

Curriculum Vitae: Sean R. Eddy, Ph.D.

Office address

Department of Molecular and Cellular Biology
Harvard University
Northwest Building, Room 429
52 Oxford Street
Cambridge, MA 02138
Email: seaneddy@fas.harvard.edu
Web: <http://eddylab.org/>

Present position

Ellmore C. Patterson Professor of Molecular & Cellular Biology and
of Applied Mathematics, Harvard University;
Investigator, Howard Hughes Medical Institute

Education and training

1986	B.S., biology Dept. of Biology, California Institute of Technology
1991	Ph.D., molecular biology Dept. of Molecular, Cellular, Developmental Biology, University of Colorado, Boulder Thesis: <i>Introns in T-even Bacteriophage</i> Advisor: Larry Gold
1991-92	Postdoctoral Fellow, NeXstar Pharmaceuticals, Boulder CO Advisor: Larry Gold
1992-95	Postdoctoral Fellow, MRC Laboratory of Molecular Biology, Cambridge UK Advisors: Richard Durbin and John Sulston

Academic positions

1995-2000	Assistant Professor, Dept. of Genetics, Washington University School of Medicine
2000-2001	Alvin Goldfarb Professor of Computational Biology, Washington U. School of Medicine
2000-2007	Adjunct Professor, Dept. of Biomedical Engineering, Washington U. (by courtesy)
2000-2007	Adjunct Professor, Dept. of Computer Science, Washington U. (by courtesy)
2000-2007	Associate Professor, Department of Genetics, Washington U. School of Medicine
2000-2006	Investigator, Howard Hughes Medical Institute
2001-2007	Alvin Goldfarb Distinguished Professor of Computational Biology, Washington University School of Medicine
2006-2015	Group leader, HHMI Janelia Research Campus
2004-	Visiting Senior Fellow, Crick-Jacobs Center for Theoretical and Computational Biology, Salk Institute
2015-	Ellmore C. Patterson Professor of Molecular & Cellular Biology and of Applied Mathematics, Harvard University
2015-	Investigator, Howard Hughes Medical Institute

University service

1997-1999	Admissions committee, Washington U. Division of Biology & Biomedical Sciences (WU DBBS)
1997-2001	Molecular Genetics Program steering committee, WU DBBS
2000-2006	Computational Biology Program steering committee, WU DBBS
2001-2003	Director, WU DBBS Genome Analysis Training Program (NIH funded)

Honors and awards

1986	Bachelor of Science degree with Honors
1988-1991	National Science Foundation Graduate Student Fellowship
1991	U. of Colorado Graduate Student Research and Creative Work Award
1992-1994	Postdoctoral Fellowship, Human Frontier Science Program
1994-1995	Postdoctoral Fellowship, National Institutes of Health
1997	Eli Lilly Biochemistry Academic Contacts Committee award
2002-	Fellow of the Academy of Science of St. Louis
2007	Benjamin Franklin Award for Open Access in the Life Sciences
2014-	Fellow of the American Association for the Advancement of Science (AAAS)

Editorial boards

2002-2013	PLOS Biology
2000-2015	BMC Bioinformatics
2002-	Nucleic Acids Research

Service and advisory roles

1998-2000	Program committee; Intelligent Systems in Molecular Biology conference
2000	Program committee; International RNA Society Meeting
2001-2003	NIH NHGRI Genome Resources and Sequencing Priorities committee
2001-2003	US National Academy of Sciences committee, <i>Community Standards for Publication-Related Data and Materials</i>
2004-2007	National Advisory Council, NIH National Human Genome Research Institute
2005-2008	External Scientific Committee, Human Cancer Genome Atlas Project, National Cancer Institute & NIH National Human Genome Research Institute
2006-2010	<i>Science and the Media</i> project, American Academy of Arts and Sciences
2008-2011	JASON Defense Advisory Panel, United States government
2009-2011	US National Academy of Sciences committee, <i>Sequence-Based Classification of Select Agents: A Brighter Line</i>
2011-2012	Advisory committee; The Field Museum of Natural History, Chicago IL
2007-2013	Advisory board; NIH National Human Genome Research Institute Intramural Sequencing Center
2007-2013	Informatics Advisory Panel; NIH National Human Genome Research Institute
2008-2013	Board of Scientific Counselors; NIH National Human Genome Research Institute
2012-2013	Advisory board; NIH National Cancer Institute, Center for Cancer Genomics
2004-2015	Advisory board; FAS Systems Biology Center, Harvard University
2009-2015	Board on Life Sciences; US National Academy of Sciences
2012-2015	Member, Wellcome Trust Interview Panel (London, UK)
2010-	Advisory board; EMBOSS Project (EBI, Cambridge UK)

2012-	SAB chairman; Galaxy Project (NIH and Penn State University)
2012-	SAB chairman; RNACentral Project (European Bioinformatics Institute)
2014-	SAB member; BioFrontiers Institute (Boulder, CO)
2015-	Co-chair, Wellcome Trust Interview Panel (London, UK)
2015-	US National Academies Intelligence Science and Technology Experts Group

Consulting relationships and corporate advising

1997-1998	Consultant; Eli Lilly & Co. (Indianapolis IN)
1997-1999	Consultant; Monsanto Life Sciences (St. Louis MO)
1998-1999	Consultant; Pangea Systems (Oakland CA)
1999-2000	Consultant; Genetics Computer Group (Madison WI)
1999-2001	SAB member; Structural GenomiX (San Diego CA); acquired by Eli Lilly
1998-2002	SAB member; Orion Genomics (St. Louis MO)
1999-2002	SAB member; Paracel (Pasadena CA); acquired by Celera Genomics
1999-2011	SAB chairman; Divergence, Inc. (St. Louis MO); acquired by Monsanto
2013-2015	Consultant; Cooley LLP (Palo Alto, CA) and Gevo, Inc. (Douglas Co., CO)
2012-	SAB member; Maverix Biomics (Santa Cruz, CA)

Invited presentations: seminars (last 5 yrs)

2011	Institute for Advanced Study, Princeton NJ
2011	Johns Hopkins University, Baltimore MD
2011	University of Maryland, College Park MD
2011	University of Pittsburgh, Pittsburgh PA
2012	Dana Farber Cancer Center, Cambridge MA
2012	Lewis-Sigler Center, Princeton University, Princeton NJ
2012	University of Missouri, Columbia MO
2013	Harvard Medical School, Boston MA
2013	University of Virginia, Charlottesville VA
2014	J. Craig Venter Institute, Rockville MD
2015	Institute for Genome Sciences, University of Maryland

Invited presentations: meetings, symposia, and courses (last 5 yrs)

2011	Speaker, <i>A Decade with the Human Genome Sequence: Charting a Course for Genomic Medicine</i> , National Institutes of Health, Bethesda MD
2011	Speaker, <i>Visions of the Future</i> , University of Washington, Seattle
2012	Speaker, co-organizer, <i>Machine Learning, Statistical Inference, and Neuroscience</i> , Janelia Farm
2012	Speaker, <i>Workshop on Computational RNA Biology</i> , Benasque Institute of Science, Benasque, Spain
2013	Speaker, co-organizer, <i>Biological Sequence Analysis and Probabilistic Models</i> , Janelia Farm
2013	Keynote, <i>Bioinformatics Open Source Conference</i> , Berlin, Germany.
2013	Invitee, <i>Workshop to Evaluate Research Directions in Quantitative Biology</i> , Allen Institute, Seattle WA
2014	Speaker, <i>BioFrontiers Symposium: Big Data, Genomics and Molecular Networks</i> ,

	BioFrontiers Institute, U. Colorado at Boulder, Boulder CO
2014	Keynote, <i>High throughput sequencing for neuroscience</i> , HHMI Janelia, Ashburn VA.
2014	Co-organizer, <i>Computational RNA Biology</i> , European Bioinformatics Institute, Cambridge UK
2015	Speaker, <i>Workshop on Computational RNA Biology</i> , Benasque Institute of Science, Benasque, Spain

Teaching

1995-1998	Lecturer; Bio5495, Computational molecular biology
1995-2002	Lecturer; Bio5491, Advanced Genetics
1996-1997	Lecturer; Bio5011, Research ethics
1998-2005	Coursemaster; Bio5495/BME 537, Computational Molecular Biology

Graduate student trainees

1995-1999	Todd M. Lowe
Ph.D., 1999	<i>Combining new computational and traditional experimental methods to identify tRNA and snoRNA gene families</i> Subsequently: postdoctoral fellow, Stanford U., with David Botstein Now: Professor, UC Santa Cruz; Founder and Chief Scientific Officer, Maverix Biomics
1997-2002	Christian M. Zmasek
Ph.D., 2002	<i>Functional analyses of proteomes by phylogenetic methods</i> Subsequently: principal investigator, Genomics Institute of the Novartis Research Foundation Now: Staff scientist, Godzik lab, Sanford-Burnham Institute, La Jolla CA
1997-2002	Zhirong Bao
Ph.D., 2002	<i>Computational identification and characterization of repeats in sequenced eukaryotic genomes</i> Subsequently: postdoctoral fellow, U. Washington Seattle, with Bob Waterston Now: Faculty, Sloan-Kettering Memorial Cancer Center, New York NY
1999-2003	Robert J. Klein (HHMI predoctoral fellowship, 1998)
Ph.D., 2003	<i>Finding noncoding RNA Genes in genomic sequences</i> Subsequently: postdoctoral fellow, Rockefeller U., with Jürg Ott Now: Faculty, Icahn School of Medicine at Mount Sinai Hospital, New York NY
1998-2004	Robin Dowell (HHMI predoctoral fellowship, 1997; Olin fellowship, 2002)
M.S., 2001	<i>A distributed annotation system</i> . Coadvised by Lincoln Stein, Michael Brent.
Ph.D., 2004	<i>RNA structural alignment using stochastic context-free grammars</i> . Subsequently: postdoctoral fellow, MIT, with David Gifford and Tommi Jaakkola. Now: Assistant professor, University of Colorado at Boulder
1999-2006	Steve Lawery Johnson
Ph.D., 2006	<i>Remote protein homology detection using hidden Markov models</i> . Now: Bioinformatics scientist at Monsanto, St. Louis, MO

- 2000-2006 Shawn L. Stricklin
Ph.D., 2006 *Noncoding RNA genes in Caenorhabditis elegans.*
Now: Bioinformatics scientist at Monsanto, St. Louis, MO
- 2000-2006 John McCutcheon
Ph.D., 2006 *Screens for noncoding RNAs and strange lifeforms.*
Subsequently: postdoctoral fellow, U. Arizona, with Nancy Moran
Now: Assistant professor, University of Montana, Missoula MT
- 2004-2009 Eric P. Nawrocki
Ph.D., 2009 *Structural RNA homology search and alignment using covariance models.*
Now: Staff scientist, Eddy lab, HHMI Janelia Research Campus
- 2004-2010 Diana L. Kolbe (NSF Graduate Fellow)
Ph.D., 2010 *Novel algorithms for structural alignment of non-coding RNAs.*
Now: Research scientist, University of Iowa
- 2004-2011 Seolkyoung Jung (Samsung LoeKunHee Fellow, South Korea)
Ph.D., 2011 *Noncoding RNA genes in Oxytricha trifallax.*
Now: Postdoctoral fellow, HHMI Janelia Research Campus, with Sean Eddy

Postdoctoral trainees

- 1997-2000 Elena Rivas (Sloan postdoctoral fellow)
Now: senior research fellow, Harvard University
- 1998-1999 Linda Lutfiyya
Now: Bioinformatics scientist at Monsanto, St. Louis, MO
- 1999-2002 Jim McCarter (coadvised with Bob Waterston)
Subsequently: Founder and chief scientific officer, Divergence Inc., St. Louis MO.
Now: Senior entrepreneur in residence, BioGenerator, St. Louis, MO
- 2002-2004 Dawn Brooks (NSF postdoctoral fellow)
Subsequently: postdoctoral fellow, U. Florida, Gainesville FL, with Steven Benner
- 2004-2008 Alex Coventry
Subsequently: postdoctoral fellow, Cornell University, Ithaca NY, with Andy Clark
- 2006-2010 Sergi Castellano
Now: Faculty, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
- 2009-2014 Travis Wheeler
Now: Assistant professor, University of Montana, Missoula

2007-2015	Fred Davis Now: Staff scientist, John O'Shea lab, NIH
2011-2015	Seolkyoung Jung Now: Postdoc, Rafael Casellas lab, NIH
2009-	Gilbert (Lee) Henry

Patents

US Patent 7250556 Transposable elements in rice and methods of use (2007).

Publications (peer reviewed)

1. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. CM Rice, EM Lenches, SR Eddy, SJ Shin, RL Sheets, JH Strauss. *Science* 229:726-733, 1985.
2. Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. C Tuerk, S Eddy, D Parma, L Gold. *J. Mol. Biol.* 213: 749-761, 1990.
3. The phage T4 *nrdB* intron: a deletion mutant of a version found in the wild. SR Eddy, L Gold. *Genes & Development* 5:1032-1041, 1991.
4. An artificial mobile DNA element constructed from the EcoRI endonuclease gene. SR Eddy, L Gold. *Proc. Natl. Acad. Sci. USA* 89:1544-1547, 1992.
5. RNA sequence analysis using covariance models. SR Eddy, R Durbin. *Nucl. Acids Res.* 22:2079-2088, 1994.
6. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. DA Shub, H Goodrich-Blair, SR Eddy. *Trends Biochem. Sci.* 19:402-404, 1994.
7. Maximum discrimination hidden Markov models of sequence consensus. SR Eddy, G Mitchison, R Durbin. *J. Comput. Biol.* 2:9-23, 1995.
8. Multiple alignment using hidden Markov models. SR Eddy. *Proceedings of the 1995 conference on Intelligent Systems in Molecular Biology*, C. Rawlings *et al.*, eds, pp. 114-120. AAAI Press, Menlo Park CA. 1995.
9. Members of the immunoglobulin superfamily in bacteria. A Bateman, SR Eddy, C Chothia. *Protein Science* 5:1939-1941, 1996.
10. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. TM Lowe, SR Eddy. *Nucl. Acids Res.* 25:955-964, 1997.
11. A member of the immunoglobulin superfamily in bacteriophage T4. A Bateman, SR Eddy, VV Mesyanzhinov. *Virus Genes* 14:163-165, 1997.

12. Pfam: a comprehensive database of protein domain families based on seed alignments. ELL Sonnhammer, SR Eddy, R Durbin. *Proteins* 28:405-420, 1997.
13. Pfam: multiple sequence alignments and HMM profiles of protein domains. ELL Sonnhammer, SR Eddy, E Birney, A Bateman, R Durbin. *Nucl. Acids Res.* 26:322-325, 1998.
14. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Genome Sequencing Consortium. *Science* 282:2012-2018, 1998.
15. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. A Bateman, E Birney, R Durbin, SR Eddy, RD Finn, ELL Sonnhammer. *Nucl. Acids. Res.* 27:260-262, 1999.
16. A dynamic programming algorithm for RNA structure prediction including pseudoknots. E Rivas, SR Eddy. *J. Mol. Biol.* 285:2053-2068, 1999.
17. A computational screen for methylation guide snoRNAs in yeast. TM Lowe, SR Eddy. *Science* 283:1168-1171, 1999.
18. The Pfam protein families database. A Bateman, E Birney, R Durbin, SR Eddy, KL Howe, ELL Sonnhammer. *Nucl. Acids Res.* 28:263-266, 2000.
19. The language of RNA: A formal grammar that includes pseudoknots. E Rivas, SR Eddy. *Bioinformatics* 16: 334-340, 2000.
20. Homologs of small nucleolar RNAs in Archaea. AD Omer, TM Lowe, AG Russell, H Ebhardt, SR Eddy, PP Dennis. *Science* 288:517-522, 2000.
21. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. E Rivas, SR Eddy. *Bioinformatics* 16:573-585, 2000.
22. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* 409:860-921, 2001.
23. ATV: Display and manipulation of annotated phylogenetic trees. CM Zmasek, SR Eddy. *Bioinformatics* 17:383-384, 2001.
24. A simple algorithm to infer gene duplication and speciation events on a gene tree. CM Zmasek, SR Eddy. *Bioinformatics* 17:821-826, 2001.
25. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. SJ Jones, DL Riddle, AT Pouzyrev, VE Velculescu, L Hillier, SR Eddy, SL Stricklin, DL Baillie, R Waterston, MA Marra. *Genome Res.* 11:1346-1352, 2001.
26. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. E Rivas, RJ Klein, TA Jones, SR Eddy. *Current Biology* 11:1369-1373, 2001.
27. The Distributed Annotation System. RD Dowell, RM Jokerst, A Day, SR Eddy, L Stein. *BMC Bioinformatics* 2:7, 2001.
28. Noncoding RNA gene detection using comparative sequence analysis. E Rivas, SR Eddy. *BMC Bioinformatics* 2:8, 2001.

29. The Pfam protein families database. A Bateman, E Birney, L Cerruti, R Durbin, L Etwiller, SR Eddy, S Griffiths-Jones, KL Howe, M Marshall, and ELL Sonnhammer. Nucl. Acids Res. 30:276-280, 2002.
30. Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus. W Speckmann, ZH Li, TM Lowe, SR Eddy, RM Terns, MP Terns. Current Biology 12:199-203, 2002.
31. RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs. CM Zmasek, SR Eddy. BMC Bioinformatics 3:14, 2002.
32. Non-coding RNA genes identified in AT-rich hyperthermophiles. RJ Klein, Z Misulovin, SR Eddy. Proc. Natl. Acad. Sci. USA 99:7542-7547, 2002.
33. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. SR Eddy. BMC Bioinformatics 3:18, 2002.
34. Automated de novo identification of repeat sequence families in sequenced genomes. Z Bao, SR Eddy. Genome Research 12:1269-1276, 2002.
35. Initial sequencing and comparative analysis of the mouse genome. Mouse Genome Sequencing Consortium. Nature 420:520-562, 2002.
36. Rfam: an RNA family database. S Griffiths-Jones, A Bateman, M. Marshall, A Khanna, SR Eddy. Nucl. Acids Res. 31:439-441, 2003.
37. An active DNA transposon family in rice. N Jiang, Z Bao, X Zhang, SR Eddy, SR McCouch, SR Wessler. Nature 421:163-167, 2003.
38. Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics. JP McCutcheon, SR Eddy. Nucl. Acids Res. 31:4119-4128, 2003.
39. The DNA sequence of human chromosome 7. LW Hillier *et al.* Nature, 424:157-164, 2003.
40. RSEARCH: Finding homologs of single structured RNA sequences. RJ Klein, SR Eddy. BMC Bioinformatics 4:44, 2003.
41. The Pfam protein families database. A Bateman, L Coin, R Durbin, RD Finn, V Hollich, S Griffiths-Jones, A Khanna, M Marshall, S Moxon, ELL Sonnhammer, DJ Studholme, C Yeats, SR Eddy. Nucl. Acids. Res. 32:D138-D141, 2004.
42. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. RD Dowell, SR Eddy. BMC Bioinformatics 5:71, 2004.
43. Circular box C/D RNAs in *Pyrococcus furiosus*. NG Starostina, S Marshburn, LS Johnson, SR Eddy, RM Terns, MP Terns. Proc. Natl. Acad. Sci USA 28:14097-14101, 2004.
44. Pack-MULEs: Transposon-mediated gene evolution in plants. N Jiang, Z Bao, X Zhang, SR Eddy, SR Wessler. Nature 30:569-573, 2004.
45. A model of the statistical power of comparative genome analysis. SR Eddy. PLoS Biology 3:e10, 2005.

46. Rfam: annotating non-coding RNAs in complete genomes. S Griffiths-Jones, S Moxon, M Marshall, A Khanna, SR Eddy, A Bateman. *Nucl. Acids Res.* 33:D121-D141, 2005.
47. Generation and annotation of the DNA sequences of human chromosomes 2 and 4. LW Hillier *et al.* *Nature* 434:724-731, 2005.
48. Kissing complex RNAs mediate interaction between the fragile-X mental retardation protein KH2 domain and brain polyribosomes. JC Darnell, CE Fraser, O Mostovetsky, G Stefani, TA Jones, SR Eddy, RB Darnell. *Genes Dev.* 19:903-918, 2005.
49. Pfam: clans, web tools and services. RD Finn, J Mistry, B Schuster-Bockler, S Griffiths-Jones, V Hollich, T Lassmann, S Moxon, M Marshall, A Khanna, R Durbin, SR Eddy, ELL Sonnhammer, A Bateman. *Nucl. Acids Res.* 34:D247-D251, 2006.
50. Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. RD Dowell, SR Eddy. *BMC Bioinformatics* 7:400, 2006.
51. Query-dependent banding (QDB) for faster RNA similarity searches. EP Nawrocki, SR Eddy. *PLoS Computational Biology* 3:e56, 2007.
52. The Pfam protein families database. RD Finn, J Tate, J Mistry, PC Coghill, SJ Sammut, H-R Hotz, G Ceric, K Forslund, SR Eddy, ELL Sonnhammer, A Bateman. *Nucl. Acids Res.* 36:D281-D288, 2008.
53. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. SR Eddy. *PLoS Comput. Biol.* 4:e1000069, 2008.
54. Probabilistic phylogenetic inference with insertions and deletions. E Rivas, SR Eddy. *PLoS Comput Biol.* 4:e1000172, 2008.
55. Rfam: Updates to the RNA families database. PP Gardner, J Daub, JG Tate, EP Nawrocki, DL Kolbe, S Lindgreen, AC Wilkinson, RD Finn, S Griffiths-Jones, SR Eddy, A Bateman. *Nucl Acids Res.* 37:D136-D240, 2009.
56. A survey of nematode SmY RNAs. TA Jones, W Otto, M Marz, SR Eddy, PF Stadler. *RNA Biol.* 6:5-8, 2009.
57. Local RNA structure alignment with incomplete sequence. DL Kolbe, SR Eddy. *Bioinformatics* 25:1236-1243, 2009.
58. Infernal 1.0: inference of RNA alignments. EP Nawrocki, DL Kolbe, SR Eddy. *Bioinformatics* 25:1335-1337, 2009.
59. A tool for identification of genes expressed in patterns of interest using the Allen Brain Atlas. FP Davis, SR Eddy. *Bioinformatics* 25:1647-1654, 2009.
60. The Pfam protein families database. RD Finn, J Mistry, J Tate, P Coghill, A Heger, JE Pollington, OL Gavin, P Gunasekaran, G Ceric, K Forslund, L Holm, ELL Sonnhammer, SR Eddy, A Bateman. *Nucleic Acids Res.* 38:D211-D222, 2010.
61. Hidden Markov model speed heuristic and iterative HMM search procedure. LS Johnson, SR Eddy, E Portugaly. *BMC Bioinformatics* 11:431, 2010.

62. Rfam: Wikipedia, clans, and the “decimal” release. PP Gardner, J Daub, J Tate, BL Moore, IH Osuch, S Griffiths-Jones, RD Finn, EP Nawrocki, DL Kolbe, SR Eddy, A Bateman. *Nucleic Acids Res.* 39:D141-D145, 2011.
63. HMMER web server: interactive sequence similarity searching. RD Finn, J Clements, SR Eddy. *Nucl. Acids Res.*, 39:W29-W37, 2011.
64. Exploiting *Oxytricha trifallax* nanochromosomes to screen for noncoding RNA genes. S Jung, EC Swart, PJ Minx, V Magrini, ER Mardis, LF Landweber, SR Eddy. *Nucl. Acids Res.*, 39:7529-7547, 2011.
65. Accelerated profile HMM searches. SR Eddy. *PLoS Comput. Biol.*, 7:e1002195, 2011.
66. Fast filtering for RNA homology search. DL Kolbe, SR Eddy. *Bioinformatics*, 27:3102-3109, 2011.
67. The Pfam protein families database. M Punta, PC Coghill, RY Eberhardt, J Mistry, J Tate, C Boursnell, N Pang, K Forslund, G Ceric, J Clements, A Heger, L Holm, EL Sonnhammer, SR Eddy, A Bateman, RD Finn. *Nucl. Acids Res.*, 40:D290-301, 2012.
68. A range of complex probabilistic models for RNA secondary structure prediction that include the nearest neighbor model and more. E Rivas, R Lang, SR Eddy. *RNA*, 18:193-212, 2012.
69. Cell type-specific genomics of *Drosophila* neurons. GL Henry, FP Davis, S Picard, SR Eddy. *Nucl. Acids Res.*, 40:9691-9704, 2012.
70. Dfam: A database of repetitive DNA based on profile hidden Markov models. TJ Wheeler, J Clements, SR Eddy, R Hubley, TA Jones, J Jurka, AFA Smit, RD Finn. *Nucl. Acids Res.*, 41:D70-D82, 2013.
71. Rfam 11.0: 10 years of RNA families. SW Burge, J Daub, R Eberhardt, L Barquist, EP Nawrocki, SR Eddy, PP Gardner, A Bateman. *Nucl. Acids Res.*, 41:D226-D232, 2013.
72. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. EC Swart, JR Bracht, V Magrini, P Minx, X Chen, Y Zhou, JS Khurana, AD Goldman, M Nowacki, K Schotanus, S Jung, RS Fulton, A Ly, S McGrath, K Haub, JL Wiggins, D Storton, JC Matese, L Parsons, WJ Chang, MS Bowen, NA Stover, TA Jones, SR Eddy, GA Herrick, TG Doak, RK Wilson, ER Mardis, LF Landweber. *PLoS Biol.*, 11:10, 2013.
73. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. J Mistry, RD Finn, SR Eddy, A Bateman, M Punta. *Nucl. Acids Res.*, 41:e121, 2013.
74. Transcription factors that convert adult cell identity are differentially Polycomb repressed. FP Davis, SR Eddy. *PLoS ONE*, 8:e63407, 2013.
75. nhmmer: DNA homology search with profile HMMs. TJ Wheeler, SR Eddy. *Bioinformatics*, 29:2487-2489, 2013.
76. Infernal 1.1: 100-fold faster RNA homology searches. EP Nawrocki, SR Eddy. *Bioinformatics*, 29:2933-2935, 2013.
77. Pfam: the protein families database. RD Finn, A Bateman, J Clements, P Coghill, RY Eberhardt, SR Eddy, A Heger, K Hetherington, L Holm, J Mistry, ELL Sonnhammer, J Tate, M Punta. *Nucl. Acids Res.* 42:D222-230, 2014.

78. Rfam 12.0: updates to the RNA families database. EP Nawrocki, SW Burge, A Bateman, J Daub, RY Eberhardt, SR Eddy, EW Floden, PP Gardner, TA Jones, J Tate, RD Finn. *Nucl. Acids Res.*, 43:D130-137, 2015.
79. Epigenomic signatures of neuronal diversity in the mammalian brain. A Mo, EA Mukamel, FP Davis, C Luo, GL Henry, S Picard, MA Urich, JR Nery, TJ Sejnowski, R Lister, SR Eddy, JR Ecker, J Nathans. *Neuron*, 86:1369–84, 2015.
80. Combinatorial DNA rearrangement facilitates the origin of new genes in ciliates. X Chen, S Jung, LY Beh, SR Eddy, LF Landweber. *Genome Biol. Evol.*, accepted pending revision, 2015.
81. Parameterizing sequence alignment with an explicit evolutionary model. E Rivas, SR Eddy. Manuscript submitted, 2015.

Other publications (reviews, chapters, commentaries, book; non-peer-reviewed)

1. *Introns in the T-Even Bacteriophages*. SR Eddy. PhD thesis, University of Colorado at Boulder, Boulder CO, USA. 1991.
2. RNA: the shape of things to come. L Gold, C Tuerk, P Allen, J Binkley, D Brown, L Green, S MacDougall, D Schneider, D Tasset, SR Eddy. In *The RNA World*, RF Gesteland and JF Atkins, eds., pp. 497-509. Cold Spring Harbor Laboratory Press, New York. 1993.
3. Effects of bacterial growth conditions and physiology on T4 infection. E Kutter, E Kellenberger, K Carlson, S Eddy, J Neitzel, L Messinger, J North, B Guttman. In *Molecular Biology of Bacteriophage T4*, J Karam, ed., pp. 406-418. American Society for Microbiology, Washington, D.C. 1994.
4. The *C. elegans* genome project. SR Eddy. In *Advances in Molecular Plant Nematology*, F Lamberti, C DeGiorgi, D McK Bird, eds., pp. 3-18. Plenum Press, New York. 1994.
5. The Human Genome Project according to Los Alamos. SR Eddy. *Trends Biochem. Sci.* 20:91-92, 1994.
6. RNA structure alignment on a massively parallel computer. H Ellingworth, SR Eddy. In *Lecture Notes in Computer Science 919: High Performance Computing and Networking*, B Hertzberger and G Serazzi, eds, pp. 502-507. Springer, Berlin. 1995.
7. Genome Maps VI: *Caenorhabditis elegans*. M Chalfie, S Eddy, MO Hengartner, J Hodgkin, Y Kohara, RHA Plasterk, RH Waterston, JG White. *Science* 270:415, 1995.
8. Is the Pope the Pope? SR Eddy, DJC MacKay. *Nature* 382:490, 1996.
9. Molecular genetics in silico: Fourth International Conference on Intelligent Systems in Molecular Biology. SR Eddy. *Trends in Genetics* 12:372, 1996.
10. Hidden Markov models. SR Eddy. *Current Opinion in Structural Biology* 6:361–365, 1996.
11. Hidden Markov models and large-scale genome analysis. SR Eddy. *Transactions of the American Crystallographic Association*, 1997.

12. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. R Durbin, SR Eddy, A Krogh, GJ Mitchison. Cambridge University Press, Cambridge UK. 1998.
13. Multiple-alignment and -sequence searches. SR Eddy. Trends Guide to Bioinformatics pp. 15-18, 1998.
14. Profile hidden Markov models. SR Eddy. Bioinformatics 14:755-763, 1998.
15. Noncoding RNA genes. SR Eddy. Current Opinion in Genetics and Development 9:695-699, 1999.
16. Shotgun coverage of human genome computing. SR Eddy. Trends Biochem. Sci. 24:124, 1999.
17. Noncoding RNA genes and the modern RNA world. SR Eddy. Nature Reviews Genetics 2:919-929, 2001.
18. Computational genomics of noncoding RNA genes. SR Eddy. Cell 109:137-140, 2002.
19. A uniform system for microRNA annotation. V Ambros, B Bartel, DP Bartel, CB Burge, JC Car-
rington, X Chen, G Dreyfuss, SR Eddy, S Griffiths-Jones, M Marshall, M Matzke, G Ruvkun and T
Tuschl. RNA 9:277-279, 2003.
20. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*.
TR Cech, SR Eddy, D Eisenberg, K Hersey, SH Holtzman, GH Poste, NV Raikhel, RH Scheller, DB
Singer, MC Waltham, RA Schoen, BKB Avila, E Ben-Ari, N Grossblatt. National Research Council
report, National Academies Press, Washington DC, 2003.
21. What is dynamic programming? SR Eddy. Nature Biotech. 22:909-910, 2004.
22. Where did the BLOSUM62 alignment score matrix come from? SR Eddy. Nature Biotech. 22:1035-
1036, 2004.
23. What is Bayesian statistics? SR Eddy. Nature Biotech. 22:1177-1178, 2004.
24. What is a hidden Markov model? SR Eddy. Nature Biotech. 22:1315-1316, 2004.
25. How do RNA folding algorithms work? SR Eddy. Nature Biotech. 22:1457-1458, 2004.
26. *C. elegans* noncoding RNA genes. SL Stricklin, S Griffiths-Jones, SR Eddy. In: *WormBook*, doi/
10.1895/wormbook.1.7.1, <http://www.wormbook.org>, 2005.
27. Antedisciplinary science. SR Eddy. PLoS Computational Biology 1:e6, 2005.
28. Total information awareness for worm genetics. SR Eddy. Science 311:381-382, 2006.
29. Computational analysis of RNAs. SR Eddy. Cold Spring Harbor Symp. Quant. Biol. 71:117-128,
2006.
30. Open revolution. SR Eddy. PLoS Biology 7:e1000078, 2009.
31. Prepublication data sharing. Toronto International Data Release Workshop Authors. Nature 461:168-
170, 2009.

32. A new generation of homology search tools based on probabilistic inference. SR Eddy. *Genome Inform.* 23:205-211, 2009.
33. *Sequence-Based Classification of Select Agents: A Brighter Line.* JW LeDuc, R Baric, RG Breeze, RM Buller, SR Eddy, S Falkow, RE Levinson, J Mulligan, AD O'Brien, F Ochoa-Corona, JS Richardson, M Riley, T Slezak. National Research Council report, National Academies Press, Washington DC, 2010.
34. The C-value paradox, junk DNA, and ENCODE. SR Eddy. *Curr. Biol.*, 22:R898-R899, 2012.
35. The ENCODE project: missteps overshadowing a success. SR Eddy. *Curr. Biol.*, 23:R259-R261, 2013.
36. Computational identification of functional RNA homologs in metagenomic data. EP Nawrocki, SR Eddy. *RNA Biol.*, 10:1170-1179, 2013.
37. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. SR Eddy. *Ann. Rev. Biophys.*, 43:433-456, 2014.
38. Homology searches for structural RNAs: from proof of principle to practical use. SR Eddy. *RNA*, 21:605-607, 2015.

Other publications from the laboratory

1. Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. N Jiang, **Z Bao**, S Temnykh, Z Cheng, J Jiang, RA Wing, SR McCouch, SR Wessler. *Genetics* 161:1293-1305, 2002.
2. Accurate multiplex polony sequencing of an evolved bacterial genome. J Shendure, GJ Porreca, NB Reppas, X Lin, **JP McCutcheon**, AM Rosenbaum, MD Wang, K Zhang, RD Mitra, GM Church. *Science* 309:1728-1732, 2005.
3. Evolutionary models for insertions and deletions in a probabilistic modeling framework. **E Rivas**. *BMC Bioinformatics* 6:63, 2005.
4. Host pathogen protein interactions predicted by comparative modeling. **FP Davis**, DT Barkan, N Eswar, JH McKerrow, A Sali. *Protein Sci.*, 16:2585-2596, 2007.
5. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. C del Val, **E Rivas**, O Torres-Quesada, N Toro, JI Jiménez-Zurdo. *Mol. Microbiol.* 66:1080-1091, 2007.

I am not automatically a coauthor on all the work done in my laboratory, especially at Janelia. People in the laboratory are encouraged to have independence and to publish independently, either in single-authored form or in collaborations they initiate and take responsibility for. This "other publications from the laboratory" section is for when my CV is used to evaluate the research of my laboratory, as opposed to just me. It includes some papers that were the continuation of research from a previous laboratory, completed in my laboratory; papers that reflect someone's own independent research agenda while they were in my lab; and papers that are aligned to my research agenda, but were independent of substantive hands-on research from me. It also includes papers from Elena Rivas, who runs her own independent laboratory, jointly and in close coordination with mine. Most but not all of these were peer-reviewed; a couple are commentaries or reviews.

6. SelenoDB 1.0: a database of selenoprotein genes, proteins and SECIS elements. **S Castellano**, VN Gladyshev, R Guigó, MJ Berry. Nucl. Acids Res. 36:D332-D338, 2008.
7. Low exchangeability of selenocysteine, the 21st amino acid, in vertebrate proteins. **S Castellano**, AM Andrés, E Bosch, M Bayes, R Guigó, AG Clark. Mol. Biol. Evol. 26:2031-2040, 2009.
8. On the unique function of selenocysteine – insights from the evolution of selenoproteins. **S Castellano**. Biochim. Biophys. Acta. 1790:1463-1470, 2009.
9. MODBASE, a database of annotated comparative protein structure models and associated resources. U Pieper, N Eswar, BM Webb, D Eramian, L Kelly, DT Barkan, H Carter, P Mankoo, R Karchin, MA Marti-Renom, **FP Davis**, A Sali. Nucl. Acids Res. 37:D347-D354, 2009.
10. The overlap of small molecule and protein binding sites within families of protein structures. **FP Davis**, A Sali. PLoS Comp. Biol. 6:e1000668, 2010.
11. Proteome-wide prediction of overlapping small molecule and protein binding sites using structure. **FP Davis**. Mol Biosyst. 7:545-557, 2011.
12. Meeting report of the RNA Ontology Consortium, January 8-9, 2011. A Birmingham, JC Clemente, N Desai, J Gilbert, A Gonzalez, N Kyrpides, F Meyer, **E Nawrocki**, P Sterk, J Stombaugh, Z Weinberg, D Wendel, NB Leontis, C Zirbel, R Knight, A Laederach. Stand. Genomic Sci. 4:252-256, 2011.
13. RNIE: genome-wide prediction of bacterial intrinsic terminators. PP Gardner, L Barquist, A Bateman, **EP Nawrocki**, Z Weinberg. Nucl. Acids Res. 39:5845-5852, 2011.
14. Gene expression analysis in the parvalbumin-immunoreactive PV1 nucleus of the mouse lateral hypothalamus. F Girard, Z Meszar, C Marti, **FP Davis**, M Celio. Eur. J. Neurosci. 34:1934-1943, 2011.
15. Phosphorylation at the interface. **FP Davis**. Structure 19:1726-1727, 2011.
16. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. G Tanifuji, NT Onodera, **TJ Wheeler**, M Dlutek, N Donaher, JM Archibald. Genome Biol. Evol. 3:44-54, 2011.
17. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. D McDonald, MN Price, J Goodrich, **EP Nawrocki**, TZ DeSantis, A Probst, GL Andersen, R Knight, P Hugenholtz. ISME J. 6:610-618, 2012.
18. Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. DL Bernick, **PP Dennis**, M Höchsmann, TM Lowe. RNA 18:420-411, 2012.
19. The four ingredients of single-sequence RNA secondary structure prediction: a unifying perspective. **E Rivas**. RNA Biol. 10:1185-1196, 2013.
20. Special focus: bioinformatics. **EP Nawrocki**, SW Burge. RNA Biol. 10:1160, 2013.
21. Skyalign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. **TJ Wheeler**, J Clements, RD Finn. BMC Bioinformatics 15:7, 2014.

22. NSF workshop report: Discovering general principles of nervous system organization by comparing brain maps across species. GF Striedter, TG Belgard, CC Chen, **FP Davis**, BL Finlay, O Güntürkün, ME Hale, JA Harris, EE Hecht, PR Hof, HA Hofmann, LZ Holland, AN Iwaniuk, ED Jarvis, HJ Karten, PS Katz, WB Kristan, ER Macagno, PP Mitra, LL Moroz, TM Preuss, CW Ragsdale, CC Sherwood, CF Stevens, MC Stüttgen, T Tsumoto, W Wilczynski. J. Comp. Neurol., in press, 2014. (Also in Brain. Behav. Evol., in press, 2014.)
23. Annotating functional RNAs in genomes using Infernal. **EP Nawrocki**. Methods Mol. Biol. 1097:163–197, 2014.

Software packages & databases (partial list)

1. HMMER - hidden Markov models of protein and DNA sequence consensus.
Lead developer: SR Eddy.
<http://hmmer.org/>
2. Infernal - profile SCFGs for homology search and alignment for RNA secondary structure.
Lead developer: EP Nawrocki.
<http://eddylab.org/software/infernal/>
3. The Pfam database of conserved protein domains.
The Pfam Consortium: European Bioinformatics Institute; Stockholm University; HHMI Janelia.
<http://pfam.xfam.org/>
4. The RFAM database of conserved RNA families.
The Rfam Consortium: European Bioinformatics Institute; Manchester University; University of Canterbury; HHMI Janelia.
<http://rfam.xfam.org/>
5. QRNA - identification of novel noncoding RNA genes by comparative genome sequence analysis.
Lead developer: E Rivas.
<http://eddylab.org/software/#qrna>
6. tRNAscan-SE - software for detection of transfer RNA genes in genome sequence.
Lead developer: TMJ Lowe.
<http://eddylab.org/software/#trnascan>