

Wyatt T. Clark

Post Doctoral Associate

266 Whitney Ave.

New Haven, CT 06511

☎ (812)-320-3951

✉ wyatt.clark@yale.edu

🌐 tinyurl.com/WyattClark

Research Experience

2013–present **Post Doctoral Associate**, *Laboratory of Mark Gerstein*.

Department of Molecular Biophysics & Biochemistry, Yale University

- *Phylogeny aware detection of unitary pseudogenes*: Unitary pseudogenes are the vestigial remnants of once functional genes in the genome. Most methods for locating these elements rely on detecting similarity with known genes. One problem with this approach is a unitary pseudogene may have accumulated a large number of mutations, making it undetectable. I am currently developing a phylogeny aware method that utilizes predicted gene loss events in reconciled gene family trees to locate unitary pseudogenes.
- *Interactions between heterogeneous vertices*: Many methods that predict vertex labels and edges in graphs are based on the concept of guilt by association. Under this framework similar vertices are more likely to be connected by an edge and connected vertices are more likely to share the same labels. I am currently investigating the strength of this assumption in protein interaction networks and developing tractable methodology for utilizing statistically significant interactions between heterogeneous vertices to predict missing edges and vertex labels.

2007–2013 **Research Assistant**, *Laboratory of Predrag Radivojac*.

School of Informatics and Computing, Indiana University

- *Protein function prediction*: Utilized multi-output neural networks and kernel methods to predict Gene Ontology function from sequence derived properties. I was also deeply involved in the *Critical Assessment of Functional Annotation*, whose results were published in *Nature Methods*
- *Testing the ortholog conjecture*: Evolutionary analysis of how speciation and gene duplication influence the rate of functional divergence between homologous sequences. Work from this project, for which I am co-first author, was published in *PLoS Computational Biology* and featured in *Nature Reviews Genetics*.
- *Classification of protein structure and function*: Developed alignment free methods for the prediction of protein function and structural class based on the kernel representations of various sequence and structure properties. The core of this method is a Vector Quantization spectral kernel.
- *Information theoretic evaluation of multi-label classification*: Utilized a bayesian network, structured according to the Gene Ontology, in order to account for dependencies between terms and provide a probabilistic basis for calculating annotation information content when a gene or protein is assigned multiple "functions" or labels.

2005–2006 **Undergraduate Research Assistant**.

- *Detection of inauthentic text*: Originated a novel method using compression profiles to detect classes of inauthentic computer generated text. This project resulted in a published paper at the 2006 SIAM Conference on Data Mining and was recognized by *Slashdot* and other media.
- *D. pulex transposable elements*: Assisted in quantifying transposable element content in the genome of *D. pulex* in the laboratory of Michael Lynch.

Education

- 2013 **Ph.D. Informatics**, *Indiana University*, Bloomington, IN.
Track: Bioinformatics
- 2006 **B.S. Informatics**, *Indiana University*, Bloomington, IN.
Minors: Biology, Philosophy

Publications

- W.T. Clark** and P. Radivojac. Vector quantization kernels for the classification of protein sequences and structures. *Pacific Symposium on Biocomputing*, 19:316–327, 2014.
- P. Radivojac, **W.T. Clark**, et al. A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3):221–227, 2013.
- W.T. Clark** and P. Radivojac. Information-theoretic evaluation of predicted ontological annotations. *Bioinformatics*, 29(13):i53–i61, 2013. (**Ian Lawson Van Toch Memorial Award for Outstanding Student Paper**).
- N.L. Nehrt, **W.T. Clark***, P. Radivojac, and M.W. Hahn. Testing the ortholog conjecture with comparative functional genomic data from mammals. *PLoS Computational Biology*, 7(6):e1002073, 2011. (*** Co-first author**).
- W.T. Clark** and P. Radivojac. Analysis of protein function and its prediction from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics*, 79(7):2086–2096, 2011.
- Y. Zhao, **W.T. Clark**, M. Mort, D.N. Cooper, P. Radivojac, and S.D. Mooney. Prediction of functional regulatory snps in monogenic and complex disease. *Human Mutation*, 32(10):1183–1190, 2011.
- M.M. Dalkilic, J.C. Costello, **W.T. Clark**, and P. Radivojac. From protein-disease associations to disease informatics. *Frontiers in Bioscience*, 13:3391–3407, 2008.
- P. Radivojac, K. Peng, **W.T. Clark**, B.J. Peters, A. Mohan, S.M. Boyle, and S.D. Mooney. An integrated approach to inferring gene-disease associations in humans. *Proteins: Structure, Function, and Bioinformatics*, 72(3):1030–1037, 2008.
- M.M. Dalkilic, **W.T. Clark**, J.C. Costello, and P. Radivojac. Using compression to identify classes of inauthentic texts. In *Proceedings of the 2006 SIAM Conference on Data Mining*, pages 604–608, April 2006.

Conference Presentations

- W.T. Clark**. *Semantic similarity, better with Bayesian networks*. ISMB ECCB, July, 2013, Berlin, Germany. (**Associated manuscript published in conference proceedings**)
- W.T. Clark**. *Information theoretic metrics for the evaluation of ontological annotations*. ISMB Automated Function Prediction SIG, July, 2012, Long Beach, CA.
- W.T. Clark**. *Using data from additional pairs of organisms to test the ortholog conjecture*. Rocky Mountain Bioinformatics Conference, December, 2011, Snow Mass, CO.
- W.T. Clark**, P. Radivojac, M.W. Hahn. *Testing the ortholog conjecture with functional data from human and mouse*. ISMB Automated Function Prediction SIG, July, 2011, Vienna, Austria.

Extracurricular activities

IU Bioinformatics Club, Co-Founder.

The Bioinformatics Club fosters undergraduate and graduate awareness of bioinformatics, as well as bring together individuals with a shared interest in the field. The group organizes talks and facility tours which focus on the three main data-generating platforms in bioinformatics: high-throughput next generation sequencing, mass-spectrometry, and DNA microarrays.

IU Caving Club, Member.

The Caving Club regularly organizes underground trips for novice and expert cavers. With an emphasis put on both recreation and conservation, member have taken an active role in the restoration of several defaced and littered caves in southern Indiana.

Academic Recognition and Awards

- 2013 **Ian Lawson Van Toch Memorial Award for Outstanding Student Paper, ISMB 2013**, Berlin, Germany.
- 2013 **Travel Fellowship, ISMB 2013**, Berlin, Germany, National Science Foundation.
- 2012 **Research Fellowship**, Center for Bioinformatics Research, Indiana University.
- 2011 **Travel Fellowship, ISMB 2011**, Vienna, Austria, National Institutes of Health.
- 2006 **Travel Fellowship, SDM 2006**, Bethesda, Maryland, Lawrence Livermore National Laboratory.

Professional Membership

- 2009–present **International Society for Computational Biology (ISCB).**

Technical Skills

- Programming **MATLAB, Perl, R, HTML, CGI, CSS.**
- Machine Learning **Multi-output Neural Networks, Decision Trees, Random Forests, Hidden Markov Models, Support Vector Machines, Natural Language Processing.**

Thesis Committee

Predrag Radivojac, *Thesis Advisor, Associate Professor*, School of Informatics and Computing, predrag@indiana.edu.

Haixu Tang, *Associate Professor*, School of Informatics and Computing, hatang@indiana.edu.

Michael Lynch, *Distinguished Professor*, Department of Biology, milynch@indiana.edu.

Matthew Hahn, *Associate Professor*, Department of Biology and School of Informatics and Computing, mwh@indiana.edu.