

Sean R. Eddy, Ph.D.

Contact information

Department of Molecular & Cellular Biology email: seaneddy@fas.harvard.edu
Biological Laboratories 1008 web: eddylab.org
16 Divinity Ave github: github.com/EddyRivasLab
Harvard University orcid: [0000-0001-6676-4706](https://orcid.org/0000-0001-6676-4706)
Cambridge, MA 02138 USA

Present positions

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

Education and training

1992-1995 Postdoctoral Fellow (HFSP and NIH support)
MRC Laboratory of Molecular Biology, Cambridge UK
Probabilistic models for biological sequence analysis
Advisors: Richard Durbin, John Sulston

1991-1992 Postdoctoral Fellow
NeXstar Pharmaceuticals, Boulder CO
Computational analysis of RNA aptamers
Advisor: Larry Gold

1991 Ph.D., Molecular, Cellular, Developmental Biology (NSF graduate research fellow)
University of Colorado, Boulder
Thesis: *Introns in T-even Bacteriophage*
Advisor: Larry Gold

1986 B.S., Biology, with honors
California Institute of Technology
Undergraduate research: Genome sequence of yellow fever virus
Advisors: Charles Rice, James Strauss

Academic positions

2015- Ellmore C. Patterson Professor of Molecular & Cellular Biology, Harvard University
2000- Investigator, Howard Hughes Medical Institute (2000-06 and 2015-present);
Group leader, HHMI Janelia Research Campus, 2006-15)
2015-2022 Co-appointment as Professor in Applied Mathematics, Harvard University
2004-2018 Visiting Senior Fellow, Crick-Jacobs Center for Theoretical and Computational
Biology, Salk Institute
2015-2017 Associate member, Broad Institute of MIT and Harvard
1995-2007 Faculty member, Department of Genetics, Washington University School of Medicine:
Alvin Goldfarb Distinguished Professor of Computational Biology (2001-07);
Alvin Goldfarb Professor (2000-2001); Assistant Professor (1995-2000);

with courtesy adjunct appointments in Computer Science (2000-07) and Biomedical Engineering (2000-07)

University service at Harvard

2015- Faculty advisor, FAS Informatics and Scientific Applications Group
2025- Member, Star-Friedman Challenge Award committee
2025 Chair, senior lecturer renewal committee
2024-2025 Member, faculty search committee, Dept. of Organismic & Evolutionary Biology
2024 Chair, associate professor promotion committee
2024 Member, tenure promotion committee
2017-2024 Member, SEAS Interdisciplinary Committee on Applied Mathematics
2023 Member, tenure promotion committee
2020-2023 Chair, Department of Molecular & Cellular Biology, Harvard University
2019-2023 Member, MCB Community Task Force on Diversity, Inclusion, and Belonging
2019-2023 Chair, FAS Research Computing faculty advisory group
2022 Member, associate professor promotion committee
2021-2022 Member, FAS Pandemic Planning and Response Group
2019-2021 Chair, MCB senior faculty search committee
2020-2021 Member, Harvard Anti-Bullying Policy Working Group
2018-2020 Chair, SEAS Applied Mathematics PhD admissions committee
2019-2020 Member, Harvard life sciences PhD program review committee
2019-2020 Chair, MCB seminar and named lectures committee
2018-2020 Member, Harvard graduate student unionization Faculty Advisory Committee
2017-2019 Member, MCB curriculum committee
2018-2019 Member, Harvard Quantitative Biology Initiative faculty search committee
2017-2018 Member, Harvard Data Science Initiative Planning Committee
2017-2018 Chair, MCB faculty search committee
2017-2018 Director of Undergraduate Studies, SEAS Applied Mathematics, Harvard
2016-2017 Member, MCB faculty search committee
2015-2016 Co-chair, FAS Center for Systems Biology faculty search committee
2015-2016 Member, Statistics faculty search committee
2015-2016 Member, Harvard Data Sciences Initiative Education committee

University service at Washington University

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

Honors and awards

2023 Google Open Source peer bonus award
2023 Harvard College Professor award
2020-2023 Voted one of Harvard's favorite professors in the senior class yearbook
2022 Fellow, International Society for Computational Biology (ISCB)

2019 Nominee, Harvard Medical School Donald O'Hara Prize for Excellence in Teaching
 2014 Fellow, American Association for the Advancement of Science (AAAS)
 2007 Benjamin Franklin Award for Open Access in the Life Sciences
 2002 Fellow, Academy of Science of St. Louis
 1997 Eli Lilly Biochemistry Academic Contacts Committee award
 1994-1995 Postdoctoral fellowship, National Institutes of Health
 1992-1994 Postdoctoral fellowship, Human Frontier Science Program
 1991 U. of Colorado Graduate Student Research and Creative Work Award
 1988-1991 National Science Foundation Graduate Research Fellowship

Editorial boards

2015-2018 eLife
 2002-2017 Nucleic Acids Research
 2000-2015 BMC Bioinformatics
 2002-2013 PLOS Biology

Professional societies

2017- RNA Society
 2017- International Society for Computational Biology

Service and advisory roles

2021- Advisory board member, NSF-funded Interdisciplinary Quantitative Biology (IQ Biology), U Colorado, Boulder
 2021- Member, NIH GCAT study section
 2020- Advisory panel member, HHMI Hanna Gray Fellows Program interviews
 2014- SAB member, BioFrontiers Institute (Boulder, CO)
 2019-2020 Ad hoc member, NIH GCAT study section
 2019 Grant review committee, Chan-Zuckerberg Initiative (San Francisco, CA)
 2015-2018 Steering committee member, Probabilistic Modeling in Genomics meetings (PROBGEN)
 2015-2018 Co-chair, Wellcome Trust Interview Panel (London, UK)
 2012-2018 SAB chairman, RNACentral Project (European Bioinformatics Institute)
 2015-2016 US National Academies Intelligence Science and Technology Experts Group
 2012-2016 SAB chairman, Galaxy Project (NIH and Penn State University)
 2012-2015 Member, Wellcome Trust Interview Panel (London, UK)
 2009-2015 Board on Life Sciences, US National Academy of Sciences
 2004-2015 Advisory board, FAS Systems Biology Center, Harvard University
 2012-2013 Advisory board, NIH National Cancer Institute, Center for Cancer Genomics
 2010-2013 Advisory board, EMBOSS Project (EBI, Cambridge UK)
 2008-2013 Board of Scientific Counselors, NIH National Human Genome Research Institute
 2007-2013 Informatics Advisory Panel, NIH National Human Genome Research Institute
 2007-2013 Advisory board, NIH NHGRI Intramural Sequencing Center
 2011-2012 Advisory committee, The Field Museum of Natural History, Chicago IL
 2009-2011 US National Academy of Sciences committee, *Sequence-Based Classification of Select Agents: A Brighter Line*
 2008-2011 JASON Defense Advisory Panel, United States government

2006-2010 *Science and the Media* project, American Academy of Arts and Sciences
 2005-2008 External Scientific Committee, Human Cancer Genome Atlas Project,
 National Cancer Institute & NIH National Human Genome Research Institute
 2004-2007 National Advisory Council, NIH National Human Genome Research Institute
 2001-2003 US National Academy of Sciences committee, *Community Standards for
 Publication-Related Data and Materials*
 2001-2003 NIH NHGRI Genome Resources and Sequencing Priorities committee
 2000 Program committee, International RNA Society Meeting
 1998-2000 Program committee, Intelligent Systems in Molecular Biology conference

Consulting relationships and corporate advising

2012-2017 SAB member, Maverix Biomics (Santa Cruz, CA)
 2013-2015 Consultant, Cooley LLP (Palo Alto, CA) and Gevo, Inc. (Douglas Co., CO)
 1999-2011 SAB chairman, Divergence, Inc. (St. Louis MO); acquired by Monsanto
 1999-2002 SAB member, Paracel (Pasadena CA); acquired by Celera Genomics
 1998-2002 SAB member, Orion Genomics (St. Louis MO)
 1999-2001 SAB member, Structural GenomiX (San Diego CA); acquired by Eli Lilly
 1999-2000 Consultant, Genetics Computer Group (Madison WI)
 1998-1999 Consultant, Pangea Systems (Oakland CA)
 1997-1999 Consultant, Monsanto Life Sciences (St. Louis MO)
 1997-1998 Consultant, Eli Lilly & Co. (Indianapolis IN)

Invited presentations: seminars (last 5 yrs)

2024 Dept of Physics, Williams College
 2024 Dept of Genetics, Yale University
 2023 University of California, Davis
 2023 Clark University, Worcester MA
 2021 University of California, Berkeley

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

2024 Speaker, Gary Stormo Symposium on Computational Biology, St. Louis MO
 2024 Speaker, Advances in Genome Biology and Technology (AGBT), Orlando FL
 2023 Keynote, CSHL *Genome Informatics* conference
 2021 Lecturer, Woods Hole Physiology course (Nicole King, Daniel Fletcher, organizers)

Teaching

2016-2022 MCB112 Biological Data Analysis (mcb112.org)
 2017-2018 Co-lecturer; LS50 Integrated Science
 1998-2005 Bio5495/BME 537, Computational Molecular Biology
 1996-1997 Co-lecturer; Bio5011, Research ethics
 1995-2002 Co-lecturer; Bio5491, Advanced Genetics
 1995-1998 Co-lecturer; Bio5495, Computational molecular biology

Graduate students

2023- Liana Merk (NSF Graduate Fellow; Biophysics program, Harvard)

- 2022- Serafina Nieves (MCO program, Harvard)
- 2021- Nick Hilgert (SSQ Bio program, Harvard)
- 2021- Arpan Sarkar (Biostatistics program, Harvard)
- 2019-2023 Mary Richardson (NSF Graduate Fellow; SSQ Bio program, Harvard)
Ph.D. 2023 *Computational Prediction of Alternative Translation Events.*
Now: Computational biologist at Moderna (Cambridge, MA)
- 2016-2022 Yekaterina Shulgina (NIH F31 fellow; SSQ Bio program, Harvard)
Ph.D. 2022 *The Evolution of Alternative Genetic Codes.*
Now: Postdoctoral fellow with Jamie Cate, UC Berkeley
- 2016-2021 Caroline M. Weisman (joint with Andrew Murray; Biophysics program, Harvard)
Ph.D. 2021 *Novelty or nuisance? where lineage-specific genes come from and why it matters.*
Now: Lewis-Sigler Fellow, Princeton University
- 2016-2021 Grey Wilburn (Physics program, Harvard)
Ph.D., 2021 *An inverse statistical physics method for biological sequence analysis.*
Now: Data scientist with the Minnesota Twins baseball organization
- 2016-2019 Harleen Saini (joint with Melissa Moore, U. Mass. Medical School, Worcester)
Ph.D., 2019 *Intron and small RNA localization in mammalian neurons.*
Now: Postdoctoral fellow with Danesh Moazed, Harvard Medical School
- 2004-2011 Seolkyoung Jung (Samsung LoeKunHee Fellow, South Korea; WashU)
Ph.D., 2011 *Noncoding RNA genes in Oxytricha trifallax.*
Subsequently: Postdoctoral fellow, Eddy lab, 2011-2015
Now: Staff Scientist, Rafael Casellas lab, NIH NIAMS
- 2004-2010 Diana L. Kolbe (NSF Graduate Fellow; WashU)
Ph.D., 2010 *Novel algorithms for structural alignment of non-coding RNAs.*
Subsequently: Postdoctoral fellow with Laura Elnitski, NIH NHGRI
Now: Director of Bioinformatics, Iowa Institute of Human Genetics, University of Iowa
- 2004-2009 Eric P. Nawrocki (WashU)
Ph.D., 2009 *Structural RNA homology search and alignment using covariance models.*
Subsequently: Senior software engineer, Eddy lab, 2009-2015
Now: Staff scientist, NIH NCBI
- 2000-2006 John P. McCutcheon (WashU)
Ph.D., 2006 *Screens for noncoding RNAs and strange lifeforms.*
Subsequently: Postdoctoral fellow with Nancy Moran, U. Arizona

Now: Professor, Arizona State University, Tempe AZ

- 2000-2006 Shawn L. Stricklin (WashU)
Ph.D., 2006 *Noncoding RNA genes in Caenorhabditis elegans.*
Now: Senior Data Architect, Joyn Bio (Woodland, CA)
- 1999-2006 Steven L. Johnson (WashU)
Ph.D., 2006 *Remote protein homology detection using hidden Markov models.*
Subsequently: Postdoctoral fellow with Skip Virgin, Washington University in St. Louis
Now: Team Lead, Bayer Crop Science (St. Louis, MO)
- 1999-2003 Robert J. Klein (HHMI fellow; WashU)
Ph.D., 2003 *Finding noncoding RNA genes in genomic sequences*
Subsequently: Postdoctoral fellow with Jürg Ott, Rockefeller University
Now: Associate professor, Icahn School of Medicine at Mount Sinai Hospital, New York
- 1998-2004 Robin D. Dowell (HHMI fellow; Olin fellow; WashU)
M.S., 2001 *A distributed annotation system.* Co-advised by Lincoln Stein, Michael Brent.
Ph.D., 2004 *RNA structural alignment using stochastic context-free grammars.*
Subsequently: Postdoctoral fellow with David Gifford and Tommi Jaakkola, MIT
Now: Associate professor, University of Colorado at Boulder
- 1997-2002 Zhirong Bao (WashU)
Ph.D., 2002 *Computational identification and characterization of repeats in sequenced eukaryotic genomes*
Subsequently: Postdoctoral fellow with Bob Waterston, U. Washington Seattle
Now: Professor, Sloan-Kettering Memorial Cancer Center, New York
- 1997-2002 Christian M. Zmasek (WashU)
Ph.D., 2002 *Functional analyses of proteomes by phylogenetic methods*
Subsequently: PI, Genomics Institute of the Novartis Research Foundation
Now: Senior Bioinformatics Engineer, J. Craig Venter Institute (La Jolla, CA)
- 1995-1999 Todd M. Lowe (WashU)
Ph.D., 1999 *Combining new computational and traditional experimental methods to identify tRNA and snoRNA gene families*
Subsequently: Postdoctoral fellow with David Botstein, Stanford
Now: Professor, UC Santa Cruz

Postdoctoral trainees

- 2024- Kumar Sarthak
- 2023- Kumaresh Krishnan
- 2018- Jenny Chen (Harvard Data Science Initiative fellow; NIH K99/R00)
Jointly advised, with Hopi Hoekstra

- 2020-2023 Samantha Petti (Harvard NSF-Simons Center fellow; Burroughs Wellcome CASI fellow)
Now: Assistant professor, Dept of Mathematics, Tufts University
- 2015-2019 Peter Koo
Now: Assistant professor, Cold Spring Harbor Laboratory, New York
- 2016-2017 Praveen Anand (visiting scholar, Harvard South Asia Institute,
Bangalore to Boston Biosciences Beginnings (B4) program)
Now: Postdoctoral fellow with Jens Lohr, Dana Farber Cancer Institute
- 2009-2016 Gilbert Lee Henry
Now: Research Investigator with Tony Zador, Cold Spring Harbor Laboratory
- 2009-2014 Travis J. Wheeler
Now: Associate professor, U. Montana, Missoula
- 2007-2015 Fred P. Davis
Subsequently: Postdoctoral fellow with John O'Shea, NIH NIAMS
Now: Senior computational biologist, Celsius Therapeutics, Cambridge MA
- 2006-2010 Sergi Castellano
Subsequently: Group leader, Max Planck Institute for Evolutionary Anthropology, Leipzig
Now: Group leader, University College London
- 2005-2008 Gene Yeo (Salk Junior Fellow, coadvised with Rusty Gage)
Now: Professor, Dept. of Cellular and Molecular Medicine, UC San Diego
- 2004-2008 Alex Coventry
Subsequently: Postdoctoral fellow with Andy Clark, Cornell University
Now: Software/protocol engineer, Legicash (Boston, MA)
- 2002-2004 Dawn J. Brooks (NSF postdoctoral fellow)
Subsequently: Postdoctoral fellow with Steven Benner, U. Florida, Gainesville
Now: Doctor of Hematology & Oncology, Berkshire Medical Center, Massachusetts
- 1999-2002 James P. McCarter (coadvised with Bob Waterston)
Subsequently: Founder and chief scientific officer, Divergence Inc., Saint Louis
Now: Medical Director, New Analyte Ventures, Abbott Diabetes Care (Alameda, CA)
- 1998-1999 Linda Lutfiyya
Now: Head of Partnership Management, Biotech/Breeding, Bayer Crop Science (St. Louis)
- 1997-2000 Elena Rivas (Sloan postdoctoral fellow)
Subsequently: Fellow, HHMI Janelia Farm, Ashburn, Virginia

Now: Senior research fellow, Harvard University

Patents

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

Publications

105. Cellular evolution of the hypothalamic preoptic area of behaviorally divergent deer mice. J Chen, PR Richardson, C Kirby, SR Eddy, HE Hoekstra. *eLife*, 2024.
<https://doi.org/10.7554/eLife.103109.1>
104. ORFeus: A computational method to detect programmed ribosomal frameshifts and other non-canonical translation events. MO Richardson, SR Eddy. *BMC Bioinformatics*, 24:471, 2023.
103. Codetta: predicting the genetic code from nucleotide sequence. Y Shulgina, SR Eddy. *Bioinformatics*, 39:btac802, 2023.
102. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. CM Weisman, AW Murray, SR Eddy. *Curr. Biol.*, 32:2632-2639, 2022.
101. Constructing benchmark test sets for biological sequence analysis using independent set algorithms. SN Petti, SR Eddy. *PLOS Comput. Biol.*, 18:e1009492, 2022.
100. A computational screen for alternative genetic codes in over 250,000 genomes. Y Shulgina, SR Eddy. *eLife*, 10:e71402, 2021.
99. Rfam 14: expanded coverage of metagenomic, viral, and microRNA families. I Kalvari, EP Nawrocki, N Ontiveros-Palacios, J Argasinska, K Lamkiewicz, M Marz, S Griffiths-Jones, C Toffano-Nioche, D Gautheret, Z Weinberg, E Rivas, SR Eddy, RD Finn, A Bateman, AI Petrov. *Nucl. Acids Res.*, 49:D192-D200, 2021.
98. Remote homology search with hidden Potts models. GW Wilburn, SR Eddy. *PLoS Comput. Biol.* 16:e1008085, 2020.
97. Many but not all lineage-specific genes can be explained by homology detection failure. CM Weisman, AW Murray, SR Eddy. *PLoS Biology* 18:e3000862, 2020.
96. Estimating the power of sequence covariation for detecting conserved RNA structure. E Rivas, J Clements, SR Eddy. *Bioinformatics*, 36:3072-3076, 2020.
95. A genetic, genomic, and computational resource for exploring neural circuit function. FP Davis, A Nern, S Picard, MB Reiser, GM Rubin, SR Eddy, GL Henry. *eLife*, 9:e50901, 2020.
94. Inferring sequence-structure preferences of RNA-binding proteins with convolutional neural networks. P Koo, P Anand, S Paul, SR Eddy. *bioRxiv*, 2018.
<https://doi.org/10.1101/418459>

93. Free circular introns with an unusual branchpoint in neuronal projections. H Saini, AA Bicknell, SR Eddy, MJ Moore. *eLife*, 8:e47809, 2019.
92. Representation learning of genomic sequence motifs with convolutional neural networks. PK Koo, SR Eddy. *PLoS Comput. Biol.* 15:e1007560, 2019.
91. The Pfam protein families database in 2019. S El-Gebali, J Mistry, A Bateman, SR Eddy, A Luciani, SC Potter, M Qureshi, LJ Richardson, GA Salazar, A Smart, ELL Sonnhammer, L Hirsh, L Paladin, D Piovesan, SCE Tosatto, RD Finn. *Nucl. Acids Res.*, 47:D427-D432, 2019.
90. Group I introns are widespread in Archaea. EP Nawrocki, TA Jones, SR Eddy. *Nucl. Acids Res.*, 46:7970-7976, 2018.
89. HMMER web server: 2018 update. SC Potter, A Luciani, SR Eddy, Y Park, R Lopez, RD Finn. *Nucl. Acids Res.*, 46:W200-W204, 2018.
88. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. I Kalvari, J Argasinska, N Quinones-Olvera, EP Nawrocki, E Rivas, SR Eddy, A Bateman, RD Finn, AI Petrov. *Nucl. Acids Res.*, 46:D335-D342, 2018.
87. Identification and characterization of a class of MALAT1-like genomic loci. B Zhang, YS Mao, SD Diermeier, IV Novikova, EP Nawrocki, TA Jones, Z Lazar, C-S Tung, W Luo, SR Eddy, KY Sanbonmatsu, DL Spector. *Cell Reports*, 19:1723-1738, 2017.
86. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. E Rivas, J Clements, SR Eddy. *Nature Methods*, 14:45-48, 2017.
85. Epigenomic landscapes of retinal rods and cones. A Mo, C Luo, FP Davis, EA Mukamel, GL Henry, JR Nery, MA Urich, S Picard, R Lister, SR Eddy, MA Beer, JR Ecker, J Nathans. *eLife*, 5:e11613, 2016.
84. The Pfam protein families database: towards a more sustainable future. RD Finn, P Coggill, RY Eberhardt, SR Eddy, J Mistry, AL Mitchell, SC Potter, M Punta, M Qureshi, A Sangrador-Vegas, GA Salazar, J Tate, A Bateman. *Nucl. Acids Res.*, 44:D279-D285, 2016.
83. The Dfam database of repetitive DNA families. R Hubley, RD Finn, J Clements, SR Eddy, TA Jones, W Bao, AFA Smit, TJ Wheeler. *Nucl. Acids Res.*, 44:D81-D89, 2016.
82. Parameterizing sequence alignment with an explicit evolutionary model. E Rivas, SR Eddy. *BMC Bioinformatics*, 16:406, 2015.
81. Combinatorial DNA rearrangement facilitates the origin of new genes in ciliates. X Chen, S Jung, LY Beh, SR Eddy, LF Landweber. *Genome Biol. Evol.*, 7:2859-2870, 2015.
80. 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*, 86:1369-1384, 2015.
79. HMMER web server: 2015 update. RD Finn, J Clements, W Arndt, BL Miller, TJ Wheeler, F Schreiber, A Bateman, SR Eddy. *Nucl. Acids Res.* 43:W30-38, 2015.

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.
77. Pfam: the protein families database. RD Finn, A Bateman, J Clements, P Coggill, 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.
76. Infernal 1.1: 100-fold faster RNA homology searches. EP Nawrocki, SR Eddy. *Bioinformatics*, 29:2933-2935, 2013.
75. nhmmer: DNA homology search with profile HMMs. TJ Wheeler, SR Eddy. *Bioinformatics*, 29:2487-2489, 2013.
74. Transcription factors that convert adult cell identity are differentially Polycomb repressed. FP Davis, SR Eddy. *PLOS ONE*, 8:e63407, 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.
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.
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.
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.
69. Cell type-specific genomics of *Drosophila* neurons. GL Henry, FP Davis, S Picard, SR Eddy. *Nucl. Acids Res.*, 40:9691-9704, 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.
67. The Pfam protein families database. M Punta, PC Coggill, 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.
66. Fast filtering for RNA homology search. DL Kolbe, SR Eddy. *Bioinformatics*, 27:3102-3109, 2011.
65. Accelerated profile HMM searches. SR Eddy. *PLoS Comput. Biol.*, 7:e1002195, 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.
63. HMMER web server: interactive sequence similarity searching. RD Finn, J Clements, SR Eddy. Nucl. Acids Res., 39:W29-W37, 2011.
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.
61. Hidden Markov model speed heuristic and iterative HMM search procedure. LS Johnson, SR Eddy, E Portugaly. BMC Bioinformatics 11:431, 2010.
60. The Pfam protein families database. RD Finn, J Mistry, J Tate, P Coggill, 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.
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.
58. Infernal 1.0: inference of RNA alignments. EP Nawrocki, DL Kolbe, SR Eddy. Bioinformatics 25:1335-1337, 2009.
57. Local RNA structure alignment with incomplete sequence. DL Kolbe, SR Eddy. Bioinformatics 25:1236-1243, 2009.
56. A survey of nematode SmY RNAs. TA Jones, W Otto, M Marz, SR Eddy, PF Stadler. RNA Biol. 6:5-8, 2009.
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.
54. Probabilistic phylogenetic inference with insertions and deletions. E Rivas, SR Eddy. PLoS Comput Biol. 4:e1000172, 2008.
53. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. SR Eddy. PLoS Comput. Biol. 4:e1000069, 2008.
52. The Pfam protein families database. RD Finn, J Tate, J Mistry, PC Coggill, SJ Sammut, H-R Hotz, G Ceric, K Forslund, SR Eddy, ELL Sonnhammer, A Bateman. Nucl. Acids Res. 36:D281-D288, 2008.
51. Query-dependent banding (QDB) for faster RNA similarity searches. EP Nawrocki, SR Eddy. PLoS Computational Biology 3:e56, 2007.
50. Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. RD Dowell, SR Eddy. BMC Bioinformatics 7:400, 2006.

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.
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.
47. Generation and annotation of the DNA sequences of human chromosomes 2 and 4. LW Hillier *et al.* *Nature* 434:724-731, 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.
45. A model of the statistical power of comparative genome analysis. SR Eddy. *PLoS Biology* 3:e10, 2005.
44. Pack-MULEs: Transposon-mediated gene evolution in plants. N Jiang, Z Bao, X Zhang, SR Eddy, SR Wessler. *Nature* 30:569-573, 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.
42. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. RD Dowell, SR Eddy. *BMC Bioinformatics* 5:71, 2004.
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.
40. RSEARCH: Finding homologs of single structured RNA sequences. RJ Klein, SR Eddy. *BMC Bioinformatics* 4:44, 2003.
39. The DNA sequence of human chromosome 7. LW Hillier *et al.* *Nature*, 424:157-164, 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.
37. An active DNA transposon family in rice. N Jiang, Z Bao, X Zhang, H Hirochika, SR Eddy, SR McCouch, SR Wessler. *Nature* 421:163-167, 2003.
36. Rfam: an RNA family database. S Griffiths-Jones, A Bateman, M. Marshall, A Khanna, SR Eddy. *Nucl. Acids Res.* 31:439-441, 2003.
35. Initial sequencing and comparative analysis of the mouse genome. Mouse Genome Sequencing Consortium. *Nature* 420:520-562, 2002.
34. Automated de novo identification of repeat sequence families in sequenced genomes. Z Bao, SR Eddy. *Genome Research* 12:1269-1276, 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.

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.
31. RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs. CM Zmasek, SR Eddy. BMC Bioinformatics 3:14, 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.
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.
28. Noncoding RNA gene detection using comparative sequence analysis. E Rivas, SR Eddy. BMC Bioinformatics 2:8, 2001.
27. The Distributed Annotation System. RD Dowell, RM Jokerst, A Day, SR Eddy, L Stein. BMC Bioinformatics 2:7, 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.
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.
24. A simple algorithm to infer gene duplication and speciation events on a gene tree. CM Zmasek, SR Eddy. Bioinformatics 17:821-826, 2001.
23. ATV: Display and manipulation of annotated phylogenetic trees. CM Zmasek, SR Eddy. Bioinformatics 17:383-384, 2001.
22. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. Nature 409:860-921, 2001.
21. Secondary structure alone is generally not statistically significant for the detection of non-coding RNAs. E Rivas, SR Eddy. Bioinformatics 16:573-585, 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.
19. The language of RNA: A formal grammar that includes pseudoknots. E Rivas, SR Eddy. Bioinformatics 16: 334-340, 2000.
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.
17. A computational screen for methylation guide snoRNAs in yeast. TM Lowe, SR Eddy. Science 283:1168-1171, 1999.
16. A dynamic programming algorithm for RNA structure prediction including pseudoknots. E Rivas, SR Eddy. J. Mol. Biol. 285:2053-2068, 1999.

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.
14. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Genome Sequencing Consortium. Science 282:2012-2018, 1998.
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.
12. Pfam: a comprehensive database of protein domain families based on seed alignments. ELL Sonnhammer, SR Eddy, R Durbin. Proteins 28:405-420, 1997.
11. A member of the immunoglobulin superfamily in bacteriophage T4. A Bateman, SR Eddy, VV Mesyanzhinov. Virus Genes 14:163-165, 1997.
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.
9. Members of the immunoglobulin superfamily in bacteria. A Bateman, SR Eddy, C Chothia. Protein Science 5:1939-1941, 1996.
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.
7. Maximum discrimination hidden Markov models of sequence consensus. SR Eddy, G Mitchison, R Durbin. J. Comput. Biol. 2:9-23, 1995.
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.
5. RNA sequence analysis using covariance models. SR Eddy, R Durbin. Nucl. Acids Res. 22:2079-2088, 1994.
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.
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.
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.
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.

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

40. Mammalian cells repress random DNA that yeast transcribes. SR Eddy. *Nature*, 628:271-273, 2024.
39. Evolution: getting something for nothing. CM Weisman, SR Eddy. *Current Biology*, 27:R661-R663, 2017.
38. Homology searches for structural RNAs: from proof of principle to practical use. SR Eddy. *RNA*, 21:605-607, 2015.
37. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. SR Eddy. *Annu. Rev. Biophys.*, 43:433-456, 2014.
36. Computational identification of functional RNA homologs in metagenomic data. EP Nawrocki, SR Eddy. *RNA Biol.*, 10:1170-1179, 2013.
35. The ENCODE project: missteps overshadowing a success. SR Eddy. *Curr. Biol.*, 23:R259-R261, 2013.
34. The C-value paradox, junk DNA, and ENCODE. SR Eddy. *Curr. Biol.*, 22:R898-R899, 2012.
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.
32. A new generation of homology search tools based on probabilistic inference. SR Eddy. *Genome Inform.* 23:205-211, 2009.
31. Prepublication data sharing. Toronto International Data Release Workshop Authors. *Nature* 461:168-170, 2009.
30. Open revolution. SR Eddy. *PLoS Biology* 7:e1000078, 2009.
29. Computational analysis of RNAs. SR Eddy. *Cold Spring Harbor Symp. Quant. Biol.* 71:117-128, 2006.
28. Total information awareness for worm genetics. SR Eddy. *Science* 311:381-382, 2006.
27. Antedisciplinary science. SR Eddy. *PLoS Computational Biology* 1:e6, 2005.
26. *C. elegans* noncoding RNA genes. SL Stricklin, S Griffiths-Jones, SR Eddy. In: *WormBook*, 2005.
<https://doi.org/10.1895/wormbook.1.7.1>
25. How do RNA folding algorithms work? SR Eddy. *Nature Biotech.* 22:1457-1458, 2004.
24. What is a hidden Markov model? SR Eddy. *Nature Biotech.* 22:1315-1316, 2004.
23. What is Bayesian statistics? SR Eddy. *Nature Biotech.* 22:1177-1178, 2004.

22. Where did the BLOSUM62 alignment score matrix come from? SR Eddy. *Nature Biotech.* 22:1035-1036, 2004.
21. What is dynamic programming? SR Eddy. *Nature Biotech.* 22:909-910, 2004.
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.
19. A uniform system for microRNA annotation. V Ambros, B Bartel, DP Bartel, CB Burge, JC Carrington, X Chen, G Dreyfuss, SR Eddy, S Griffiths-Jones, M Marshall, M Matzke, G Ruvkun and T Tuschl. *RNA* 9:277-279, 2003.
18. Computational genomics of noncoding RNA genes. SR Eddy. *Cell* 109:137-140, 2002.
17. Noncoding RNA genes and the modern RNA world. SR Eddy. *Nature Reviews Genetics* 2:919-929, 2001.
16. Shotgun coverage of human genome computing. SR Eddy. *Trends Biochem. Sci.* 24:124, 1999.
15. Noncoding RNA genes. SR Eddy. *Current Opinion in Genetics and Development* 9:695-699, 1999.
14. Profile hidden Markov models. SR Eddy. *Bioinformatics* 14:755-763, 1998.
13. Multiple-alignment and -sequence searches. SR Eddy. *Trends Guide to Bioinformatics* pp. 15-18, 1998.
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.
11. Hidden Markov models and large-scale genome analysis. SR Eddy. *Transactions of the American Crystallographic Association*, 1997.
10. Hidden Markov models. SR Eddy. *Current Opinion in Structural Biology* 6:361-365, 1996.
9. Molecular genetics in silico: Fourth International Conference on Intelligent Systems in Molecular Biology. SR Eddy. *Trends in Genetics* 12:372, 1996.
8. Is the Pope the Pope? SR Eddy, DJC MacKay. *Nature* 382:490, 1996.
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.
6. RNA structure alignment on a massively parallel computer. H Elingworth, 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.
5. The Human Genome Project according to Los Alamos. SR Eddy. *Trends Biochem. Sci.* 20:91-92, 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.
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.
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.
1. *Introns in the T-Even Bacteriophages*. SR Eddy. PhD thesis, University of Colorado at Boulder, Boulder CO, USA. 1991.

Other publications from the laboratory*

30. Inferring sparse structure in genotype-phenotype maps. **S Petti**, G Reddy, MM Desai. *Genetics*, 225:iyad127, 2023.
29. End-to-end learning of multiple sequence alignments with differentiable Smith-Waterman. **S Petti**, N Bhattacharya, R Rao, J Dauparas, N Thomas, J Zhou, AM Rush, PK Koo, S Ovchinnikov. *Bioinformatics*, 39:btac724, 2023.
28. Discovery of 17 conserved structural RNAs in fungi. W Gao, **TA Jones**, E Rivas. *Nucl. Acids Res.*, 49:6128–6143, 2021.
27. Synthetic genome recoding: new genetic codes for new features. J Kuo, F Stirling, YH Lau, **Y Shulgina**, JC Way, PA Silver. *Curr. Genet.*, 64:327-333, 2018.
26. Inferring functional neural connectivity with deep residual convolutional networks. **TW Dunn**, **PK Koo**. *bioRxiv*, 2017. <https://doi.org/10.1101/141010>
25. C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. **PP Dennis**, V Tripp, L Lui, T Lowe, L Randau. *BMC Genomics* 16:632, 2015.
24. High-performance probes for light and electron microscopy. S Viswanathan, ME Williams, EB Bloss, TJ Stasevich, CM Speer, A Nern, BD Pfeiffer, BM Hooks, WP Li, BP English, T Tian, **GL Henry**, JJ Macklin, R Patel, CR Gerfen, X Zhuang, Y Wang, GM Rubin, LL Looger. *Nature Methods* 12:568–576, 2015.

* I am not a coauthor on all the work done in my laboratory. People in the laboratory often have independent projects and publish independently, either in single-authored form or in collaborations they initiate and take lead responsibility for. This “other publications from the laboratory” section is here for when my CV is being used as a summary of research supported by my laboratory, as opposed to just me. It includes some papers that were the continuation of research from a previous laboratory that was completed in my laboratory; papers that reflect someone’s own independent research agenda while they were in my lab; and other papers that are independent of substantive hands-on research from me. It also includes papers until 2020 from Elena Rivas, who shared the laboratory with me for many years but often worked and published independently; Elena separated her laboratory from mine in 2020. Most but not all of these were peer-reviewed; a couple are commentaries or reviews.

23. Annotating functional RNAs in genomes using Infernal. **EP Nawrocki**. *Methods Mol. Biol.* 1097:163–197, 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.* 522:1445–1453, 2014. (Also *Brain. Behav. Evol.* 83:1–8, 2014.)
21. Skylign: 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.
20. Special focus: bioinformatics. **EP Nawrocki**, SW Burge. *RNA Biol.* 10:1160, 2013.
19. The four ingredients of single-sequence RNA secondary structure prediction: a unifying perspective. **E Rivas**. *RNA Biol.* 10:1185-1196, 2013.
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.
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.
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.
15. Phosphorylation at the interface. **FP Davis**. *Structure* 19:1726-1727, 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.
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.
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.
11. Proteome-wide prediction of overlapping small molecule and protein binding sites using structure. **FP Davis**. *Mol. Biosyst.* 7:545-557, 2011.
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.

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.
8. On the unique function of selenocysteine – insights from the evolution of selenoproteins. **S Castellano**. Biochim. Biophys. Acta. 1790:1463-1470, 2009.
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.
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.
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.
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.
3. Evolutionary models for insertions and deletions in a probabilistic modeling framework. **E Rivas**. BMC Bioinformatics 6:63, 2005.
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.
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.

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://eddylib.org/infernal/>
3. The Pfam database of conserved protein domains.
The Pfam Consortium: European Bioinformatics Institute; Stockholm University; Harvard University.
<http://pfam.xfam.org/>

4. The Rfam database of conserved RNA families.
The Rfam Consortium: European Bioinformatics Institute; Manchester University; University of Canterbury; Harvard University.
<http://rfam.xfam.org/>
5. The Dfam database of mobile repetitive DNA elements.
The Dfam Consortium: University of Montana; Institute for Systems Biology, Seattle; Harvard University; and European Bioinformatics Institute.
<http://dfam.org>
6. tRNAscan-SE - software for detection of transfer RNA genes in genome sequence.
Lead developer: TMJ Lowe.
<http://eddylab.org/software/#trnascan>
7. QRNA - identification of novel noncoding RNA genes by comparative genome sequence analysis.
Lead developer: E Rivas.
<http://eddylab.org/software/#qrna>
8. R-scape - RNA structural covariation above phylogenetic expectation.
Lead developer: E Rivas.
<http://eddylab.org/R-scape/>