

David A. Knowles

Stanford University School of Medicine
300 Pasteur Drive
Stanford, CA 94305-5105
email: knowles84@gmail.com
URL: <http://cs.stanford.edu/~davidknowles/>
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 ultra-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 Scholarship, *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

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

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

Publications

JOURNAL ARTICLES

- 2017 **David A. Knowles**, Joe R. Davis, Hilary Edgington, Anil Raj, Marie-Julie Favé, Xiaowei Zhu, James B. Potash, Myrna M. Weissman, Jianxin Shi, Douglas F. Levinson, Philip Awadalla, Sara Mostafavi, Stephen B. Montgomery, Alexis Battle. Allele-specific expression reveals interactions between genetic variation and environment. *Nature Methods*.
- 2017 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 Po-Yuan Tung, John D Blischak, Chiaowen Joyce Hsiao, **David A Knowles**, Jonathan E Burnett, Jonathan K Pritchard, Yoav Gilad. Batch effects and the effective design of single-cell gene expression studies. *Nature Scientific Reports*.
- 2017 Emily K Tsang, Nathan S Abell, Xin Li, Vanessa Anaya, Konrad J Karczewski, **David A Knowles**, Raymond G Sierra, Kevin S Smith, Stephen B Montgomery. Small RNA Sequencing in Cells and Exosomes Identifies eQTLs and 14q32 as a Region of Active Export. *G3: Genes | Genomes | Genetics*.
- 2016 Yang I. Li, Bryce van de Geijn, Anil Raj, **David A. Knowles**, Allegra A. Petti, David Golan, Yoav Gilad, Jonathan K. Pritchard. RNA splicing is a primary link between genetic variation and disease. *Science*.
- 2016 Kimberly R. Kukurba, Princy Parsana, Kevin S. Smith, Zachary Zappala, **David A. Knowles**, Marie-Julie Fave, Xin Li, Xiaowei Zhu, James B. Potash, Myrna M. Weissman, Jianxin Shi, Anshul Kundaje, Douglas F. Levinson, Philip Awadalla, Sara Mostafavi, Alexis Battle, Stephen B. Montgomery. Impact of the X chromosome and sex on regulatory variation. *Genome Research*.
- 2016 Joe R. Davis, Laure Fresard, **David A. Knowles**, Mauro Pala, Carlos D. Bustamante, Alexis Battle, Stephen B. Montgomery. An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *American Journal of Human Genetics*.
- 2014 **David A. Knowles** and Zoubin Ghahramani. Pitman-Yor Diffusion Trees for Bayesian hierarchical clustering. *IEEE TPAMI Special Issue on Bayesian Nonparametrics*.
- 2014 Konstantina Palla, **David A. Knowles** and Zoubin Ghahramani. Relational learning and network modelling using infinite latent attribute models. *IEEE TPAMI Special Issue on Bayesian Nonparametrics*.
- 2014 Xin Li, Alexis Battle, Konrad J. Karczewski, Zach Zappala, **David A. Knowles**, Kevin S. Smith, Kim R. Kukurba, Eric Wu, Noah Simon, Stephen B. Montgomery. Transcriptome Sequencing of a Large Human Family Identifies the Impact of Rare Noncoding Variants. In *American Journal of Human Genetics*.
- 2014 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, Stephen B. Montgomery (2014). Allelic Expression of Deleterious Protein-Coding Variants across Human Tissues. *PLoS Genetics*.
- 2013 Tim Salimans, **David A. Knowles**. Fixed-Form Variational Posterior Approximation through Stochas-

tic Linear Regression. *Bayesian Analysis* 8(4) [Winner of the 2014 Lindley Prize].

- 2013 Daniel Glass, Ana Vinuela, Mathew 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, Paola Di Meglio, Frank O Nestle, Mina Ryten, The UK Brain Expression consortium, Muther Consortium, Richard Durbin, Mark I McCarthy, Panagiotis Deloukas, Emmanouil T Dermitzakis, Mike E Weale, Veronique Bataille & Tim D Spector. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biology*.
- 2012 Elin Grundberg, Kerrin S Small, Åsa 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, Sue Ring, Karen Ho, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteindottir, Chrysanthi Ainali, Antigone S Dimas, Neelam Hassanali, Catherine Ingle, **David A. Knowles**, 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, Kouros R Ahmadi, Eric E Schadt, Kari Stefansson, George Davey Smith, Mark I McCarthy, Panos Deloukas, Emmanouil T Dermitzakis, Tim D Spector & The Multiple Tissue Human Expression Resource (MuTHER) Consortium. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*.
- 2011 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, Roger Foo. Distinct epigenomic features in human cardiomyopathy. *Circulation, American Heart Association*.
- 2011 Cornelia Schone, Anne Venner, **David A. Knowles**, Mahesh M Karnani, Denis Burdakov. Dichotomous cellular properties of mouse orexin/hypocretin neurons. In *The Journal of Physiology*.
- 2011 **David A. Knowles** and Zoubin Ghahramani. Nonparametric Bayesian Sparse Factor Models with application to Gene Expression modelling. In *Annals of Applied Statistics*.
- 2010 Daniel Glass, Leopold Parts, **David A. Knowles**, Abraham Aviv, and Tim D. Spector. No Correlation Between Childhood Maltreatment and Telomere Length. In *Biological Psychiatry*.

PEER—REVIEWED CONFERENCE PAPERS

- 2017 Konstantina Palla, **David A. Knowles** and Zoubin Ghahramani. A birth-death process for feature allocation. *Proceedings of the 34th International Conference on Machine Learning (ICML)*.
- 2015 Amar Shah, **David A. Knowles**, Zoubin Ghahramani. Stochastic Variational Inference Algorithms for the Beta Bernoulli Process. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*.
- 2015 Kien Nguyen, **David A. Knowles**, Joerg Bredno. Nuclei classification in histology images: use of contextual information. *International Symposium on Biomedical Imaging (ISBI)*.
- 2014 Konstantina Palla, **David A. Knowles** and Zoubin Ghahramani. A reversible infinite HMM using normalised random measures. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*.
- 2014 Creighton Heaukulani, **David A. Knowles** and Zoubin Ghahramani. Beta Diffusion Trees. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*.
- 2013 Novi Quadrianto, Viktoriia Sharmanska, **David A. Knowles**, Zoubin Ghahramani. The Supervised IBP: Neighbourhood Preserving Infinite Latent Feature Models. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013)*.

- 2012 **David A. Knowles***, Konstantina Palla* and Zoubin Ghahramani. A nonparametric variable clustering model. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*.
- 2012 Konstantina Palla, **David A. Knowles** and Zoubin Ghahramani. An Infinite Latent Attribute Model for Network Data. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*.
- 2012 Andrew Wilson, **David A. Knowles** and Zoubin Ghahramani. Gaussian Process Regression Networks. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*.
- 2011 **David A. Knowles** and Thomas P. Minka. Non-conjugate Variational Message Passing for Multinomial and Binary Regression. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*.
- 2011 **David A. Knowles** and Zoubin Ghahramani. Pitman-Yor Diffusion Trees. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*.
- 2011 **David A. Knowles**, Jurgen Van Gael, and Zoubin Ghahramani. Message Passing Algorithms for the Dirichlet Diffusion Tree. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*.
- 2009 Finale Doshi*, **David A. Knowles***, Shakir Mohamed* and Zoubin Ghahramani. Large Scale Non-parametric Inference: Data Parallelisation in the Indian Buffet Process. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*.
- 2007 **David A. Knowles** and Zoubin Ghahramani. Infinite Sparse Factor Analysis and Infinite Independent Components Analysis. In *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation (ICA 2007)*.

WORKSHOP ABSTRACTS AND PAPERS

- 2016 **David A. Knowles**, Yang Li, Jonathan Pritchard. LeafCutter: Annotation-free quantification and prediction of RNA splicing. *American Society of Human Genetics 66th Annual Meeting*.
- 2015 Yang Li, Bryce van de Geijn, Allegra Petti, Anil Raj, **David A. Knowles**, John Blischak, Yoav Gilad, Jonathan Pritchard. The effects of human genetic variation on the gene regulatory cascade. *American Society of Human Genetics 65th Annual Meeting*.
- 2015 **David A. Knowles**, Joe R. Davis, Stephen B. Montgomery, Alexis Battle. Detecting gene-by-environment interactions using allele specific expression. *The Biology of Genomes Meeting (CSHL)*.
- 2014 Emily K. Tsang, Xin Li, Vanessa Anaya, Konrad J. Karczewski, **David A. Knowles**, Kevin S. Smith, Stephen B. Montgomery. Dissecting the genetic regulation of exosome RNA cargo in a large family. *American Society of Human Genetics 64th Annual Meeting*.
- 2014 Joe R. Davis, **David A. Knowles**, Stephen B. Montgomery, Alexis Battle. Rare variation and the genomic context of allele-specific expression. *American Society of Human Genetics 64th Annual Meeting*.
- 2013 **David A. Knowles**, Alexis Battle, Daphne Koller. Discovering latent cancer characteristics predictive of drug sensitivity. *RECOMB/ISCB Conference on Regulatory & Systems Genomics (selected for oral presentation)*.
- 2013 Alexis Battle*, **David A. Knowles***, Sara Mostafavi, Xiaowei Zhu, James B. Potash, Myrna M. Weissman, Courtney McCormick, Christian D. Haudenschield, Kenneth B. Beckman, Jianxin Shi, Rui Mei, Alexander E. Urban, Douglas F. Levinson, Daphne Koller, Stephen B. Montgomery. The relationship between common environmental and genetic effects on human gene splicing and expression. *American Society of Human Genetics 63rd Annual Meeting*.

*These authors contributed equally.

- 2011 **David A. Knowles**, Leopold Parts, Daniel Glass and John M. Winn. Inferring an individual's "physiological" age from multiple ageing-related phenotypes. *Cambridge Statistics Initiative Special One-Day Meeting*.
- 2011 **David A. Knowles**, Jurgen Van Gael and Zoubin Ghahramani. Message Passing Algorithms for the Dirichlet Diffusion Tree. *Eighth Workshop on Bayesian Nonparametrics*.
- 2010 **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 on Predictive Models in Personalized Medicine*.
- 2009 **David A. Knowles** and Susan Holmes. Statistical tools for ultra-deep pyrosequencing of fast evolving viruses. *NIPS Computational Biology Workshop*.

WORKING PAPERS/UNDER SUBMISSION

- 2017 Diego Calderon, Anand Bhaskar, **David A. Knowles**, David Golan, Towfique Raj, Audrey Fu, Jonathan K. Pritchard. Inferring Relevant Cell Types For Complex Traits Using Single-Cell Gene Expression.
<http://biorxiv.org/content/early/2017/05/10/136283>.
- 2017 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 CJ 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, Francis S Collins. Interactions between genetic variation and cellular environment in skeletal muscle gene expression.
<http://biorxiv.org/content/early/2017/02/03/105429>.
- 2016 Yang I Li*, **David A Knowles***, Jonathan K Pritchard. LeafCutter: Annotation-free quantification of RNA splicing.
<http://biorxiv.org/content/early/2016/03/16/044107>.
- 2015 David A. Knowles. Stochastic gradient variational Bayes for gamma approximating distributions.
<http://arxiv.org/abs/1509.01631>
- 2014 Tim Salimans, **David A. Knowles**. On Using Control Variates with Stochastic Approximation for Variational Bayes and its Connection to Stochastic Linear Regression.
<http://arxiv.org/abs/1401.1022> [stat.ML]
- 2013 **David A. Knowles***, Konstantina Palla* and Zoubin Ghahramani. A dependent partition-valued process for multitask clustering and time evolving network modelling.
<http://arxiv.org/abs/1303.3265> [stat.ML]

Last updated: August 29, 2017 • Typeset in X_YTeX

<http://cs.stanford.edu/~davidknowles/>