelaez, Wenwu Cui, Grace B Schwartz, Eric D Chow, Efstathios Kanterakis, Hong Gao, Amirali Kia, Serafim Batzoglou, Stephan J Sanders, and Kyle Kai-How Farh. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* (2019).
- [47] **David A Knowles**, Gina Bouchard, and Sylvia K Plevritis. Sparse discriminative latent characteristics for predicting cancer drug sensitivity from genomic features. *PLoS Computational Biology* (2019).
- [48] Michael Wainberg, Nasa Sinnott-Armstrong, Nicholas Mancuso, Alvaro N Barbeira, **David A Knowles**, David Golan, Raili Ermel, Arno Ruusalepp, Thomas Quertermous, Ke Hao, Johan L M Björkegren, Hae Kyung Im, Bogdan Pasaniuc, Manuel A Rivas, and Anshul Kundaje. Opportunities and challenges for transcriptome-wide association studies. *Nature Genetics* (2019).
- [49] **David A Knowles**^{*}, Courtney K Burrows^{*}, John D Blischak, Kristen M Patterson, Daniel J. Serie, Nadine Norton, Carole Ober, Jonathan K Pritchard, and Yoav Gilad. Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *eLife* (2018). ^{*}Equal contribution.
- [50] D Leland Taylor, **David A Knowles**, Laura J Scott, Andrea H Ramirez, Francesco Paolo Casale, Brooke N Welford, Li Guan, Arushi Varshney, Ricardo D’oliveira Albanus, Stephen C J Parker, Narisu Narisu, Peter S Chines, Michael R Erdos, Ryan P Welch, Leena Kinnunen, Jouko Saramies, Jouko Sundvall, Timo A Lakka, Markku Laakso, Jaakko Tuomilehto, Heikki A Koistinen, Oliver Stegle, Michael Boehnke, Ewan Birney, and Francis S Collins. Interactions between genetic variation and cellular environment in skeletal muscle gene expression. *PLoS One* (2018).
- [51] Lindsay A. Becker, Brenda Huang, Gregor Bieri, Rosanna Ma, **David A. Knowles**, Paymaan Jafar-Nejad, James Messing, Hong Joo Kim, Armand Soriano, Georg Auburger, Stefan M. Pulst, J. Paul Taylor, Frank Rigo, and Aaron D. Gitler. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nature* (2017).
- [52] Diego Calderon, Anand Bhaskar, **David A Knowles**, David Golan, Towfique Raj, Audrey Q Fu, and Jonathan K Pritchard. Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *American Journal of Human Genetics* (2017).
- [53] Emily K. Tsang, Nathan S. Abell, Xin Li, Vanessa Anaya, Konrad J. Karczewski, **David A. Knowles**, Raymond G. Sierra, Kevin S. Smith, and Stephen B. Montgomery. Small RNA sequencing in cells and exosomes identifies eQTLs and 14q32 as a region of active export. *G3 Genes/Genomes/Genetics* (2017).
- [54] Po-Yuan Tung, John D. Blischak, Chiaowen Joyce Hsiao, **David A. Knowles**, Jonathan E. Burnett, Jonathan K. Pritchard, and Yoav Gilad. Batch effects and the effective design of single-cell gene expression studies. *Scientific Reports* (2017).
- [55] Joe R. Davis, Laure Fresard, **David A. Knowles**, Mauro Pala, Carlos D. Bustamante, Alexis Battle, and Stephen B. Montgomery. An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *The American Journal of Human Genetics* (2016).

- [56] Kimberly R. Kukurba, Princy Parsana, Brunilda Balliu, Kevin S. Smith, Zachary Zappala, **David A. Knowles**, Marie-Julie Favé, Joe R. Davis, Xin Li, Xiaowei Zhu, James B. Potash, Myrna M. Weissman, Jianxin Shi, Anshul Kundaje, Douglas F. Levinson, Philip Awadalla, Sara Mostafavi, Alexis Battle, and Stephen B. Montgomery. Impact of the X chromosome and sex on regulatory variation. *Genome Research* (2016).
- [57] Yang I Li, Bryce van de Geijn, Anil Raj, **David A Knowles**, Allegra A Petti, David Golan, Yoav Gilad, and Jonathan K Pritchard. RNA splicing is a primary link between genetic variation and disease. *Science* (2016).
- [58] Kimberly R. Kukurba, Rui Zhang, Xin Li, Kevin S. Smith, **David A. Knowles**, Meng How Tan, Robert Piskol, Monkol Lek, Michael Snyder, Daniel G. MacArthur, Jin Billy Li, and Stephen B. Montgomery. Allelic Expression of Deleterious Protein-Coding Variants across Human Tissues. *PLoS Genetics* (2014).
- [59] Xin Li, Alexis Battle, Konrad J. Karczewski, Zach Zappala, **David A. Knowles**, Kevin S. Smith, Kim R. Kukurba, Eric Wu, Noah Simon, and Stephen B. Montgomery. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *American Journal of Human Genetics* (2014).
- [60] Daniel Glass, Ana Viñuela, Matthew N Davies, Adaikalavan Ramasamy, Leopold Parts, **David A. Knowles**, Andrew A Brown, Asa K Hedman, Kerrin S Small, Alfonso Buil, Elin Grundberg, Alexandra C Nica, Paoladi Meglio, Frank O Nestle, Mina Ryten, Richard Durbin, Mark I McCarthy, Panagiotis Deloukas, Emmanouil T Dermitzakis, Michael E Weale, Veronique Bataille, and Tim D Spector. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome biology* (2013).
- [61] Elin Grundberg, Kerrin S Small, Asa K Hedman, Alexandra C Nica, Alfonso Buil, Sarah Keildson, Jordana T Bell, Tsun-Po Yang, Eshwar Meduri, Amy Barrett, James Nisbett, Magdalena Sekowska, Alicja Wilk, So-Youn Shin, Daniel Glass, Mary Travers, Josine L Min, **David A. Knowles**, Sue Ring, Karen Ho, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteindottir, Chrysanthi Ainali, Antigone S Dimas, Neelam Hassanali, Catherine Ingle, Maria Krestyaninova, Christopher E Lowe, Paola Di Meglio, Stephen B Montgomery, Leopold Parts, Simon Potter, Gabriela Surdulescu, Loukia Tsaprouni, Sophia Tsoka, Veronique Bataille, Richard Durbin, Frank O Nestle, Stephen O’Rahilly, Nicole Soranzo, Cecilia M Lindgren, Krina T Zondervan, Kourosh R Ahmadi, Eric E Schadt, Kari Stefansson, George Davey Smith, Mark I McCarthy, Panos Deloukas, Emmanouil T Dermitzakis, and Tim D. Spector. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* (2012).
- [62] Mehregan Movassagh, Mun-Kit Choy, **David A Knowles**, Lina Cordeddu, Syed Haider, Thomas Down, Lee Siggins, Ana Vujic, Ilenia Simeoni, Chris Penkett, Martin Goddard, Pietro Lio, Martin Bennett, and Roger Foo. Distinct Epigenomic Features in End-Stage Failing Human Hearts. *Circulation, American Heart Association* (2011).
- [63] Cornelia Schöne, Anne Venner, **David A. Knowles**, Mahesh M Karnani, and Denis Burdakov. Dichotomous cellular properties of mouse orexin/hypocretin neurons. *The Journal of Physiology* (2011).
- [64] Daniel Glass, Leopold Parts, **David Knowles**, Abraham Aviv, and Tim D Spector. No correlation between childhood maltreatment and telomere length. *Biological psychiatry* (2010).

Journal articles (machine learning)

- [65] **David A. Knowles** and Zoubin Ghahramani. Pitman Yor Diffusion Trees for Bayesian hierarchical clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2015).

- [66] Konstantina Palla, **David A. Knowles**, and Zoubin Ghahramani. Relational learning and network modelling using infinite latent attribute models. *IEEE Transactions on Pattern Analysis and Machine Intelligence Special Issue on Bayesian Nonparametrics* (2015).
- [67] Tim Salimans and **David A. Knowles**. Fixed-form variational posterior approximation through stochastic linear regression. *Bayesian Analysis* (2013). Winner of the International Society for Bayesian Analysis Lindley Prize.
- [68] **David A. Knowles** and Zoubin Ghahramani. Nonparametric Bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics* (2011).

Peer-reviewed conference papers

- [69] **Arghamitra Talukder**, Nicholas Keung, Itsik Pe'er†, and **David A Knowles**†. CLADES-contrastive learning augmented DifferEntial splicing with orthologous positive pairs. *Research in Computational Molecular Biology (RECOMB)*. 2026. †Co-corresponding.
- [70] **Andrew Stirn** and **David A Knowles**. The VampPrior Mixture Model. *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2025.
- [71] **Jiayu Su**, **David A Knowles**, and Raul Rabadan. Disentangling interpretable factors with Supervised Independent Subspace Principal Component Analysis. *Advances in Neural Information Processing Systems*. 2024.
- [72] **Keren Isaev** and **David A Knowles**. Investigating RNA splicing as a source of cellular diversity using a binomial mixture model. *Proceedings of Machine Learning Research: Machine Learning in Computational Biology*. 2023.
- [73] **Andrew Stirn**, Hans-Hermann Wessels, **Megan Schertzer**, **Laura Pereira**, Neville E Sanjana, and **David A Knowles**. Faithful Heteroscedastic Regression with Neural Networks. *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2023.
- [74] **Andrew Stirn**, Tony Jebara, and **David A. Knowles**. A New Distribution on the Simplex with Auto-Encoding Applications. *Advances in Neural Information Processing Systems*. 2019.
- [75] Konstantina Palla*, **David A. Knowles***, and Zoubin Ghahramani. A birth-death process for feature allocation. *Proceedings of the 34th International Conference on Machine Learning*. 2017. *Equal contribution.
- [76] Amar Shah, **David A Knowles**, and Zoubin Ghahramani. An Empirical Study of Stochastic Variational Inference Algorithms for the Beta Bernoulli Process. *Proceedings of the 32nd International Conference on Machine Learning*. 2015.
- [77] Creighton Heaukulani, **David A. Knowles**, and Zoubin Ghahramani. Beta Diffusion Trees. *Proceedings of the 31st International Conference on Machine Learning*. 2014.
- [78] **David A. Knowles**, Konstantina Palla, and Zoubin Ghahramani. A reversible infinite HMM using normalised random measures. *Proceedings of The 31st International Conference on Machine Learning*. 2014.
- [79] Novi Quadrianto, Viktoriia Sharmanska, **David A Knowles**, and Zoubin Ghahramani. The Supervised IBP: Neighbourhood Preserving Infinite Latent Feature Models. *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*. 2013.
- [80] Konstantina Palla*, **David A. Knowles***, and Zoubin Ghahramani. A nonparametric variable clustering model. *Advances in Neural Information Processing Systems*. 2012. *Equal contribution.

- [81] Konstantina Palla*, **David A. Knowles***, and Zoubin Ghahramani. An Infinite Latent Attribute Model for Network Data. *Proceedings of the 29th International Conference on Machine Learning*. 2012. *Equal contribution.
- [82] Andrew Gordon Wilson, **David A. Knowles**, and Zoubin Ghahramani. Gaussian Process Regression Networks. *Proceedings of the 29th International Conference on Machine Learning*. 2012.
- [83] **David A. Knowles**, Jurgen Van Gael, and Zoubin Ghahramani. Message Passing Algorithms for the Dirichlet Diffusion Tree. *Proceedings of the 28th International Conference on Machine Learning*. 2011.
- [84] **David A. Knowles** and Zoubin Ghahramani. Pitman-Yor Diffusion Trees. *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*. 2011.
- [85] **David A. Knowles** and Tom Minka. Non-conjugate Variational Message Passing for Multinomial and Binary Regression. *Advances in Neural Information Processing Systems*. 2011.
- [86] Finale Doshi-Velez*, Shakir Mohamed*, **David A. Knowles***, and Zoubin Ghahramani. Large Scale Nonparametric Bayesian Inference: Data Parallelisation in the Indian Buffet Process. *Advances in Neural Information Processing Systems*. 2009. *Equal contribution.
- [87] **David A. Knowles** and Zoubin Ghahramani. Infinite Sparse Factor Analysis and Infinite Independent Components Analysis. *7th International Conference on Independent Component Analysis and Signal Separation*. 2007.

Peer-reviewed workshop papers

- [88] **Anjali Das**, **Chirag Lakhani**, Towfique Raj, and **David A. Knowles**. A unified burden and overdispersion framework for Bayesian functionally informed rare variant association testing. *RECOMB-Genetics*. 2025.
- [89] **Arghamitra Talukder**, Shree Thavarekere, Madison Mehlferber, Gloria M Sheynkman, and **David A Knowles**. Multi-sample, multi-platform isoform quantification using empirical Bayes. *RECOMB-Seq* (2025).
- [90] **Sei Chang**, **Zaiqian Chen**, Bianca Dumitrascu, and **David A Knowles**. CellFlows: Inferring Splicing Kinetics from Latent and Mechanistic Cellular Dynamics. *ICML Workshop on ML for Life and Material Science*. 2024.
- [91] **David A. Knowles**, Alexis Battle, and Daphne Koller. Discovering latent cancer characteristics predictive of drug sensitivity. *RECOMB/ISCB Conference on Regulatory & Systems Genomics (selected for oral presentation)* (2013).
- [92] **David A. Knowles**, Leopold Parts, Daniel Glass, and John M. Winn. Modeling skin and ageing phenotypes using latent variable models in Infer.NET. *NeurIPS Workshop: Computational Biology* (2010).
- [93] **David A. Knowles** and Susan Holmes. Statistical tools for ultra-deep pyrosequencing of fast evolving viruses. *NeurIPS Workshop: Computational Biology* (2009).

Working papers

- [94] **Brielin C. Brown** and **David A. Knowles**. Phenome-scale causal network discovery with bidirectional mediated Mendelian randomization. *bioRxiv* (2020).
- [95] **Udai G. Nagpal** and **David A. Knowles**. Active Learning in CNNs via Expected Improvement Maximization. *arXiv* (2020).

- [96] **Andrew Stirn** and **David A Knowles**. Variational Variance: Simple and Reliable Predictive Variance Parameterization. *arXiv* (2020).
- [97] **David A Knowles**. Stochastic gradient variational Bayes for gamma approximating distributions. *arXiv* (2015).
- [98] Tim Salimans and **David A Knowles**. On using control variates with stochastic approximation for variational Bayes and its connection to stochastic linear regression. *arXiv* (2014).
- [99] Konstantina Palla*, **David A Knowles***, and Zoubin Ghahramani. A dependent partition-valued process for multitask clustering and time evolving network modeling. *arXiv* (2013).
*Equal contribution.

Typeset in X_YTEX
<https://daklab.github.io/>