

Big Data on small organisms: From whole-cell models to clinical decision support systems



Ilias Tagkopoulos

Assistant Professor,
Department of Computer Science and
Genome Center



**UCD – LLNL meeting
December 2014**

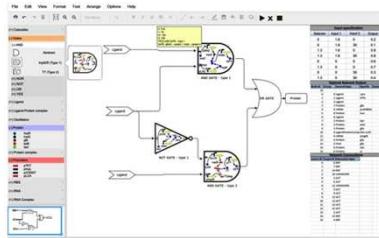


Systems and Synthetic Biology



SYSTEMS
ENGINEERING

Circuit design automation



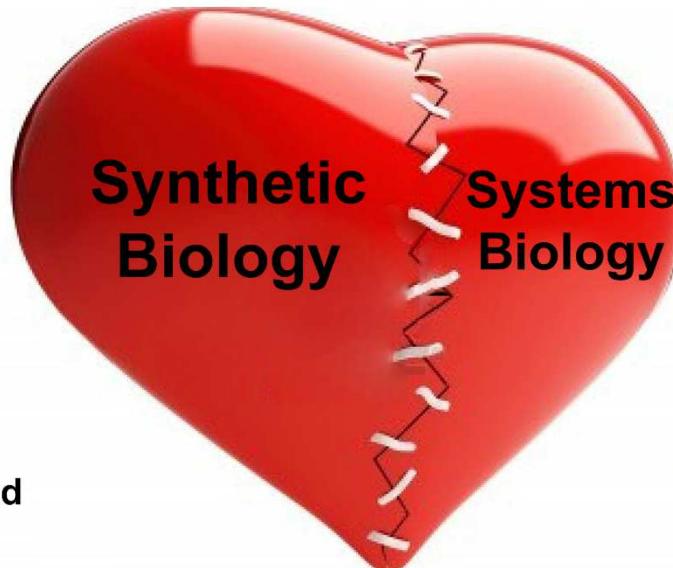
Part construction and Characterization



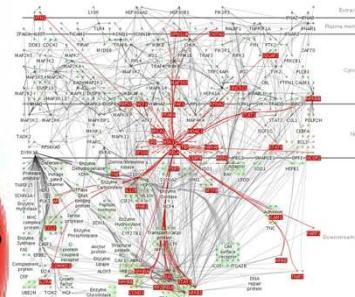
Machine Learning Algorithms

Integration and Inference

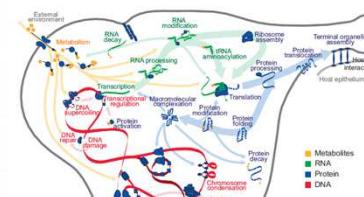
questions
algorithm
complex
objects
relations
new
different
give people
interactions
many
online
theory
machine
algorithms
data
statistical
learning
graphs
clustering
use amounts
social extract
analysis
properties guarantees
analyze becomes
urgent



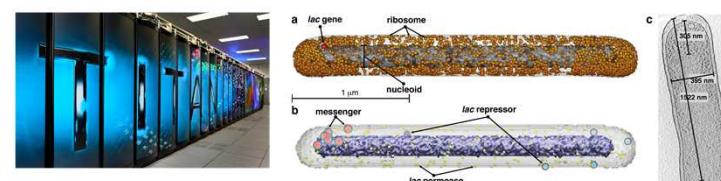
Biological Networks



Whole cell modeling

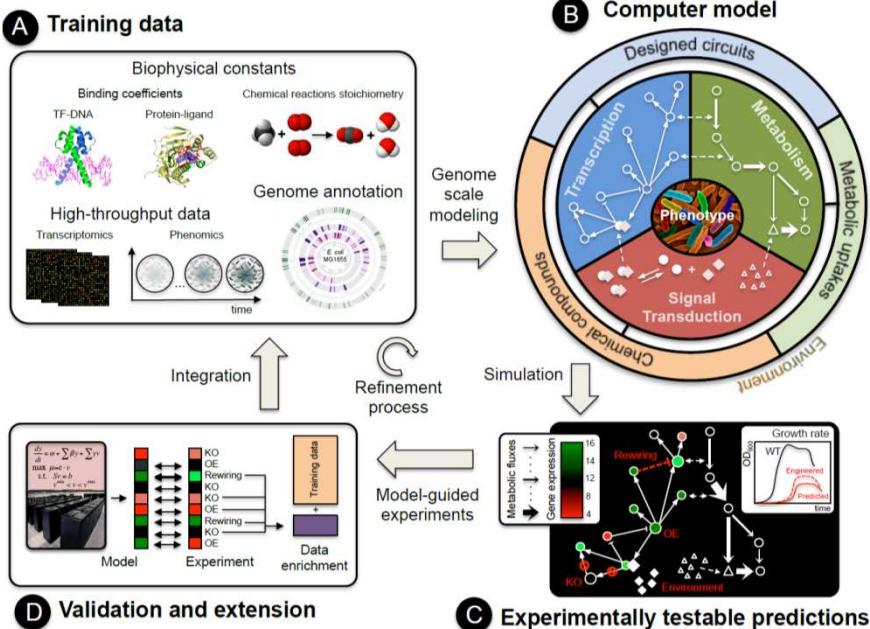


Multi-scale simulation



SCIENCE

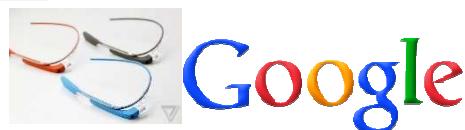
Some of our projects



Decision support tools for agriculture and medicine



MONSANTO



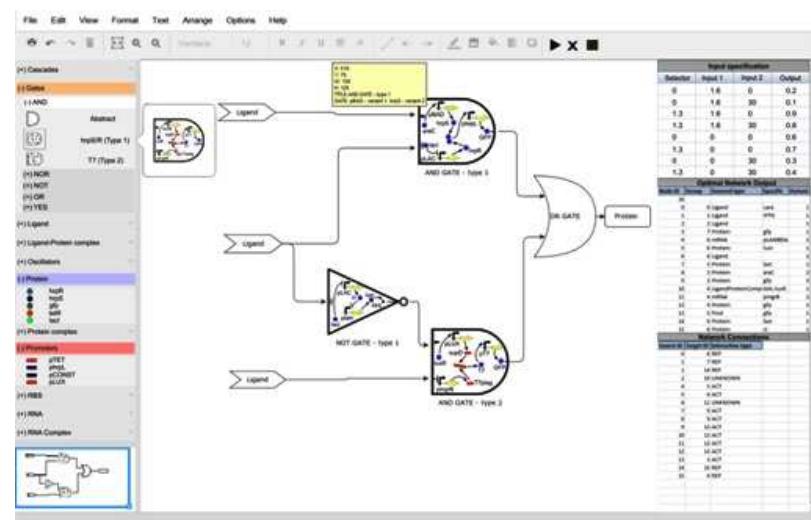
Data-driven, multi-scale modeling and HPC simulation

XSEDE

Extreme Science and Engineering Discovery Environment



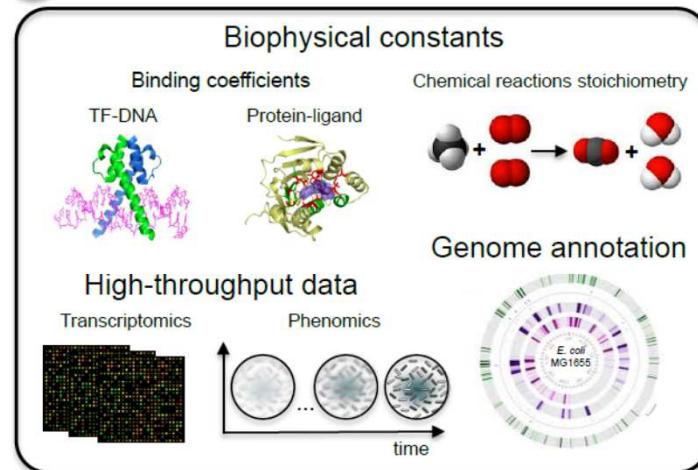
Algorithmic foundations in Synthetic Biology



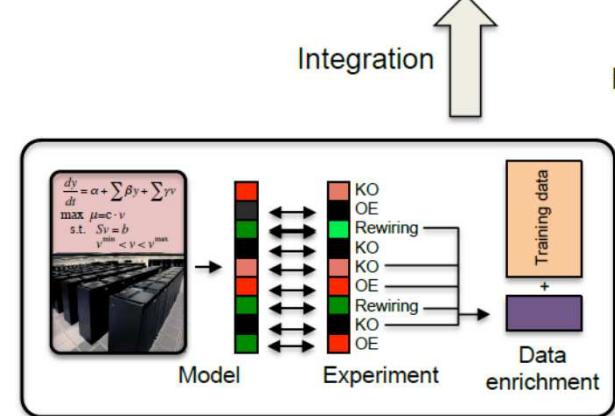
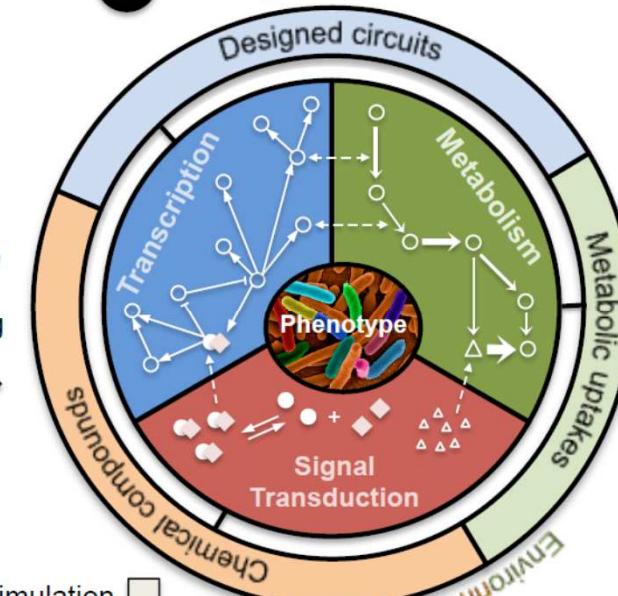
A genome-scale predictive framework for bacterial organisms



A Training data



B Computer model



D Validation and extension

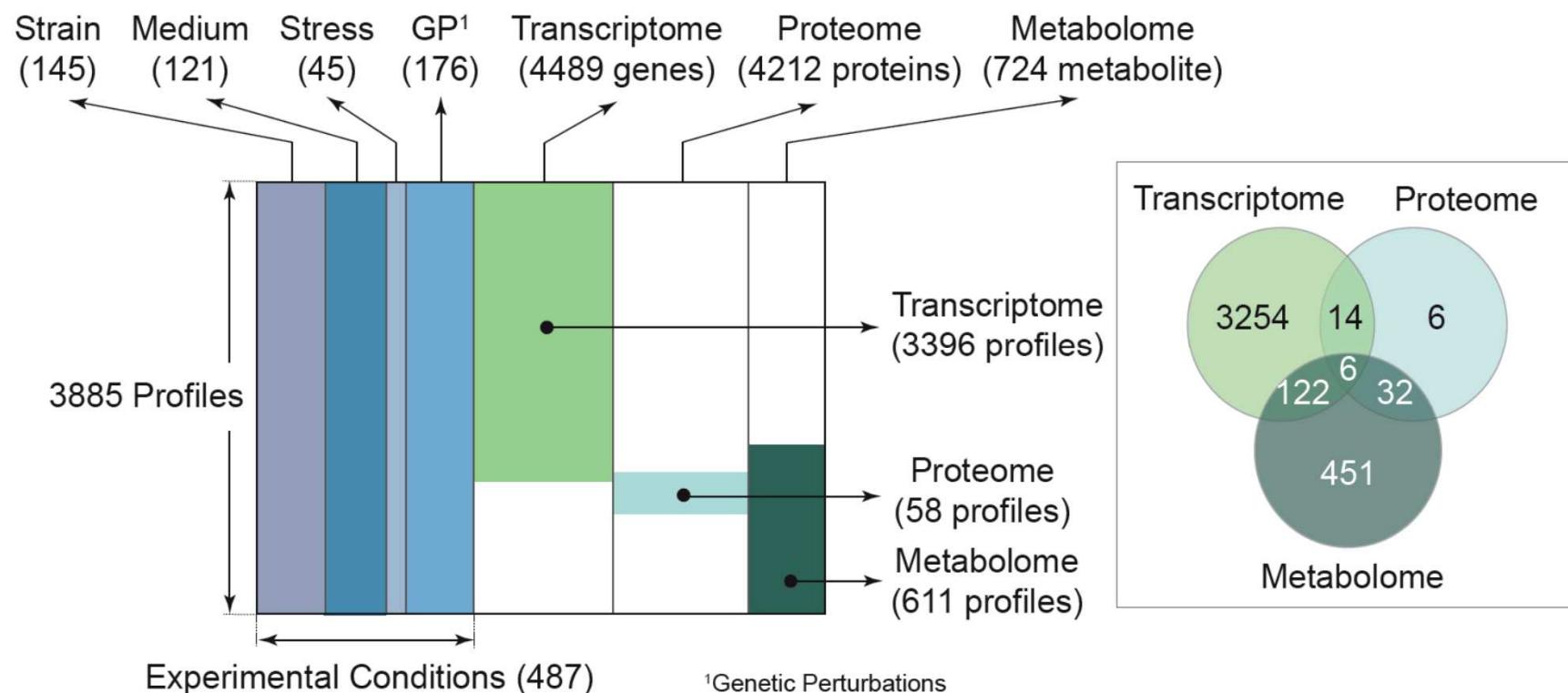
C Experimentally testable predictions

J. Carrera, R. Estrela, J. Luo, N. Rai, A. Tsoukalas, and Tagkopoulos, I. (2014) An integrative, genome-scale model reveals the phenotypic landscape of *Escherichia coli*. *Molecular Systems Biology*, 10:375, doi:10.15252/msb.20145108

A genome-scale predictive framework for bacterial organisms



Training Data: A Multi-omics compendium for data mining and learning

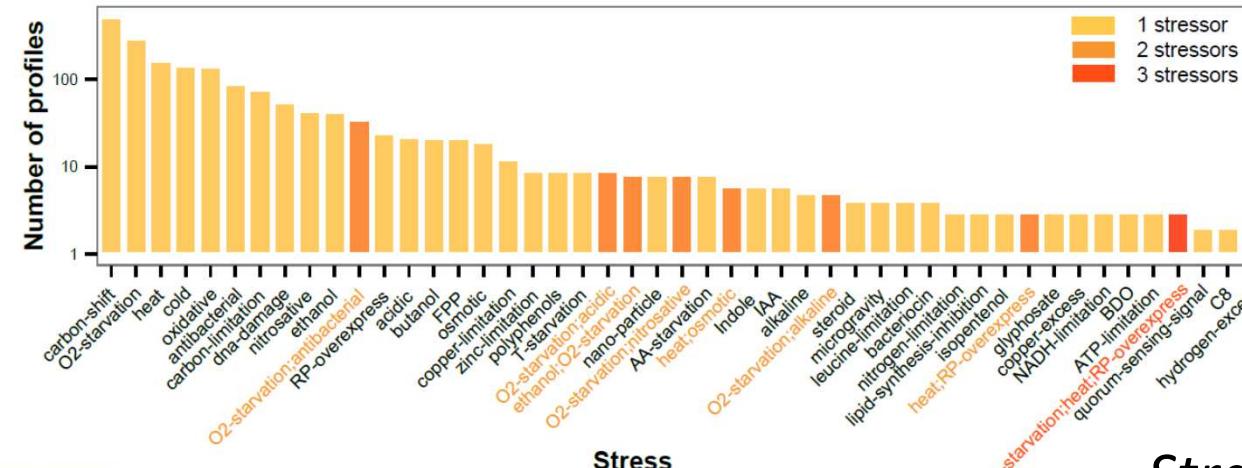


A genome-scale predictive framework for bacterial organisms

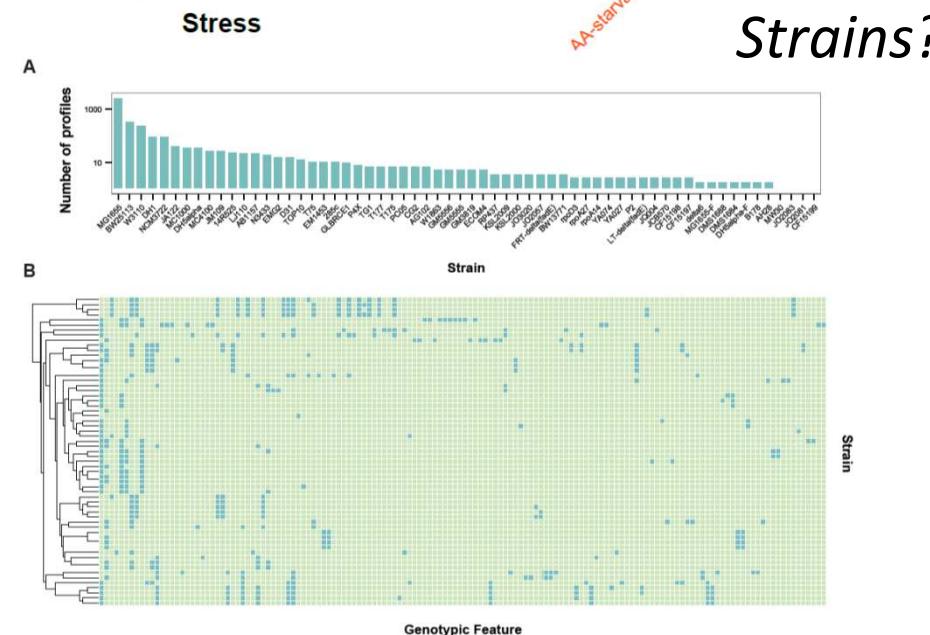
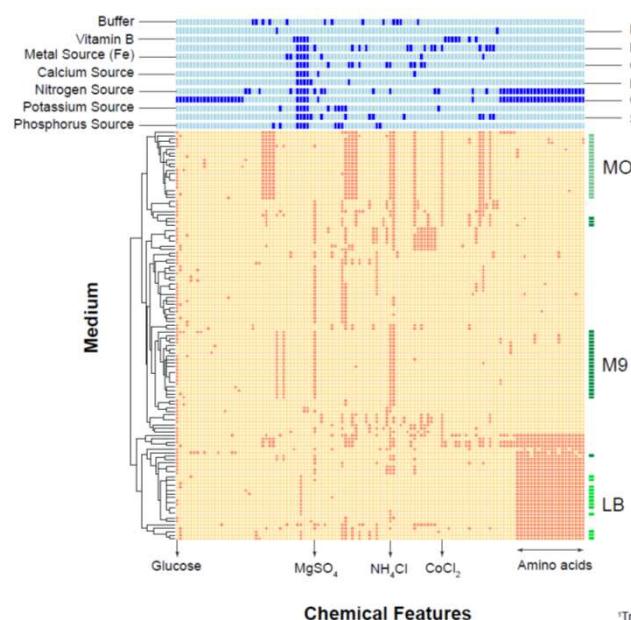


Eye-opening meta-data and analysis: What are we lacking?

*Knowledge on
stresses ?*



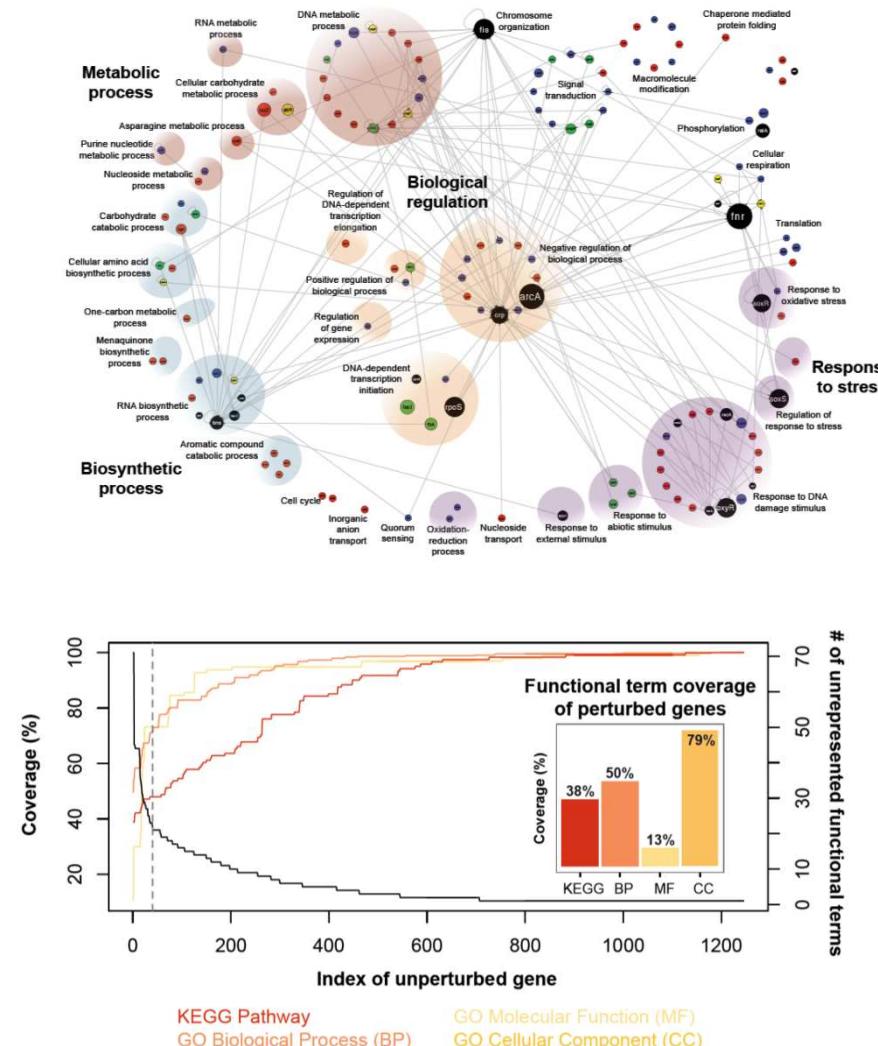
*Media and
chemicals?*



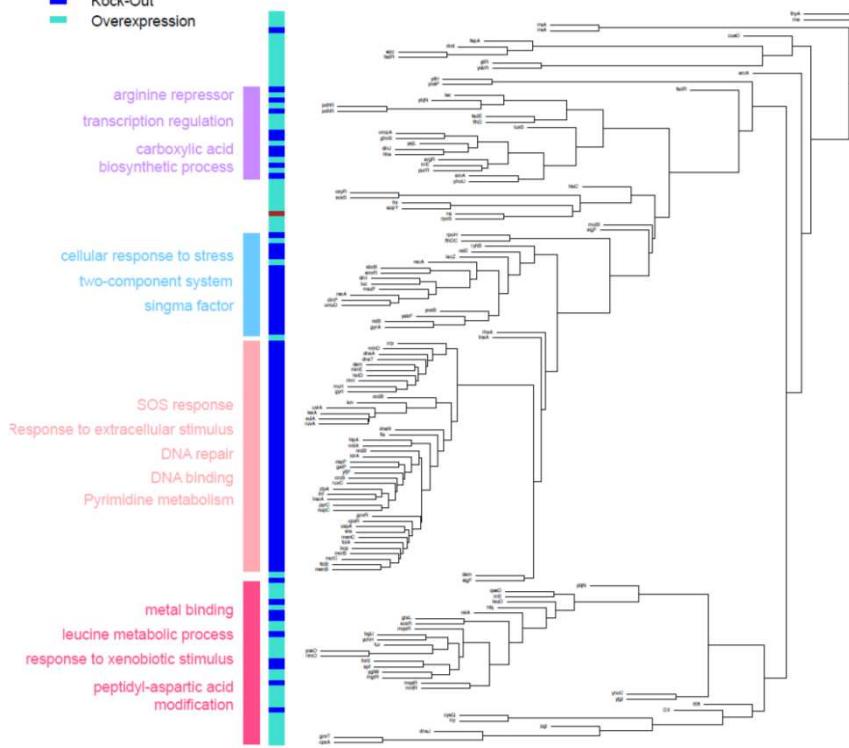
A genome-scale predictive framework for bacterial organisms



Redefining and extending knowledge



Data-driven ontology for genes and environments



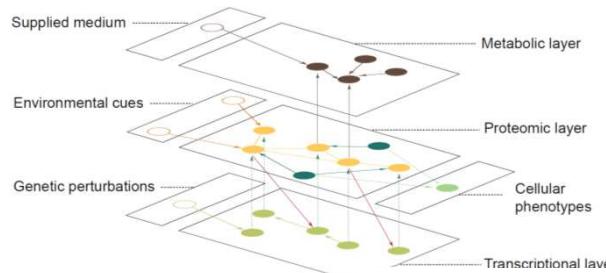
Profiling just 20 genes will increase Ontology coverage by ~15%

A genome-scale predictive framework for bacterial organisms

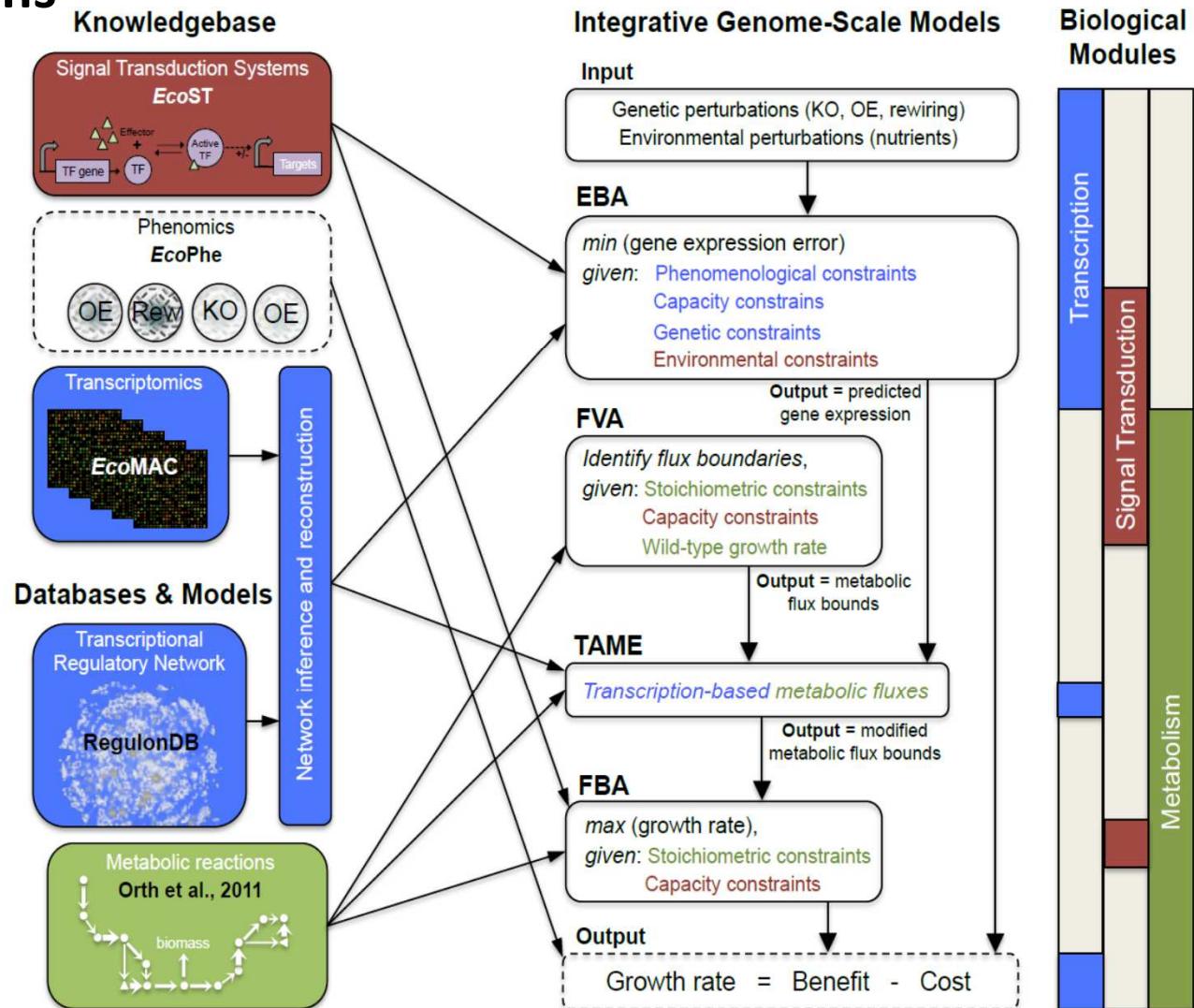
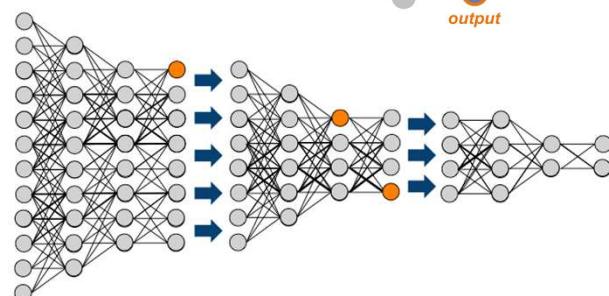


Modeling and Algorithms

Exact and approximate optimization methods



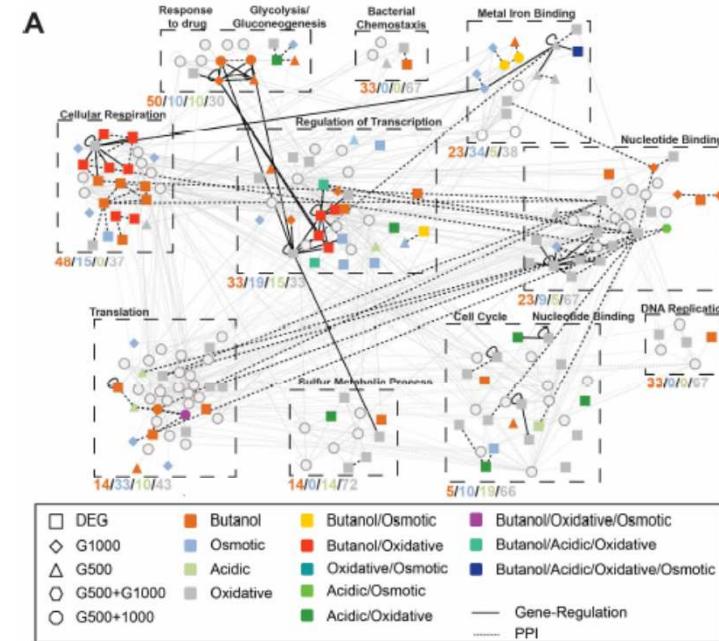
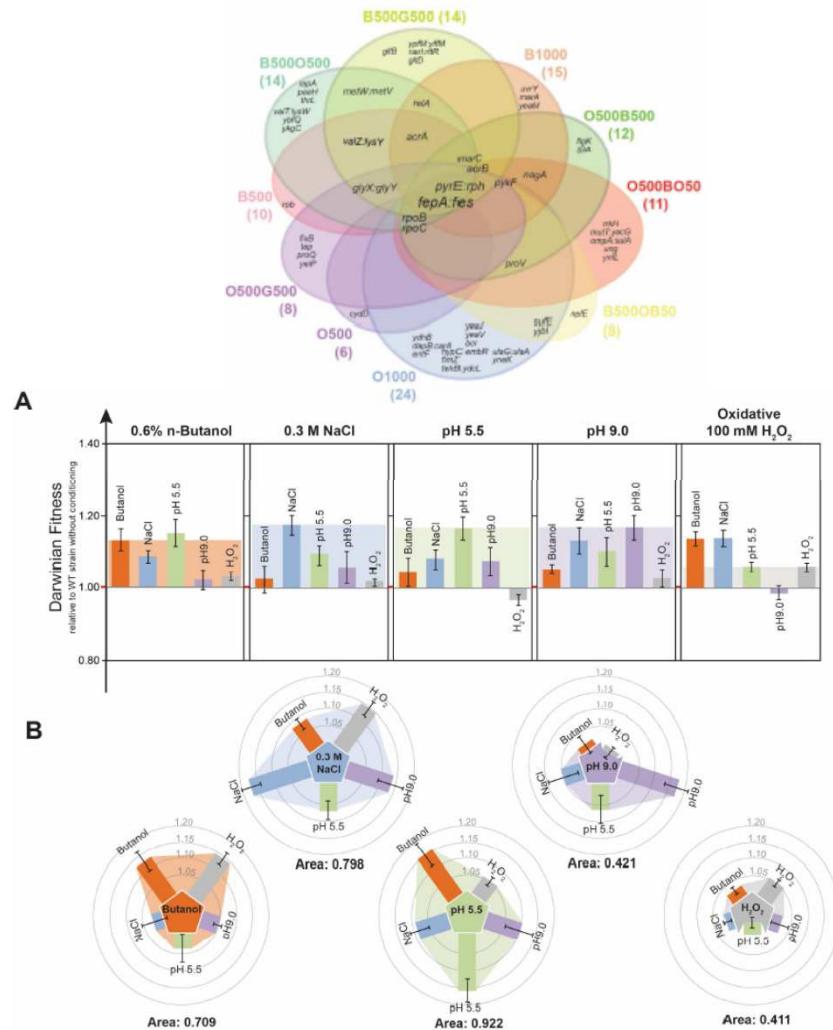
Deep learning across *omics* layers



A genome-scale predictive framework for bacterial organisms



Targeted experimentation (*E. coli* cross-stress behavior)



Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., and Tagkopoulos, I. (2013) Evolutionary potential, cross-stress dependencies, and the genetic basis of acquired stress resistance in *E. coli*. *Molecular Systems Biology*, 9:643. doi:10.1038/msb.

A genome-scale predictive framework for bacterial organisms



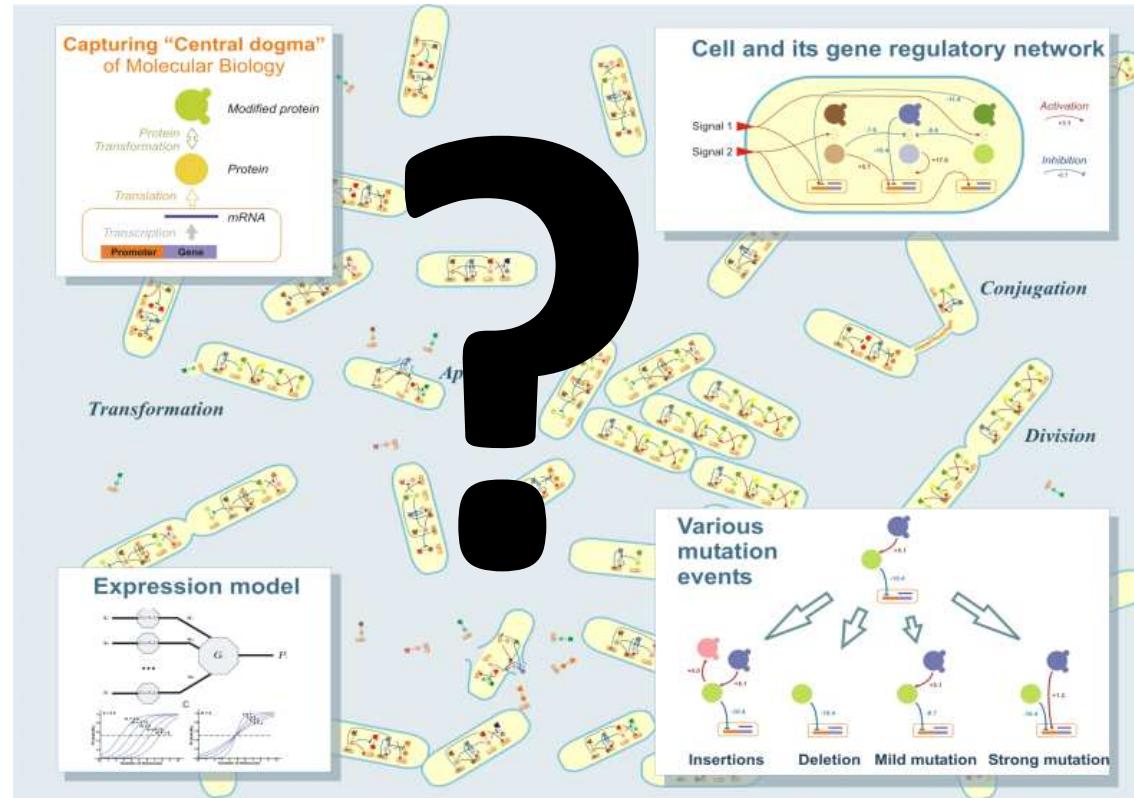
Predicting evolution

Data ?

Models ?

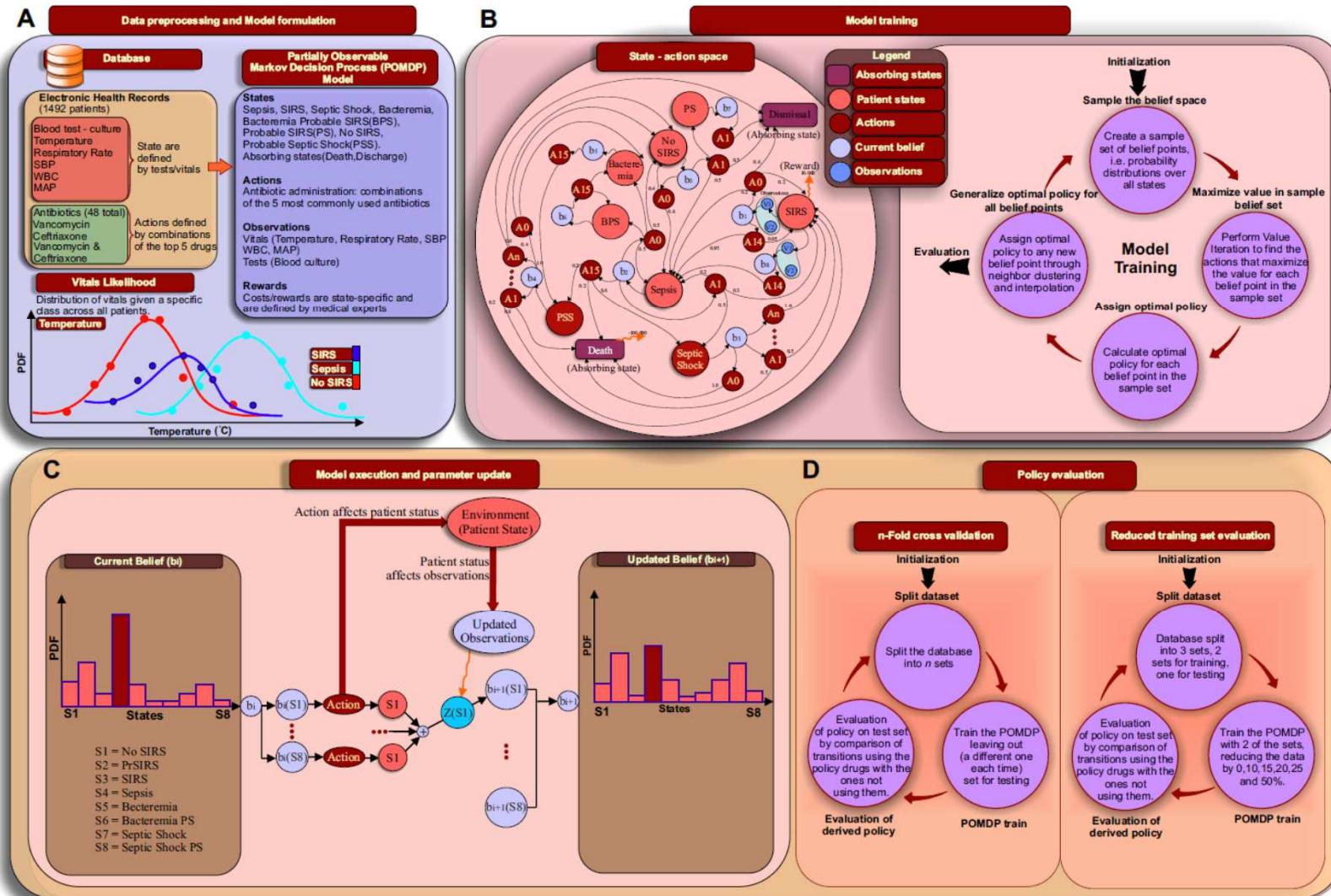
Algorithms?

Evaluation?



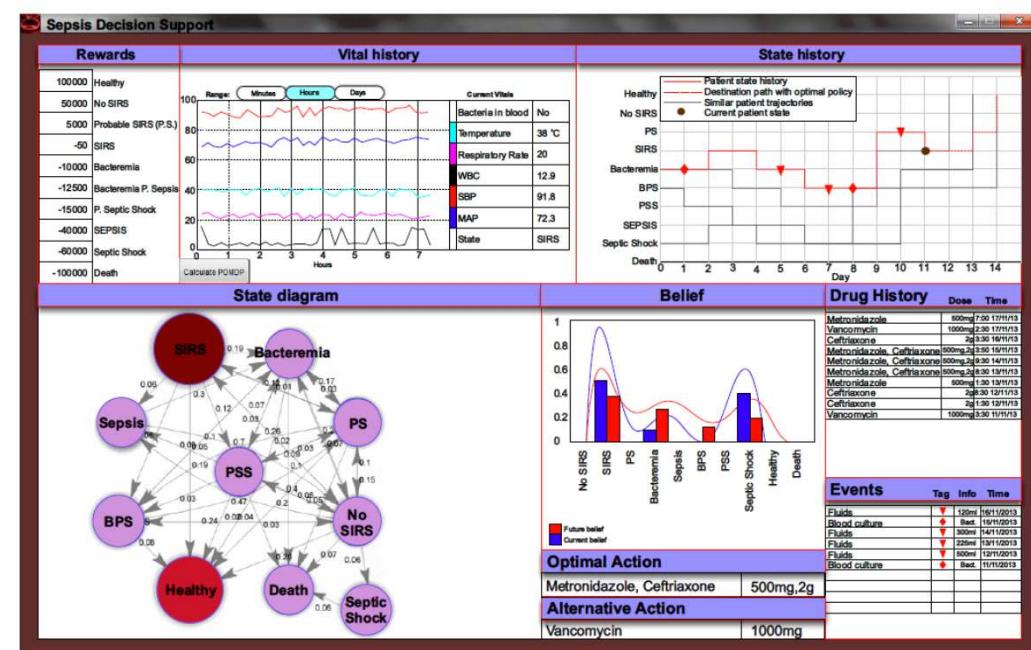
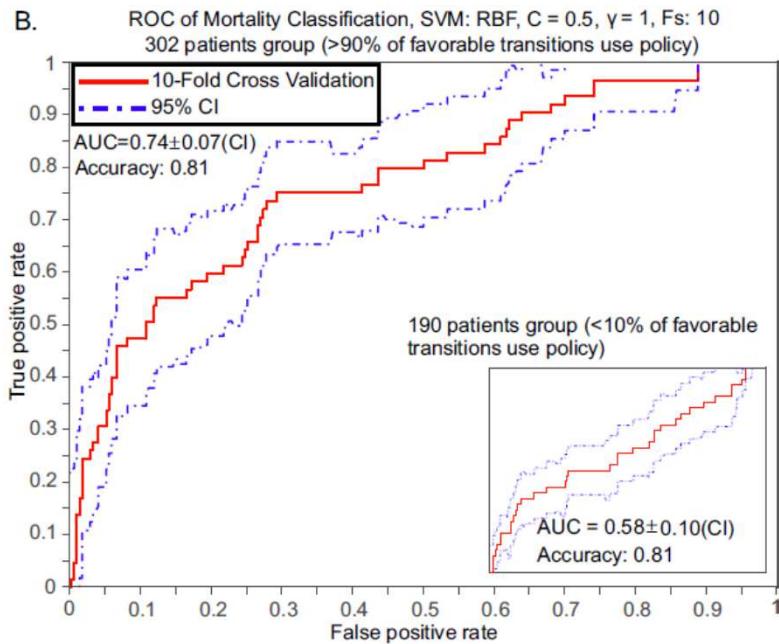
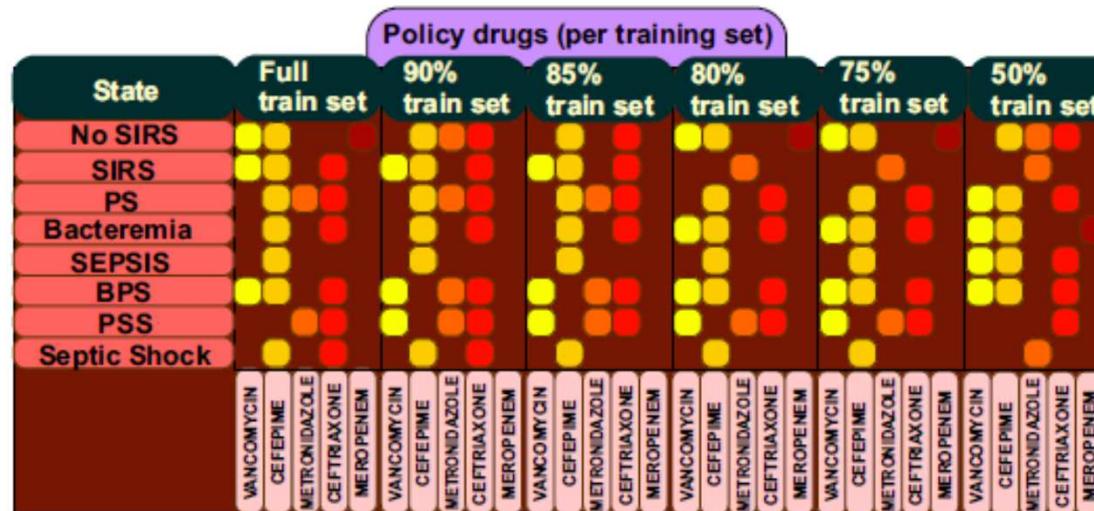
Tagkopoulos, I., Liu, Y., and Tavazoie, S. (2008) Predictive behavior within microbial genetic networks. *Science*, 320: 1313-7.

Medical Informatics: Mining the Electronic Health Records



Gultepe E., Green J., Nguyen H., Adams J., Albertson T., Tagkopoulos I. (2014) From vital signs to clinical outcomes for patients with sepsis: A machine learning basis for a clinical decision support system. *Journal of American Medical Informatics Association (JAMIA)*, 21(2):315-25, doi:10.1136/amiajnl-2013-1815

Medical Informatics: Mining the Electronic Health Records



Another example: agricultural informatics



1498 crop fields in CA San Joaquin valley, 1997 to 2008
Optimal pest management (*Lygus Hesperus*)

Only ~82% of the farmers use the data-predicted optimal management

	5-11Jun	12-18Jun	19-25Jun	26Jun-2Jul	3-9Jul	10-16Jul	17-23Jul
Low	0.97	0.98	0.95	0.97	0.99	0.99	1.00
Medium	0.08	0.89	0.89	0.96	0.98	0.97	0.97
High	0.18	0.67	0.74	0.83	0.88	0.90	0.87

Cost from sub-optimal pest management per acre (yield-related cost only)

	cotton.						
	5-11Jun	12-18Jun	19-25Jun	26Jun-2Jul	3-9Jul	10-16Jul	17-23Jul
Low	\$14.50	\$17.51	\$20.81	\$19.79	\$4.04	\$54.70	\$35.50
Medium	\$45.59	\$2.79	\$18.75	\$18.55	\$22.46	\$17.95	\$45.61
High	\$33.81	\$2.64	\$20.06	\$12.96	\$23.89	\$19.99	\$42.99



“Data! Data! Data!” he cried impatiently. ‘I can’t make bricks without clay.’

Sherlock Holmes - Arthur Conan Doyle, *The Adventure of the Copper Beeches*

- *Data-driven, whole-cell, population-level tools will revolutionize Biology and Bioengineering*
- *Clear and present need to adopt, adapt and develop computational and computer science methods*



Lab members:

Postdocs

- Violeta Zorraquino
- Navneet Rai
- Nasos Tsoukalas

Graduate students

- Linh Huynh
- Minseung Kim
- Xiaokang Wang
- Ameen Etemadi
- Beatriz Pereira

Acknowledgements

Funding



BLUE WATERS PROJECT



Center for Information Technology
Research in the Interest of Society



Links

<http://tagkopouloslab.ucdavis.edu>

Thank you