

MICHAEL O. DUFF

Staff Scientist

Program of Biostatistics, Bioinformatics and Epidemiology

Vaccine and Infectious Disease Division

Fred Hutchinson Cancer Research Center

Seattle WA

moduff@fredhutch.org

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

*Machine Learning, Bioinformatics/Computational Biology/Computational Genomics,
Reinforcement Learning Theory, Bayesian methods, Applied Statistics/Mathematics,
Deep-Learning, Multi-armed bandit problems, Optimization and Operations Research*

EDUCATION

Ph.D., Computer Science, University of Massachusetts at Amherst, 2002. Concentration: Machine Learning, Decision Theory, Applied Probability, Reinforcement Learning for Adaptive Prediction and Control (Andrew Barto, advisor).

Master of Science, Systems Engineering, University of Arizona. Concentration: Applied Mathematics, Optimization Theory and Control Theory.

Bachelor of Science, Systems Engineering, University of Arizona. Concentration: Operations Research and Numerical Analysis.

AT&T Bell Laboratories Course Work: Digital Signal Processing, Quantization Theory.

EMPLOYMENT

Program of Biostatistics, Bioinformatics and Epidemiology, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle WA—As a Staff Scientist, I’m currently conducting mathematical modelling and bioinformatic analyses relating to B-cell receptor repertoire dynamics in response to vaccination. Dec 1, 2018 - present.

Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington CT—As an Assistant Professor in Residence, I served as the principal bioinformatics researcher/analyst for the department. Efforts were directed toward bringing computational approaches, statistical modeling methods, and bioinformatics tools to bear upon problems posed by the analysis of Next-Generation-Sequencing (NGS) data. This includes the standard repertoire of techniques drawn from computational genomics: genome assembly, sequence alignment, gene-expression analysis, motif finding, phylogenetic analysis, and identification of regulatory sequences and networks—as well as the development of special-purpose ad-hoc computational tools for particular applications and novel techniques for data visualization. Projects include analysis of RNAseq data for the modENCODE and ENCODE projects, application of machine learning techniques for gene network inference and quantitation of differential expression from time-series data, DNA/RNA/ChIP/CLIP/RIP-seq data analysis, RNAseq data modeling and analysis relating to alternative-splicing, transplicing, human stem-cell differentiation, non-coding RNAs, Archaeal CRISPR-Cas systems, Polyomavirus editing & transcript expression profiling, RNA interactions and functional intronic domains, Genome-Wide Association Studies (longevity), and new courses developed and taught on Bioinformatics and Machine Learning for Genomics. 2007-2017.

Bioinformatics Research Laboratory, Department of Computer Science, University of Massachusetts, Amherst—As a Senior Research Scientist, I investigated the application of machine learning classification approaches to variation detection (micro-array DNA resequencing) data and optimal probe design. June 2006-June 2007.

Institute for Genomics & Bioinformatics, and Department of Computer Science, University of California, Irvine—As a Visiting Research Fellow, I collaborated with biologists at CalTech and Machine Learning theorists at UC Irvine on the general problem of principled statistical inference of biologically-plausible genetic regulatory structures from sparse data (and collateral efforts to extend standard hidden Markov models to incorporate smooth models of regime change); these efforts appeal to Bayesian graphical modeling techniques, and Markov Chain Monte Carlo computational methods. May 2005-May 2006.

Trinity College, Visiting Assistant Professor, Computer Science Department: Courses taught: Machine Learning, Computer Architecture, Intro to Internet Programming, August 2004-May 2005.

Gatsby Computational Neuroscience Unit, University College London: As a Senior Research Fellow, I developed theories and algorithms relating to probabilistic modeling, Bayesian learning, and Monte Carlo computation. I also investigated connections between Reinforcement Learning Theory and the dopamine system, Dec 2002-Dec 2003.

Department of Physiology, University of Connecticut Health Center—I worked as a developer contributing software to the Virtual Cell, a simulation tool for computational cell biologists. I also conducted analysis relating to stochastic modeling of membrane channels and filtering of intracellular transport imagery data, Spring-Summer 2002.

PUBLICATIONS

Bioinformatics / Computational Biology / Computational Genomics

Van Nostrand, E.L., et al (2019) A Large-Scale Binding and Functional Map of Human RNA Binding Proteins (To appear in *Nature*). <https://www.biorxiv.org/content/10.1101/179648v2>

Chang, H.H., Eibi, G., et al (2017) Incidence of Pancreatic Cancer is Dramatically Increased by a High Fat, High Calorie Diet in KrasG12D Mice. *PLoS ONE* 12(9).

Pilling, L.C., et al (2017) Red Blood Cell Distribution Width: genetic evidence for aging pathways in 116,666 volunteers. *PLoS ONE* 12(9).

Duff, M.O., Olson, S, Wei, X., Osman, A., Plocik, A., Bolisetty, M. Celniker, S. & Graveley, B.R. (2015) Genome-wide Identification of Zero Nucleotide Recursive Splicing in *Drosophila*. *Nature* 521, May 21, 376-9.

Garren, S., Duff, M.O., & Carmichael, G. (2015) Global analysis of mouse polyomavirus infection reveals dynamic regulation of viral and host gene expression and promiscuous viral RNA editing. *PLoS Pathogens*, 11(9), September 25.

Brooks, A.N., Duff, M.O., May, G., Yang, L., Landolin, J., Wan, K., Sandler, J., Celniker, S.E., Graveley, B.R. and Brenner, S.E. (2015) Regulation of alternative splicing in *Drosophila* by 56 RNA binding proteins. *Genome Research*, published online Aug 20.

Gerstein, M.B., et al (2014) Comparative analysis of the transcriptome across distant species. *Nature* 512, 445-448.

Chen, Z.X. et al (2014) Comparative validation of the *D. melanogaster* modENCODE Transcriptome Annotation. *Genome Research*. July;24(7):1209-23.

Brown, J.B. et al (many authors) (2014) Diversity and dynamics of the *Drosophila* transcriptome. *Nature*, published online March 16.

Eipper-Mains JE et al (2013) Effects of cocaine and withdrawal on the mouse nucleus accumbens transcriptome. *Genes Brain Behav.*12:21-33

Smibert P., et al. (2012) Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Reports*, Feb 23.

Yang L., Duff MO, Graveley BR, Carmichael GG and Chen L-L. (2011) Genomewide characterization of long nonpolyadenylated RNAs. *Genome Biology*, published online Feb 16

Roy, S. *et al* (many authors) (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330:1787-1797.

Graveley, B. *et al* (M.O. Duff co-first author) (2010) The developmental transcriptome of *Drosophila melanogaster*. *Nature*, Dec 22.

Cherbas, L., *et al* (many authors) (2010) The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Research*, Dec 22.

Brooks, A.N., Yang, L., Duff, M.O., Hansen, K.D., Dudoit, S., Brenner, S.E., Graveley, B.R. (2010) Conservation of an RNA Regulatory Map between *Drosophila* and Mammals. *Genome Research*, Oct 4.

McManus, C.J., Duff, M.O., Eipper-Mains, J., Graveley, B.R. (2010) Global Analysis of Transplicing in *Drosophila*. *Proc Natl Acad Sci U S A*. Jul 1.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R. and Wittkopp, P.J. (2010) Regulatory Divergence in *Drosophila* Revealed by mRNA-Seq. *Genome Research*. 20(6):816-25.

Hale, C.R, Zhao, P., Olsen, S., Duff, M., Graveley, B., Wells, L., Terns, R., & Terns, M., (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex, *Cell*, vol 139, Nov 25, pp 45-956.

Machine Learning & Reinforcement Learning

Duff, M.O. Chudova, D. Wold, B., Smyth, P., & Mjolsness, E. (2005) Statistical inference of biologically-plausible dynamic regulatory networks with core-leaf topology. *International Conference on Systems Biology*.

Niv, Y., Duff, M.O., Dayan, P. (2005) Dopamine, uncertainty, and TD-learning. *Behavioral and Brain Functions*, May 4 1:6 doi10.1186/1744-9081-1-6

Duff, M.O. (2003) Design for an optimal probe. *Proceedings of the 20th International Conference on Machine Learning*, 131-138.

Duff, M.O. (2003) Diffusion approximation for Bayesian Markov chains. *Proceedings of the 20th International Conference on Machine Learning*, 139-146.

Duff, M.O. (2002) Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes. *Ph.D. Thesis*, Dept. of Computer Science, Univ. of Massachusetts, Amherst.

Duff, M.O. (2001) Monte-Carlo Algorithms for the Improvement of Finite-State Stochastic Controllers: Application to Bayes-Adaptive Markov Decision Processes. Eighth International Workshop on Artificial Intelligence and Statistics, Key West FL.

Duff, M.O. & Barto, A. (1997) Local bandit approximation for optimal learning problems. *Advances in Neural Information Processing Systems —9*: 1019-1025.

Duff, M.O. (1995) Q-learning for bandit problems. *Proceedings of the 12th International Conference on Machine Learning*, 209-217.

Bradke, S. & Duff, M.O. (1995) Reinforcement learning methods for continuous-time Markov decision processes. *Advances in Neural Information Processing Systems—7*, 393-400.

Duff, M.O. (1994) Solving Bellman’s equation by the method of continuation. *Proceedings of the American Control Conference*, 2671-2682.

Barto A. & Duff, M.O. (1994) Monte-Carlo matrix inversion and reinforcement learning. *Advances in Neural Information Processing Systems*, vol 6, 687-694.

PRESENTATIONS

Gordon Research Conference, Post-Transcriptional Gene Regulation RNA: From Single Molecule to Transcriptome, July 10-15, 2016 Stowe, VT. “A ‘Bayes-balance’ estimate of conditional transcript localization probability from fractionated RNAseq data—assessing the localizing-influence of RNA binding-proteins”

Genome Informatics Meeting, Cold Spring Harbor October 27-30, 2009—“SPA, EXON_HITTER & WIG_INTEGRATOR: from RNA-seq to expression-feature fold-change and alternative-splicing analysis”

Janelia Farm conference: Improving the Toolkit for Drosophila Neurogenetics. Oct 4-7, 2009—“Developmental time-course of Drosophila neural gene-expression at single nucleotide resolution”

Janelia Farm conference: The Logic of Gene Regulation. May 25-28, 2008—“Computational identification of conserved long-range RNA structures involved in alternative splicing”

International Conference on Systems Biology, Boston, 2005—“Statistical inference of biologically-plausible dynamic regulatory networks with core-leaf topology”

AI and Statistics Conference, Key West, 2001—“Monte Carlo algorithms for the improvement of finite-state stochastic controllers”

OTHER

Gatsby fellowship, 2002-2003

NIH/BIT fellowship, 2005-2006

Program Committee member, 2007 International Conference on Machine Learning

Journal paper submission reviews: Bioinformatics, Nature Methods, Genome Biology, Journal of Machine Learning Research, PLoS ONE.

COURSES TAUGHT: Algorithms (*3 times*), Machine Learning, Bioinformatics Practicum (*3 times*), Machine Learning for Genomics (*twice*), Computer Architecture, Internet Programming, Great Ideas in Computer Science (*125 students*).