

Wyatt T. Clark

Post Doctoral Associate

Molecular Biophysics and Biochemistry,
Yale University

266 Whitney Ave.
New Haven, CT 06511

Phone: 812-320-3951

email: wyatt.clark@yale.edu

URL: <http://tinyurl.com/WyattClark>

~ Areas of specialization ~

Computational Biology • Machine Learning • Bioinformatics • Data Mining

~ Education ~

2013 PH.D. in Informatics, Indiana University
2006 B.S. in Informatics, Indiana University

~ Technical skills ~

Programming MATLAB, Perl, Python, R, HTML, Web Scripting, CSS
Machine Learning Multi-output Neural Networks, Bayesian Networks, Decision Trees, Random Forests, Hidden Markov Models, Support Vector Machines, Natural Language Processing, Cross Validation

~ Research experience ~

- 2013-present **Post Doctoral Associate**, *Laboratory of Mark Gerstein*.
Department of Molecular Biophysics and Biochemistry, Yale University
- *The influence of recombination on selection*: analyzed how variance in recombination rates influences the distribution rare and common SNV's throughout the genome, in addition to its effect on the localization of genes and non-functional elements such as pseudogenes.
 - *Heterogeneous labels in interaction networks*: improved gene function and disease association prediction by detecting statistically significant co-occurrences of heterogeneous labels in interaction graphs.
- 2007-2013 **Research Assistant**, *Laboratory of Predrag Radivojac*.
School of Informatics and Computing, Indiana University
- *Bayesian networks and information content*: originated novel metrics grounded in interpreting the structure of an ontology as a Bayesian network in order to more accurately calculate information content and evaluate ontological annotations.
 - *Protein function prediction*: created a high accuracy method for the multi-label classification task of predicting protein function. Its basis was a multi-output neural network that can model the high rate of dependence between labels.
- 2005-2006 **Undergraduate Research Assistant**.
- *Detection of inauthentic text*: exploited the unintuitive nature of the relationship between compressibility and information content in natural text to distinguish between human and computer generated fake journal articles.
 - *Daphnia transposable elements*: assisted in the quantification of transposable elements in the genome of *Daphnia pulex*, a crustacean that often reproduces through parthenogenesis.

~ Honors and awards ~

- 2014 **Travel Fellowship**, National Library of Medicine/National Institutes of Health
PSB 2014, Hawaii
- 2013 **Ian Lawson Van Toch Memorial Award for Outstanding Student Paper**
ISMB 2013, Berlin, Germany
- 2013 **Travel Fellowship**, National Science Foundation
ISMB 2013, Berlin, Germany
- 2013 **Research Fellowship**, Center for Bioinformatics Research, Indiana University
- 2011 **Travel Fellowship**, National Institutes of Health
ISMB 2011, Vienna, Austria
- 2006 **Travel Fellowship**, Lawrence Livermore National Laboratory
SDM 2006, Bethesda, Maryland

~ Professional membership ~

2009-present **International Society for Computational Biology**

~ Publications ~

- 2014 **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.
- 2013 P. Radivojac, **W.T. Clark**, et al. A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3):221–227, 2013.
- 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**).
- 2011 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**).
- 2011 **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.
- 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.
- 2008 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.
- 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.
- 2006 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.

~ Talks ~

- 2014 **W.T. Clark**. *Vector quantization kernels for the classification of protein sequences and structures*. PSB, The Big Island of Hawaii. (**Associated manuscript published in conference proceedings**)
- 2013 **W.T. Clark**. *Semantic similarity, better with Bayesian networks*. ISMB ECCB, Berlin, Germany. (**Associated manuscript published in conference proceedings**)
- 2012 **W.T. Clark**. *Information theoretic metrics for the evaluation of ontological annotations*. ISMB Automated Function Prediction SIG, Long Beach, CA.
- 2011 **W.T. Clark**. *Using data from additional pairs of organisms to test the ortholog conjecture*. Rocky Mountain Bioinformatics Conference, Snow Mass, CO.
- 2011 **W.T. Clark**, P. Radivojac, M.W. Hahn. *Testing the ortholog conjecture with functional data from human and mouse*. ISMB Automated Function Prediction SIG, Vienna, Austria.

~ Extracurricular activities ~

2013-2013 **IU Bioinformatics Club, Co-Founder**
2005-2012 **IU Caving Club, Member**