EROL S. KAVVAS

eskavvas@ucdavis.edu

EDUCATION

University of California, San Diego 2015 - 2020
Ph.D candidate, Department of Bioengineering
Thesis: *Biologically-Interpretable Machine Learning for Microbial Genomics*

University of California, Davis 2015
B.S., Department of Civil and Environmental Engineering GPA: 3.70

SKILLS

Computational skills: Statistical analysis, Machine Learning (SVM, linear models, tree-based methods, ICA, PCA), NGS data analysis (genomics, transcriptomics, metabolomics, ^{13}C-fluxomics), Genome-scale metabolic modeling

Programming languages/software: Python, R, SQL, Fortran, Linux, Affinity Designer, Matlab, Git, Ableton

EXPERIENCE

Systems Biology Research Group September 2015 - June 2020
PhD candidate, Advisor: Bernhard Ø. Palsson UCSD, San Diego

- **Project 1**: Identification of genetic signatures underlying *M. tuberculosis* antimicrobial resistance (AMR) using a large genomics dataset
  - Constructed a pan-genome representation of the 1,595 drug-tested genome sequences to enable an unbiased, reference-agnostic perspective of genetic variation
  - Developed a machine learning approach (L1-regularized SVM with boostrapping) that recovered 33 known AMR genes and provided 24 novel candidates
  - Study received an F1000 recommendation, UCSD press release, and lead to two invited conference talks
- **Project 2**: Characterization of the metabolic basis of *M. tuberculosis* AMR using metabolic networks and genomics
  - Updated and standardized a genome-scale metabolic model of *M. tuberculosis* that outperforms previous models in gene essentiality predictions
  - Developed a biochemically-interpretable machine learning classifier that recapitulated AMR mechanisms for three drugs and linked alleles to specific biochemical effects
- **Project 3**: Elucidation of mutational logic in *E. coli* adaptive evolution using a deep multi-omics dataset of six strains
  - Perfomed statistical tests on the fluxomics data to identify convergent and divergent metabolic adaptations
  - Applied independent component analysis (ICA) and statistics to identify six conserved transcriptomic strategies and five regulatory tradeoffs governing *E. coli* adaptation
  - Developed a statistical approach leveraging experimental design that identified four mutation-flux correlates and eight mutation-transcriptome correlates, thereby linking mutations to multi-scale evolutionary constraints

Sinopia Biosciences October 2019 - January 2020
Machine learning consultant JLABS, San Diego

- Normalized and batch corrected a metabolomics dataset describing $1,000$+ drug-treated red blood cells
- Applied machine learning to identify metabolic signatures distinguishing drug action

INVITED TALKS

Artificial Intelligence in Genomics September 27, 2019
Association of Gene Diagnostics meeting Potsdam, Germany

Tuberculosis and Machine Learning November 20, 2019
4th Turning the Tide of Antimicrobial resistance meeting Oslo, Norway

PUBLICATIONS

1. **ES. Kavvas**, L. Yang, JM. Monk, D. Heckmann, BO. Palsson. (2020). A biochemically-interpretable machine learning classifier for microbial GWAS. *Nature Communications* 11 (2580).
2. **ES. Kavvas**, E. Catoui, N. Mih, JT. Yurkovich, Y. Seif, N. Dillon, D. Heckmann, A. Anand, L. Yang, C. Nizet, JM. Monk and BO. Palsson. (2018). Machine learning and structural analysis of *Mycobacterium tuberculosis* pangenome identifies genetic signatures of antibiotic resistance. *Nature Communications* 9 (4306).
3. **ES. Kavvas**, MR. Antoniewicz, C. Long, Y. Ding, JM. Monk, BO. Palsson, A. Feist. (2020). Laboratory evolution of multiple E. coli strains reveals unifying principles of adaptation but diversity in driving genotypes. *bioRxiv* DOI:10.1101/2020.05.19.104992.
4. **ES. Kavvas**, Y. Seif, JT. Yurkovich, C. Norsigian, S. Poudel, WW. Greenwald, S. Ghatak, BO. Palsson and JM. Monk. (2018). Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Systems Biology* 12 (25).
5. JC. Hyun, **ES. Kavvas**, JM. Monk, BO. Palsson. (2020). Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. *PLoS computational biology* 16 (3), e1007608.
6. Y. Seif, **ES. Kavvas**, JC. Lachance, JT. Yurkovich, SP. Nuccio, X. Fang, E. Catoiu, M. Raffatellu, BO. Palsson, JM Monk. (2018). Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. *Nature Communications* 9:3771.
7. CJ. Norsigian, **ES. Kavvas**, Y. Seif, BO. Palsson, JM. Monk. (2018). iCN718, an updated and improved genome-scale metabolic network reconstruction of Acinetobacter baumannii AYE. *Frontiers in genetics* 9, 121.
8. B. Du, DC. Zielinski, **ES. Kavvas**, A. Drager, J. Tan, Z. Zhang, KE. Ruggiero, GA. Arzumanyan and BO. Palsson. (2016). Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *BMC Systems Biology* 10:40.
9. Y Seif, JM Monk, H Machado, **ES. Kavvas**, BO Palsson. (2019). Systems Biology and Pangenome of Salmonella O-Antigens. *mBio* 10 (4), e01247-19.
10. KS. Choudhary, N. Mih, JM. Monk, **ES. Kavvas**, JT. Yurkovich, G. Sakoulas, BO. Palsson. (2018). The Staphylococcus aureus two-component system AgrAC displays four distinct genomic arrangements that delineate genomic virulence factor signatures. *Frontiers in microbiology* 9, 1082.
11. N. Mih, E. Brunk, K. Chen, E. Catoiu, A. Sastry, **ES. Kavvas**, JM. Monk, Z. Zhang, BO. Palsson. (2018). ssbio: a Python framework for structural systems biology. *Bioinformatics* 34 (12), 2155-2157.
12. X. Fang, JM. Monk, N. Mih, B. Du, AV. Sastry, **ES. Kavvas**, Y. Seif, L. Smarr, BO. Palsson. (2018). Escherichia coli B2 strains prevalent in inflammatory bowel disease patients have distinct metabolic capabilities that enable colonization of intestinal mucosa. *BMC systems biology* 12 (1), 66.

LEADERSHIP & COMMUNITY ENGAGEMENT

UCSD Department of Bioengineering January 2016 - March 2018
Teaching assistant La Jolla, CA

- Lead discussion sections, created and graded assignments for three courses: Systems Biology and Bioengineering II, Dynamic Simulation in Bioengineering, and Modeling and Computation in Bioengineering