Eric E. Snyder, Ph.D.

(540) 230-5225 (mobile)
Mesa, AZ
eesnyder@eesnyder.org
http://www.linkedin.com/in/eesnyder

Summary

Senior bioinformatics researcher and software engineer with experience in industry, academia and government. Trained in molecular biology, biochemistry, pharmacology and bioinformatics. Expertise in software development, especially as applied to sequence analysis for genome annotation, phylogenetics and biomarker development. Experience building online databases for genomic and post-genomic data in support of biomedical research.

Seeking position where I can use these skills with the goal of developing diagnostics and therapeutics for human diseases and understanding the origins and diversity of terrestrial life.

Core Competencies

1. **Bioinformatics**
   a) Nucleic acid and protein **sequence analysis,** eu- and prokaryotic **genome analysis** and **annotation**
   b) Workflows for NGS data handling, genome assembly and alignment, RNAseq
   c) Algorithms for **gene prediction, motif recognition, sequence alignment**
   d) **Microarray analysis**, biomarker discovery using machine learning
   e) Phylogenetics and statistical genetics (linkage analysis, association studies)
   f) Ontologies (OBO, Sequence Ontology)

2. **Software development**
   a) **Programming** in **Python** (*incl.* NumPy, SciPy, BioPython), **Perl** (*incl.* BioPerl, CGI), **R** (*incl.* BioConductor), csh, sed, awk, C
   b) **Algorithm design** (dynamic programming, HMMs)
   c) **High-performance/cluster computing** (Oracle/Sun Grid Engine, AWS)
   d) **Database development** & design; **RDBMS** (Postgres, Oracle, mySQL); BioSQL
   e) **Requirements** elicitation, analysis and specification
   f) **UNIX system administration** (Linux: Open-SuSE, CentOS; IRIX, Solaris)
   g) **Agile software development** (rapid prototyping, iterative development, customer feedback)
   h) **Service-Oriented Architecture** (web services, ServiceMix, NCBI eUtilities)

3. **Biology**
   a) **Gene expression**, regulation, pre-mRNA processing/alternative splicing
   b) **Protein structure**, receptor/ligand interaction, GPCRs
   c) **Functional RNA** (ribozymes, rRNAs, siRNA, *etc.*)

4. **Statistics**
   a) **Classification, correlation and prediction,** regression analysis, hypothesis testing
   b) **Machine Learning,** artificial neural networks, support vector machines, naïve Bayesian classifiers

Professional Experience

Bioinformatics Consultant **Mar 2018 – May 2018**
Inscripta, Inc., Computational Sciences and Engineering Group
Boulder, CO 80301

- Inscripta possesses proprietary enzymes and plasmid constructs for performing targeted mutagenesis using the CRISPR-CAS9 system.
- Developed a Python application to integrate the laboratory information management system (LIMS) with their sequence analysis pipeline with the goal of initiating the analysis as soon as the sequence data comes off the instrumentation.

Senior Research Scientist **Feb 2017 – Nov 2017**
Biodesign Institute, Center for Innovations in Medicine, Arizona State University
Tempe, AZ

- Analyzed peptide microarrays for characterizing and differentiating immune response between various infectious diseases. Identified peptide motifs as putative epitopes targeted by immune system.
- Examined immune response to various neoplasms including breast and pancreatic cancer.
- Developed Array Runner, a Python application which automates the steps required to make a predictive (diagnostic) tool based on microarray data. Given a vector of feature "scores" (typically intensity values from an array scanner) for each study sample, previously grouped into two or more mutually-exclusive categories (*e.g.*, "control" or "affected"), the scores are normalized, rescaled, and ranked by t-test, then used to train an SVM or Bayesian classifier using a range of top-ranked features. Classifier generalization performance is evaluated by leave-one-out or k-fold cross validation. Samples identified as blinded or unknown are also predicted and can be scored if their unblinded category names are available. Datasets consisting of more than one-thousand samples and 300,000 array features can be processed in a few minutes with commonly available desktop hardware.

Senior Business Analyst **Sep 2015 – Oct 2016**
Contractor[1] for the Centers for Disease Control and Prevention; Office of Public Health Scientific Services; Center for Surveillance, Epidemiology, and Laboratory Services; Division of Health Informatics and Surveillance (HHS/CDC/OPHSS/CSELS/DHIS)
Atlanta, GA 30329

- Collected, organized and documented a detailed software requirements specification (SRS) for EPI INFO™ version 7.2, a suite of programs to support epidemiological studies. The application enables: the creation of sophisticated data collection questionnaires, suitable for field use on smartphones and other portable electronics; data storage and integration using RDBMS platforms and common file formats; data processing, recoding and analysis using descriptive and inferential statistics including multiple linear and logistic regression, KM survival curves, proportional hazards, complex sampling strategies and social (interaction) network analysis; integration with geographical information allowing subjects and other relevant features to be located on dynamic maps and overlaid with demographic, socio-economic or other reference data.
- The SRS, based on the ISO/IEEE-830 and -29148 standards, is aimed at promoting and facilitating collaborative and 3rd-party development of the open source/GPL EPI INFO™ software. Requirements are published online using Atlassian Confluence and integrated with JIRA, an Agile-oriented task-tracking system.

[1] Employed by CSRA Inc. (acquired on April 2, 2018, by General Dynamics Information Technology, Falls Church, VA.)

Senior Software Engineer, Bioinformatics **Jul 2013 – Sep 2015**
Contractor[2] for the Centers for Disease Control and Prevention, Office of Infectious Diseases, National Center for Immunization and Respiratory Diseases, Influenza Division, Office of Director
Atlanta, GA 30329

- Lead developer responsible for design and implementation of sequence analysis pipeline services for Influenza Surveillance Data Management System (ISDMS), an in-house database and analysis resource for influenza sequence, annotation and metadata.
- Primary annotation program used a database of 92 HMMs based on SME-derived, lineage-specific segment groups to identify influenza type (A or B), segment (1-8) and lineage of sequence, then performs NA-to-AA spliced alignment to known segment proteins to determine precise CDS coordinates. Also created BLAST-based programs for identifying segments from 2009 H1N1 pandemic strain and recently-identified H1 and H3 variants.
- Conceived of and designed JSON Feature Format (JFF), a standardized web-service-friendly format for sequence features and metadata, used by all ISDMS pipeline programs. Wrote Perl libraries to manipulate, validate and load JFF data into BioSQL databases (such as ISDMS).
- Development mechanism for synchronizing ISDMS with GenBank using NCBI eUtilities. Parsed free-text host data and mapped to NCBI taxonomy. Wrote script to query ISDMS and efficiently build non-redundant AA & NT BLAST databases using MD5 hash keys.

Senior Scientist, Bioinformatics **May 2011 – Feb 2013**
Contractor[2] for the Data Coordination Center (DCC) of the Cancer Genome Atlas project (TCGA), a joint program of the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI)
Rockville, MD 20852-4902

- Program goal: elucidate the molecular basis of cancers using high-throughput *omics technologies.
- Acted as POC for the Biospecimen Core Resource centers to coordinate the development and evolution of the XML schema representing clinical information and biospecimens.
- Wrote and documented software to assemble, validate and submit data to the DCC and other NIH databases; supplied detailed requirements for the 2nd-generation DCC database and UI.
- Represented the DCC on the Analysis Working Groups (AWGs) for acute myeloid leukemia (AML), stomach and esophageal carcinoma and ovarian serous cystadenocarcinoma, managing data freezes and ensuring availability of all clinical and molecular data. Also contributed to publications on the comprehensive molecular profiling of breast tumors and squamous cell lung cancers [24-29].

Senior Scientist, Bioinformatics (special research faculty) **Oct 2004 – Sep 2010**
Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University (VaTech)
Blacksburg, VA 24061-0447.

- Conceived and documented requirements for the NIAID-funded PATRIC Bioinformatics Resource Center website and analytical tools contained therein. PATRIC is an online database of bacterial 'omics data and literature aimed at supporting researchers developing vaccines, diagnostics and therapeutics.
- Developed software to implement high-throughput analyses using Perl and/or R. Analyses include genome annotation functions, ortholog identification, phylogenetics, gene expression analysis and k-tuple frequency fingerprinting.
- Conducted comparative genomics studies of Brucella [19, 21] and Rickettsia [20].
- Built large-scale pipeline for protein collection, ortholog identification, multiple alignment and phylogenetic tree construction and applied it to the γ-proteobacteria [22].
- Wrote research papers and technical reports to document novel methods and findings [17–22].
- Co-authored successful proposal for the competitive 5-year renewal of PATRIC contract, worth $27M, the largest in Virginia Tech history.

[2] Employed by SRA International, Inc. (which became CSRA Inc. through a merger with the Government Services division of Computer Sciences Corporation of Tysons Corner, VA, in November 2017).

Associate Professor (research) **Apr 2000 – Oct 2004**
Pennington Biomedical Research Center, Louisiana State University
Baton Rouge, LA 70808-4124

- Used linkage analysis and association studies to identify QTLs (and genes) associated with human obesity.
- Wrote software to automate SNP calling in TaqMan assays using various statistical models.
- Conceived and implemented the Obesity Gene Map Database, a web resource containing data and literature on candidate genes and QTLs for obesity-related phenotypes.
- Created SAGEparser, a program to identify, tabulate and compare populations of sequence tags from Serial Analysis of Gene Expression (SAGE), a sequence-based method for measuring differential gene expression [14].

Director of Computational Biology **Jun 1997 – Jul 1999**
Genomica Corporation
Boulder, CO 80303

- Developed requirements for company's flagship product, Discovery Manager (DM), enterprise software for genomics targeted at the pharmaceutical industry.
- Evaluated sequence analysis algorithms for inclusion in DM's analytical suite.
- Tested and documented DM's genetic and physical map functionality, reported bugs, *etc.*
- Worked with potential customers to assess needs and demonstrate DM's functionality.

Director of Genome Analysis **Apr 1994 – Jul 1997**
Sequana Therapeutics
La Jolla, CA 92129

- Recruited and led the Genome Analysis Group which provided bioinformatics support to the Sequencing and Physical Mapping groups.
- Designed, implemented and maintained Sequana's annotation pipeline for genomic sequence and physical mapping database.
- Used pipeline to identify gene responsible for X-linked lymphoproliferative disease [7].

Education

09/90 – 05/94	**Ph.D.**	**University of Colorado, Boulder, CO; Department of Molecular, Cellular and Developmental Biology**
09/88 – 05/90	B.A.	University of Colorado, Boulder, CO; Double Major: Biochemistry and Molecular Cellular and Developmental Biology
09/84 – 12/86		Johns Hopkins University, Baltimore, MD; Major: Chemistry

Training

Department of Molecular, Cellular and Developmental Biology **Sep 1991 – May 1994**
University of Colorado (Mentor: Prof. Gary D. Stormo)
Boulder, CO 80309

- Designed and implemented novel algorithm for eukaryotic gene prediction based on dynamic programming and optimized using a back-propagation (artificial neural) network [4, 6].
- Developed novel content-based measures to characterize and discriminate between intronic, exonic and inter-genic DNA.
- Prepared position-weight matrices for splice sites and other gene-associated motifs.
- Performed RNA selection experiments on 30S ribosomal subunit with and without S1 to compare sequences of optimum ligands [5].
- Trained neural networks to identify ribosome binding sites and compared performance to weight matrices with zero- and first-order positional interactions.

Department of Chemistry and Biochemistry **Jun 1989 – Aug 1990**
University of Colorado (Mentor: Prof. Joseph J. Falke)
Boulder, CO 80309

- Performed fluorometric binding assays of metal ions to probe structure-affinity relationship in the *E. coli* galactose receptor, a model EF-hand protein [1].
- Created site-directed mutants of the E. coli galactose receptor to test hypotheses concerning optimal binding site configuration for maximum affinity and specificity [3].

Department of Molecular, Cellular and Developmental Biology **Sep 1988 – May 1989**
University of Colorado (Mentors: Dr. Ravi G. Menon & Prof. Richard G. Ham)
Boulder, CO 80309

- Screened λ gt11 libraries with degenerate oligonucleotides to clone the human κ-casein gene.

Merrell-Dow Fellow, National Institute on Drug Abuse (NIDA) **May 1986 – Aug 1986**
Addiction Research Center, Francis Scott Key Medical Center (Mentor: Dr. Tsutomu Suzuki)
Baltimore, MD 21224

- Conducted ethanol and etonitazene self-administration experiments in Lewis and Fischer rats to examine the effect of genetic background on dependence liability to alcohol and opiates, respectively.

Department of Psychology **Sep 1984 – Apr 1986**
The Johns Hopkins University (Mentors: Dr. Priscilla Kehoe & Prof. Elliott M. Blass)
Baltimore, MD 21218

- Used the separation-induced distress vocalization paradigm to study opioid-mediation of mother-infant bonding in neonatal rats.

Publications

Original Manuscripts

1. **Snyder, E. E.**, Buoscio, B. W. & Falke, J. J. Calcium(II) site specificity: effect of size and charge on metal ion binding to an EF-hand-like site. Biochemistry 29: 3937-3943 (1990). PMID: 2162201.
2. **Snyder, E. E.** & Fall, R. R. Western blotting with a concanavalin A-horseradish peroxidase conjugate. Biochem. Educ. 18: 147-148 (1990).
3. Falke, J. J., **Snyder, E. E.**, Thatcher, K. C. & Voertler, C. S. Quantitating and engineering the ion specificity of an EF-hand-like Ca+2 binding. Biochemistry 30: 8690-8697 (1991). PMID: 1653605.
4. **Snyder, E. E.** & Stormo, G. D. Identification of coding regions in genomic DNA sequences: an application of dynamic programming and neural networks. Nucleic Acids Res 21: 607-613 (1993). PMID: 8441672.
5. Ringquist, S., Jones, T., **Snyder, E. E.**, Gibson, T., Boni, I. & Gold, L. High-affinity RNA ligands to Escherichia coli ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. Biochemistry 34: 3640-3648 (1995). PMID: 7534475.
6. **Snyder, E. E.** & Stormo, G. D. Identification of protein coding regions in genomic DNA. J Mol Biol 248: 1-18 (1995). PMID: 7731036.
7. Nichols, K. E., Harkin, D. P., Levitz, S., Krainer, M., Kolquist, K. A., Genovese, C., Bernard, A., Ferguson, M., Zuo, L., **Snyder, E.**, Buckler, A. J., Wise, C., Ashley, J., Lovett, M., Valentine, M. B., Look, A. T., Gerald, W., Housman, D. E. & Haber, D. A. Inactivating mutations in an SH2 domain-encoding gene in X-linked lymphoproliferative syndrome. Proc Natl Acad Sci U S A 95: 13765-13770 (1998). PMID: 9811875.
8. Ukkola, O., Ravussin, E., Jacobson, P., **Snyder, E. E.**, Chagnon, M., Sjostrom, L. & Bouchard, C. Mutations in the preproghrelin/ghrelin gene associated with obesity in humans. J Clin Endocrinol Metab 86: 3996-3999 (2001). PMID: 11502844.

9. Jacobson, P., Ukkola, O., Rankinen, T., **Snyder, E. E.**, Leon, A. S., Rao, D. C., Skinner, J. S., Wilmore, J. H., Lonn, L., Cowan, G. S., Jr., Sjostrom, L. & Bouchard, C. Melanocortin 4 receptor sequence variations are seldom a cause of human obesity: the Swedish Obese Subjects, the HERITAGE Family Study, and a Memphis cohort. J Clin Endocrinol Metab 87: 4442-4446 (2002). PMID: 12364415.
10. Rankinen, T., Perusse, L., Weisnagel, S. J., **Snyder, E. E.**, Chagnon, Y. C. & Bouchard, C. The human obesity gene map: the 2001 update. Obes Res 10: 196-243 (2002). PMID: 11886943.
11. Chagnon, Y. C., Rankinen, T., **Snyder, E. E.**, Weisnagel, S. J., Perusse, L. & Bouchard, C. The human obesity gene map: the 2002 update. Obes Res 11: 313-367 (2003). PMID: 12634430.
12. Yoshioka, M., Tanaka, H., Shono, N., **Snyder, E. E.**, Shindo, M. & St-Amand, J. Serial analysis of gene expression in the skeletal muscle of endurance athletes compared to sedentary men. FASEB J 17: 1812-1819 (2003). PMID: 14519660.
13. **Snyder, E. E.**, Walts, B., Perusse, L., Chagnon, Y. C., Weisnagel, S. J., Rankinen, T. & Bouchard, C. The human obesity gene map: the 2003 update. Obes Res 12: 369-439 (2004). PMID: 15044658.
14. Dinel, S., Bolduc, C., Belleau, P., Boivin, A., Yoshioka, M., Calvo, E., Piedboeuf, B., **Snyder, E. E.**, Labrie, F. & St-Amand, J. Reproducibility, bioinformatic analysis and power of the SAGE method to evaluate changes in transcriptome. Nucleic Acids Res 33: e26 (2005). PMID: 15716308.
15. Hance, M. E., Czar, M. J., Azad, A., Purkayastha, A., **Snyder, E. E.**, Crasta, O. R., Setubal, J. C. & Sobral, B. W. The pathogen resource integration center: implications for Rickettsial research. Ann N Y Acad Sci 1063: 459-465 (2005). PMID: 16481560.
16. Perusse, L., Rankinen, T., Zuberi, A., Chagnon, Y. C., Weisnagel, S. J., Argyropoulos, G., Walts, B., **Snyder, E.** & Bouchard, C. The human obesity gene map: the 2004 update. Obes Res 13: 381-490 (2005). PMID: 15833932.
17. Baker, S. C., Jukneliene, D., Purkayastha, A., **Snyder, E. E.**, Crasta, O. R., Czar, M. J., Setubal, J. C. & Sobral, B. W. Developing bioinformatic resources for coronaviruses. Adv Exp Med Biol 581: 395-398 (2006). PMID: 17037566.
18. **Snyder, E. E.**, Kampanya, N., Lu, J., Nordberg, E. K., Karur, H. R., Shukla, M., Soneja, J., Tian, Y., Xue, T., Yoo, H., Zhang, F., Dharmanolla, C., Dongre, N. V., Gillespie, J. J., Hamelius, J., Hance, M., Huntington, K. I., Jukneliene, D., Koziski, J., Mackasmiel, L., Mane, S. P., Nguyen, V., Purkayastha, A., Shallom, J., Yu, G., Guo, Y., Gabbard, J., Hix, D., Azad, A. F., Baker, S. C., Boyle, S. M., Khudyakov, Y., Meng, X. J., Rupprecht, C., Vinje, J., Crasta, O. R., Czar, M. J., Dickerman, A., Eckart, J. D., Kenyon, R., Will, R., Setubal, J. C. & Sobral, B. W. PATRIC: the VBI PathoSystems Resource Integration Center. Nucleic Acids Res 35: D401-406 (2007). PMID: 17142235.
19. Yu, G. X., **Snyder, E. E.**, Boyle, S. M., Crasta, O. R., Czar, M., Mane, S. P., Purkayastha, A., Sobral, B. & Setubal, J. C. A versatile computational pipeline for bacterial genome annotation improvement and comparative analysis, with Brucella as a use case. Nucleic Acids Res 35: 3953-3962 (2007). PMID: 17553834.
20. Gillespie, J. J., Williams, K., Shukla, M., **Snyder, E. E.**, Nordberg, E. K., Ceraul, S. M., Dharmanolla, C., Rainey, D., Soneja, J., Shallom, J. M., Vishnubhat, N. D., Wattam, R., Purkayastha, A., Czar, M., Crasta, O., Setubal, J. C., Azad, A. F. & Sobral, B. S. Rickettsia phylogenomics: unwinding the intricacies of obligate intracellular life. PLoS One 3: e2018 (2008). PMID: 19194535.
21. Wattam, A. R., Williams, K. P., **Snyder, E. E.**, Almeida, N. F., Jr., Shukla, M., Dickerman, A. W., Crasta, O. R., Kenyon, R., Lu, J., Shallom, J. M., Yoo, H., Ficht, T. A., Tsolis, R. M., Munk, C., Tapia, R., Han, C. S., Detter, J. C., Bruce, D., Brettin, T. S., Sobral, B. W., Boyle, S. M. & Setubal, J. C. Analysis of ten Brucella genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. J Bacteriol 191: 3569-3579 (2009). PMID: 19346311.
22. Williams, K. P., Gillespie, J. J., Sobral, B. W., Nordberg, E. K., **Snyder, E. E.**, Shallom, J. M. & Dickerman, A. Phylogeny of gamma-proteobacteria. J Bacteriol 192: 2305-2314 (2010). PMID: 20207755.

23. Gillespie, J.J., Wattam, A.R., Cammer, S.A., Gabbard, J., Shukla, M.P., Dalay, O., Driscoll, T., Hix, D., Mane, S.P., Mao, C., Nordberg, E.K., Scott, M., Schulman, J.R., **Snyder, E.E.**, Sullivan, D.E., Wang, C., Warren, A., Williams, K.P., Xue, T., Yoo, H.S., Zhang, C., Zhang, Y., Will, R., Kenyon, R.W., Sobral, B.W. PATRIC: The Comprehensive Bacterial Bioinformatics Resource with a Focus on Human Pathogenic Species. Infect Immun. 79(11):4286-98 (2011). PMID: 21896772.
24. Cancer Genome Atlas Network (incl. **Snyder, E. E.**) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487(7407):330-7 (2012). PMID: 22810696.
25. Cancer Genome Atlas Network (incl. **Snyder, E. E.**) Comprehensive genomic characterization of squamous cell lung cancers. Nature 489(7417):519-25 (2012). PMID: 22960745.
26. Cancer Genome Atlas Network (incl. **Snyder, E. E.**) Comprehensive molecular portraits of human breast tumours. Nature 490(7418):61-70 (2012). PMID: 23000897.
27. Cancer Genome Atlas Network (incl. **Snyder, E. E.**) Comprehensive molecular characterization of acute myeloid leukemia. N Engl J Med, 368:2059-2074 (2013). PMID: 23634996.
28. Cancer Genome Atlas Network (incl. **Snyder, E. E.**) Integrated genomic characterization of endometrial carcinoma. Nature 497(7447):67-73 (2013). PMID: 23636398.
29. Cancer Genome Atlas Research Network (incl. **Snyder, E. E.**) Diversity of lung adenocarcinoma revealed by integrative molecular profiling. Nature 511(7511):543-50 (2014). PMID: 25079552.

Certifications

Certified SAFe Practitioner, Scaled Agile Framework (paradigm for enterprise software development), Dec 2015.

Professional Societies

International Society for Computational Biology, 2003 – current.

Service

Ad Hoc Manuscript Reviewer

BCM Bioinformatics
Bioinformatics
Computer Applications in Biological Sciences
Journal of Molecular Biology
Nucleic Acids Research
Obesity Research

Teaching

Introductory Biology,	Fall, 1991	University of Colorado, Boulder, CO
Molecular Cell Biology	Spring, 1992	University of Colorado, Boulder, CO