

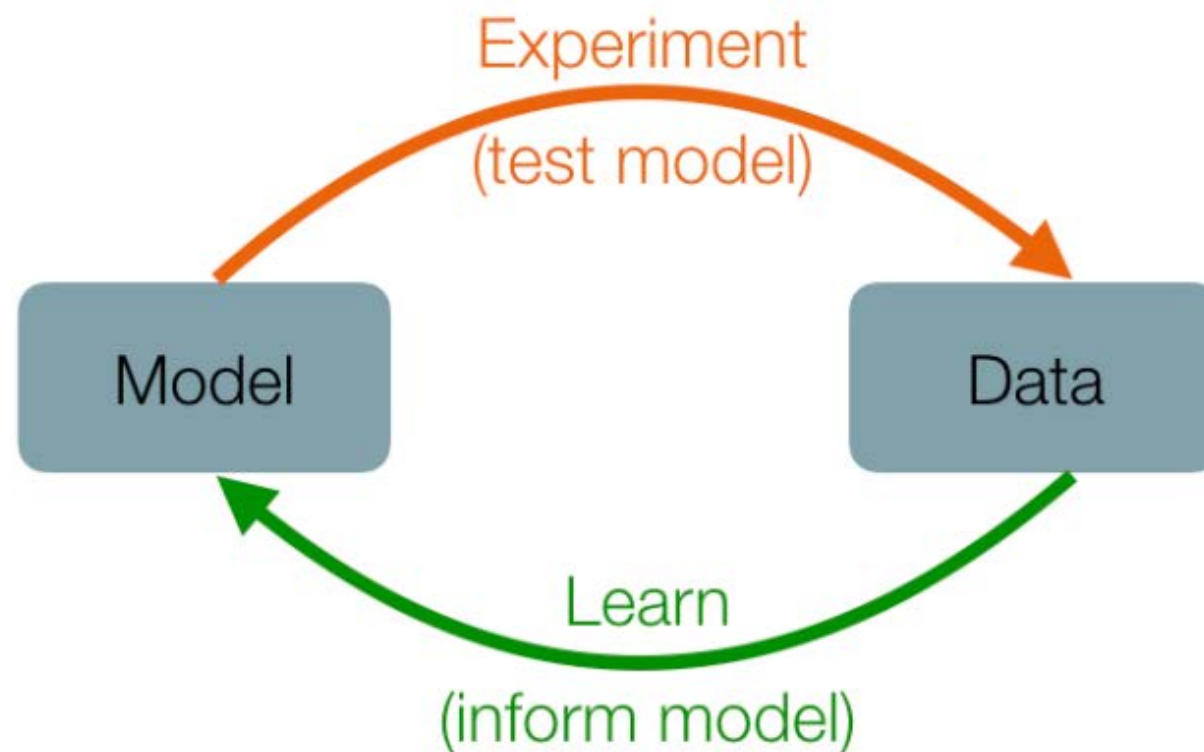
Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics

Mayank Baranwal^{1,2†}, Ryan L Clark^{3†}, Jaron Thompson⁴, Zeyu Sun⁵,
Alfred O Hero^{5,6,7*}, Ophelia S Venturelli^{3,4,8*}

¹Department of Systems and Control Engineering, Indian Institute of Technology, Bombay, India; ²Division of Data & Decision Sciences, Tata Consultancy Services Research, Mumbai, India; ³Department of Biochemistry, University of Wisconsin-Madison, Madison, United States; ⁴Department of Chemical & Biological Engineering, University of Wisconsin-Madison, Madison, United States; ⁵Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, United States; ⁶Department of Biomedical Engineering, University of Michigan, Ann Arbor, United States; ⁷Department of Statistics, University of Michigan, Ann Arbor, United States; ⁸Department of Bacteriology, University of Wisconsin-Madison, Madison, United States

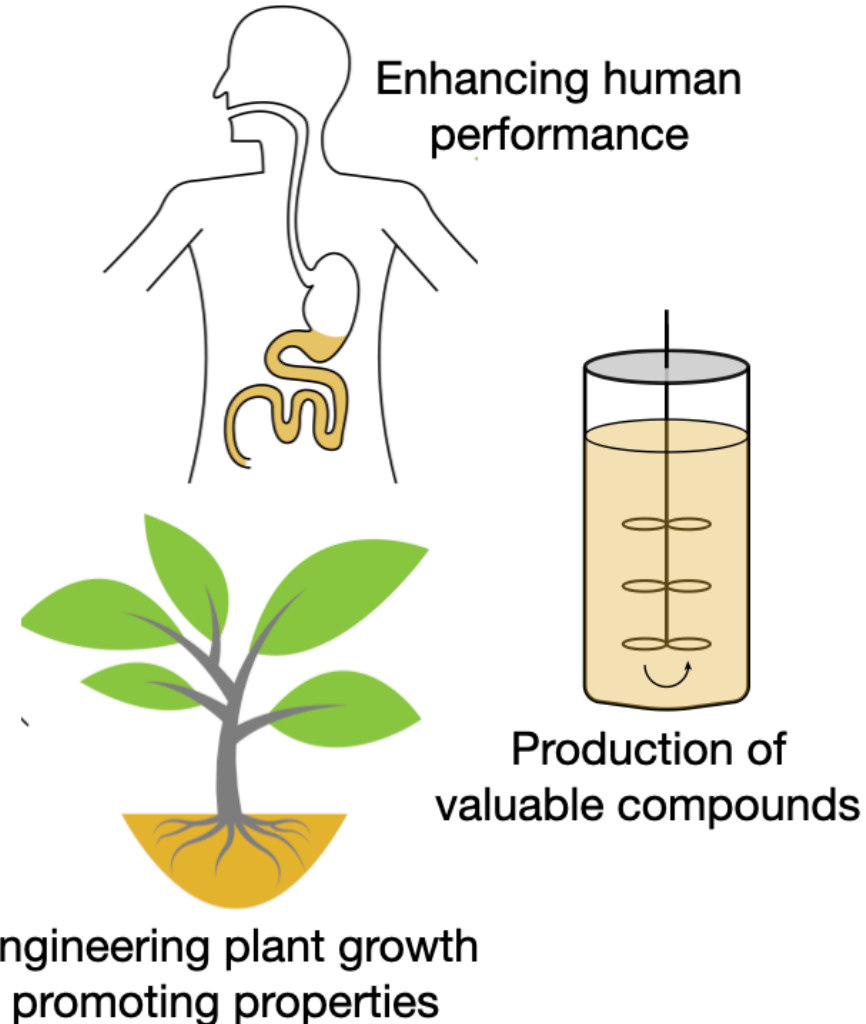
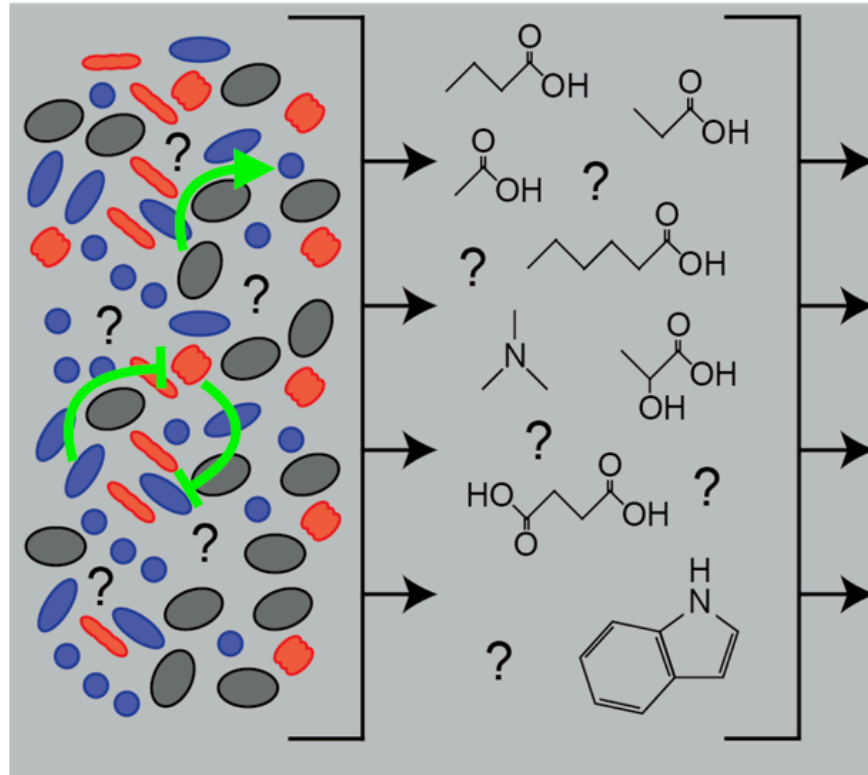
Takeaways

1. Our ML model accelerates discovery of novel microbial communities
2. ML capable of learning relations btwn composition, growth, and metabolic function from very limited data
 - Behavior of 33M communities predicted from a handful (413) of experiments
 - Full metabolic-yield dynamics predicted from a single timepoint (endpoint)
 - Discovery of design principles: growth vs metabolic specialization of bacteria
3. We accomplish this by embedding ML into scientific discovery process
 - Stage 1: Perform handful of experiments, “seed” the ML predictor
 - Stage 2: Use outcomes to perform additional experiments, “update” predictor
 - Generalization: temporal profile of metabolic yield predicted from static data



- D. Atkins, *et al*, “Automated Research Workflows For Accelerated Discovery: Closing the Knowledge Discovery Loop,” Report on a consensus study from the National Academies of Science, Engineering and Medicine, 2022.

Molecules produced by gut microbiota shape host health and ecological dynamics

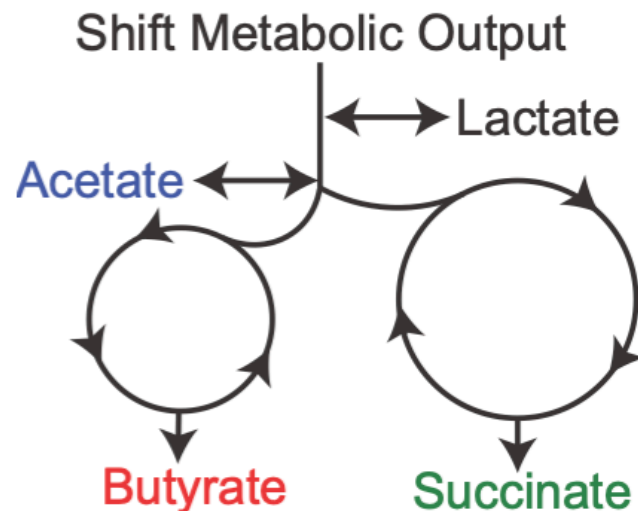
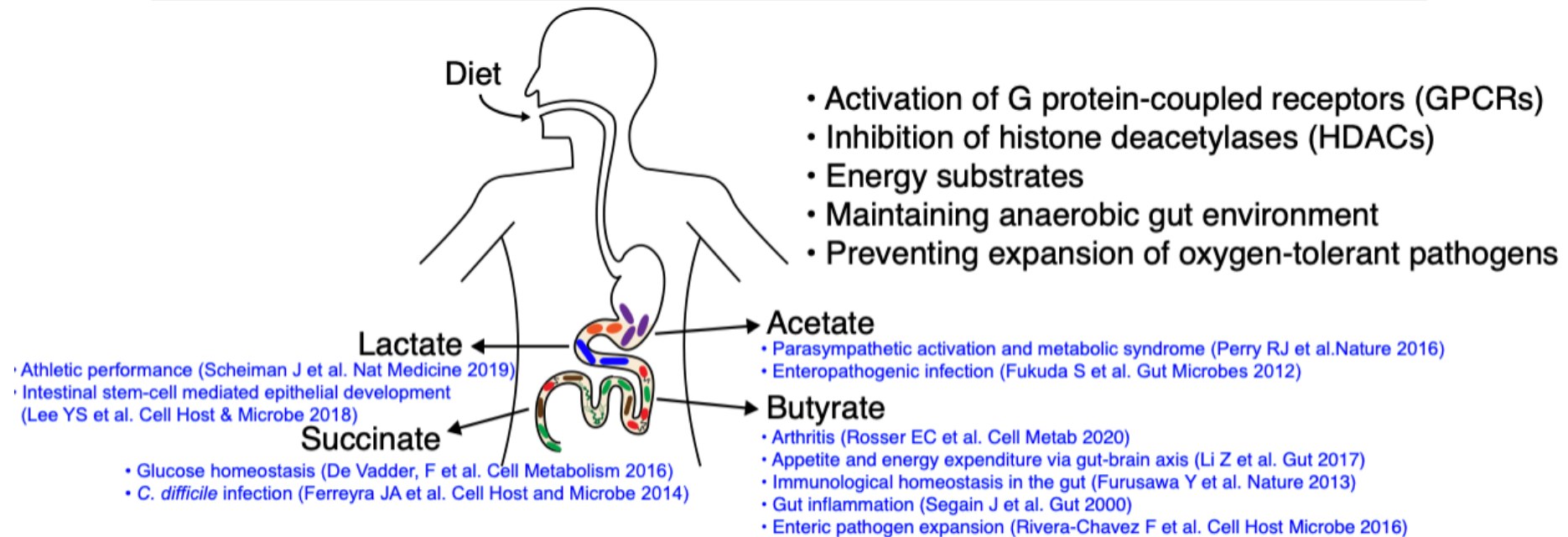


Community

Function

Impact

Microbial fermentation end products are associated with human health and disease



- Predict metabolite profiles of complex communities
- Understand the multidimensional functional landscape of complex communities
- Control metabolite dynamics

Gut microbiome

1. Function: perform chemical and physical transformations
 - Species composition (abundance distribution)
 - Metabolite ingestion and production
2. Role: maintaining human health
 - Mediation of digestive processes
 - human health and disease
3. Gut microbiome is a dynamic community
 - Microbial interactions shape community assembly, growth & function
 - Interactions may be pairwise or higher order
4. Microbiome engineering:
 - Design gut microbiomes to have particular multifunctional profiles
 - Maximally impact health maintenance, disease treatment or performance enhancement

Microbiome community design space is huge

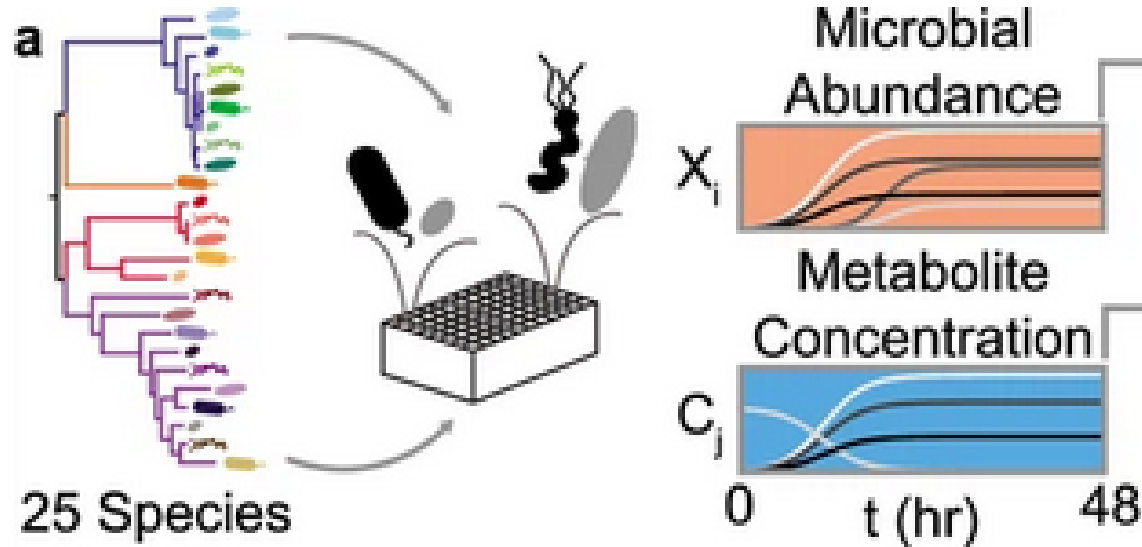


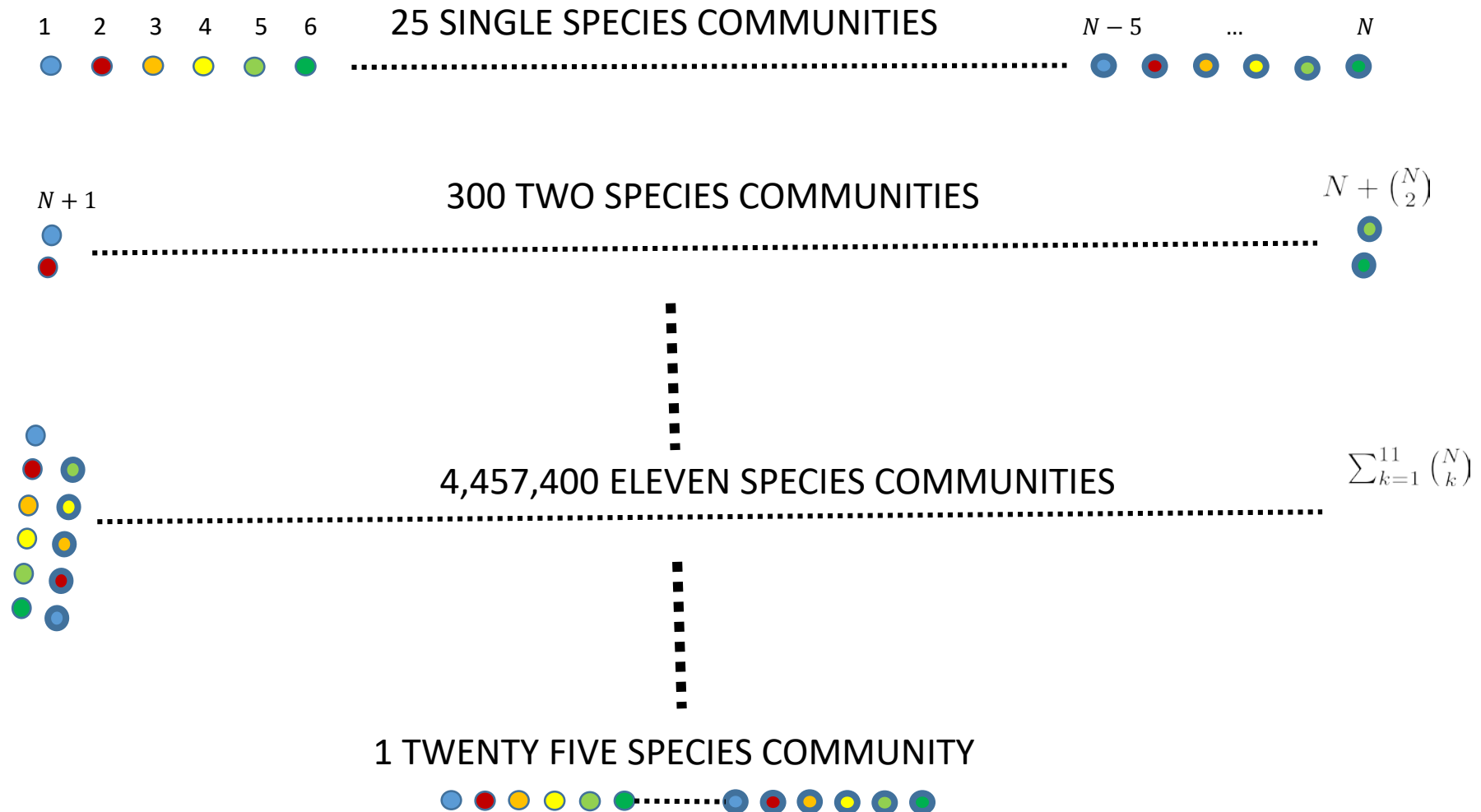
Fig. 3a from our 2022 paper.

- Main challenge: combinatorial explosion of design space

With N species can make $2^N - 1$ distinct communities

➤ Impractical to experimentally test every one of these possibilities

With $N=25$ species one can make
33,554,431 distinct communities



Alternative to exploring entire design space

- Extrapolate from small subset of tests to tests on the entire design space

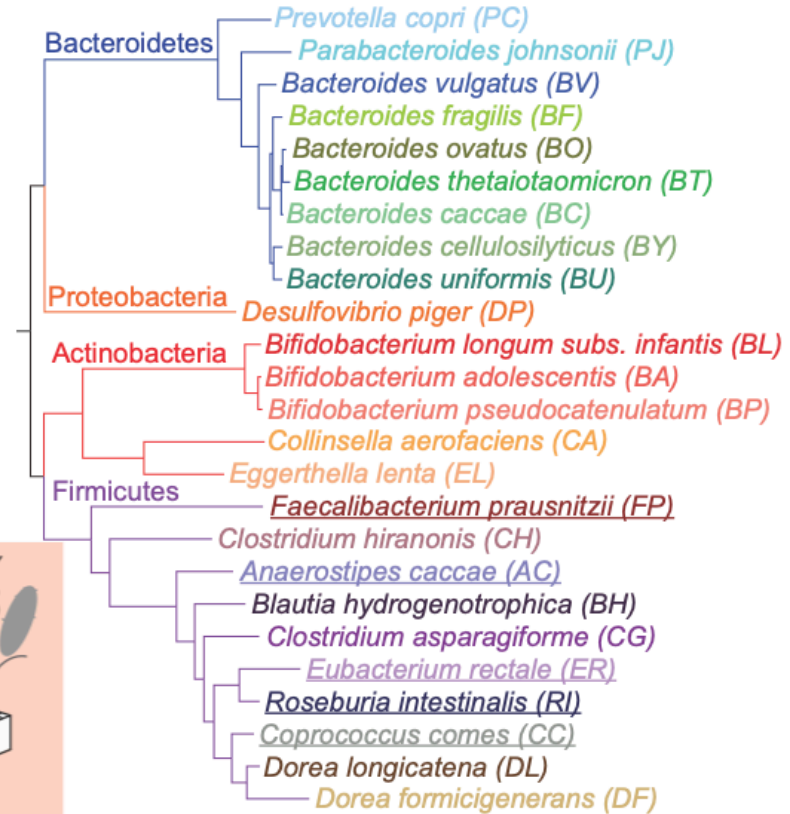
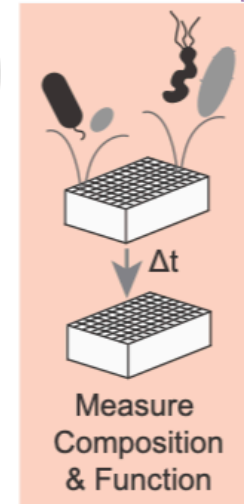
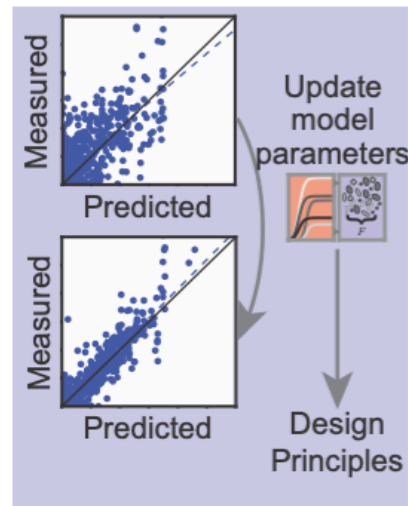
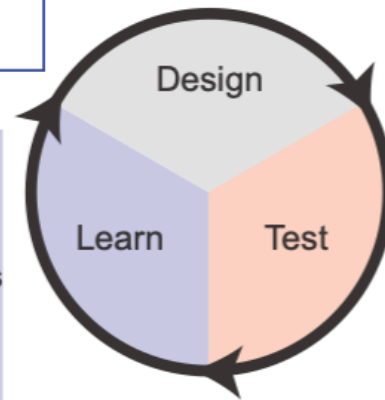
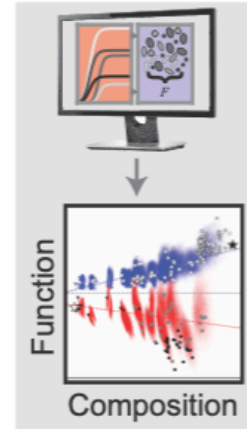
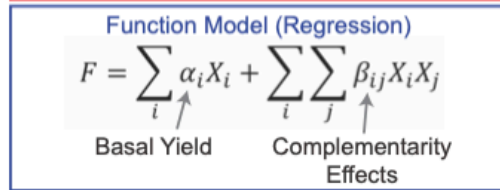
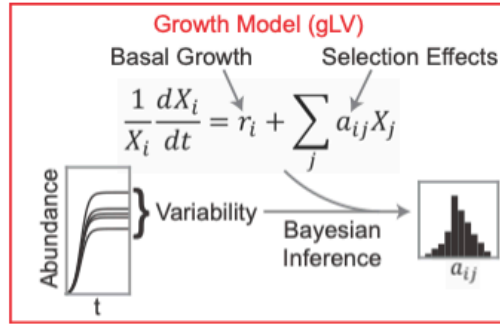


- Main challenge: what predictor to use and how to fit it to subset of MCs?

Design-test-learn cycle to explore the structure-function landscape for a synthetic human gut microbiome



Predator/Prey Model



Clark RL et al. *Nat Commun* 2021

Improving on prior art in predictive modeling

1. Ecological growth models (Venturelli 2018, Clark 2021):

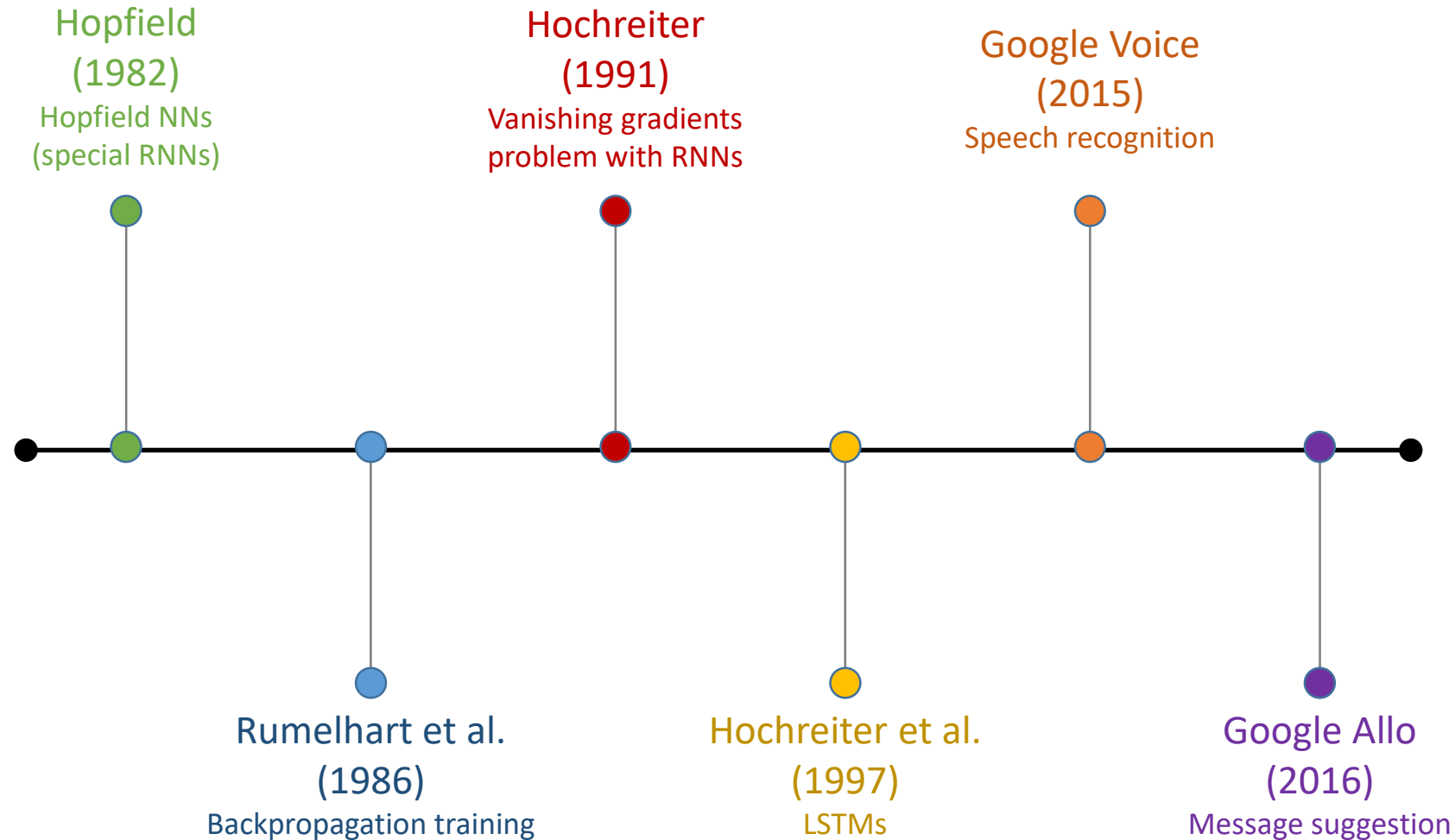
- Generalized Lotke-Volterra model: predicts composition over time but not function
- Composite gLV model (gLV+Regression): restricted to a single function at endpoint
- These models only exploit pairwise interactions between microbial species
- Training the model is computationally heavy

2. This paper: a unified predictor model based on machine learning

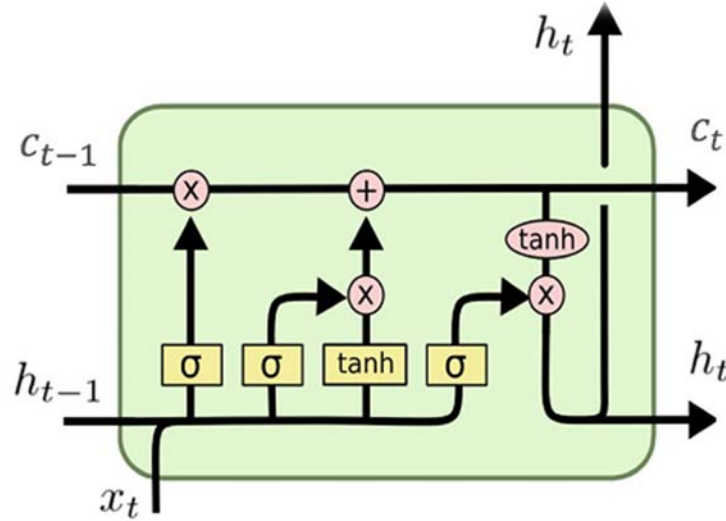
- We use a recurrent neural network (RNN) model to predict composition+function
- The RNN model exploits higher order interactions
- The RNN model predicts multiple community-level functions at any time point

3. Our model results in much higher accuracy than previous models

Recurrent neural networks



RNN/LSTM improves on the gLV ecology model



LSTM
(Long-Short Term Memory)

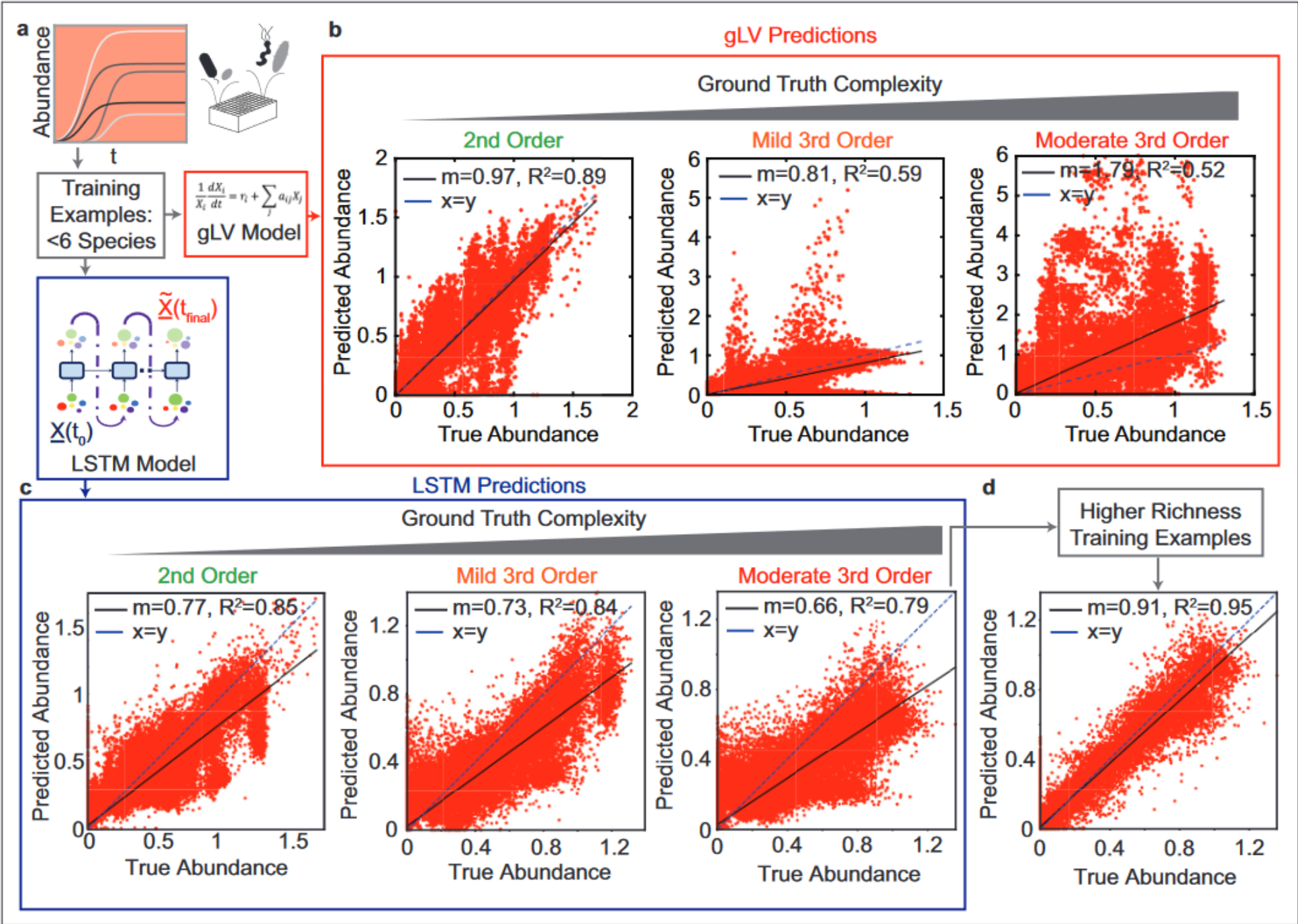
VS



gLV

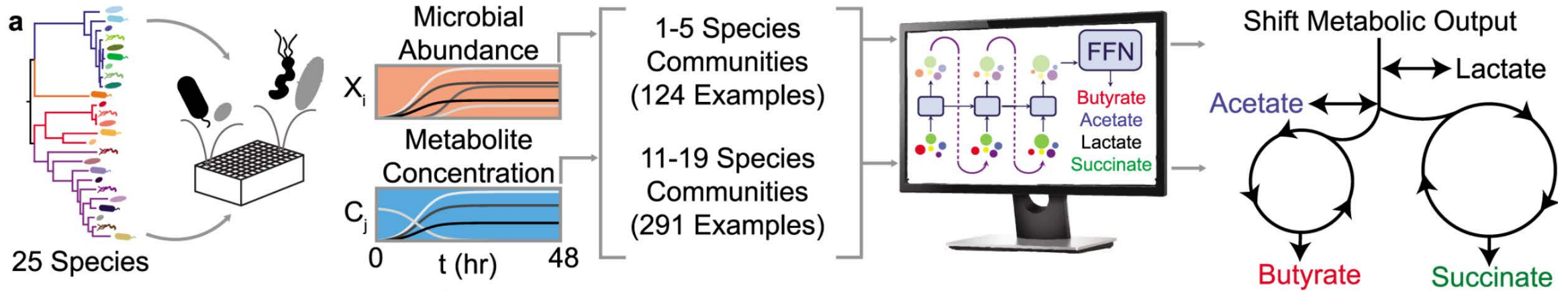
Our findings:

1. LSTM as good as gLV when there are only 2nd order interactions
2. LSTM better than gLV when there are higher order interactions



Fig, 1 from our 2022 paper.

Experiment 1: conceptual diagram



- Experiment creates data for 413 of the possible 33M communities
- LSTM is trained on composition and function of the 413 communities
- LSTM is used to predict function of the remaining possible communities
- These predictions are used to inform a 2nd validation experiment

Experiment 1: all-community prediction

Alfred Hero, RNN microbiome. eLife June 23 2022

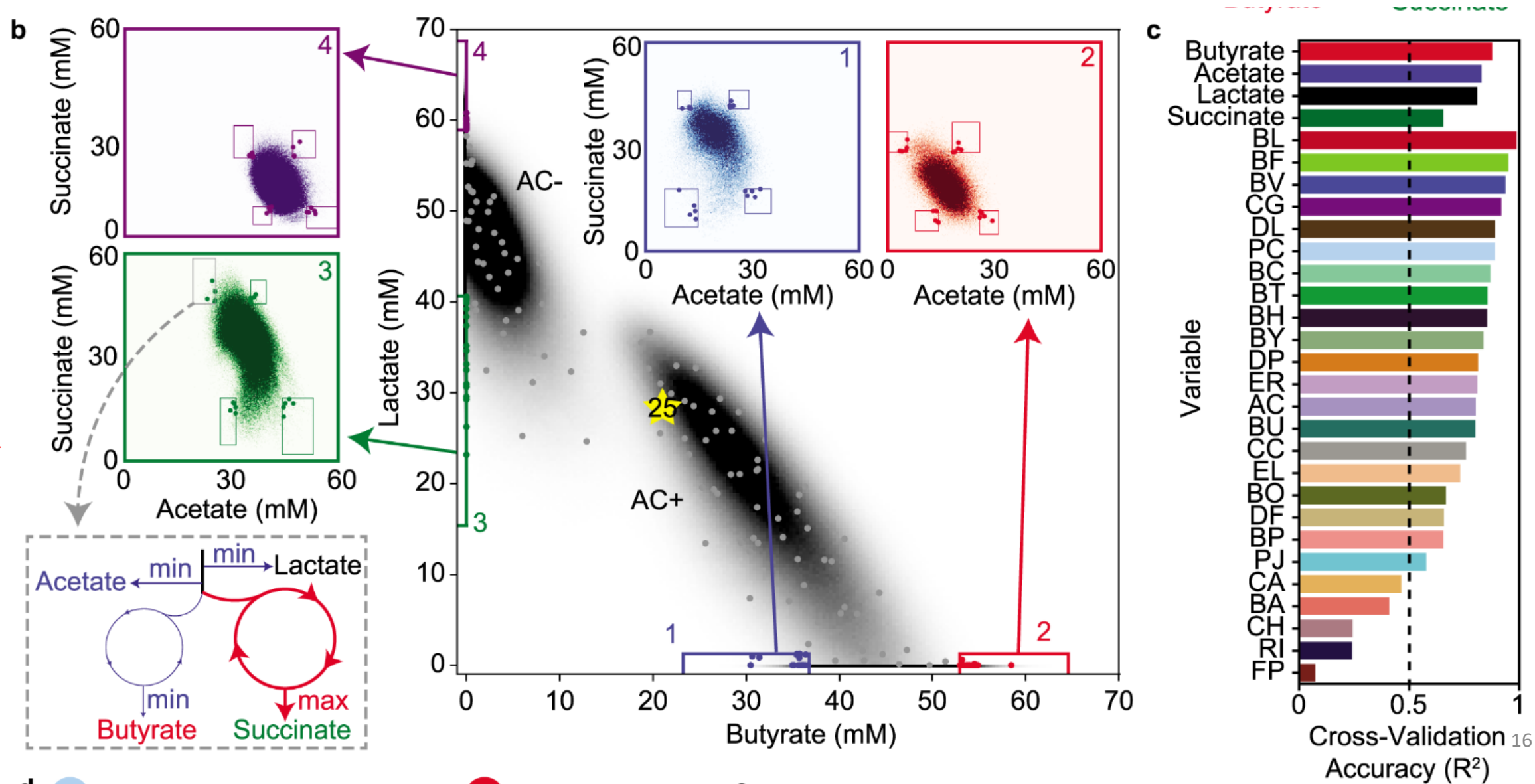
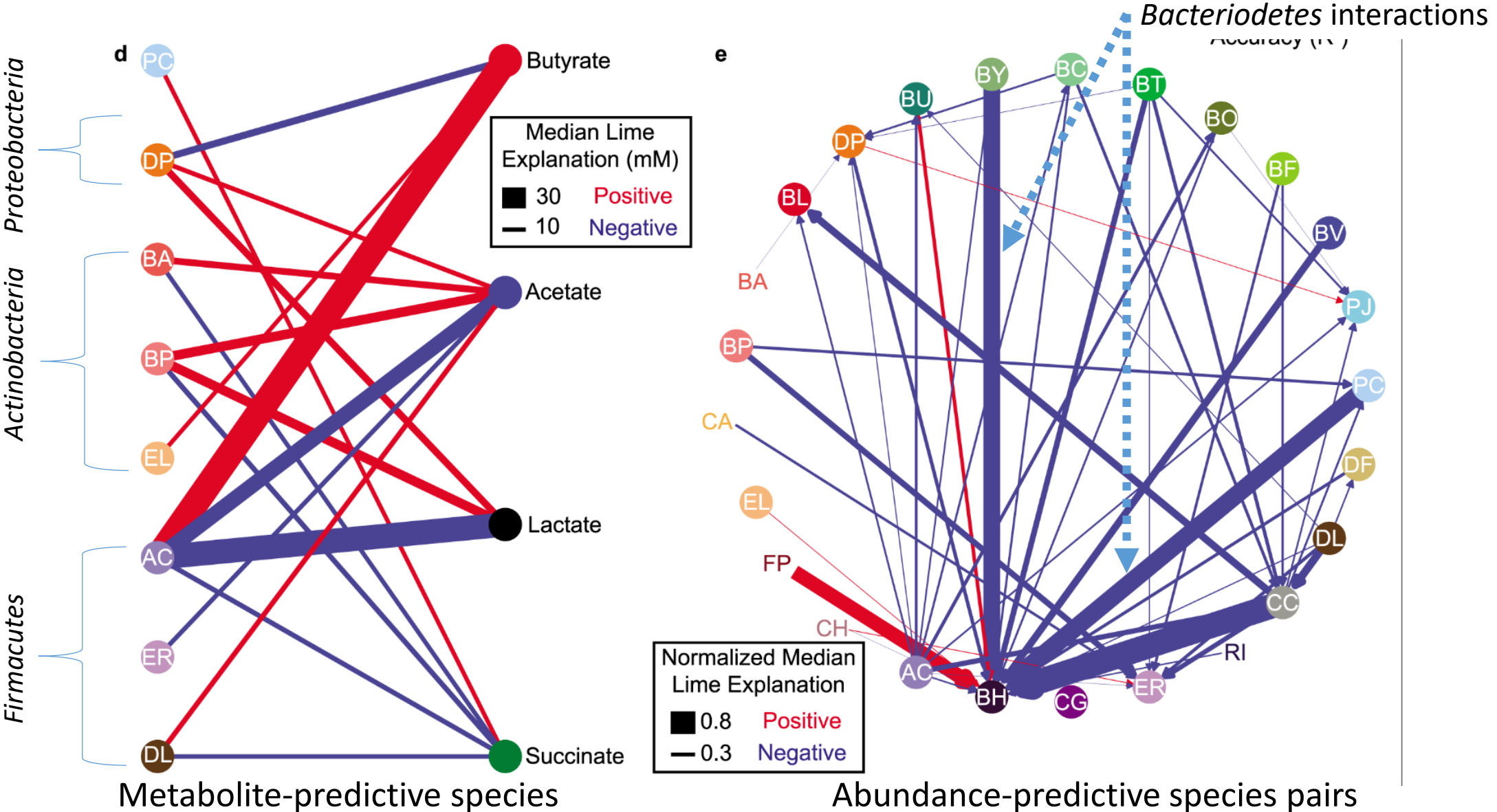


Fig. 2b,c from our 2022 paper.

Experiment 1: model explains species' roles

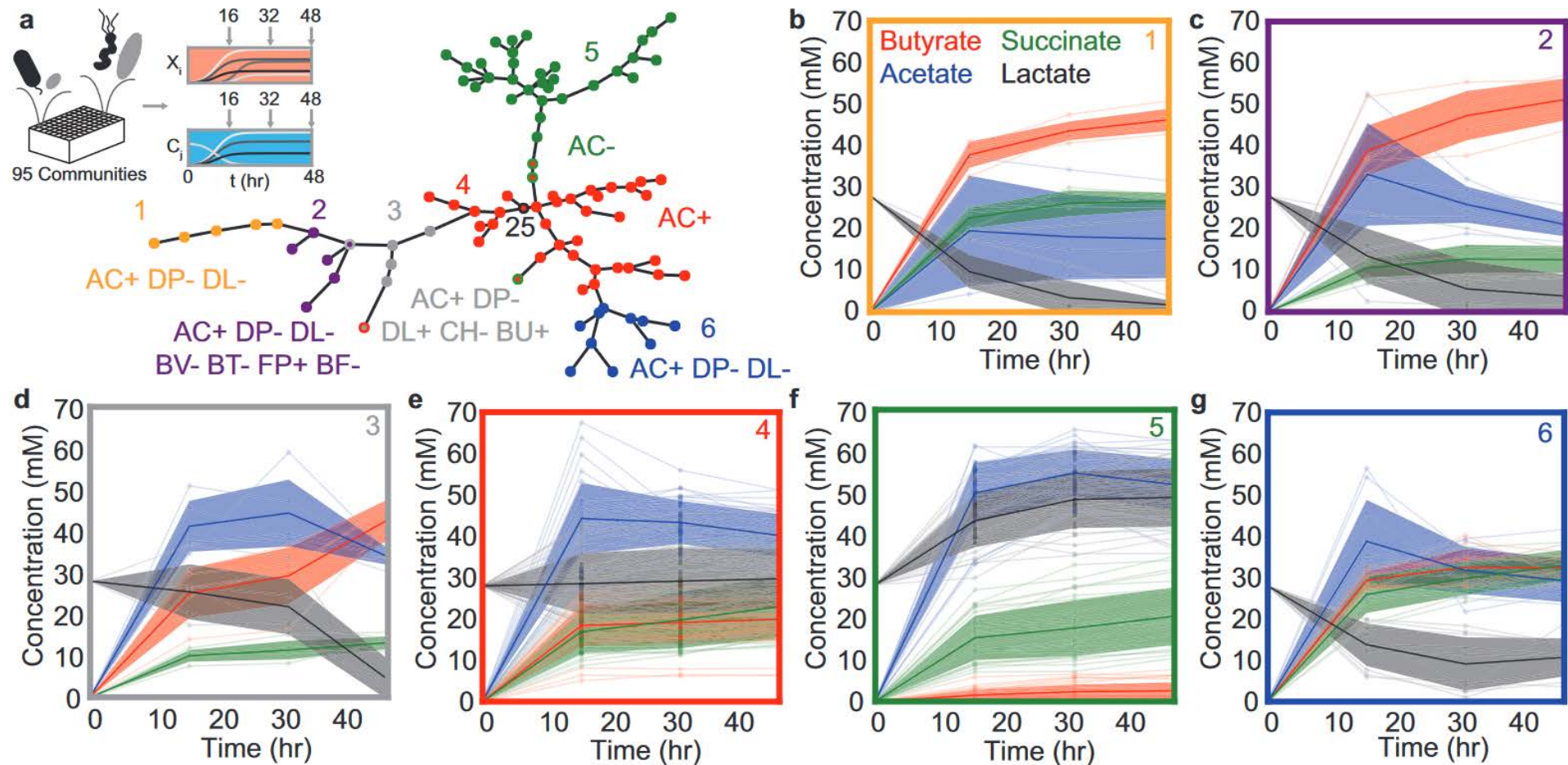


Fig, 2d,e from our 2022 paper.

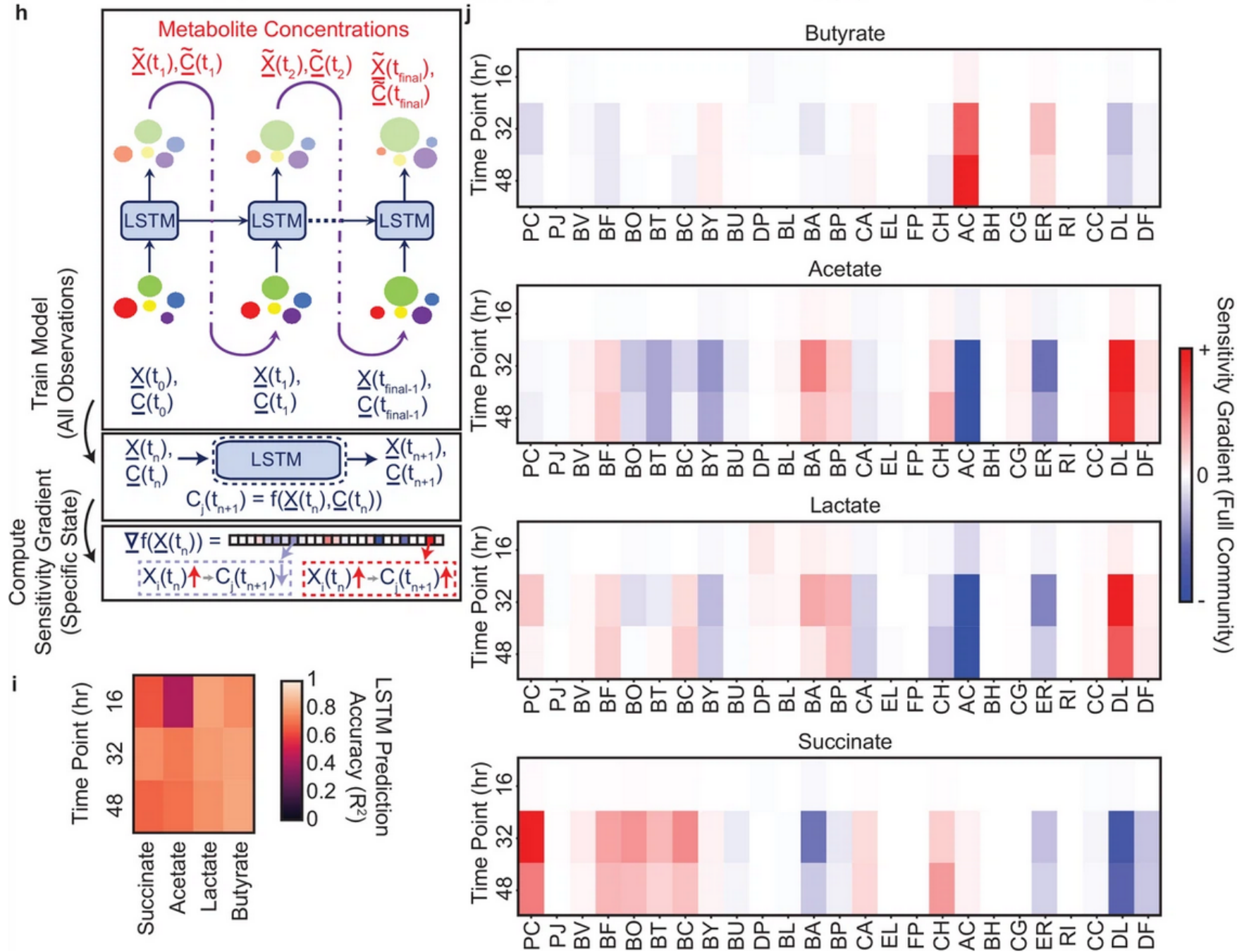
Alfred Hero, RNN microbiome. eLife June 23 2022

Experiment 2: validation on 95 novel communities

1. In Experiment 2 we tested 95 a representative set of communities predicted by Experiment 1
2. Added measurements of full dynamic metabolomic profiles in addition to species growth profiles
3. Corner cases identified in Exp 1, validated with near zero Butyrate or Lactate at 48hr endpoint
4. Model updated (next slide) with



Fig, 5a-g from our 2022 paper.



Fig, 5h-j from our 2022 paper.

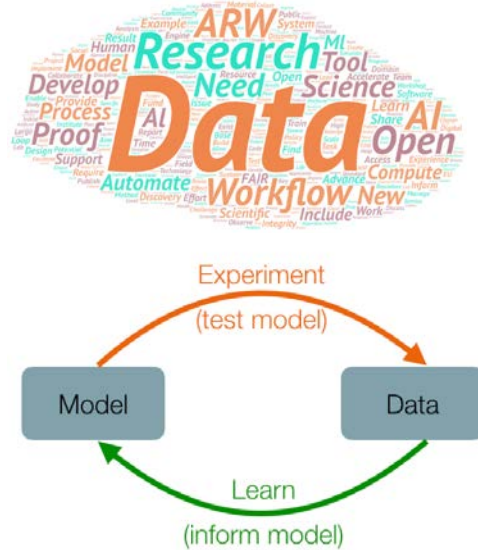
Conclusions

1. Our embedded ML model accelerates scientific discovery
 - Reduces number of experiments (cost)
 - Improves quality of predictions (accuracy)
2. Our embedded ML model is interpretable - explainable-AI:
 - Insights into microbe-metabolite interactions driving metabolic yield
 - Insights into microbe-microbe interactions driving community growth
3. Useful design principles have emerged from our study
 - Drivers of metabolic yield (butyrate, acetate, lactate):
 - *Firmacutes* (AC, ER, DL), *Actinobacteria* (BA, BP, EL), *Proteobacteria* (DP)
 - Drivers of growth dynamics
 - *Bacteroids* (BH, PC)

Future directions

1. Validation on *in vivo* gut microbiome experiments
2. Design environment and growth media for optimizing yield
3. Application to different ecological microbiomes: plants, antibiotics
4. Deeper embedding of ML into design cycle – ARW systems

Automated Research Workflows For Accelerated Discovery: Closing the Knowledge Discovery Loop (2022)



[Link to news release.](#)

Principal recommendations

1. ARW design principles should

- Facilitate openness, reproducibility, and transparency.
- Incorporate principles of responsible AI.
- Prioritize reuse and sustainability.
- Be driven and controlled by the research community.

2. Infrastructure, code and data sustainability for ARWs should be developed that support open standards & interfaces, reproducibility, and FAIR data sharing

3. Human resource investments by funders, schools, and professional societies to develop educational programs and career tracks in ARW.

4. Culture and incentives for encouraging behaviors central to realizing the potential of ARW.

5. Preservation of privacy must be robustly addressed in the ARW world.

US National Academies Study - Committee

Dan Atkins (Chair) – University of Michigan

Ilkay Altintas – San Diego Supercomputer Center

Shreyas Cholia – Lawrence Berkeley National Laboratory

Merce Crosas – Government of Catalunya

Alfred Hero – University of Michigan

Rebecca Lawrence – F100 Research Ltd, London

Bradley Mailin – Vanderbilt University

Lara Magravite – Sage Bionetworks

Bian Nosek – Center for Open Science

Tapio Schneider – California Institute of Technology