Simon TAVARÉ

BIOGRAPHICAL

Citizenship: England. Permanent Resident of USA.
email `st3193@columbia.edu`
website `https://tavarelab.cancerdynamics.columbia.edu`

Current Research Interests

Statistics: stochastic computation, data science, computational biology, statistical bioinformatics. *Cancer:* computational cancer genomics, evolutionary approaches to cancer, molecular annotation of tumors at single cell resolution in 3D. *Population and evolutionary genetics:* coalescent theory, statistical inference in molecular biology, human genetics, molecular evolution, paleontology. *Imaging:* microscopy, tracking from videos. *Probability:* probabilistic combinatorics.

Education

University of Sheffield, UK	Probability and Statistics	BSc 1974
University of Sheffield, UK	Probability and Statistics	MSc 1975
University of Sheffield, UK	Probability and Statistics	PhD 1979

Some results for Markov processes with applications to genetic models
Supervisor: Professor C. Cannings

Professional Experience

Columbia University:

2020 – present	Member, Columbia Data & Society Council
2018 – present	Founding Director, Herbert and Florence Irving Institute for Cancer Dynamics
2018 – present	Professor, Departments of Statistics and Biological Sciences
2018 – present	Full Member, Cancer Genomics and Epigenomics Program, Herbert Irving Comprehensive Cancer Center
2018 – present	Core Member, Data Science Institute
2018 – present	Member, Program for Mathematical Genomics
2018 – present	Senior Associate Core Member, New York Genome Center

University of Cambridge:

2019 – 2023	Director of Research, Cancer Research UK Cambridge Institute
2013 – 2018	Director, CRUK Cambridge Institute
2006 – 2019	Senior Group Leader, CRUK Cambridge Institute
2003 – 2019	Professor of Cancer Research, Department of Oncology
2003 – 2019	Professorial Fellow, Christ's College
2003 – 2019	Professor, Department of Applied Mathematics and Theoretical Physics

University of Southern California:

2006 – 2014	Research Professor, Department of Biological Sciences
1989 – 2006	Professor, Department of Biological Sciences
1999 – 2006	Professor, Department of Preventive Medicine, Keck School of Medicine
1989 – 2006	Professor, Department of Mathematics

Earlier:

1984 – 1989	Associate Professor of Mathematics, University of Utah
1983 – 1985	Associate Professor of Statistics, Colorado State University
1981 – 1983	Assistant Professor of Statistics, Colorado State University
1978 – 1981	Instructor of Mathematics, University of Utah
1978 – 1979	Research Fellow, Medical Biophysics and Computing, University of Utah
1977 – 1978	Lecturer in Probability and Statistics, University of Sheffield.

Visiting Appointments

Senior Scientist, New York Genome Center. 4/2016–9/2018
Adjunct Research Professor, Department of Biological Sciences, USC. 2014–2018
Professor-at-Large, Keck Graduate Institute of Applied Life Sciences, 2000–2006
SAMSI, North Carolina. 9-10/2002
Université Blaise Pascal, Clermont-Ferrand, France. 7/2001
Institute for Pure and Applied Mathematics, UCLA. 9–12/2000
Isaac Newton Institute, University of Cambridge, England. 9/1998
Department of Mathematical Statistics, Chalmers University of Technology, Sweden. 5/1998
Department of Applied Mathematics, University of Zürich, Switzerland.
  5–6/1997, 4–7/1996 (Gastprofessor), 8/1993, 12/1992, 5–6/1992.
Department of Mathematics, Monash University, Australia.
  7/1997, 7/1994, 3/1993, 4–5/1990, 7–12/1980.
Department of Mathematics, Queen Mary and Westfield College, London. 8/1989, 9–12/1988
Department of Statistical Science, University College, London. 7–8/1988
Department of Mathematics, University of Southern California. 8–12/1987, 10–11/1984.
Mathematics Department, Stanford University. 6–8/1982, 6–8/1981, 6–8/1979.

Honors and Awards

2023	DSc *honoris causa*, University of St Andrews
2020 – present	Elected as Honorary Fellow, Christ's College, Cambridge
2018 – present	Herbert and Florence Irving Director, Irving Institute for Cancer Dynamics
2018	Professor Emeritus of Quantitative and Computational Biology, USC
2018	Elected as International Member, National Academy of Sciences
2017	Elected Fellow, American Mathematical Society (Class of 2018)
2017	DSc *honoris causa*, University of Sheffield
2015 – 2017	President, London Mathematical Society (2014–2015 President-Designate)
2015	Elected as Member of EMBO
2015	Invited Lecture, 8th International Congress on Industrial and Applied Mathematics (ICIAM)
2015	American Mathematical Society Einstein Public Lecture in Mathematics

2014	Doctor of Technology *honoris causa*, Chalmers University of Technology, Sweden.
2011	Elected as Fellow, Royal Society of London (FRS)
2010	Jubilee Professor, Chalmers University of Technology, Sweden
2010	Fellow by nomination, Institute of Mathematics and its Applications (FIMA)
2009	Fellow by nomination, Royal Society of Biology (FRSB)
2009	Elected as Fellow, Academy of Medical Sciences (FMedSci)
2004	Elected as Fellow, American Statistical Association
2003 – 2009	Royal Society/Wolfson Research Merit Award
2003	G. C. Steward Visiting Fellowship in Mathematics, Gonville and Caius College, Cambridge (Jan–July)
2001	Special Invited Paper, Institute of Mathematical Statistics
1998 – 2014	George and Louise Kawamoto Chair in Biological Sciences, USC
1998	Elected as Fellow, American Association for the Advancement of Science
1993	Special Invited Paper, Institute of Mathematical Statistics
1992	Elected as Fellow, Institute of Mathematical Statistics

Cambridge University Graduate Programmes

2004 – 2013	Director, MPhil, DAMTP *Computational and Systems Biology*
2010 – 2017	Director, Wellcome Trust 4-year PhD, DAMTP *Mathematical Genomics and Medicine*

PUBLICATIONS

Submitted Articles

† denotes equivalent contributions, †† denotes joint senior authors.

1. da Silva PH, Jamshidpey A & Tavaré S. Markov chains arising from biased random derangements. arXiv version: 2211.13831, 2022.

Books

1. Deonier RC, Tavaré S & Waterman MS. *Computational Genome Analysis: An Introduction*, 2005. Springer Verlag, New York. 540pp. ISBN: 0-387-98785-1

2. Arratia R, Barbour AD & Tavaré S. *Logarithmic Combinatorial Structures: A Probabilistic Approach.* EMS Monographs in Mathematics, **1**, 2003. European Mathematical Society Publishing House, Zuerich. 352pp.

3. Donnelly P & Tavaré S. (eds) *Progress in Population Genetics and Human Evolution.* IMA Volumes in Mathematics and its Applications, **87**, 1997. Springer Verlag, Berlin. 330pp.

Journal Articles

All articles refereed, ** denotes review

Unsubmitted papers on arXivs

4. Dinh KN, Vázquez-García I, Chan A, Malhotra R, Weiner A, McPherson A & Tavaré S. CINner: modeling and simulation of chromosomal instability in cancer at single-cell resolution. biorXiv, `10.1101/2024.04.03.587939`, 2024.

In press

5. Ng AWT, McClurg DP, Wesley B, Zamani SA, Miremadi A, Giger O, ten Hoopen R, Devonshire G, Redmond AM, Grehan N, Blasko AG, Li X, Aparicio S, Tavaré S, Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium, Nowicki-Osuch K & Fitzgerald RC. Disentangling oncogenic amplicons in esophageal adenocarcinoma. *Nature Communications*, in press, 2024.

2024

6. da Silva PH, Jamshidpey A & Tavaré S. Another view of sequential sampling in the birth process with immigration. *Journal of Mathematical Biology*, **88**, art. 27, 2024. `https://doi.org/10.1007/s00285-023-02041-0`. arXiv version: 2210.07307, 2022.

2023

7. Melamed D, Choi A, Reilein A, Tavaré S & Kalderon D. Spatial regulation of Drosophila ovarian Follicle Stem Cell division rates and cell cycle transitions. *PLoS Genetics*, **19**, e1010965, 2023.

8. Kunes RZ, Yin M, Land M, Haviv D, Pe'er D & Tavaré S. Gradient estimation for binary latent variables via gradient variance clipping. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37(7)**, 8405–8412, 2023.

9. González-Solares EA, Dariush A, González-Fernández C, Küpcü Yoldaş A, Molaeinezhad A, Al Sa'd M, Smith L, Whitmarsh T, Millar N, Chornay N, Falciatori I, Fatemi A, Goodwin D, Kuett L, Mulvey CM, Páez Ribes M, Qosaj F, Roth A, Vázquez-García I, Watson S, Windhager J, Aparicio S, Bodenmiller B, Boyden E, Caldas C, Harris O, Shah SP, Tavaré S, CRUK IMAXT Grand Challenge Team, Bressan D, Hannon G & Walton NA. The Imaging and Molecular Annotation of Xenografts and Tumours (IMAXT) high throughput data and analysis infrastructure. *Biological Imaging*, **3**, e11, 2023.

10. Nowicki-Osuch K, Zhuang L, Cheung TS, Black EL, Masqué-Soler N, Devonshire G, Redmond AM, Freeman A, di Pietro M, Pilonis N, Januszewicz W, O'Donovan M, Tavaré S, Shields J & Fitzgerald RC. Single-cell RNA sequencing unifies developmental programs of Esophageal and Gastric Intestinal Metaplasia. *Cancer Discovery*, **13**, 1346–1363, 2023.

11. da Silva PH, Jamshidpey A, McCullagh P & Tavaré S. Fisher's measure of variability in repeated samples. *Bernoulli*, **29**, 1166–1194, 2023.

2022

12. da Silva PH, Jamshidpey A & Tavaré S. The Feller Coupling for random derangements. *Stochastic Processes and their Applications*, **150**, 1139–1164, 2022.

13. Ng AWT, Contino G, Killcoyne S, Devonshire G, Hsu R, Abbas S, Su J, Redmond A, Weaver JMJ, Eldridge MD, Tavaré S, Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium, Edwards PAW & Fitzgerald RC. Rearrangement processes and structural variations show evidence of selection in oesophageal adenocarcinomas. *Communications Biology*, **5**, art 335, 2022.

14. Tavaré S. A note on the Screaming Toes game. *J. Appl. Probab.*, **59**, 118–130, 2022. [Original version: *arXiv*, 2006.04805, 2020].

2021

15. Tavaré S. The magical Ewens Sampling Formula. *Bulletin of the London Mathematical Society*, **53**, 1563–1582, 2021.

16. Williams EH, Flint TR, Connell CM, Giglio D, Lee H, Ha T, Gablenz E, Bird N, Weaver JMJ, Potts H, Whitley CT, Bookman MA, Lynch AG, Meyer HV, Tavaré S & Janowitz T. CamGFR v2: a new model for estimating the glomerular filtration rate from standardized or non-standardized creatinine in patients with cancer. *Clinical Cancer Research*, **27**, 1381–1390, 2021.

17. Touloumis A, Marioni JC & Tavaré S. Hypothesis testing for the covariance matrix in high-dimensional transposable data with Kronecker product dependence structure. *Statistica Sinica*, **31**, 1309–1329, 2021. [Original version: *arXiv*:1404.7684v2, 2014.]

2020

18. Biasci D, Thaventhiran J† & Tavaré S†. Fibroblastic reticular cells predict response to immune checkpoint inhibitors. *biorXiv*, `doi.org/10.1101/2020.02.19.955666`, 2020.

19. Perner J, Abbas S, Nowicki-Osuch K, Devonshire G, Eldridge MD, Tavaré S & Fitzgerald R. The mutREAD method detects mutational signatures from low quantities of cancer DNA. *Nature Communications*, **11**, art. 3166, 2020.

20. Jammula S, Katz-Summercorn AC, Li X, Linossi C, Smyth E, Killcoyne S, Biasci D, Subash VV, Abbas S, Blasko A, Devonshire G, Grantham A, Wronowski F, O'Donovan M, Grehan N, Eldridge M, Tavaré S, the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium & Fitzgerald RC. Identification of subtypes of Barrett's esophagus and esophageal adenocarcinoma based on DNA methylation profiles and integration of transcriptome and genome data. *Gastroenterology*, **158**, 1682–1697, 2020.

21. Dinh KN, Jaksik R, Kimmel M, Lambert A & Tavaré S. Statistical inference for the evolutionary history of cancer genomes. *Statistical Science*, **35**, 129–144, 2020.

2019

22. Williams EH, Connell CM, Weaver JMJ, Beh I, Potts H, Whitley CT, Bird N, Al-Sayed T, Fehr M, Cathomas R, Bertelli G, Quinton A, Lewis P, Shamash J, Wilson P, Dooley M, Poole S, Mark PB, Bookman MA, Earl H, Jodrell D, Tavaré S, Lynch AG & Janowitz T. Multicentre validation of the CamGFR model for estimated glomerular filtration rate. *JNCI Cancer Spectrum*, **3**, pkz068, 2019.

23. Spencer S, Bal SK, Egner W, Allen HL, Raza SI, Ma CA, Gürel M, Zhang Y, Sun G, Sabroe RA, Greene D, Rae W, Shahin T, Kania K, Ardy RC, Thian M, Staples E, Pecchia-Bekkum A, Worrall WPM, Stephens J, Brown M, Tuna S, York M, Shackley F, Kerrin D, Sargur R, Condliffe A, Tipu HN, Kuehn HS, Rosenzweig SD, Turro E, Tavaré S, Thrasher AJ, Jodrell DI, Smith KGC, Boztug K, Milner JD & Thaventhiran JED. Loss of the interleukin-6 receptor causes immunodeficiency, atopy, and abnormal inflammatory responses. *Journal of Experimental Medicine*, **216**, 1986–1998, 2019.

24. Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, Abbas S, Perner J, Bower L, Devonshire G, Ococks E, Grehan N, Mok J, O'Donovan M, MacRae S, Eldridge MD, Tavaré S, Fitzgerald RC & the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium. The landscape of selection in 551 Esophageal Adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet*, **51**, 506–516, 2019.

25. Tavaré S. The linear birth-death process: an inferential retrospective. In *Branching and Applied Probability. Papers in honour of Peter Jagers.* Editors Asmussen S, Klebaner F, Nerman O & Vatutin V. *Adv Appl Prob*, **50A**, 253–269, 2019.

2018

26. Griffiths RC & Tavaré S. Ancestral inference from haplotypes and mutations. *Theoret Popul Bio*, **122**, 12–21, 2018.

27. Arratia R, Barbour AD, Ewens WJ & Tavaré S. Simulating the component counts of combinatorial structures. *Theoret Popul Biol*, **122**, 5–11, 2018.

28. Nicholson AM, Olpe C, Hoyle A, Thorsen A-S, Rus T, Colombé M, Brunton-Sim R, Kemp R, Marks K, Quirke P, Malhotra S, ten Hoopen R, Ibrahim A, Lindskog C, Myers MB, Parsons B, Tavaré S, Wilkinson M, Morrissey E & Winton DJ. Fixation and spread of somatic clones in adult human colonic epithelium. *Cell Stem Cell*, **22**, 909–918, 2018.

29. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, Zamora J, Butler A, Whitaker H, Kote-Jarai Z, Alexandrov LB, Van Loo P, Massie CE, Dentro S, Warren AY, Verrill C, Berney DM, Dennis N, Merson S, Hawkins S, Howat W, Lu YJ, Lambert A, Kay J, Kremeyer B, Karaszi K, Luxton HJ, Camacho N, Marsden L, Edwards S, Matthews L, Bo V, Leongamornlert D, McLaren S, Ng A, Yu Y, Zhang H, Dadaev T, Thomas S, Easton D, Ahmed M, Bancroft E, Fisher C, Livni N, Nicol D, Tavaré S, Gill P, Greenman C, Khoo V, Van As N, Kumar P, Ogden C, Cahill D, Thompson A, Mayer E, Rowe E, Dudderidge T, Gnanapragasam VJ, Shah N, Raine K, Jones D, Menzies A, Stebbings L, Teague J, Hazell S, Corbishley C, CAMCAP study group, de Bono J, Attard G, Isaacs W, Visakorpi T, Fraser M, Boutros PC, Bristow RG, Workman P, Sander C, The TCGA consortium, Hamdy FC, Futreal A, McDermott U, Al-Lazikani B, Lynch A, Bova GS, Foster CS, Brewer DS, Neal DE,

Cooper C & Eeles R. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet*, **50**, 682–692, 2018.

30. Reilein A, Melamed D, Tavaré S & Kalderon D. Division-independent differentiation mandates proliferative competition among stem cells. *Proc Natl Acad Sci USA*, **115**, E3182–E3191, 2018.

2017

31. Asim M, Tarish F, Zecchini HI, Sanjiv K, Gelali E, Massie CE, Baridi A, Warren AY, Zhao W, Ogris C, McDuffus L-A, Mascalchi P, Shaw G, Dev H, Wadhwa K, Wijnhoven P, Forment JV, Lyons SR, Lynch AG, O'Neill C, Zecchini VR, Rennie PS, Baniahmad A, Tavaré S, Mills IG, Galanty Y, Crosetto N, Schultz N, Neal D, Helleday T. Synthetic lethality between Androgen receptor signaling and the PARP pathway in prostate cancer. *Nat Comm*, **8**, art 374, 2017.

32. Janowitz T[†], Williams EH[†], Marshall A, Ainsworth N, Thomas PB, Sammut SJ, Shepherd S, White J, Mark PB, Lynch AG, Jodrell D, Tavaré S & Earl H. A new model for estimating glomerular filtration rate in patients with cancer. *J Clinical Oncology*, **35**, 2798–2805, 2017.

33. Noorani A[†], Bornschein J[†], Lynch AG[†], Secrier M, Achilleos A, Eldridge M, Bower L, Weaver JMJ, Crawte J, Ong CAJ, Shannon N, MacRae S, Grehan N, Nutzinger B, O'Donovan M, Hardwick R, Tavaré S & Fitzgerald RC, on behalf of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium. A comparative analysis of whole genome sequencing of oesophageal adenocarcinoma pre- and post-chemotherapy. *Genome Research*, **27**, 902–912, 2017.

2016

34. Secrier M, Li X, de Silva N, Eldridge MD, Contino G, Bornschein J, MacRae S, Grehan N, O'Donovan M, Miremadi A, Yang T-P, Bower L, Chettouh H, Crawte J, Galeano-Dalmau N, Grabowska A, Saunders J, Underwood T, Waddell N, Barbour AP, Nutzinger B, Achilleos A, Edwards PAW, Lynch AG, Simon Tavaré & Fitzgerald RC on behalf of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium. Mutational signatures in esophageal adenocarcinoma reveal etiologically distinct subgroups with therapeutic relevance. *Nature Genetics*, **48**, 1131–1141, 2016.

35. Touloumis A, Marioni JC & Tavaré S. HDTD: Analyzing multi-tissue gene expression data. *Bioinformatics*, **32**, 2193–2195, 2016.

36. Arratia R, Barbour AD & Tavaré S. Exploiting the Feller coupling for the Ewens sampling formula. *Statistical Science*, **31**, 27–29, 2016.

37. Asim M, Massie CE, Orafidiya F, Pértega-Gomes N, Warren AY, Esmaeili M, Selth LA, Zecchini HI, Luko K, Qureshi A, Baridi A, Menon S, Madhu B, Escriu C, Lyons S, Vowler SL, Zecchini VR, Shaw G, Hessenkemper W, Russell R, Mohammed H, Stefanos N, Lynch AG, Grigorenko E, D'Santos C, Taylor C, Lamb A, Sriranjan R, Yang J, Stark R, Dehm SM, Rennie PS, Carroll JS, Griffiths JR, Tavaré S, Mills IG, McEwan IJ, Baniahmad A, Tilley WD & Neal DE. Choline kinase alpha as an androgen receptor chaperone and prostate cancer therapeutic target. *Journal of the National Cancer Institute*, **108**, djv371, 2016.

2015

38. Madhu B, Narita M, Jauhiainen A, Menon S, Stubbs M, Tavaré S, Narita M & Griffiths JR. Metabolomic changes during cellular transformation monitored by metabolite-metabolite correlation analysis and correlated gene expression. *Metabolomics*, **11**, 1848–1863, 2015.

39. Paterson AL, Weaver JMJ, Eldridge MD, Tavaré S, Fitzgerald RC & Edwards PAW, on behalf of the OCCAMs Consortium. Mobile element insertions are frequent in oesophageal adenocarcinomas and can mislead paired-end sequencing analysis. *BMC Genomics*, **16**, 473, 2015.

40. Greenwood AK, Ardekani R, McCann SR, Dubin ME, Sullivan A, Bensussen S, Tavaré S & Peichel CL. Genetic mapping of natural variation in schooling tendency in the threespine stickleback. *G3*, **5**, 761–769, 2015.

41. Touloumis A, Tavaré S & Marioni JC. Testing the mean matrix in high-dimensional transposable data. *Biometrics*, **71**, 157–166, 2015.

42. Kirschner K[†], Samarajiwa SA[†], Cairns JM, Menon S, Pérez-Mancera PA, Tomimatsu K, Bermejo-Rodriguez C, Ito Y, Chandra T, Narita M, Lyons SK, Lynch AG, Kimura H, Ohbayashi T, Tavaré S & Narita M. Phenotype specific analyses reveal distinct regulatory mechanism for chronically activated p53. *PLoS Genetics,* **11**, e1005053, 2015.

43. Josephidou M, Lynch AG & Tavaré S. multiSNV: a probabilistic approach for improving detection of somatic point mutations from multiple related tumour samples. *Nucleic Acids Research*, **43**, e61, 2015.

44. Gubernator M, Slater SC, Spencer HL, Spiteri I, Sottoriva A, Riu F, Rowlinson J, Avolio E, Katare R, Mangialardi G, Oikawa A, Reni C, Campagnolo P, Spinetti G, Touloumis A, Tavaré S, Prandi F, Pesce M, Hofner M, Klemens V, Emanueli C, Angelini G & Madeddu P. The epigenetic profile of human adventitial progenitor cells correlates with therapeutic outcomes in a mouse model of limb ischemia. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **35**, 675–688, 2015.

45. Massie CE, Spiteri I, Ross-Adams H, Luxton H, Kay J, Whitaker H, Dunning MJ, Lamb AD, Ramos-Montoya A, Brewer D, Cooper CS, Eeles R, Warren A, Tavaré S, Neal DE & Lynch AG. HES5 silencing is an early and recurrent change in prostate tumourigenesis. *Endocrine-Related Cancer*, **22**, 131–144, 2015.

46. Piccirillo SGM, Spiteri I[†], Sottoriva A[†], Touloumis A, Ber S, Price SJ, Heywood R, Francis N-J, Howarth KD, Collins VP, Venkitaraman AR, Curtis CN, Marioni JC, Tavaré S & Watts C. Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone. *Cancer Research*, **75**, 194–202, 2015.

2014

47. Silva A-L, Dawson SN, Arends MJ, Guttula K, Hall N, Cameron EA, Huang TH-M, Brenton JD, Tavaré S, Bienz M & Ibrahim AEK. Boosting Wnt activity during colorectal cancer progression through selective hypermethylation of Wnt signaling antagonists. *BMC Cancer*, **14**: 891, 2014.

48. Cairns J, Lynch AG & Tavaré S. Quantifying the impact of inter-site heterogeneity on the distribution of ChIP-seq data. *Frontiers in Genetics*, **5**: 399, 2014.

49. Weaver JMJ†, Ross-Innes CS†, Shannon N†, Lynch AG†, Forshew T, Barbera M, Murtaza M, Ong C-AJ, Lao-Sirieix P, Dunning M, Smith L, Smith ML, Anderson CM, Carvalho B, O'Donovan M, Underwood TJ, May AP, Grehan N, Hardwick R, Davies J, Oloumi A, Aparicio S, Caldas C, Eldridge MD, Edwards PAW, Rosenfeld N, Tavaré S, Fitzgerald RC, and the OCCAMS consortium. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nature Genetics*, **46**, 837–843, 2014.

50. Jauhiainen A, Basetti M, Narita M, Narita, M, Griffiths J & Tavaré S. Normalization of metabolomics data with applications to correlation maps. *Bioinformatics*, **30**, 2155–2161, 2014.

51. Ostrow AZ, Nellimoottil T, Knott SRV, Fox CA, Tavaré S & Aparicio OM. Fkh1 and Fkh2 bind multiple chromosomal elements in the *S. cerevisiae* genome with distinct specificities and cell cycle dynamics. *PLoS ONE*, **9**, e87647, 2014.

52. Turro E, Astle WJ & Tavaré S. Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics*, **30**, 180–188, 2014.

2013

53. Ardekani R, Greenwood AK, Peichel CL & Tavaré S. Automated quantification of the schooling behaviour of sticklebacks. *EURASIP Journal on Image and Video Processing*, Art 61, 2013. doi:10.1186/1687-5281-2013-61.

54. Vermeulen L†, Morrissey E†, van der Heijden M, Nicholson AM, Sottoriva A, Buczacki S, Kemp R, Tavaré S & Winton DJ. Defining stem cell dynamics in models of intestinal tumor initiation. *Science*, **342**, 995–998, 2013.

55. Fu QA, Russell S, Bray S & Tavaré S. Bayesian clustering of replicated time-course gene expression data with weak signals. *Annals of Applied Statistics*, **7**, 1334–1361, 2013.

56. ** Ardekani R, Tavaré S & Tower J. Assessing senescence in Drosophila using video tracking. *Methods Mol Bio*, **965**, 501–526, 2013.

57. Kozar SK†, Morrissey E†, Nicholson AM, van der Heijden M, Kemp R, Tavaré S, Vermeulen L& Winton DJ. Continuous clonal labeling defines the functional stem cell compartment in intestinal crypts and adenomas. *Cell Stem Cell*, **13**, 626–633, 2013.

58. Newman S, Howarth KD, Greenman CD, Bignell GR, Tavaré S & Edwards PAW. The relative timing of mutations in a breast cancer genome. *PLoS ONE*, **8**, e64991, 2013.

59. Zhong Y, Nellimoottil T, Peace JM, Knott SRV, Villwock SK, Yee JM, Jancuska JM, Rege S, Tecklenburg M, Sclafani RA, Tavaré S & Aparicio OM. The level of origin firing inversely affects the rate of replication fork progression. *Journal of Cell Biology*, **201**, 373–383, 2013.

60. Sottoriva A†, Spiteri I†, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, Curtis CN, Watts C†† & Tavaré S††. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci USA*, **110**, 4009–4014, 2013.

61. Housden BE, Fu AQ, Krejci A, Bernard F, Fischer B, Tavaré S, Russell S & Bray SJ. Transcriptional dynamics elicited by a short pulse of Notch activation involves feed-forward regulation by E(spl)/Hes genes. *PLoS Genetics*, **9**: e1003162, 2013.

62. Ardekani R, Biyani A, Dalton JE, Saltz JB, Arbeitman MN, Tower J, Nuzhdin S & Tavaré S.Three-dimensional tracking and behaviour monitoring of multiple fruit flies. *Journal of the Royal Society Interface*, **10**:20120547, 2013.

63. Sottoriva A, Spiteri I, Shibata D, Curtis C & Tavaré S. Single molecule genomic data delineate patient-specific tumor profiles and cancer stem cell organization. *Cancer Research*, **73**, 41–49, 2013.

2012

64. Lynch AG, Chin S-F, Dunning MJ, Caldas C, Tavaré S & Curtis C. Calling sample mixups in cancer population studies. *PLoS ONE*, **7**; e41815, 2012.

65. Ardekani R, Huang YM, Sancheti P, Stanciauskas P, Tavaré S & Tower J. Using GFP video to track 3D movement and conditional gene expression in free-moving flies. *PLoS ONE*, **7**, e40506, 2012.

66. Chandra T, Kirschner K, Thuret J-Y, Pope BD, Ryba T, Newman S, Ahmed K, Samarajiwa SA, Salama R, Carroll T, Stark R, Janky R, Narita M, Xue L, Chicas A, Nuñez S, Janknecht R, Hayashi-Takanaka Y, Wilson MD, Marshall A, Odom DT, Babu MM, Bazett-Jones DP, Tavaré S, Edwards PAW, Lowe SW, Kimura H, Gilbert DM & Narita M. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Molecular Cell,* **47**, 203–214, 2012.

67. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale A-L, Brenton J, Tavaré S, Caldas C & Aparicio S. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352, 2012.

68. Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavaré S & Aparicio OM. Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell*, **148**, 99–111, 2012.

2011

69. Manolopoulou I, Legarreta L, Emerson B, Brooks SP & Tavaré S. A Bayesian approach to phylogeographic clustering. *Interface Focus*, **1**, 909–921, 2011.

70. Speed D & Tavaré S. Sparse partitioning: nonlinear regression with binary or tertiary predictors, with application to association studies. *Ann Appl Stats*, **5**, 873–893, 2011.

71. Siegmund KD, Marjoram P, Tavaré S & Shibata D. High DNA methylation pattern intratumoral diversity implies weak selection in many human colorectal cancers. *PLoS ONE* **6**: e21657, 2011.

72. Sottoriva A, Vermeulen L & Tavaré S. Modeling evolutionary dynamics of epigenetic mutations in hierarchically organized tumors. *PLoS Comp Bio*, **7**: e1001132, 2011.

73. Narita M, Young ARJ, Arakawa S, Samarajiwa SA, Nakashima T, Yoshida S, Hong SK, Berry LS, Reichelt S, Ferreira M, Tavaré S, Inoki K, Shimizu S & Narita M. Spatial coupling of mTOR and autophagy augments secretory phenotypes. *Science*, **332**, 966-970, 2011.

74. Ibrahim AEK, Arends MJ, DeSilva AL, Wyllie AH, Gerger L, Ito Y, Vowler SL, Huang TH-M, Tavaré S, Murrell A & Brenton JD. Progressive DNA methylation changes are associated with DNMT3B2 overexpression in colorectal neoplastic development. *Gut*, **60**, 499–508, 2011.

75. Cairns J, Spyrou S, Stark R, Smith ML, Lynch AG & Tavaré S. BayesPeak - An R package for analysing ChIP-seq data. *Bioinformatics*, **27**, 713–714, 2011.

76. Wilkinson RD, Steiper M, Soligo C, Martin R, Yang Z & Tavaré S. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology*, **60**, 16–31, 2011.

2010

77. Ivakhno S, Royce T, Cox AJ, Evers D, Cheetham KR & Tavaré S. CNAseg – a novel framework for identification of copy number changes in cancer from second-generation sequencing data. *Bioinformatics*, **26**, 3051–3058, 2010.

78. Lynch AG, Hadfield J, Dunning MJ, Osborne M, Thorne NP & Tavaré S. The cost of reducing starting RNA quantity for Illumina BeadArrays: A bead-level dilution experiment. *BMC Genomics*, **11**: 540, 2010.

79. Ritchie ME, Forrest MS, Dimas AS, Daelemans C, Dermitzakis ET, Deloukas P & Tavaré S. Data analysis issues for allele-specific expression using Illumina's GoldenGate assay. *BMC Bioinformatics*, **11**: 280, 2010.

80. Ivakhno S & Tavaré S. CNAnova: a new approach for finding recurrent copy number abnormalities in cancer SNP microarray data. *Bioinformatics*, **26**, 1395–1402, 2010.

81. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, Campino S, Clark TG, Stanier P, Kwiatkowski D, Deloukas P, Dermitzakis ET, Tavaré S, Moore GE & Dunham I. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genetics*, **11**: 25, 2010.

82. Smith ML, Dunning MJ, Tavaré S & Lynch AG. Identification and correction of previously unreported spatial phenomena using raw Illumina BeadArray data. *BMC Bioinformatics*, **11:** 208, 2010.

83. Clark AM, Goldstein LD, Tevlin M, Tavaré S, Shaham S & Miska EA. The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Research*, **38**, 3780–3793, 2010.

84. Barbosa-Morais NL, Dunning MJ, Samarajiwa S, Darot JFJ, Ritchie ME, Lynch AG & Tavaré S. A reannotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Research*, **38**: e17, 2010.

2009

85. Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, Chin S-F, Brenton J, Tavaré S & Caldas C. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics*, **10**: 588, 2009.

86. Grover D & Tavaré S. Finding behavioral motifs in fly trajectories. *Communications in Information and Systems*, **9**, 283–294, 2009.

87. Avril-Sassen S, Goldstein L, Stingl J, Blenkiron C, Le Quesne J, Spiteri I, Karagavriilidou K, Watson CJ, Tavaré S, Miska EA & Caldas C. Characterization of microRNA expression in post-natal mouse mammary gland development. *BMC Genomics*, **10**: 548, 2009.

88. Waskar M, Landis GN, Shen J, Curtis CN, Abdueva D, Skvortsov D, Tavaré S & Tower J. *Drosophila melanogaster* p53 has developmental stage-specific and sex-specific effects on adult life span indicative of sexual antagonistic pleiotropy. *Aging*, **1**, 903–936, 2009.

89. Grover D, Ford D, Brown C, Hoe N, Erdem E, Tavaré S & Tower J. Hydrogen peroxide stimulates activity and alters behavior in *Drosophila melanogaster*. *PLoS ONE*, **4**: e7580, 2009.

90. Knott S, Viggiani CJ, Aparicio OM & Tavaré S. Strategies for analyzing highly enriched IP-chip datasets. *BMC Bioinformatics*, **10:** 305, 2009.

91. Spyrou C, Stark R, Lynch AG & Tavaré S. BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, **10:** 299, 2009.

92. Fernando H, Sewitz S, Darot J, Tavaré S, Huppert JL & Balasubramanian S. Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Research*, **37**, 6716–6722, 2009.

93. Ritchie ME, Carvalho BS, Hetrick KN, Tavaré S & Irizarry RA. R/Bioconductor software for Illumina's Infinium whole-genome genotyping BeadChips. *Bioinformatics*, **25**, 2621–2623, 2009.

94. ** Siegmund KD, Marjoram P, Tavaré S & Shibata D. Many colorectal cancers are "flat" clonal expansions. *Cell Cycle*, **8**, 2187–2193, 2009.

95. Tiemann-Boege I, Curtis CN, Shinde DN, Goodman DB, Tavaré S & Arnheim N. Product length, dye choice, and detection chemistry in the bead-emulsion amplification of millions of single DNA molecules in parallel. *Analytical Chemistry*, **81**, 5770–5776, 2009.

96. Knott S, Viggiani CJ, Tavaré S & Aparicio OM. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. *Genes & Dev*, **23**, 1077–1090, 2009.

97. Grover D, Yang J, Ford D, Tavaré S & Tower J. Simultaneous tracking of movement and gene expression in multiple *Drosophila melanogaster* flies using GFP and DsRED fluorescent reporter transgenes. *BMC Research Notes*, **2**: 58, 2009.

98. Wilkinson RD & Tavaré S. Estimating primate divergence times by using conditioned birth-and-death processes. *Theoret Popul Biol*, **75**, 278–285, 2009.

99. Young ARJ, Narita M, Ferreira M, Kirschner K, Sadaie M, Darot JFJ, Tavaré S, Arakawa S, Shimizu S, Watt FM & Narita M. Autophagy mediates the mitotic-senescence transition. *Genes & Dev*, **23**, 798–803, 2009.

100. Siegmund KD, Marjoram P, Woo Y-J, Tavaré S & Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc Natl Acad Sci USA*, **106**, 4828–4833, 2009.

101. Shen J, Curtis C, Tavaré S & Tower J. A screen of apoptosis and senescence regulatory genes for life span effects when over-expressed in *Drosophila*. *Aging*, **1**, 191–211, 2009.

102. Woo Y-J, Siegmund KD, Tavaré S & Shibata D. Older individuals appear to acquire mitotically older colorectal cancers. *J Path*, **217**, 483–488, 2009.

103. Jiang R, Tavaré S & Marjoram P. Population genetic inference from resequencing data. *Genetics*, **181**, 187–197, 2009. Corrigendum: **190**, 829–830, 2012.

2008

104. Grover D, Yang J, Tavaré S & Tower J. Simultaneous tracking of fly movement and expression using GFP. *BMC Biotechnology*, **8**: 93, 2008. (With press release)

105. Dimas AS, Stranger BE, Beazley C, Finn RD, Ingle CE, Forrest MS, Ritchie ME, Deloukas P, Tavaré S & Dermitzakis ET. Modifier effects between regulatory and protein-coding variation. *PLoS Genetics*, **4** (10): e1000244, 2008.

106. Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VLJ, Fisher EMC, Tavaré S & Odom DT. Species-specific transcription in mice carrying human chromosome 21. *Science*, **322**, 434–438, 2008.

107. Hull ML, Rangel Escareno C, Godsland J, Doig J, Johnson C, Butt RP, Phillips S, Smith SK, Tavaré S, Print CG & Charnock-Jones DS. Endometrial-peritoneal interactions during endometriotic lesion establishment. *Am J Path*, **173**, 700–715, 2008.

108. Marioni JC, White M, Tavaré S & Lynch AG. Hidden copy number variation in the HapMap trios. *Proc Natl Acad Sci USA*, **105**, 10067–10072, 2008.

109. Szyjka SJ, Aparicio JG, Viggiani CJ, Knott S, Xu W, Tavaré S & Aparicio OM. Rad53 regulates replication fork restart after DNA damage in *Saccharomyces cerevisiae*. *Genes & Dev*, **22**, 1902–1920, 2008.

110. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, Tomazou EM, Bäckdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJP, Carter NP, Tavaré S & Beck S. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*, **18**, 1518–1829, 2008.

111. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, Matthews N, Berezikov E, Ketting RF, Tavaré S & Miska EA. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Molecular Cell*, **31**, 79–90, 2008.

112. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Bäckdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJP, Durbin R, Tavaré S & Beck S. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, **26**, 779–785, 2008.

113. Dunning MJ, Ritchie ME, Barbosa-Morais NL, Tavaré S & Lynch AG. Spike-in validation of an Illumina-specific variance-stabilizing transformation. *BMC Research Notes*, **1**: 18, 2008.

114. Palmer RD, Barbosa-Morais NL, Gooding EL, Muralidhar B, Thornton CM, Pett MR, Roberts I, Schneider DT, Thorne N, Tavaré S, Nicholson JC & Coleman N; On behalf of The Children's Cancer and Leukaemia Group (CCLG). Pediatric malignant germ cell tumours show characteristic transcriptome profiles. *Cancer Research*, **68**: 4239–4247, 2008.

115. Chu M, Siegmund KD, Hao Q-L, Crooks GM, Tavaré S & Shibata D. Inferring relative numbers of human leukocyte genome replications. *British Journal of Haematology*, **141**: 862–871, 2008.

116. Grover D, Tower J & Tavaré S. O fly, where art thou? *J R Soc Interface*, **5**: 1181–1191, 2008.

117. Dunning MJ, Barbosa-Morais NL, Lynch AG, Tavaré S & Ritchie ME. Statistical issues in the analysis of Illumina data. *BMC Bioinformatics*, **9**: 85, 2008.

2007

118. Curtis C, Landis GN, Folk D, Wehr NB, Hoe N, Waskar M, Abdueva D, Skvortsov D, Ford D, Luu A, Badrinath A, Levine RL, Bradley TJ, Tavaré S & Tower J. Transcriptional profiling of MnSOD-mediated lifespan extension in Drosophila reveals a species-general network of aging and metabolic genes. *Genome Biology*, **8(12)**: R262, 2007.

119. Marioni JC, Thorne NP, Valsesia A, Fitzgerald T, Redon R, Fiegler H, Andrews DT, Stranger BE, Lynch AG, Dermitzakis ET, Carter NP, Tavaré S & Hurles ME. Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biology*, **8(10)**: R228, 2007.

120. Conboy CM, Spyrou C, Thorne NP, Wade EJ, Barbosa-Morais N, Bhattacharjee A, Young RA, Tavaré S, Lees JA & Odom DT. Cell cycle genes are the evolutionarily conserved targets of mammalian DREAM member E2F4. *PLoS ONE*, **2**: e1061, 2007.

121. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa J, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavaré S, Brenton JD, Ylstra B & Caldas C. High-resolution array-CGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biology*, **8 (10)**: R215, 2007.

122. Blenkiron C, Goldstein LD, Thorne NP, Spiteri MI, Chin SF, Dunning M, Barbosa-Morais NL, Tavaré S, Caldas C & Miska EA. MicroRNA expression profiling of human breast cancer identifies new markers of tumour subtype. *Genome Biology*, **8 (10)**: R214, 2007.

123. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle C, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P & Dermitzakis ET. Population genomics of human gene expression. *Nat Genet*, **39**, 1217–1224, 2007.

124. ** Shibata D & Tavaré S. Stem cell chronicles: Autobiographies within genomes. *Stem Cell Reviews*, **3**, 94–103, 2007.

125. Martin RD, Soligo C & Tavaré S. Primate origins: Implications of a Cretaceous ancestry. *Folia Primatol*, **78**, 277–296, 2007.

126. Chu MW, Siegmund KD, Eckstam CL, Kim JY, Yang AS, Kanel GC, Tavaré S & Shibata D. Lack of increases in methylation at three CpG-rich genomic loci in non-mitotic adult tissues during aging. *BMC Medical Genetics*, **8**:50, 2007.

127. Dunning MJ, Smith ML, Ritchie ME & Tavaré S. *beadarray*: R classes and methods for Illumina bead-based data. *Bioinformatics*, **23**, 2183–2184, 2007.

128. Skvortsov D, Abdueva D, Curtis CN, Schaub B & Tavaré S. Explaining differences in saturation levels for Affymetrix GeneChip arrays. *Nucleic Acids Research*, **35**, 4154–4163, 2007.

129. Skvortsov D, Abdueva D, Stitzer M, Finkel SA & Tavaré S. Using expression arrays for copy number detection: an example from *E. coli*. *BMC Bioinformatics*, **8**:203, 2007.

130. Chin S-F, Wang Y, Thorne NP, Teschendorff AE, Pinder SE, Vias M, Naderi A, Roberts I, Barbosa-Morais NL, Garcia M, Iyer NG, Kranjac T, Robertson J, Ruffalo T, Aparicio SFR, Tavaré S, Ellis I, Brenton JD, Caldas C. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene*, **26**, 1959–1970, 2007.

131. Nicolas P, Kim K-M, Shibata D & Tavaré S. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Computational Biology*, **3**, 364–374, 2007.

132. Stranger BE, Forrest MS, Dunning M, Thorne NP, Ingle CE, Beazley C, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME & Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853, 2007.

2006

133. ** Dunning M, Smith M, Thorne NP, Tavaré S. beadarray: an R package to analyse Illumina BeadArrays. *R News*, **6**(5), 17–23, 2006.

134. Xu W, Aparicio JG, Aparicio OM & Tavaré S. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and precise identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics*, **7**: 276, 2006.

135. Ibrahim AEK, Thorne NP, Baird K, Barbosa-Morais NL, Tavaré S, Collins VP, Wyllie AH, Arends MJ & Brenton JD. MMASS: an optimised array-based method for assessing CpG island methylation. *Nucleic Acids Research*, **34**: e136, 2006.

136. Abdueva D, Skvortsov D & Tavaré S. Non-linear analysis of GeneChip arrays. *Nucleic Acids Research*, **34**, e105, 2006.

137. ** Marjoram P & Tavaré S. Modern computational approaches for analysing molecular-genetic variation data. *Nat Rev Genet*, **7**, 759–770, 2006.

138. Jiang R, Marjoram P, Borevitz JO & Tavaré S. Inferring population parameters from single-feature polymorphism data. *Genetics*, **173**, 2257–2267, 2006.

139. ** Shibata D & Tavaré S. Counting divisions in a human somatic cell tree: how, what and why. *Cell Cycle*, **5**, 610–614, 2006.

140. Millman K, Black CM, Stamm WE, Jones RB, Hook EW, Martin DH, Bolan G, Tavaré S & Dean D. Population-based genetic epidemiologic analysis of *Chlamydia trachomatis* serotypes and lack of associations between *ompA* polymorphisms and clinical phenotypes. *Microbes and Infection*, **8**, 604–611, 2006.

141. Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C & Nordborg M. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Mol Biol Evol*, **23**, 1217–1231, 2006.

142. Marioni JC, Thorne NP & Tavaré S. BioHMM: a heterogeneous hidden Markov model for segmenting array CGH data. *Bioinformatics*, **22**, 1144–1146, 2006.

143. Dunning M, Thorne NP, Camilier I, Smith ML & Tavaré S. Quality control and low-level statistical analysis of Illumina BeadArrays. *REVSTAT*, **4**, 1–30, 2006.

144. Arratia R, Barbour AD & Tavaré S. A tale of three couplings: Poisson-Dirichlet and GEM approximations for random permutations. *Combin Probab Comput*, **15**, 31–62, 2006.

145. Kim JY, Tavaré S & Shibata D. Human hair genealogies and stem cell latency. *BMC Biology*, **4**:2, 2006.

2005

146. Stranger BE, Forrest MS, Clark AG, Minichiello M, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P & Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet*, **1**, 695–704, 2005.

147. Kim JY, Tavaré S & Shibata D. Counting human somatic cell replications: Methylation mirrors human endometrial stem cell divisions. *Proc Natl Acad Sci USA*, **102**, 17739–17744, 2005.

148. Clark RM, Tavaré S & Doebley J. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol Biol Evol*, **22**, 2304–2312, 2005.

149. Calabrese P, Mecklin JP, Järvinen HJ, Aaltonen LA, Tavaré S & Shibata D. Numbers of mutations to different types of colorectal cancer. *BMC Cancer*. **5**:126, 2005.

150. Kim JY, Siegmund KD, Tavaré S & Shibata D. Age-related human small intestine methylation: evidence for stem cell niches. *BMC Medicine*. **3**:10, 2005.

151. Arratia R, Barbour AD & Tavaré S. A probabilistic approach to analytic arithmetic on algebraic function fields. *Math Proc Camb Phil Soc*, **139**, 1–26, 2005.

152. Innan H, Zhang K, Marjoram P, Tavaré S & Rosenberg NA. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics*, **169**, 1763–1777, 2005.

2004

153. Tavaré S. Ancestral inference in population genetics. In *Lectures on Probability Theory and Statistics. Ecole d'Etés de Probabilité de Saint-Flour XXXI – 2001.* Ed. J. Picard. Lecture Notes in Mathematics, **1837**, 1–188, 2004. Springer Verlag, New York.

154. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, Tavaré S & Tower J. Similar gene expression patterns characterize aging and oxidative stress in *Drosophila melanogaster. Proc Natl Acad Sci USA,* **101**, 7663–7668, 2004.

155. Millman KL, Black CM, Johnson R, Stamm W, Jones R, Hook E, Martin D, Bolan G, Tavaré S, & Dean D. Population-based genetic and evolutionary analysis of *Chlamydia trachomatis* urogenital strain variation in the United States. *J Bacteriol*, **186**, 2457–2465, 2004.

156. Calabrese P, Tavaré S & Shibata D. Pre-tumor progression: clonal evolution of human stem cell populations. *Am J Pathol*, **164**, 1337–1346, 2004.

157. Calabrese P, Tsao J-L, Yatabe Y, Salovaara R, Mecklin J-P, Järvinen HJ, Aaltonen LA, Tavaré S & Shibata D. Colorectal pre-tumor progression before and after loss of DNA mismatch repair. *Am. J. Pathol.*, **164**, 1447–1453, 2004.

158. Kim K-M, Calabrese P, Tavaré S & Shibata D. Enhanced stem cell survival in familial adenomatous polyposis. *Am. J. Pathol.*, **164**, 1369–1377, 2004.

2003

159. Marjoram P, Molitor J, Plagnol V & Tavaré S. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, **100**, 15324–15328, 2003.

2002

160. Tavaré S, Marshall CR, Will O, Soligo C & Martin RD. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, **416**, 726–729, 2002.

161. ** Nordborg M & Tavaré S. Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, **18**, 83–90, 2002.

2001

162. Millman KL, Tavaré S & Dean D. Recombination in the *ompA* gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance and persistence of the organism. *Journal of Bacteriology*, **183**, 5997–6008, 2001.

163. Yatabe Y, Tavaré S & Shibata D. Investigating stem cells in human colon by using methylation patterns. *Proc. Natl. Acad. Sci. USA,* **98**, 10839–10844, 2001. (With editorial.)

164. Tavaré S, Adams DC, Fedrigo O & Naylor GJP. A model for phylogenetic inference using structural and chemical covariates. *Pacific Symposium on Biocomputing*, **6**, 215–225, 2001.

2000

165. Arratia R, Barbour AD & Tavaré S. Limits of logarithmic combinatorial structures. *Ann. Probab.*, **28**, 1620–1644, 2000.
166. Markovtsova L, Marjoram P & Tavaré S. The effects of rate variation on ancestral inference in the coalescent. *Genetics*, **156**, 1427–1436, 2000.
167. Markovtsova L, Marjoram P & Tavaré S. The age of a unique event polymorphism. *Genetics*, **156**, 401–409, 2000.
168. Arratia R, Barbour AD & Tavaré S. The number of components in a logarithmic combinatorial structure. *Ann. Appl. Prob.*, **10**, 331–361, 2000.
169. Schbath S, Bossard N & Tavaré S. The effect of non-homogeneous clone length distribution on the progress of an STS mapping project. *J. Comp. Biol.*, **7**, 47–57, 2000.
170. Tsao JL, Yatabe Y, Salovaara R, Jarvinen HJ, Mecklin JP, Aaltonen LA, Tavaré S & Shibata D. Genetic reconstruction of individual colorectal tumor histories. *Proc. Natl. Acad. Sci. USA*, **97**, 1236–1241, 2000.

1999

171. Jaruzelska J, Zietkiewicz E, Batzer M, Cole D, Moisan J-P, Scozzari R, Tavaré S & Labuda D. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics*, **152**, 1091-1101, 1999.
172. Fan F, Liu C, Tavaré S & Arnheim N. Polymorphisms in the human DNA repair gene XPF. *Mutation Research Genomics*, **406**, 115–120, 1999.
173. Leeflang E, Tavaré S, Marjoram P, Neal COS, Srinidhi J, MacFarlane H, MacDonald ME, Gusella JF, de Young M, Wexler N & Arnheim N. Analysis of germline mutation spectra at the Huntington Disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.*, **8**, 173–183, 1999.
174. Tsao JL, Tavaré S, Salovaara R, Jass JR, Aaltonen LA & Shibata D. Colorectal adenoma and cancer divergence: evidence of multi-lineage progression. *Am. J. Pathol.*, **154**, 1815–1824, 1999.
175. Griffiths RC & Tavaré S. The ages of mutations in gene trees. *Ann. Appl. Prob.*, **9**, 567–590, 1999.
176. Arratia R, Barbour AD & Tavaré S. The Poisson-Dirichlet distribution and the scale-invariant Poisson process. *Combin. Probab. Comput.*, **8**, 407–416, 1999.
177. Arratia R, Barbour AD & Tavaré S. On Poisson-Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.*, **8**, 193–208, 1999.

1998

178. Griffiths RC & Tavaré S. The age of a mutation in a general coalescent tree. *Stochastic Models*, **14**, 273–295, 1998.

1997

179. Arratia R, Barbour AD & Tavaré S. Random combinatorial structures and prime factorizations. *Notices of the American Mathematical Society*, **44**, 903–910, 1997.

180. Tavaré S, Balding DJ, Griffiths RC & Donnelly P. Inferring coalescence times from molecular sequence data. *Genetics*, **145**, 505–518, 1997.

1996

181. Shenkar R, Navidi W, Tavaré S, Dang MH, Chomyn A, Attardi G, Cortopassi G & Arnheim N. The mutation rate of the human mitochondrial DNA deletion mtDNA^{4977}. *Amer. J. Hum. Gen.*, **59**, 772–780, 1996. (With editorial.)

182. Donnelly P, Tavaré S, Balding DJ & Griffiths RC. Estimating the age of the common ancestor of men from the ZFY intron. *Science*, **272**, 1357–1359, 1996.

183. Griffiths RC & Tavaré S. Monte Carlo inference methods in population genetics. *Mathl. Comput. Modelling*, **23**, 141–158, 1996.

1995

184. Donnelly P & Tavaré S. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.*, **29**, 401–421, 1995.

185. Leeflang EP, Zhang L, Tavaré S, Hubert R, Srinidhi J, MacDonald ME, Myers RM, de Young M, Wexler NS, Gusella JF & Arnheim N. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum. Molec. Genet.*, **4**, 1519–1526, 1995.

186. Arratia R, Stark D & Tavaré S. Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies. *Ann. Probab.*, **23**, 1347-1388, 1995.

187. Joyce P & Tavaré S. The distribution of rare alleles. *J. Math. Biol.*, **33**, 602–618, 1995.

188. Griffiths RC & Tavaré S. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.*, **127**, 77–98, 1995.

1994

189. Griffiths RC & Tavaré S. Ancestral inference in population genetics. *Statistical Science*, **9**, 307–319, 1994.

190. Griffiths RC & Tavaré S. Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. Lond. B*, **344**, 403–410, 1994.

191. Barbour AD & Tavaré S. A rate for the Erdős-Turán law. *Combin. Probab. Comput.*, **3**, 167–176, 1994.

192. Griffiths RC & Tavaré S. Simulating probability distributions in the coalescent. *Theor. Popn. Biol.*, **46**, 131–159, 1994.

193. Arratia R & Tavaré S. Independent process approximations for random combinatorial structures. *Adv. Math.*, **104**, 90–154, 1994.

194. Raftery AE & Tavaré S. Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution (MTD) model. *Applied Statistics*, **43**, 179–199, 1994.

1993

195. Arratia R, Barbour AD & Tavaré S. On random polynomials over a finite field. *Math. Proc. Camb. Phil. Soc.*, **114**, 347-368, 1993.

1992

196. Lundstrom R, Tavaré S & Ward RH. Modelling the evolution of the human mitochondrial genome. *Math. Biosci.*, **112**, 319–336, 1992.

197. Lundstrom R, Tavaré S & Ward RH. Estimating substitution rates from molecular data using the coalescent. *Proc. Natl. Acad. Sci., USA*, **89**, 5961–5965, 1992.

198. Arratia R & Tavaré S. Limit theorems for combinatorial structures via discrete process approximations. *Rand. Struct. Alg.*, **3**, 321–345, 1992.

199. Arratia R, Barbour AD & Tavaré S. Poisson process approximations for the Ewens Sampling Formula. *Ann. Appl. Probab.*, **2**, 519–535, 1992.

200. Arratia R & Tavaré S. The cycle structure of random permutations. *Ann. Probab.*, **20**, 1567–1591, 1992.

201. Joyce PJ & Tavaré S. A convergence theorem for symmetric functionals of random partitions. *J. Appl. Prob.*, **29**, 280–290, 1992.

1991

202. Donnelly P, Kurtz TG & Tavaré S. On the functional central limit theorem for the Ewens sampling formula. *Ann. Appl. Probab.*, **1**, 539–545, 1991.

203. Arratia R, Lander ES, Tavaré S & Waterman MS. Genomic mapping by anchoring random clones: a mathematical analysis. *Genomics*, **11**, 806–827, 1991.

1990

204. Joyce PJ & Tavaré S. Random permutations and neutral evolution models. *Stoch. Proc. and Applns.*, **36**, 245–262, 1990.

1989

205. Tavaré S, Ewens WJ & Joyce P. Is knowing the age-order of alleles useful in testing for selective neutrality? *Genetics*, **122**, 705–711, 1989.

206. Tavaré S & Song B. Codon preference and primary sequence structure in protein coding regions. *Bull. Math. Biol.*, **51**, 95–115, 1989.

1987

207. Joyce PJ & Tavaré S. Cycles, permutations and the structure of the Yule process with immigration. *Stoch. Proc. and Applns.*, **25**, 309–314, 1987.
208. Donnelly P & Tavaré S. The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.*, **25**, 381–391, 1987.
209. Tavaré S. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.*, **25**, 161–168, 1987.

1986

210. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86, 1986.
211. Donnelly P & Tavaré S. The ages of alleles and a coalescent. *Adv. Appl. Prob.*, **18**, 1–19, 1986. Correction:**18**, 1023.

1984

212. Tavaré S. Lines of descent and genealogical processes, and their application in population genetics models. *Theoret. Popn. Biol.*, **26**, 119–164, 1984.
213. Saunders IW, Tavaré S & Watterson GA. On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob.*, **16**, 471–491, 1984.

1983

214. Karlin S, Ghandhour G, Ost F, Tavaré S & Korn LJ. New approaches for computer analysis of nucleic acid sequences. *Proc. Natl. Acad. Sci. USA*, **80**, 5660–5664, 1983.
215. Ethier SN & Tavaré S. The proportional bettor's return on investment. *J. Appl. Prob.*, **20**, 563–573, 1983.
216. Seneta E & Tavaré S. Some stochastic models for plasmid copy number. *Theoret. Popn. Biol.*, **23**, 241–256, 1983.
217. Seneta E & Tavaré S. A note on models using the branching process with immigration stopped at zero. *J. Appl. Prob.*, **20**, 11–18, 1983.
218. Tavaré S. Serial dependence in contingency tables. *J. Royal Statist. Soc. B*, **45**, 100–106, 1983.
219. Tavaré S & Altham PME. Serial dependence of observations leading to contingency tables, with corrections to χ^2 statistics. *Biometrika*, **70**, 139–144, 1983.
220. Karlin S & Tavaré S. A class of diffusion processes with killing arising in population genetics. *SIAM J. Appl. Math.*, **43**, 31–41, 1983.

1982

221. Karlin S & Tavaré S. Linear birth and death processes with killing. *J. Appl. Prob.*, **19**, 477–487, 1982.

222. Karlin S & Tavaré S. Detecting particular genotypes in populations under non-random mating. *Math. Biosci.*, **59**, 57–75, 1982.

223. Karlin S & Tavaré S. A diffusion with killing: the time to formation of recurrent deleterious mutant genes. *Stoch. Proc. and Applns.*, **13**, 249–261, 1982.

1981

224. Pakes AG & Tavaré S. Comments on the age distribution of Markov processes. *Adv. Appl. Prob.*, **13**, 681–703, 1981.

225. Karlin S & Tavaré S. The detection of particular genotypes in finite populations. II. The effects of partial penetrance and family structure. *Theoret. Popn. Biol.*, **19**, 215–229, 1981.

226. Karlin S & Tavaré S. The detection of particular genotypes in finite populations. I. Natural selection effects. *Theoret. Popn. Biol.*, **19**, 187-214, 1981.

227. Karlin S & Tavaré S. The detection of a recessive visible gene in finite populations. *Genetical Research*, **37**, 22–46, 1981.

1980

228. Tavaré S. Time reversal and age distributions. I. Discrete-time Markov chains. *J. Appl. Prob.*, **17**, 33–46, 1980.

1979

229. Tavaré S. Dual diffusions, killed diffusions and the age distribution problem in population genetics. *Theoret. Popn. Biol.*, **16**, 253–266, 1979.

230. Tavaré S. A note on finite continuous-time Markov chains. *Biometrics*, **35**, 831–834, 1979.

231. Tavaré S. Sojourn times for conditioned Markov chains in genetics. *Theoret. Popn. Biol.*, **15**, 108–112, 1979.

232. Oakley JR & Tavaré S. Another approach to the assessment of growth in early infancy. *Developmental Medicine and Child Neurology*, **21**, 186–193, 1979.

Newsletter Articles

233. Dinh K, Tavaré S & Xiang Z. ABC and distributional random forests. IICD Newsletter, February 2024.
`cancerdynamics.columbia.edu/news/approximate-bayesian-computation-and-distributional-random-forests`
`acrobat.adobe.com/id/urn:aaid:sc:VA6C2:c0ec8dc1-4c77-4f22-a7fa-835f93e028f1`

234. Tavaré S. Computing the exponential of a matrix. IICD Newsletter, September 2020.
`cancerdynamics.columbia.edu/news/iicd-newsletter-september-2020`

235. Tavaré S. The complex step method for numerical differentiation. IICD Newsletter, August 2020. `cancerdynamics.columbia.edu/news/iicd-newsletter-august-2020`

Book Chapters

† denotes refereed article, ** denotes review

236. †Lynch AG, Smith ML, Eldridge MD & Tavaré S on behalf of the OCCAMS consortium. PCR duplicate proportion estimation and consequences for DNA copy number calculations. pp. 259–279 in *Recent Developments in Statistics and Data Science*, eds. Bispo R, Henriques-Rodrigues L, Alpizar-Jara R & de Carvalho M. Vol 398, Springer Proceedings in Mathematics & Statistics, 2022.

237. †Tavaré S. The history of ABC. Chapter 2 in *Handbook of Approximate Bayesian Computation*. Eds. Sisson SA, Fan Y & Beaumont MA. Taylor and Francis, 2018.

238. Sottoriva A & Tavaré S. Population genetics of neoplasms. Pp 31–42 in *Frontiers in Cancer Research: Evolutionary Foundations, Revolutionary Directions*, eds. Maley CC & Greaves M. Springer, New York, 2016.

239. Andrews DJ, Lynch AG & Tavaré S. Using methylation patterns for reconstructing cell division dynamics: assessing validation experiments. Chapter 1, pp 3–15 in *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology – Systems & Applications*. Tran Q-N & Arabnia HR, editors. Morgan Kaufmann, 2016.

240. ** Ivakhno S, Frese K, Tavaré S, Iacobuzio-Donahue C & Tuveson D. Pancreatic cancer. Chapter 26, pp 409–420 in *Systems Biology of Cancer*, ed. Thiagalingam S. Cambridge University Press, 2015.

241. †Cairns JC, Lynch AG & Tavaré S. Statistical aspects of ChIP-seq analysis. Pp 138–169 in *Advances in Statistical Bioinformatics: Models and Integrative Inference for High-Throughput Data*, eds. Do K, Qin ZS & Vannucci M. Cambridge University Press, 2013.

242. ** Marjoram P & Tavaré S. The coalescent. Pp. 54–57 in *Brenner's Encyclopedia of Genetics, 2nd Edition,* Volume 2, eds. Maloy S & Hughes K, Academic Press, San Diego, 2013.

243. †Ardekani R, Greenwood AK, Peichl CL & Tavaré S. Using non-local background modeling to quantify the schooling behaviour of sticklebacks. *VAIB12: Workshop on Visual Observation and Analysis of Animal and Insect Behavior 2012*. In conjunction with *21st International Conference on Pattern Recognition (ICPR 2012).*

244. †Sottoriva A & Tavaré S. Integrating approximate Bayesian computation with complex agent-based models for cancer research. In *COMPSTAT 2010 – Proceedings in Computational Statistics*, eds. Saporta G & Lechevallier Y. Springer, Physica Verlag, pp. 57–66, 2010.

245. †Barbour AD & Tavaré S. Assessing molecular variability in cancer genomes. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, eds. Bingham N & Goldie CM. Cambridge, Cambridge University Press, pp. 91–111, 2010.

246. †** Thorne NP, Marioni JC, Rakyan V, Ibrahim AEK, Massie C, Curtis C, Brenton JD, Murrell A & Tavaré S. DNA methylation arrays: Methods and analysis. In *Microarray Innovations: Technology and Experimentation in Drug Discovery and Biomedical Research*, ed. G. Hardiman. CRC Press/Taylor and Francis, Ch. 13, pp. 175–206, 2009.

247. †Journée M, Teschendorff AE, Absil P-A, Tavaré S & Sepulchre R. Geometric optimization methods for the analysis of gene expression data. In *Principal Manifolds for Data Visualization and Dimension Reduction*, eds. A. Gorban, B. Kegl, D. Winsch & A. Zinovyev. LNCSE, Springer, pp. 271–292, 2007.

248. †Soligo C, Will O, Tavaré S, Marshall CR & Martin RD. New light on the dates of primate origins and divergence. In *Primate Origins: Adaptations and Evolution*. Eds. MJ Ravosa & M Dagosto. Springer Verlag, New York, pp. 29–49, 2007. ISBN: 0-387-30335-9

249. †Tavaré S. Ancestral inference for branching processes. In *Branching Processes in Biology: Variation, Growth, Extinction*, eds. P Haccou, P Jagers & V Vatutin. Cambridge University Press, pp. 208–217, 2005.

250. †Plagnol V & Tavaré S. Approximate Bayesian computation and MCMC. In *Monte Carlo and Quasi-Monte Carlo Methods 2002*, ed. H Niederreiter, Springer-Verlag, pp. 99–114, 2004.

251. †Navidi WC, Tavaré S & Arnheim N. The roles of mutation rate and selective pressure on observed levels of the human mitochondrial DNA deletion mtDNA^{4977} In *Science and Statistics: A Festschrift for Terry Speed*, ed. DR Goldstein, IMS Lecture Notes – Monograph Series, Volume 40, pp. 247–258, 2003.

252. †** Tavaré S. Coalescent theory. In *Nature Encyclopedia of the Human Genome*, Vol. 1, ed. Cooper DN, pp. 836–845, 2003. Nature Publishing Group, London.

253. †Griffiths RC & Tavaré S. The genealogy of a neutral mutation. In *Highly Structured Stochastic Systems*. Eds. Green P, Hjort N and Richardson S, pp. 393–412, 2003. Oxford Statistical Science Series, Vol. 27, Oxford University Press.

254. ** Neuhauser C & Tavaré S. The coalescent. *Encyclopedia of Genetics*, Vol. I. Eds. Brenner S and Miller J, pp. 392–397, 2001. Academic Press, New York.

255. Barbour AD & Tavaré S. A rate for the Erdős-Turán law. In *Combinatorics, geometry and probability. A tribute to Paul Erdős*. Eds. Bollobás B and Thomason A, pp. 71–80, 1998. Cambridge University Press. (A reprinting of article 191 above)

256. Leeflang EP, Tavaré S, Marjoram P, Grewal R, Neal COS & Arnheim N. Human germline mutation analysis by single genome PCR: application to dynamic mutations. Chapter 37 in *Genetic Instabilities and Hereditary Neurological Diseases*. Eds. Wells RD and Warren ST, pp. 543–560, 1998. Academic Press, New York.

257. †** Ewens WJ & Tavaré S. The Ewens Sampling Formula. In *Encyclopedia of Statistical Science*, Vol. 2 update. Eds. Kotz S, Read CB and Banks DL, pp. 230–234, 1998. Wiley, New York.

258. †Tavaré S. Ancestral inference from DNA sequence data. Chapter 5 in *Case Studies in Mathematical Modeling: Ecology, Physiology, and Cell Biology*. Eds. Othmer HG, Adler FR, Lewis MA and Dallon J, pp. 81–96, 1997. Prentice-Hall.

259. †** Tavaré S & Ewens WJ. Multivariate Ewens distribution. Chapter 41 in *Discrete Multivariate Distributions*. Eds. Johnson NS, Kotz S and Balakrishnan N, pp. 232–246, 1997. Wiley, New York.

260. †Griffiths RC & Tavaré S. Computational methods for the coalescent. Chapter 10 in *Progress in Population Genetics and Human Evolution.* Eds. P. Donnelly P and Tavaré S. IMA Volumes in Mathematics and its Applications, **87**, 165–182, 1997. Springer Verlag, Berlin.

261. †Tavaré S. Calibrating the clock: using stochastic processes to measure the rate of evolution. Chapter 5 in *Calculating the Secrets of Life*, Eds. Lander ES and Waterman MS, pp. 114–152, 1995. National Academy Press, Washington.

262. Tavaré S. The genealogy of the birth, death and immigration process. Chapter 3 in *Mathematical Evolutionary Theory.* Ed. Feldman MW, pp. 41–56, 1989. Princeton University Press.

263. †Tavaré S & Giddings BW. Some statistical aspects of the primary structure of nucleotide sequences. *Mathematical Methods for DNA Sequences.* Ed. Waterman MS, pp. 117–132, 1988. CRC Press.

264. Donnelly P & Tavaré S. A genealogical description of the infinitely-many neutral alleles model. *Stochastic Methods in Biology.* Eds. Kimura M, Kallianpur G, Hida T. Lecture Notes in Biomathematics, **70**, 27–35, 1987. Springer-Verlag.

265. †Williamson JA, Bishop DT, Tavaré CJ & Tavaré S. On the independence structure that exists within a pedigree with an application to testing for a major gene. In *Progress in Clinical and Biological Research.* Eds. Sing CF and Skolnick MH. **32**, 271–276, 1979. AR Liss, New York.

Letters to the Editor

266. Dunning MJ, Curtis CN, Barbosa-Morais NL, Caldas C, Tavaré S & Lynch AG. The importance of platform annotation in interpreting microarray data: A response to "Breast cancer molecular profiling with single sample predictors: a retrospective analysis". *The Lancet Oncology*, **11**, 717, 2010.

267. Lynch AG, Marioni JC & Tavaré S. Numbers of CNVs and false negative rates will be underestimated if we do not account for the dependence between repeated experiments. *Amer J Hum Genet*, **81**, 418–420, 2007.

268. Markovtsova L, Marjoram P & Tavaré S. On a test of Depaulis and Veuille. *Mol. Biol. Evol.*, **18**, 1132–1133, 2001.

Book Reviews

269. Arratia R & Tavaré S. *Probability approximations via the Poisson clumping heuristic* by Aldous D, *Poisson approximation* by Barbour AD, Holst L and Janson S. *Ann. Probab.*, **21**, 2269–2279, 1993.

270. Tavaré S. *Mathematical Methods of Applied Probability*, by Hunter JJ. *J. A. S. A.*, **80**, 1071–1072, 1985.

271. Tavaré S. *Genealogical and Genetic Structure*, by Cannings C and Thompson EA. *J. A. S. A.*, **79**, 743–744, 1984.

Consortium Papers

272. Abujudeh S, Zeki SS, van Lanschot MCJ, Pusung M, Weaver JMJ, Li X, Noorani A, Metz AJ, Bornschein J, Bower L, Miremadi A, Fitzgerald RC, Morrissey ER & Lynch AG. Low-cost and clinically applicable copy number profiling using repeat DNA. *BMC Genomics*, **23**:599, 2022.

273. Kuett L, Catena C, Özcan A, Plüss A, Cancer Grand Challenges IMAXT Consortium, Schraml P, Moch H, de Souza N & Bodenmiller B. Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment. *Nature Cancer*, **3**, 122–133, 2022.

274. Au L, Hatipoglu E, . . . Jamal-Hanjani M on behalf of the PEACE Consortium, . . . , Swanton C, Quezada SA, Turajlic S and on behalf of the TRACERx Renal Consortium. Determinants of anti-PD-1 response and resistance in clear cell renal cell carcinoma. *Cancer Cell*, **39**, 1497–1518, 2021.

275. Georgopoulou D, Callari M, Rueda OM, Shea A, Martin A, Giovannetti A, Qosaj F, Dariush A, Chin S-F, Carnevalli LS, Provenzano E, Greenwood W, Lerda G, Esmaeilishirazifard E, O'Reilly M, Serra V, Bressan D, IMAXT Consortium, Mills GB, Ali HR, Cosulich SS, Hannon GJ, Bruna A & Caldas C. Landscapes of cellular phenotypic diversity in breast cancer xenografts and their impact on drug response. *Nat Comm*, **12**, art. 1998, 2021.

276. Alon S, Goodwin DR, Sinha A et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656, 2021.

277. Knight WRC, McEwan R, Byrne BE, Habib W, Bott R, Zylstra J, Mahadeva U, Gossage JA & OCCAMS Consortium. Endoscopic tumour morphology impacts survival in adenocarcinoma of the oesophagus. *European Journal of Surgical Oncology*, **46**, 2257–2261, 2020.

278. Litchfield K, Stanislaw S, Spain L, Gallegos LL, Rowan A, Schnidrig D, Rosenbaum H, Harle A, Au L, Hill SM, Tippu Z, Thomas J, Thompson L, Xu H, Horswell S, Barhoumi A, Jones C, Leith KF, Burgess DL, Watkins TBK, Lim E, Birkbak NJ, Lamy P, Nordentoft I, Dyrskjøt L, Pickering L, Hazell S, Jamal-Hanjani M, PEACE Consortium, Larkin J, Swanton C, Alexander NR & Turajlic S. Representative sequencing: unbiased sampling of solid tumor tissue. *Cell Reports*, **31**, 107550, 2020.

279. Rahman SA,Walker RC, Lloyd MA, Grace BL, van Boxel GI, Kingma BF, Ruurda JP, van Hillegersberg R, Harris S, Parsons S, Mercer S, Griffiths EA, O'Neill JR, Turkington R, Fitzgerald RC & Underwood TJ, on behalf of the OCCAMS Consortium. Machine learning to predict early recurrence after oesophageal cancer surgery. *Br J Surg*, **107**, 1042–1052, 2020.

280. Ali HR, Jackson HW, Zanotelli VRT, Danenberg E, Fischer JR, Bardwell H, Provenzano H, CRUK IMAXT Grand Challenge Team, Rueda OM, Chin S-F, Aparicio S, Caldas C & Bodenmiller B. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nature Cancer*, **1**, 163–175, 2020.

281. Laks E, McPherson A, Zahn H, Lai D, Steif A, Brimhall J, Biele J, Wang B, Masud T , Ting J, Grewal D, Nielsen C, Leung, S, Bojilova V, Smith M, Golovko O, Poon S, Eirew P,

Kabeer F, Ruiz de Algara T, Lee SR, Taghiyar MJ, Huebner C, Ngo J, Chan T, Vatrt-Watts S, Walters P, Abrar N, Chan S, Wiens M, Martin, L Scott RW, Underhill TM, Chavez E, Steidl C, Da Costa D, Ma Y, Coope RJN, Corbett R, Pleasance S, Moore R, Mungall AJ, Mar C, Cafferty F, Gelmon K, Chia S, CRUK IMAXT Grand Challenge Team, Marra MA, Hansen C, Shah SP & Aparicio S. Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell*, **179**, 1207–1221, 2019.

282. Mourikis TP, Benedetti L, Foxall E, Temelkovski D, Nulsen J, Perner J, Cereda M, Lagergren J, Howell M, Yau C, Fitzgerald RC, Scaffidi P, The Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium & Ciccarelli FD. Patient-specific cancer genes contribute to recurrently perturbed pathways and establish therapeutic vulnerabilities in esophageal adenocarcinoma. *Nature Communications*, **10**, art 3101, 2019.

283. Bornschein J, Wernisch L, Secrier M, Miremadi A, Perner J, MacRae S, O'Donovan M, Newton R, Menon S, Bower L, Eldridge MD, Devonshire G, Cheah C, Turkington R, Hardwick RH, Selgrad M, Venerito M, Malfertheiner P, OCCAMS Consortium & Fitzgerald RC. Transcriptomic profiling reveals three molecular phenotypes of adenocarcinoma at the gastro-esophageal junction. *Int J Cancer*, **145**, 3389–3401, 2019.

284. Turkington RC, Knight LA, Blayney JK, Secrier M, Douglas R, Parkes EE, Sutton EK, Stevenson L, McManus D, Halliday S, Mccavigan AM, Logan GE, Walker SM, Steele CJ, Perner J, Bornschein J, Macrae S, Miremadi A, Mccarron E, McQuaid S, Arthur K, James JA, Eatock MM, O'Neill R, Noble F, Underwood TJ, Harkin P, Salto-Tellez M, Fitzgerald RC, Kennedy RD on behalf of the Oesophageal cancer clinical and Molecular Stratification (OccaMS) Study group. Immune activation by DNA damage predicts response to chemotherapy and survival in oesophageal adenocarcinoma. *Gut*, **68**, 1918–1927, 2019.

285. Sawas T, Killcoyne S, Iyer PG, Wang KK, Smyrk TC, Kisiel JB, Qin Y, Ahlquist DA, Rustgi AK, Costa RJ, Gerstung M, Fitzgerald RC & Katzka DA for the OCCAMS Consortium. Identification of prognostic phenotypes of esophageal adenocarcinoma in two independent cohorts. *Gastroenterology*, **155**, 1720–1728, 2018.

286. Noble F, Lloyd MA, Turkington R, Griffiths E, O'Donovan M, O'Neill JR, Mercer S, Parsons SL, Fitzgerald RC, Underwood TJ & OCCAMS consortium. Multicentre cohort study to define and validate pathological assessment of response to neoadjuvant therapy in oesophagogastric adenocarcinoma. *Br J Surg*, **104**, 1816–1828, 2017.

287. Fels Elliott DR, Perner J, Li X, Symmons MF, Verstak B, Eldridge M, Bower L, O'Donovan M, Gay NJ, OCCAMS Consortium & Fitzgerald RC. Impact of mutations in Toll-like receptor pathway genes on esophageal carcinogenesis. *PLoS Genet*, 13:e1006808, 2017.

288. Britton E, Rogerson C, Mehta S, Li Y, Li X, OCCAMS consortium, Fitzgerald RC, Ang YS & Sharrocks AD. Open chromatin profiling identifies AP1 as a transcriptional regulator in oesophageal adenocarcinoma. *PLoS Genet*, **13**:e1006879, 2017.

289. Garcia E, Hayden A, Birts C, Britton E, Cowie A, Pickard K, Mellone M, Choh C, Derouet M, Duriez P, Noble F, White MJ, Primrose JN, Strefford JC, Rose-Zerilli M, Thomas GJ, Ang Y, Sharrocks AD, Fitzgerald RC, Underwood TJ & OCCAMS consortium. Authentication and characterisation of a new oesophageal adenocarcinoma cell line: MFD-1. *Sci Rep*, **6**, 32417, 2016. See correction at `https://www.nature.com/articles/s41598-018-37591-7`

290. Behjati S, Gundem G, Wedge DC, Roberts ND, Tarpey PS, Cooke SL, Van Loo P, Alexandrov LB, Ramakrishna M, Davies H, Nik-Zainal S, Hardy C, Latimer C, Raine KM, Stebbings L, Menzies A, Jones D, Shepherd R, Butler AP, Teague JW, Jorgensen M, Khatri B, Pillay N, Shlien A, Futreal PA, Badie C, ICGC Prostate Group, McDermott U, Bova GS, Richardson AL, Flanagan AM, Stratton MR & Campbell PJ. Mutational signatures of ionizing radiation in second malignancies. *Nat Commun*, **7**, 12605, 2016.

291. Ross-Innes CS, Becq J, Warren A, Cheetham RK, Northen H, O'Donovan M, Malhotra S, di Pietro M, Ivakhno S, He M, Weaver JMJ, Lynch AG, Kingsbury Z, Ross M, Humphray S, Bentley D, Fitzgerald RC for the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Study Group. Whole-genome sequencing defines new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nature Genetics*, **47**, 1038–1046, 2015.

292. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, Annala M, Kivinummi K, Goody V, Latimer C, O'Meara S, Dawson KJ, Isaacs W, Emmert-Buck MR, Nykter M, Foster C, Kote-Jarai Z, Easton D, Whitaker HC, ICGC Prostate UK Group, Neal DE, Cooper CS, Eeles RA, Visakorpi T, Campbell PJ, McDermott U, Wedge DC & Bova GS. The evolutionary history of lethal metastatic prostate cancer. *Nature*, **520**, 353–357, 2015.

293. Vollan HKM, Rueda OM, Chin S-F, Curtis C, Turashvili G, Shah S, Lingjærde AC, Yuan Y, Ng CK, Dunning MJ, Dicks E, Provenzano E, Sammut S, McKinney S, Ellis IO, Pinder S, Purushotham A, Murphy LC, Kristensen VN, METABRIC Group, Brenton JD, Pharoah PDP, Børresen-Dale A-L, Aparicio S, Caldas C. A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Molecular Oncology*, **9**, 115–127, 2015.

294. Papatheodorou I, Crichton C, Morris L, Maccallum P, METABRIC Group, Davies J, Brenton JD & Caldas C. A metadata approach for clinical data management in translational genomics studies in breast cancer. *BMC Medical Genomics*, **2**: 66, 2009.

Other Manuscripts

295. ** Lynch AG, Smith ML, Dunning MJ, Cairns JM, Barbosa-Morais NL & Tavaré S. *beadarray*, BASH and HULK – tools to increase the value of Illumina BeadArray experiments. In *Statistical Tools for Challenges in Bioinformatics*, eds. Gusnato A, Mardia KV, & Fallaize CJ. Leeds, Leeds University Press, pp. 33–37, 2009.

296. Lynch AG, Curtis CN & Tavaré S. Correcting for probe-design in the analysis of gene-expression microarrays. In *Systems Biology & Statistical Bioinformatics*, eds. Barber S, Baxter PD & Mardia KV. Leeds, Leeds University Press, pp. 83–86, 2007.

297. Tavaré S, Nicolas P & Shibata D. Stochastic models for the evolution of stem cells in colon crypts. In *Systems Biology & Statistical Bioinformatics*, eds. Barber S, Baxter PD & Mardia KV. Leeds, Leeds University Press, pp. 61–64, 2007.

298. Marjoram P, Markovtsova L, & Tavaré S. "I see dead people:" gene mapping via ancestral inference. Draft for *Genetic Analysis Workshop 12*, 2000.
Available at `www.damtp.cam.ac.uk/user/st321/allstpapers.html`

299. Marjoram P, Markovtsova L & Tavaré S. Estimation in ancestral recombination graphs using Markov chain Monte Carlo. Research Report, October 2000.
Available at `www.cmb.usc.edu/people/stavare/allstpapers.html`

300. Markovtsova L, Marjoram P & Tavaré S. Discussion of 'Inference in molecular population genetics' by Stephens and Donnelly. *J. Royal Statist. Soc. B*, **62**, 647, 2000.

301. Martin RD, Soligo C, Tavaré S, Will O & Marshall CR. New light on the dates of primate origins and divergence. *Folia Primatol.*, **71**, 358–359, 2000.

302. Tavaré S. Random trees in molecular genetics. *Bull. I.S.I.*, **52**, 269–272, 1999.

303. Tavaré S & Feng Y. Reconstructing phylogenetic trees when sites are dependent. Pp. 55–57 in *Proceedings of Phylogeny Workshop*, DIMACS Technical Report 95-48, ed. S. Tavaré, 1995. [See `http://dimacs.rutgers.edu/Workshops/Phylogeny/`]

304. Tavaré S. International Conference on Random Mappings, Partitions, and Permutations. Report, *Adv. Appl. Prob.*, **24**, 761–777, 1992.

305. Tavaré S. Mathematical models in population genetics. *Composite Function*, 1991. (A reprinting of article 313 below)

306. Tavaré S & Lundstrom R. Comments on 'Sampling strategies for distances between DNA sequences', by Weir and Basten. *Biometrics*, **46**, 577–582, 1990.

307. Tavaré S & Doerge RW. An application of computer algebra to likelihood analysis. Manuscript, 19pp, 1988.

308. Donnelly P & Tavaré S. The genealogy of the infinite alleles model. In *Proceedings, First World Congress of the Bernoulli Society, 1986.* VNU Sci. Press, Utrecht, 1987, pp. 701–704.

309. Tavaré S. A genealogical view of some stochastic models in population genetics. *Stoch. Proc. Applns.*, **19**, 10, 1985.

310. Tavaré S & Janzen T. On estimating substitution rates from pairs of homologous nucleotide sequences. Manuscript, 28pp, 1985.

311. Tavaré S. The estimation of substitution rates and divergence times from DNA sequence data. *Third Rocky Mountain Regional Conference on Medical Applications of Statistics*, Preprint Volume, 89–96, 1985.

312. Seneta E & Tavaré S. Stochastic models for plasmid copy number. *First Rocky Mountain Regional Conference on Medical Applications of Statistics*, Preprint Volume, 27–33, 1982.

313. Tavaré S. Mathematical models in population genetics. *Function*, **4**, 26-32, 1980.

314. Tavaré S. Time reversion and age distributions. *Adv. Appl. Prob.*, **11**, 10–12, 1979.

315. Tavaré S. Conference on ecological statistics - a summary report. *J. Royal Statist. Soc. C*, **27**, 344–345, 1978.

316. Tavaré S. Age distributions for Markov chains in genetics. *Adv. Appl. Prob.*, **10**, 17–19, 1978.

317. Tavaré S. Sojourn times for conditional Markov chains in genetics. *Adv. Appl. Prob.*, **8**, 645–647, 1976.

GRANT SUPPORT

Active:

MacMillan Foundation, sub-award via NYGC. *Inferring the role of the non-coding genome and epigenome in the regulation of gene expression within extrachromosomal DNA (ecDNA) in specific cancer types.* 01/23 – 12/26. Co-I (K. Nowicki-Osuch PI)

PROFESSIONAL SERVICE

Advisory Boards

Member, Expert Panel 2 (Informatics and Mathematics), Singapore Academic Research Council (2020 – 2025)

External Advisory Board member, Training Program in Biostatistics for Cancer Research (T32 CA096520-14A1), M. Kimmel, Rice University, PI (2019 – present)

Member, International Advisory Board, Bosch Health Campus, Stuttgart (2019 – present)

Consultant, Kallyope Inc. (2019 – 2022)

Science Advisory Board, Totient Inc. (2020 – 2021)

Member, Scientific Council, International Agency for Research on Cancer (IARC) (2016 – 2020)

Science Advisory Board, IPSEN (2018 – 2020)

Advisory Board, BioConductor (Chair; 2006 – 2016)

Scientific Steering Committee, Isaac Newton Institute, Cambridge (2016 – 2018)

International Scientific Advisory Committee, MRC Biostatistics Unit (2015 – 2018)

Science Advisory Board, Max F. Perutz Laboratories, Vienna. (2015 – 2017)

Advisory Board, Vienna Graduate School of Population Genetics (2010 – 2017)

Advisory Board, Gothenburg Mathematical Modelling Centre, Sweden (2006 – 2011)

Science Advisory Board, BioDiscovery, Inc. (2003 – 2010)

Scientific Advisory Board, Gene Ontology Consortium (2006 – 2010, 2012–2017.)

Steering Committee for Statistics, Eurandom, Holland. (2004 – 2008)

Advisory Board, Cornell Research Training Group in Probability (2005 – 2008)

Science Advisory Board, Institute for Pure and Applied Mathematics (IPAM), UCLA (2002 – 2008)

Senior Advisory Board, NESCent, North Carolina. (2004 – 2007)

International Advisory Board, Swegene, Sweden (2001 – 2006)

Scientific Advisory Panel, The Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, New Zealand. (2003 – 2005)

Science Advisory Board, Amersham plc. (2001 – 2004)

Scientific Advisory Board, Chalmers University of Technology Bioscience Initiative (1999 – 2001)

Scientific Advisory Board, GeneData AG (1997 – 2002)

Society Membership

International Society for Computational Biology (ISCB) (2019 – present)

National Academy of Sciences (2018 – present)

London Mathematical Society (2014 – present)

  President (Nov 2015 – Nov 2017)

  President Designate (Nov 2014 – Nov 2015)

American Mathematical Society (2014 – present)

Royal Society of London (2011 – present)

  Member, Newton International Fellowships Committee (2016 – 2018)

  Member, Sectional Committee 9 (2012 – 2015)

  Member, URF Panel Bii (2011 – 2014)

  Member, RS Wolfson Laboratory Refurbishment Panel (2014 – 2015)

  Journal of the Royal Society Interface (Editorial Board, 2003 – 2009)

Academy of Medical Sciences (2009 – present)

  Member, Sectional Committee 1 (2010 – 2012)

  Member, Council (2011 – 2014)

  Member, Fellowship Committee (2017)

Institute of Mathematics and its Applications (2009 – present)

Royal Society of Biology (2009 – present)

American Association for the Advancement of Science (1991 – present)

  Member, Electorate Nominating Committee, Statistics (2015 – 2018)

American Statistical Association (1992 – present)

Institute of Mathematical Statistics (1978 – present)

  Member, Council (2016 – 2019)

  Annals of Applied Statistics (Associate Editor, 2007 – 2011)

  Annals of Probability (Associate Editor, 2000 – 2002)

  Member, Committee on Fellows (1996 – 1999)

  Annals of Applied Probability (Associate Editor, 1993 – 1999)

Royal Statistical Society (1974 – present)

Editorial and Refereeing Services

Statistics Surveys (IMS Associate Editor, 2007 – present)

Evolutionary Bioinformatics (Editorial Board, 2005 – present)

Journal of Computational Biology (Associate Editor, 1997 – present)

ISCB/Springer Series in Computational Biology (Advisory Board, 1998 – present)

Human Genomics (Editorial Board, 2003 – 2024)

Science Translational Medicine (Advisory Board, 2012 – 2015)

MathematicS In Action (Editorial Board, 2008 – 2013)

Cancer Prevention Research (Editorial Board, 2008 – 2014)

Annals of Human Genetics (Associate Editor, 2003 – 2013)

Genetics (Associate Editor, 1997 – 2003)

Journal of Mathematical Biology (Editorial Board, 2000 – 2003)

Bernoulli (Associate Editor, 2000 – 2003)

Combinatorics, Probability and Computing (Editorial Board, 1996 – 2001)

Stochastic Models (Editorial Board, 1991 – 2001)

Mathematical Biosciences (Editorial Board, 1990 – 2001)

Stochastic Processes and their Applications (Associate Editor, 1989 –1993)

Theoretical Population Biology (Associate Editor, 1985 –1987)

Reviewer for 30+ journals in statistics, probability, combinatorics, biology and medicine

Member, Biological Sciences Funding Committee, CRUK (2008 – 2012)

Referee for numerous grant proposals (e.g. ERC, CRUK, BBSRC, EPSRC, NSF, NIH, NSA, Wellcome Trust, Keck, Swiss Nationalsfond)

LAB MEMBERS, PAST & PRESENT

Postdoctoral Research Associates

I. Vázquez García (Columbia, 2018 – present; joint with S. Shah, MSKCC)

B. Trippe (Columbia, 2022 – present)

M. de Barros (Columbia, 2024 – present; joint with K. Nowicki-Osuch, DKFZ)

Oesophageal Adenocarcinoma ICGC

V. Devonshire (Bioinformatics support, 2016–present)

Research Students

R. Kunes (PhD 2019–2024; NSF GRF)

B. Wesley (PhD 2020–2025; joint with Jellert Gaublomme)

Interns

H. Mohammed-Yassine (École Polytechnique, Apr - Jul 2024)

J. Engel (Columbia College, Sep - Dec 2022)

F. Hanau (Columbia CS, Jan - Dec 2022)

E. Bergomi (École Polytechnique, Mar - Jul 2022)

Former Postdoctoral Fellows

M.S. McPeek (1994–1995) Professor of Statistics, University of Chicago.

P. Marjoram (1995–1998) Professor of Research Preventive Medicine, USC Keck School of Medicine.

J. K. Lee (1995–1997) Senior Member, Moffitt Cancer Centre.

S. Schbath (1996) Directrice de Recherche 1‘ere classe, Unit Applied Mathematics and Computer Science, from Genomes to the Environnement (MaIAGE), INRAE, Jouy-en-Josas, France.

A. Pluzhnikov (1996–1998) Research Associate, Department of Neurology, University of Chicago.

S. Kruglyak (1998–2000) VP of Informatics, Element Biosciences, San Diego.

D. Golinelli (2000–2002) Research Professor, University of Pennsylvania, School of Nursing, Center for Health Outcomes and Policy Research, Philadelphia, PA.

N. Rosenberg (2001–2004) Stanford Professorship in Population Genetics and Society, Department of Biology, Stanford University

P. Calabrese (2001–2004) Associate Professor (Teaching), Quantitative and Computational Biology, USC.

10. P. Nicolas (2004–2005) Researcher in MaIAGE laboratory, INRAE, Jouy-en-Josas, France

C. Rangel Escareño (2003–2006) Principal Investigator, Computational and Integrative Biology, Instituto Nacional de Medicina, Mexico.

A. E. Teschendorff (Cambridge, 2005–2006) CAS-MPG Computational Biology Institute & Shanghai Institute of Nutrition and Health, and UCL Cancer Institute.

N. P. Thorne (Cambridge, 2004–2008) Lead, Genomic Information Technology, Melbourne Genomics Health Alliance.

I. Tiemann-Boege (Cambridge, 2007–2008) Associate Professor and Group Leader, Institute of Biophysics, Johannes Kepler University, Linz, Austria.

M. E. Ritchie (Cambridge, 2006–2008) Laboratory Head, Molecular Medicine, WEHI, Melbourne.

N. Barbosa-Morais (Cambridge, 2006–2010) Research Group Leader, IMM, Lisbon, Portugal.

C. N. Curtis (Cambridge, 2007–2010) Professor of Genetics and Medicine, Stanford University.

Q. A. Fu (Cambridge, 2008–2011) Associate Professor, Department of Mathematics and Statistical Science, University of Idaho.

N. Marko (AANS Van Wagenen Fellow. Cambridge, 2011–2012) Director of Neurosurgery at LewisGale Regional Health System, Roanoke, Virginia.

20. E. Turro (Cambridge, 2011–2012) Associate Professor, Genetics and Genomic Sciences, Mount Sinai Health System, New York.

B. Carvalho (Cambridge, 2010–2012) Professor, Computational Biology and Biostatistics Laboratory, University of Campinas, Sao Paulo, Brazil

N. Shannon (Cambridge, 2011–2013; joint with Fitzgerald ICGC group) General Surgery Resident, SingHealth, Singapore.

I. Spiteri (Senior Scientific Officer; Cambridge, 2009–2014) Roche.

S. Samarajiwa (Cambridge, 2008–2014). Senior Lecturer in Genomics, Department of Metabolism, Digestion and Reproduction, Imperial.

C. Massie (2013–2014; joint with Neal ICGC group) Group Leader, CRUK Cambridge Centre Early Detection Programme, Cambridge University. † 5.5.23

A. Achilleos (2013–2015; joint with Fitzgerald ICGC group) Associate Senior Bioinformatician, NIPD Genetics, Cyprus.

A. Touloumis (Cambridge, 2011–2015; joint with J. Marioni, EBI) Principal Lecturer, School of Architecture, Technology and Engineering, University of Brighton.

M. L. Smith (Cambridge, 2013–2015; FP7 grant) Senior Scientific Software Developer, EMBL.

M. Secrier (2014 – 2016; joint with Fitzgerald ICGC group) Associate Professor, Department of Genetics, Evolution and Environment, UCL.

30. E. Morrissey (Cambridge, 2011 – 2016) Director, Data Sciences and Quantitative Biology, AstraZeneca, Cambridge.

A. G. Lynch (Associate Scientist, CI, 2006 – 2017) Professor of Statistics in Bioscience, School of Mathematics and Statistics, University of St Andrews.

J. Truszkowski (Cambridge, 2014 – 2017; joint with N. Goldman, EBI) Researcher, Department of Biology and Environmental Sciences, University of Gothenburg.

A. Larionov (Cambridge, 2013 – 2018; joint with M. Tischkowitz. ERC grant) Lecturer in Bioinformatics, School of Water, Energy and Environment, Cranfield University, UK.

J. Perner (2015 – 2018; joint with Fitzgerald ICGC group) Senior Principal, Novartis Institutes for BioMedical Research (NIBR), Basel.

D. Biasci (Cambridge, 2015 – 2018) Group Leader (Innovation Investigator track), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal, University of Oxford. Sciences

S. Field (Cambridge, 2015–2019) Research Associate, Metzakopian Lab, UK Dementia Research Centre, University of Cambridge.

S. Jammula (2017–2020; joint with Fitzgerald ICGC group) Associate Director, Bioinformatics. MedGenome Labs Ltd., Bengaluru, Karnataka, India

M. Yin (Columbia, 2020 – 2022; joint with D. Blei. DSI Fellowship) Assistant Professor, Marketing Department, Warrington Business School, University of Florida.

A. Ng (2019–2022; joint with R. Fitzgerald) President's Fellowship, NTU, Singapore.

40. A. Jamshidpey (Columbia, 2020 – 2022) Associate Research Scientist, IICD, Columbia University.

K. Dinh (Columbia, 2019 – 2023) Associate Research Scientist, IICD, Columbia University.

P. da Silva (Columbia, 2020 – 2023) Associate Research Scientist, IICD, Columbia University.

Former Research Students

P. J. Joyce (PhD 1988) *Age-ordered distributions for population genetics models.* Dean of College of Science and Professor of Mathematics and Statistics, University of Idaho. † 26.9.58 – 22.4.16

R. S. Lundstrom (PhD 1990) *Stochastic models and statistical methods for DNA sequence data.* Senior Software Engineer in Data Science, T.D. Williamson, Salt Lake City.

L. Markovtsova (PhD 2000) *Markov chain Monte Carlo methods in population genetics.* Senior Director, Early Development Analytics, Novartis. Basel, Switzerland

O. Will (PhD 2001) *Statistical inference in the fossil record.* Cerner Enviza.

J. Wu (PhD 2001) *Statistical inference for molecular data: man, motifs and microarrays.*

K. Millman (MD; PhD 2005) *Molecular and genetic epidemiology of* Chlamydia trachomatis *in the United States.* The Millman Clinic, San Jose.

D. Abdueva (PhD 2006) *Probe-target hybridization modeling and its application to the analysis of microarrays.* Founder & CEO, Aqtual Inc, CA.

D. Skvortsov (PhD 2006) *Detection of deleted and duplicated genomic DNA using HMM analysis of GeneChip data.* Product Development Scientist, ABB Group, CA.

W. Xu (PhD 2006) *Searching for and beyond replication origins.* Visiting Professor, Biochemistry - Genome Center, Stanford University.

10. O. Sargsyan (PhD 2006) *Analytical and simulation results for the general coalescent.* Senior Bioinformatics Scientist, Centrillion Biosciences Inc., Palo Alto, CA.

V. Plagnol (PhD 2006) *Ancestral inference from molecular and paleontological data.* Head of Precision Health, Genomics plc

R. Jiang (PhD 2006) *Modeling and analyzing novel genetic data: single feature polymorphisms and polony sequences.* Biostatistic Director, GRAIL Inc.

R. Alaghband (PhD 2007) *Inference for stochastic models of molecular data.* Senior Staff Lead, Data Science, Google, Los Angeles.

C. Curtis (PhD 2007) *Analysis of high-density oligonucleotide gene expression data for dissecting aging pathways.* Professor of Genetics and Medicine, Stanford University.

J.-C. Marioni (PhD 2007) *Statistical methods for array CGH and copy number variation experiments.* SVP and Head of Computation, gRED, Genentech, San Francisco.

R. Wilkinson (PhD 2007) *Bayesian inference of primate divergence times.* Professor of Statistics, School of Mathematical Sciences, University of Nottingham.

M. Dunning (PhD 2008) *Genome-wide analyses using bead-based microarrays.* Bioinformatics Core Director, Sheffield Institute for Translational Neuroscience, University of Sheffield.

T. Hardcastle (PhD 2009) *Model-driven analysis of high-throughput genomic data in late-stage ovarian cancer.* [Joint with J. Brenton] Machine Learning Researcher, BIOS. † 18.3.20

I. Manolopoulou (PhD 2009) *A Bayesian approach to Nested Clade Analysis.* Professor, Department of Statistical Science, UCL.

20. D. Grover (PhD 2009) *Real-time tracking and analysis of Drosophila behavior and gene expression.* Assistant Researcher, Kavli Institute for Brain and Mind, UCSD.

L. Goldstein (PhD 2010) *Statistical analysis of microRNA expression and related data.* Deputy Head, Data Science Platform, Garvan Institute of Medical Research, Sydney, Australia.

C. Spyrou (PhD 2010) *Development and application of Bayesian methodology for some missing data problems in biology.* Principal Statistician, Nielsen Holdings, Cyprus.

S. Ivakhno (PhD 2011) *A statistical framework for the analysis of copy number aberrations in high-throughput cancer genomic data.* ML Tech lead and Engineering Manager, Captify, London.

S. Knott (PhD 2011) *Measuring, modeling and identifying factors that influence eukaryotic DNA replication.* [Joint with O. Aparicio] Assistant Professor, Biomedical Sciences, Cedars Sinai Medical Institute.

D. Speed (PhD 2011) *Exploring non-linear regression methods, with application to association studies.* Professor, Center for Quantitative Genetics and Genomics, Aarhus University, Denmark.

J. Woolford (PhD 2011) *Statistical analysis of small RNA high-throughput sequencing data.* [Joint with E. A. Miska]

A. Sottoriva (PhD 2012) *Spatial cell ancestry inference: determining in vivo cancer dynamics from patient molecular data.* Head of Computational Biology Research Centre, Human Technopole, Milan.

R. Ardekani (PhD 2013) *Computer vision approaches to the analysis of animal behavior.* Machine Learning/Computer Vision, Apple.

M. L. Smith (PhD 2013) *Low-level artefacts affecting microarrays and next-generation sequencing in a cancer genomics environment.* Data Management Coordinator, EMBL.

30. T. Nellimoottil (PhD 2013) *Computational analysis of factors affecting timing of replication and fork dynamics in S. cerevisiae* [Joint with O. Aparicio] Director of Data Science, ML Platform/GTM/Strategic Partnerships, Signifyd, Los Angeles.

J. Cairns (PhD 2014) *Identification of functional binding elements through integrative ChIP-seq and expression analysis.* Associate Principal Statistician, AstraZeneca.

D. Andrews (PhD 2015) *Statistical models of PCR for quantification of target DNA by sequencing.* Staff Bioinformatics Scientist, Illumina Inc.

M. Josephidou (PhD 2017) *Statistical models and algorithms for inferring single nucleotide variants from multi-sample cancer genomic data.* Machine Learning Software Engineer, Meta.

M. Abbasi (PhD 2017; USC, joint with P. Marjoram) *Automatic tracking of flies and the analysis of fly behavior.* Associate Director, Exact Sciences, San Diego.

A. Wendler (PhD 2019; joint with C. Watts) *The evolution of glioblastoma: a single case study.* Research Scientist (Oncology & Cell Biology), Vernalis, Cambridge.

E. Williams (PhD 2020) *Statistical analysis of renal function in patients with cancer.* Power Analyst, Nanook Energy Advisors LLP, London.

M. Gürel (PhD 2020) *Intratumoral B and T cell receptors: Reconstruction and analysis.* Senior Bioinformatics Data Scientist, Benevolent AI, London.

Supervised 26 MS thesis students (1985 – 2006)

Conference Organization (since 2016)

Co-organizer (with M. Kimmel) Minisymposium "Stochastic models of cancer: An update of theory and data". SMB 2021 Annual Meeting. June 14 2021.

Organizing Committee, CNIO La Caixa Frontiers Meetings 2019, "Heterogeneity and Evolution in Cancer." Madrid, September 23–25 2019.

INVITED CONFERENCE LECTURES

Lecture Series (since 2018; from 19 since 1990)

An introduction to Approximate Bayesian Computation (ABC). Statistics Department, Columbia University. (10 hours) 6/4–18/2019.

Qbio2019. Rice University. (4 hours) 7/16–18/2019.

Invited Conference Lectures (since 2018; from 170 since 1980)

Invited Speaker, Modeling and Theory in Population Biology BIRS Hybrid Thematic Program Symposium. 5/8/2024.

Modeling and simulation of cancer evolution in single cells. Invited lecture, Annual Congress of the Ecole de l'INSERM Liliane Bettencourt. Paris. 3/18–19/2023.

A probabilistic model for shallow whole-genome single cell DNA sequencing. Invited lecture, Extremes and Time Series: A Workshop on the Occasion of Richard Davis' 70th Birthday. Columbia University, 1/20–21/2023.

Modeling and simulation of cancer evolution in single cells. Keynote speaker, USC Computational Biology Conference 2022. Los Angeles. 5/19–21/2022.

Modeling and simulation of cancer evolution in single cells. Conference on "Evolution of cancer – reconstructing the past, predicting the future". Bad Honnef, Germany. 3/21–25/2022.

Predicting response to cancer immunotherapy. BETRNet Research Center 1 Annual Meeting. 5/6/2021.

Predicting response to immunotherapy. 5th Annual Engineering in Medicine Symposium, Columbia University. 2/18/2021.

Old data, new tricks: fibroblastic reticular cells predict response to cancer immunotherapy independently of T and B cells. New York Stem Cell Foundation Conference. 10/20–21/2020.

Chris Cannings and the coalescent. Chris Cannings Memorial Symposium, School of Mathematics, University of Sheffield. 11/16/2019.

Studying tumors in 3D. New perspectives on cancer through the lens of artificial intelligence and data science. Evening Talks series, New York Genome Center. 11/13/2019.

Computational cancer dynamics. Keynote Lecture, IMMUNORAD 4.0 (4th International Conference on Radiotherapy Immunotherapy Combinations. Weill-Cornell Medicine. New York, 9/14/2019.

Invited panelist on "Interdisciplinary research and leadership: how to make an impact in the data science age." JSM, Denver. 7/31/2019.

Imagining cancer. International Conference on Mathematics and Culture. Venice, 3/29–31/2019.

Some statistical problems in cancer genomics. Murray and Adylin Rosenblatt Endowed Lecture in Applied Mathematics. UC San Diego, 2/7/2019.

The combinatorics of spaghetti hoops. Cambridge Philosophical Society Lecture. Cambridge, 11/19/2018.

Studying tumours in 3.5D. Gliwice Scientific Meetings 2018. Gliwice, 11/17/2018.

Computational cancer genomics: NYGC and beyond. New York Cancer Genomics Research Network Meeting, NYGC. 6/11/2018.

Cancer at the crossroads. Statistics at a Crossroads: Challenges and Opportunities in the Data Science Era. NSF Workshop, Arlington VA. 10/15–17/2018.

Spaghetti hoops, Chinese restaurants and the magical Ewens Sampling Formula. Keynote address, CCBI Annual Symposium, DAMTP, Cambridge. 5/16/2018.

Inferring modes of cancer evolution from DNA sequencing data. Symposium on Cancer Genomics and Mathematical Data Analysis. Columbia University, NY. 2/7–8/2018.

Contributed Conference Lectures or Departmental Seminars

37 (since 2018; from 288 since 1980) in Australia, Austria, Canada, China, Denmark, England, France, Germany, Ireland, Portugal, Singapore, Sweden, Switzerland, UK, USA.

April 6, 2024