David A. Knowles

Stanford University School of Medicine
300 Pasteur Drive
Stanford, CA 94305-5105
email: knowles84@gmail.com
URL: https://davidaknowles.github.io/
Nationality: British

Education

2008-2012 PhD Engineering (Machine Learning)
University of Cambridge
Thesis: *Bayesian non-parametric models and inference for sparse and hierarchical latent structure*
Advisor: Prof. Zoubin Ghahramani

2007-2008 MSc Bioinformatics and Systems Biology - Distinction
Imperial College London
Thesis: *Statistical tools for ulta-deep pyrosequencing of fast evolving viruses.*
Thesis advisor: Prof. Susan Holmes, Statistics Department, Stanford University.

2003-2007 MEng Engineering - Distinction
Thesis: *A non-parametric extension to Independent Components Analysis.*
Thesis advisor: Prof. Zoubin Ghahramani.
BA Natural Sciences (Physics) - First Class
University of Cambridge

Academic Positions

2014-ongoing 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

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 for *undergraduates adjudged to be the most worthy for intellectual qualifications or moral conduct and practical activities*
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, *Wolfson College, University of Cambridge*, funded by Microsoft Research

2006 Summer Undergraduate Research Fellow, *California Institute of Technology*

Industry Positions

2009-2012 Contract Software Engineer
Microsoft Research Cambridge
Extending Infer.NET.

2009 Research Intern
Microsoft Research Cambridge
Non-conjugate Variational Message Passing. Supervisor: Thomas P. Minka.

2005 Equity Research Intern
UBS Investment Bank, London
Developed a financial model of Belgacom Telecom. Declined offer.

2004 Software Engineering Intern
Data Connection Limited, London
Developed an automatic build system using Unix shell scripts.

Teaching & Mentoring

2017 CBIO 244: Lecture Series in Cancer Systems Biology
2013- Statistical, Mathematical, and Computational Consulting (SMACC) hosted by Stanford ICME
2013 Advising five groups of students for graduate machine learning (CS229) course projects
2013 Advised undergraduate student for CURIS Undergraduate Research Internship

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

Reviewing & Service

2017 Advances in Neural Information Processing Systems (NIPS) Best Reviewer Award

Journals

Machine Learning, PLOS Genetics, PLOS Computational Biology, Bioinformatics, Genome Research, Journal of the American Statistical Association, Statistics and Computing (Springer), Bayesian Analysis, 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

Conferences

Pacific Biocomputing Symposium (PBS), International Conference on Artificial Intelligence and Statistics (AISTATS), International Joint Conferences on Artificial Intelligence (IJCAI), Advances in Neural Information Processing Systems (NIPS), International Conference on Machine Learning (ICML)

Workshops

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

Professional affiliations

Early Stage Investigators in Cancer Systems Biology Steering Committee, International Society of Bayesian Analysis (ISBA) Chair of the Continuing Education Committee

Selected Talks

2018 UCLA Computational Genomics Winter Institute
2017 American Society of Human Genetics Annual Meeting
2017 Statistical and Computational Challenges in Large Scale Molecular Biology. Banff International Research Station for Mathematical Innovation and Discovery
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 for Mathematical Innovation and Discovery
2015 The Biology of Genomes meeting at Cold Spring Harbor Laboratory
2014 NIPS Variational Inference Workshop (invited speaker)
2013 RECOMB/ISCB Conference on Regulatory & Systems Genomics
2012 Stanford Biostatistics Workshop
2012 BayLearn, the Bay Area Machine Learning Symposium
2012 Collegio Carlo Alberto Statistics Seminar
2011 Cambridge Statistics Initiative One Day Special Meeting
2011 27th Conference on Uncertainty in Artificial Intelligence (UAI)
2011 NIPS Workshop on Predictive Models in Personalized Medicine
2011 RAD Lab, University of California at Berkeley
2011 28th International Conference on Machine Learning (ICML)
2011 Eighth Workshop on Bayesian Nonparametrics; Veracruz, Mexico

2010 Twins Research Department, King's College London
2009 NIPS Workshop on Computational Biology

Highlighted research

Using allelic specific expression to detect GxE interaction effects. Detecting gene-by-environment (GxE) effects on the transcriptome is challenging in observational data. I developed a computational method, EAGLE, which leverages allele-specific expression as a controlled, with-in individual test of the influence of environment factors on different genetic backgrounds. Code is available at https://github.com/davidaknowles/eagle.

[1] David A Knowles, Joe R Davis, Hilary Edgington, Anil Raj, Marie-Julie Favé, Xiaowei Zhu, James B Potash, Myrna M Weissman, Jianxin Shi, Doug Levinson, Philip Awadalla, Sara Mostafavi, Stephen B Montgomery, and Alexis Battle. "Allele-specific expression reveals interactions between genetic variation and environment". *Nature Methods* (2017). DOI: 10.1038/nmeth.4298.

Doxorubicin response QTL mapping and implications for anthracycline-cardiotoxicity (ACT). ACT is a common side-effect of chemotherapy, but its genetic basis and molecular mechanism remain unclear. We measured transcriptomic and cellular damage response to doxorubicin in a panel of iPSC-derved cardiomyocytes from 45 individuals. Using a novel, efficient linear mixed model, `suez`, we identified hundreds of loci that interact with ACT response. These response-QTLs are significantly enriched in ACT GWAS.

[2] 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). DOI: https://doi.org/10.7554/eLife.33480. *These authors contributed equally to this work.

Quantifying splicing variation through intron-excision events. We developed `LeafCutter` to identify, quantify and test variable intron splicing events, obviating the need for accurate transcript annotations and circumventing the challenges in estimating relative isoform abundance. A paper describing the method has been accepted at Nature Genetics[3], an early version of `LeafCutter` was used in our study linking complex disease and splicing[4], and we actively maintain code at https://github.com/davidaknowles/leafcutter.

[3] 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). DOI: 10.1038/s41588-017-0004-9. *These authors contributed equally to this work.

[4] 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* 352.6285 (2016), pp. 600–4. DOI: 10.1126/science.aad9417.

Journal articles (genetics)

[5] D Leland Taylor, David A Knowles, Laura J Scott, Andrea H Ramirez, Francesco Paolo Casale, Brooke N Wolford, 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* 13.4 (Apr. 2018), e0195788. DOI: 10.1371/journal.pone.0195788.

[6] 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* 7 (2017), p. 39921. DOI: 10.1038/srep39921.

[7] 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". en. *American Journal of Human Genetics* (2017).

[8] 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* 7.1 (2017), pp. 31–39. DOI: 10.1534/g3.116.036137.

[9] 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* 544.7650 (2017), pp. 367–371. DOI: 10.1038/nature22038.

[10] 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* 98.1 (2016), pp. 216–224. DOI: 10.1016/j.ajhg.2015.11.021.

[11] 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* 26.6 (2016), pp. 768–777. DOI: 10.1101/gr.197897.115.

[12] 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* 10.5 (2014), e1004304. DOI: 10.1371/journal.pgen.1004304.

[13] 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* 95.3 (2014), pp. 245–56. DOI: 10.1016/j.ajhg.2014.08.004.

[14] 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* 14.7 (2013), R75. DOI: 10.1186/gb-2013-14-7-r75.

[15] 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* 44.10 (2012), pp. 1084–9. DOI: 10.1038/ng.2394.

[16] 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* 589.Pt 11 (2011), pp. 2767–79. DOI: 10.1113/jphysiol.2011.208637.

[17] Mehregan Movassagh, Mun-Kit Choy, David A Knowles, Lina Cordeddu, Syed Haider, Thomas Down, Lee Siggens, 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* 135 (2011). DOI: 10.1161/CIRCULATIONAHA.111.040071.

[18] Daniel Glass, Leopold Parts, David Knowles, Abraham Aviv, and Tim D Spector. "No correlation between childhood maltreatment and telomere length." *Biological psychiatry* 68.6 (2010), e21–2. DOI: 10.1016/j.biopsych.2010.02.026.

Journal articles (machine learning)

[19] David A. Knowles and Zoubin Ghahramani. "Pitman Yor Diffusion Trees for Bayesian hierarchical clustering". English. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.2 (2015), pp. 271–289. DOI: 10.1109/TPAMI.2014.2313115.

[20] 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* 37.2 (2015), pp. 462–474. DOI: 10.1109/TPAMI.2014.2324586.

[21] Tim Salimans and David A. Knowles. "Fixed-form variational posterior approximation through stochastic linear regression". *Bayesian Analysis* 8.4 (2013), pp. 837–882. DOI: 10.1214/13-BA858. Winner of the International Society for Bayesian Analysis Lindley Prize.

[22] David A. Knowles and Zoubin Ghahramani. "Nonparametric Bayesian sparse factor models with application to gene expression modeling". *The Annals of Applied Statistics* 5.2B (2011), pp. 1534–1552. DOI: 10.1214/10-AOAS435.

Peer-reviewed conference papers (machine learning)

[23] 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. *These authors contributed equally to this work.

[24] 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, pp. 1594–1603.

[25] 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.

[26] Creighton Heaukulani, David A. Knowles, and Zoubin Ghahramani. "Beta Diffusion Trees". *Proceedings of the 31st International Conference on Machine Learning*. 2014, pp. 1809–1817.

[27] 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.

[28] Konstantina Palla*, David A. Knowles*, and Zoubin Ghahramani. "A nonparametric variable clustering model". *Advances in Neural Information Processing Systems*. Vol. 5. 2012, pp. 2987–2995. *These authors contributed equally to this work.

[29] 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, pp. 1607–1614. *These authors contributed equally to this work.

[30] Andrew Gordon Wilson, David A. Knowles, and Zoubin Ghahramani. "Gaussian Process Regression Networks". *Proceedings of the 29th International Conference on Machine Learning*. 2012, pp. 599–606.

[31] 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, pp. 721–728.

[32] David A. Knowles and Tom Minka. "Non-conjugate Variational Message Passing for Multinomial and Binary Regression". *Advances in Neural Information Processing Systems*. 2011, pp. 1701–1709.

[33] David A. Knowles and Zoubin Ghahramani. "Pitman-Yor Diffusion Trees". *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*. 2011, pp. 410–418.

[34] 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, pp. 1294–1302. *These authors contributed equally to this work.

[35] 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. DOI: 10.1007/978-3-540-74494-8.

Working papers/Under submission

[36] David A Knowles, Gina Bouchard, and Sylvia K Plevritis. "Sparse discriminative latent characteristics for predicting cancer drug sensitivity from genomic features". *In revision* (2018).

[37] Michael Wainberg, Nasa Sinnott-Armstrong, David A Knowles, David Golan, Raili Ermel, Arno Ruusalepp, Thomas Quertermous, Ke Hao, Johan L. M. Bjorkegren, Manuel A. Rivas, and Anshul Kundaje. "Vulnerabilities of transcriptome-wide association studies". *bioRxiv* (2017). DOI: 10.1101/206961.

[38] David A Knowles. "Stochastic gradient variational Bayes for gamma approximating distributions". *arXiv* (2015), p. 1509.01631.

[39] 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), p. 1401.1022.

[40] Konstantina Palla*, David A Knowles*, and Zoubin Ghahramani. "A dependent partition-valued process for multitask clustering and time evolving network modeling". *arXiv* (2013), p. 1303.3265. *These authors contributed equally to this work.

Peer-reviewed workshop papers

[41] 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).

[42] David A. Knowles, Leopold Parts, Daniel Glass, and John M. Winn. "Modeling skin and ageing phenotypes using latent variable models in Infer.NET". *NIPS Workshop: Computational Biology* (2010).

[43] David A. Knowles and Susan Holmes. "Statistical tools for ultra-deep pyrosequencing of fast evolving viruses". *NIPS Workshop: Computational Biology* (2009), pp. 1–9.

Last updated: May 25, 2018 • Typeset in XeLaTeX
https://davidaknowles.github.io/