Kevin Kaichuang Yang

614 260 9454 yangkky@gmail.com https://yangkky.github.io

EDUCATION

PhD in Chemical Engineering	California Institute of Technology	2018 (projected)
B.S. in Chemical Engineering	Ohio State University	2011

SKILLS

- I develop machine-learning algorithms to explain and predict biological data.
- I design, implement, and train machine-learning models, usually in Keras or PyTorch
- I write robust, tested, reproducible, version-controlled code, usually in Python.
- I clean, manipulate, and visualize biological data.
- I have a strong background in Gaussian processes, deep learning, representation learning, Bayesian statistics, and graphical modeling.
- I effectively communicate my findings to diverse audiences with a variety of scientific backgrounds.

RESEARCH EXPERIENCE

Graduate Research Assistant, Professor Frances Arnold's Group

California Institute of Technology, Pasadena, CA August 2014 – Present

- Used Gaussian process models (github.com/yangkky/gpmodel) to design channelrhodopsins
- Designed embedded representations of protein sequences based on doc2vec to streamline machine-learning pipelines (github.com/fha_lab/embeddings_reproduction)
- Built neural machine translation models in PyTorch to predict signal peptides from protein sequences

Computational Intern

Ambry Genetics, Aliso Viejo, CA June 2017 – September 2017

- Developed and implemented neural network models in Keras and PyTorch to predict outcomes of genetic variation by transferring information across paralogous proteins
- Incorporated model into a pipeline that finds paralogs for variants of interest and then predicts outcomes

PUBLICATIONS

Invited Talks

1. "Learning the language of proteins." Gray-Hill Lecture at Occidental College. 29 June 2018.

Peer-Reviewed Papers

3. **Yang, K. K**., Wu, Z., Bedbrook, C. N., Arnold, F.H. "Protein embeddings for machine learning." *Bioinformatics*. 23 March 2018. doi.org/10.1093/bioinformatics/bty178
2. *Bedbrook, C. N., ***Yang, K. K.**, Rice, A. J., Gradinaru, V., Arnold, F.H. "Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization". *PLOS Comp. Bio.* 23 Oct 2017. doi.org/10.1371/journal.pcbi.1005786
1. Bedbrook, C. N., Rice, A. J., **Yang, K. K.**, Ding, X., Chen, S., LeProust, E. M., Gradinaru, V., Arnold, F.H. Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins. *PNAS.* 10 Mar 2017. doi/10.1073/pnas.170026911

[*co-first authors]

Conference presentations

2. Poster Presentation: "Learning the language of proteins." Intelligent Systems for Molecular Biology, Chicago, IL, 9 July 2018.
1. Oral Presentation**:** "Machine Learning to Predict Eukaryotic Expression and Plasma Membrane Localization of an Integral Membrane Protein." Proteins Gordon Research Seminar, Holderness, NH. 17 June 2017.

HONORS, AWARDS, AND FELLOWSHIPS

Caltech Chemistry and Chemical Engineering Teaching Assistantship Award	2017
NIH Biotechnology Leadership Pre-doctoral Training Program	2015
National Science Foundation Research Experience for Teachers Grant	2013
Green Dot Public Schools Rising Star Teacher	2013
National Science Foundation Graduate Research Fellowship	2011

TEACHING AND MENTORING

Teaching Assistantships

Caltech ChE/BE 163 Introduction to biomolecular engineering	2018
Caltech ChE/BE 163 Introduction to biomolecular engineering	2016
Caltech ChE 101 Chemical reaction engineering	2015

Physics and Math Teacher

Green Dot Public Schools, Los Angeles, CA August 2011 – August 2014

- Planned and implemented curricula for algebra 1 and physics
- Received highest evaluation based on test scores and observations