White-Box Machine Learning
Insights into Antibiotic Lethality

Jason H. Yang, Ph.D.

Assistant Professor and Chancellor Scholar
Microbiology, Biochemistry and Molecular Genetics
Center for Emerging and Re-Emerging Pathogens
Rutgers New Jersey Medical School

January 21, 2020
DIGITAL ID2020 Conference
DEEP LEARNING APPROACHES FOR ANTIBIOTIC DRUG DISCOVERY

Directed message passing neural network

Large scale predictions (upper limit $10^8$+)

Chemical landscape

10^8

10^7

10^6

10^5

10^4

Conventiona antibiotic screening

Chemical screening (upper limit $10^6$ - $10^8$)

Hit validation (1 - 3% hit rate)

Training set ($10^4$ molecules)

Machine learning

Small scale predictions & molecule validation

Growth

[antibiotic]

Lead identification & optimization

STOKES JM, CELL, ACCEPTED
Model-based reasoning can uniquely provide causal mechanistic understanding.
I used to think correlation implied causation.

Then I took a statistics class. Now I don’t.

Sounds like the class helped. Well, maybe.
DCell: Neural Networks Embedded in Gene Ontology (GO)

Visible neural network

Genotype

Small complexes, subunits, reactions
Large complexes, signaling pathways
Organelles, broad processes

Combinatorial gene disruptions

Double-strand break repair
Base-excision repair
Mismatch repair

DNA repair

Predicted growth & genetic interaction
**DO-Calculus: Counterfactual Analysis**

![Diagram showing the DO-Calculus framework for counterfactual analysis in causal inference. The diagram illustrates the transformation of a causal model into an intervention model through observable and intervention joint distributions.]
“What does it mean to say that event X caused outcome Y in biology? Explaining the central structure underlying the dynamic function of living systems is a central goal of biology...”
Q: **WHAT IS A “MECHANISM”?**

**mechanism**  
*me-ke-ni-zem*  

**Definition of mechanism**  

1. a: a piece of machinery (see MACHINERY sense 1)  
   // The camera’s shutter mechanism is broken.
2. b: a process, technique, or system for achieving a result  
   // the mechanisms of peace  
   — F. D. Roosevelt
3. : mechanical operation or action : WORKING sense 2  
   // he acknowledges nothing besides matter and motion; so that all must be performed either by mechanism or by accident  
   — Richard Bentley
4. : a doctrine that holds natural processes (as of life) to be mechanically determined and capable of complete explanation by the laws of physics and chemistry
5. : the fundamental processes involved in or responsible for an action, reaction, or other natural phenomenon
   // meteorologists believe that this pressure jump is the mechanism responsible for storms and tornadoes  
   — *Think*
   — compare DEFENSE MECHANISM

**constituent parts**

**sequence of activities**

**first principles**

**causal, predictive**
A call for a better understanding of causation in cell biology

Morten Bizzarri1, Douglas E. Brash2, Jacques Brucot3, Verenica A. Griesenbeck1, Claudia D. Stern4 and Michael Levin1,2,5,*

What does it mean to say that event X caused outcome Y in biology? Explaining the causal structure underlying the dynamic function of living systems is a central goal of biology. Transformational advances in regenerative medicine and synthetic biomaterials will require efficient strategies to cause desired system-level outcomes. We present a perspective on the need to move beyond the classical necessary and sufficient approach to biological causality.

From genes to processes in development

Causal theory and the development of specific tissues have been identified and organized into pathways. This has increased our understanding of how organs are created and what goes wrong in disorders and diseases, but in a very small biological processes tend to be driven by genes with simple, fixed and specific functions within biological pathways, in which many regulatory nodes are driven developmentally. That, to understand and machinery, and provide conceptual insight into how and why processes occur, a shift in attention is required from genes to patterns and dynamics of the causation between components.

This change in perspective requires an interdisciplinary mixture of theory and experiment. The approach results in new models and strategies that involve the interaction between the genetic and environmental factors that contribute to the outcome. Biological processes are not separate entities but are interdependent and continuous, and thus the interactions between them are dynamic and regulatory.

Biophysical properties as causes

Specifying distributed physical factors, of which standard parameters, system mechanics, and interfacial organization, microsystem biology has become crucial for understanding how these factors interact with each other and with the environment. Through the processes described above, proteins in the extracellular matrix, intracellular signaling, and microenvironmental factors contribute to the development of specific tissues. When these interactions are sufficiently complex, they can lead to the development of new tissues and organs.

An appreciation of new kinds of instructive influence from network science and physics must be coupled with hypothesis testing, focused on models of functional processes. Developing methodology for the rigorous identification and efficient control of causal processes in complex biological systems is very much a nascent field that is ready to be integrated with molecular approaches to growth and form."

“What does it mean to say that event X caused outcome Y in biology? Explaining the causal structure underlying the dynamic function of living systems is a central goal of biology ..."
**Mechanistic Discovery is Difficult by Conventional Approaches**

**INFORMATICS-Driven Approaches**
- Chemogenomic library
- Disease-relevant phenotypic screen
- Target ‘hits’
- Pathway analysis, e.g., causal reasoning
- Testable hypotheses

**BARRIER 1:** Statistical power limitations

**DATA-Driven Approaches**
- ML Input Data
- ML Output Data

**BARRIER 2:** Mechanistically uninterpretable

**Counter-screens, e.g., cytotoxicity and fluorescence**18,46(450,233,890,404)

**JONES LH, NAT REV DRUG DISCOVERY 2017**

**YANG JH, CELL 2019**
A MODEL-DRIVEN “WHITE-BOX” MACHINE LEARNING APPROACH

ML Inputs: simulated network states

ML Outputs: measured phenotypes

pathway mechanisms

biological features
MECHANISMS OF ANTIBIOTIC LETHALITY

- Antibiotics
- Membrane regeneration
- Transcription
- Translation
- DNA replication
- Cell death
- Altered cell metabolism

REFERENCES:

YANG JH*, BENING SC*, Curr Opin Microbiol 2017
MEYLAN S, CELL CHEM BIOL 2017
TAKAHASHI N*, GRUBER CC*, PNAS 2017
LOBRITZ MA*, BELENKY P*, PNAS 2015
DWYER DJ*, BELENKY P*, YANG JH*, PNAS 2014
LOPATKIN AJ, NAT MICROBIOL 2019
A WHITE-BOX MACHINE LEARNING APPROACH FOR DISCOVERING ANTIBIOTIC MECHANISMS OF ACTION

BARRIER 1: Statistical power limitations

≤ 3% HIT RATE

This approach expands the biological information content from screening data.
A WHITE-BOX MACHINE LEARNING APPROACH FOR DISCOVERING ANTIBIOTIC MECHANISMS OF ACTION

This approach enables us to interpret experimental data through the mechanistic lens of prior knowledge.

BARRIER 2: Mechanistically uninterpretable

White-Box ML Predictive Model:

ML Inputs:
- Network Simulations
- metabolite conditions
- metabolic fluxes

Pathway Mechanisms

ML Outputs:
- Screening Data
- antibiotic lethality measurements

YANG JH, CELL 2019
ANTIBIOTIC RESPONSES TO METABOLISM ARE PATHWAY-SPECIFIC
CENTRAL CARBON METABOLISM CONTRIBUTES TO ANTIBIOTIC LETHALITY

Superpathway of phenylethylamine degradation
Superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass
Superpathway of glyoxylate bypass and TCA
TCA cycle I (prokaryotic)
Inosine-5'-phosphate biosynthesis I
Superpathway of sulfate assimilation and cysteine biosynthesis
Superpathway of histidine, purine, and pyrimidine biosynthesis
Superpathway of purine nucleotides de novo biosynthesis II
Arginine biosynthesis I (via L-ornithine)
Superpathway of glycolysis and the Entner-Doudoroff pathway
Pentose phosphate pathway
Superpathway of 5-aminomidazole ribonucleotide biosynthesis
Salvage pathways of pyrimidine ribonucleotides
H: Deletion of purine biosynthesis genes will decrease AMP or CIP lethality and increase GENT lethality.
PURINE BIOSYNTHESIS PARTICIPATES IN ANTIBIOTIC LETHALITY
PURINE BIOSYNTHESIS PARTICIPATES IN ANTIBIOTIC LETHALITY

YANG JH, CELL 2019
Model-based reasoning can uniquely provide causal mechanistic understanding.
YANG SYSTEMS BIOLOGY LAB AT RUTGERS NJMS

WE’RE RECRUITING!
VISION: WE CAN LEARN DISEASE MECHANISMS FROM CLINICAL SPECIMENS

Clinical Isolates from TBTC Study 22

SAEZ-RODRIGUEZ J, MOL SYS BIOL 2020
COLANGELI R, NEJM 2018
ACKNOWLEDGMENTS

MIT / Broad Institute
James Collins
Miguel Alcantar
Meagan Hamblin
Jonathan Stokes
Sarah Wright

Barnard College
Allison Lopatkin

UC San Diego
Bernhard Palssön
Douglas McCloskey
Lars Schrübbers

Rutgers NJMS
David Alland
Roberto Colangeli

Funding / Support
NIH (NIGMS, NIAID)
K99 GM118907
U19 AI111276

DTRA
HDTRA1-15-1-0051

Website: www.jasonyanglab.org
Email: jason.y@rutgers.edu
Twitter: jasonhyang