

# A Comprehensible Framework to Active Learning Genome-Scale Metabolic Networks

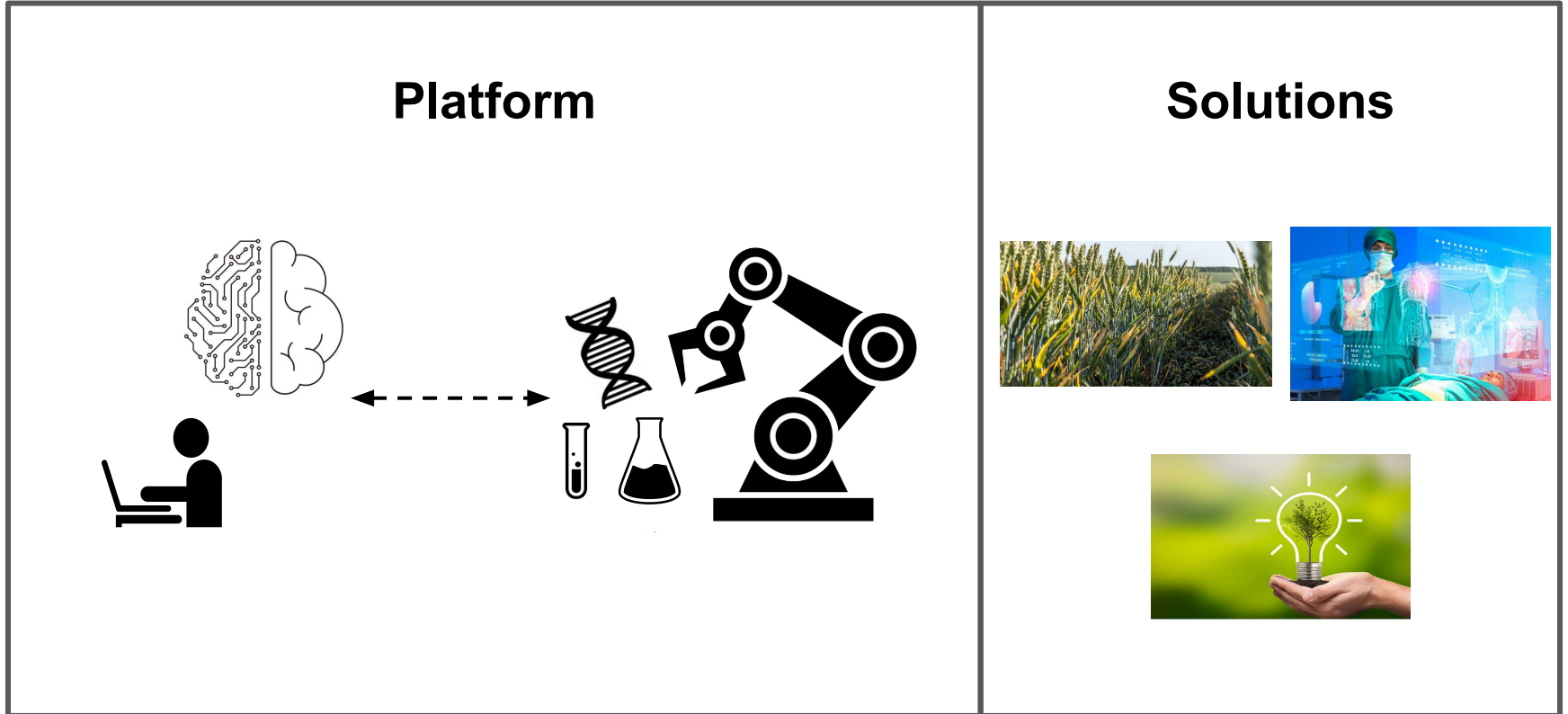
Lun Ai<sup>1</sup>□, Shishun Liang<sup>2</sup>, Stephen H. Muggleton<sup>1</sup>, Geoff S. Baldwin<sup>2</sup>

<sup>1</sup>Department of Computing, Imperial College London, UK

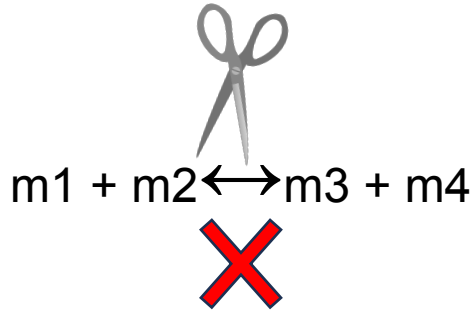
<sup>2</sup>Department of Life Science, Imperial College London, UK

□lun.ai15@imperial.ac.uk

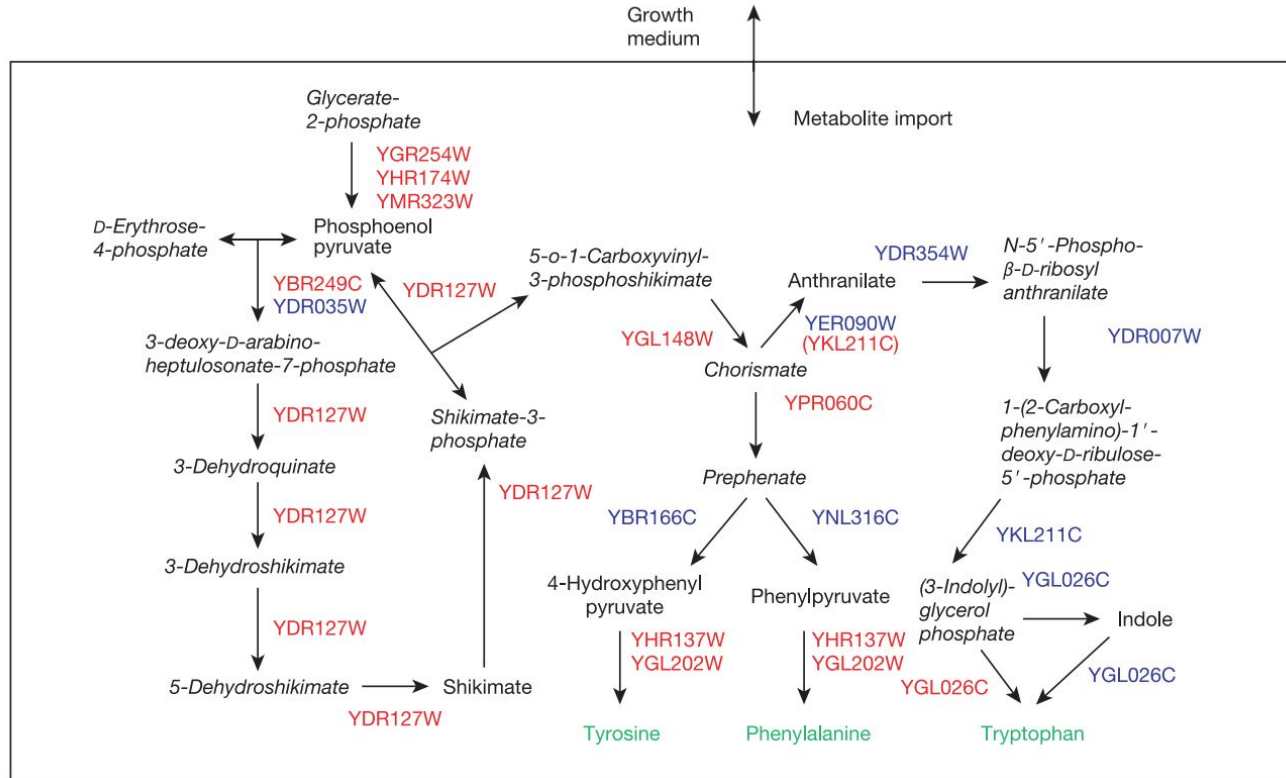
# How do we enhance Synthetic Biology platform for real-world solutions?



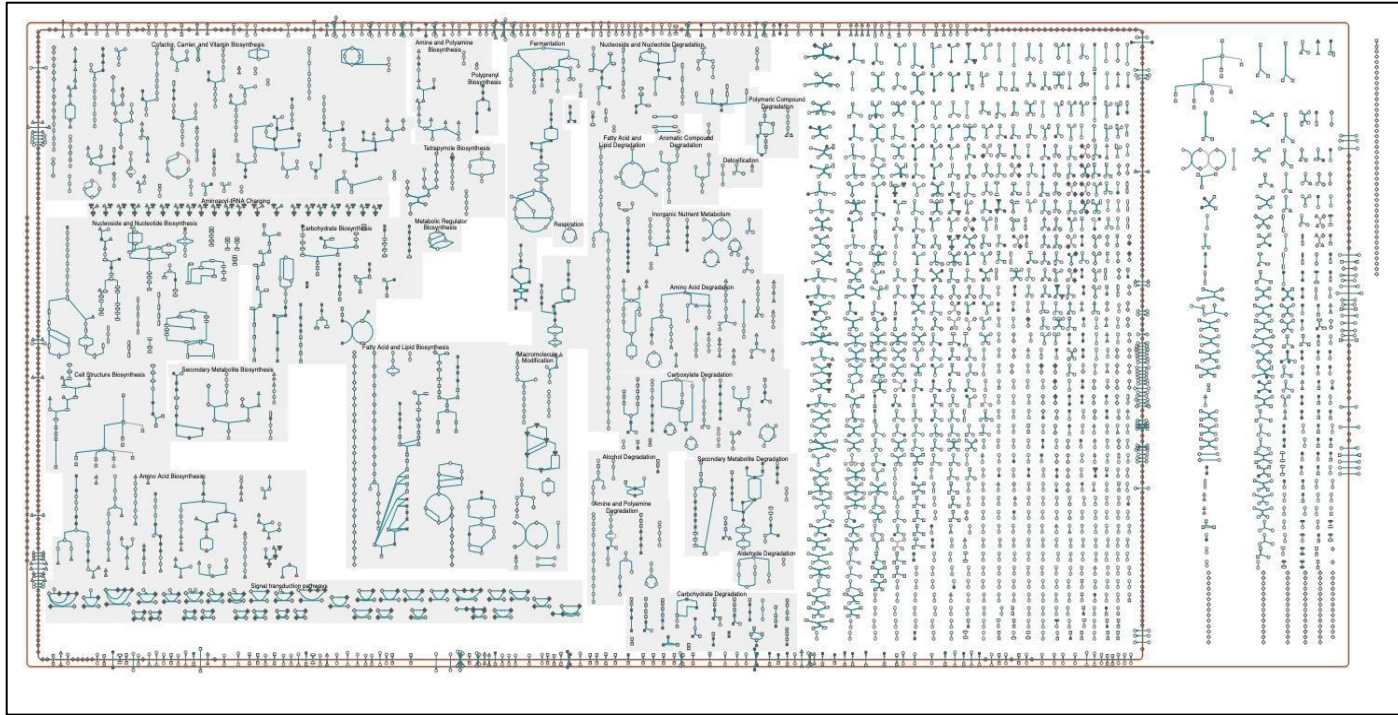
# Traditional approaches are **not scalable**



Aromatic amino acid  
metabolic  
Network (17 genes)

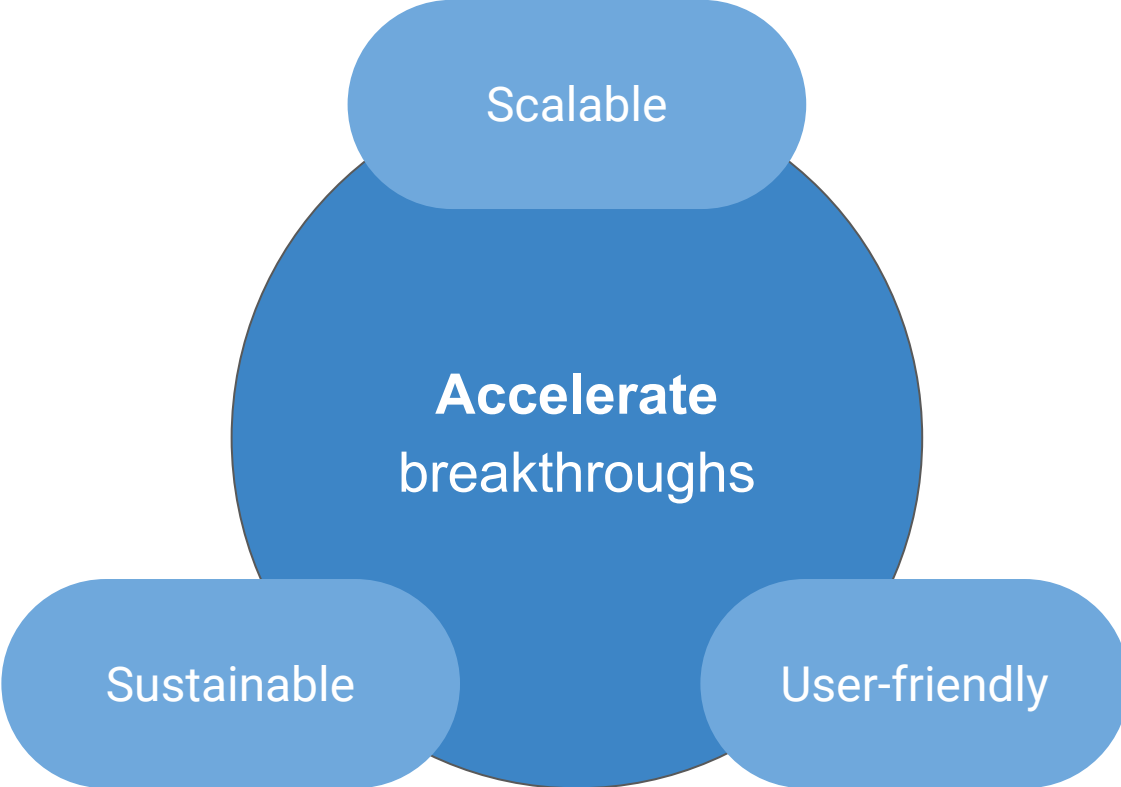


# Can we efficiently **navigate** and **learn** genome-scale metabolic networks?

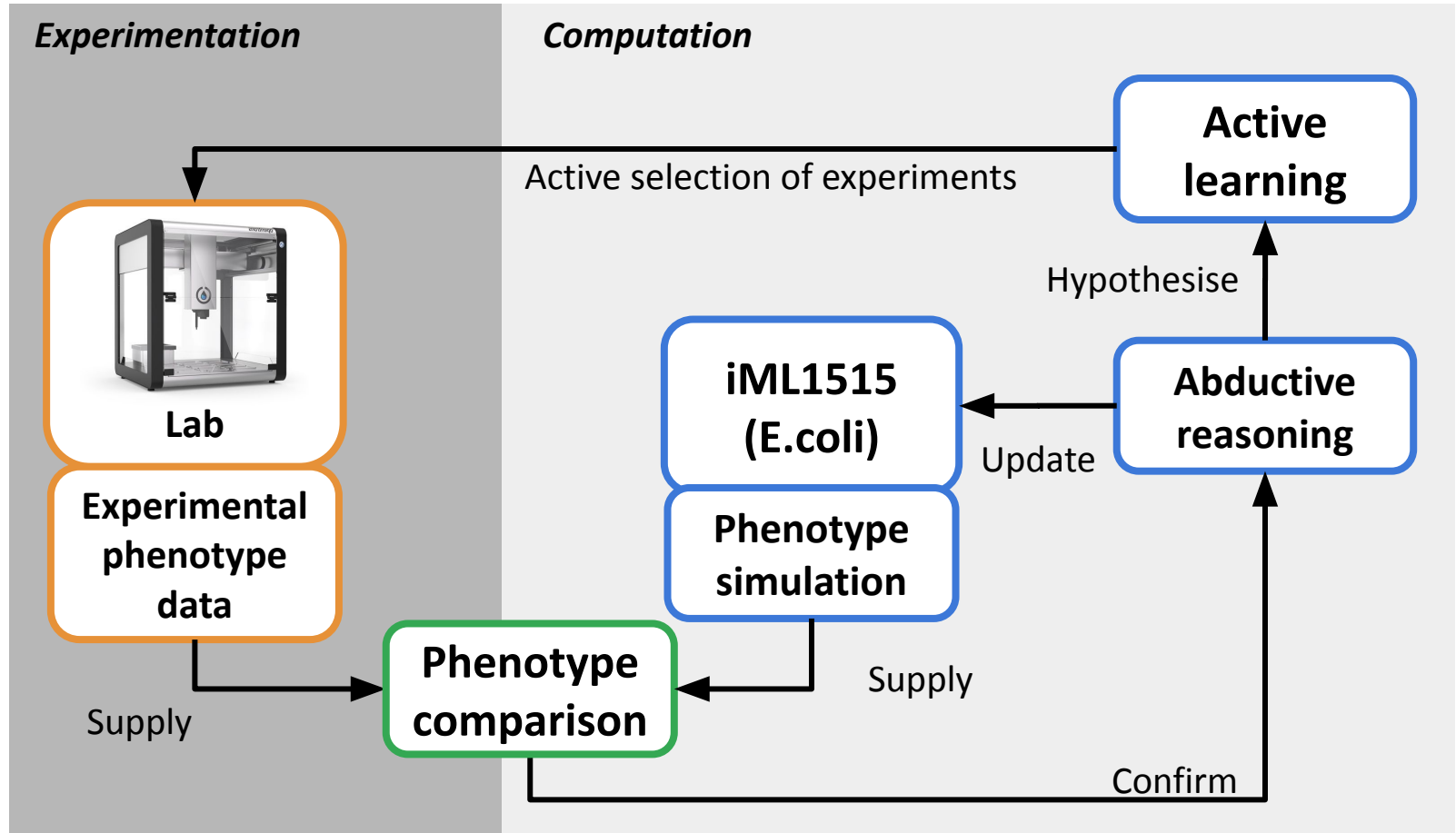


iML1515 (Monk et al. 2017), **1515** genes + **2719** reactions **with room to improve**

# Foundation for **next generation** Synthetic Biology

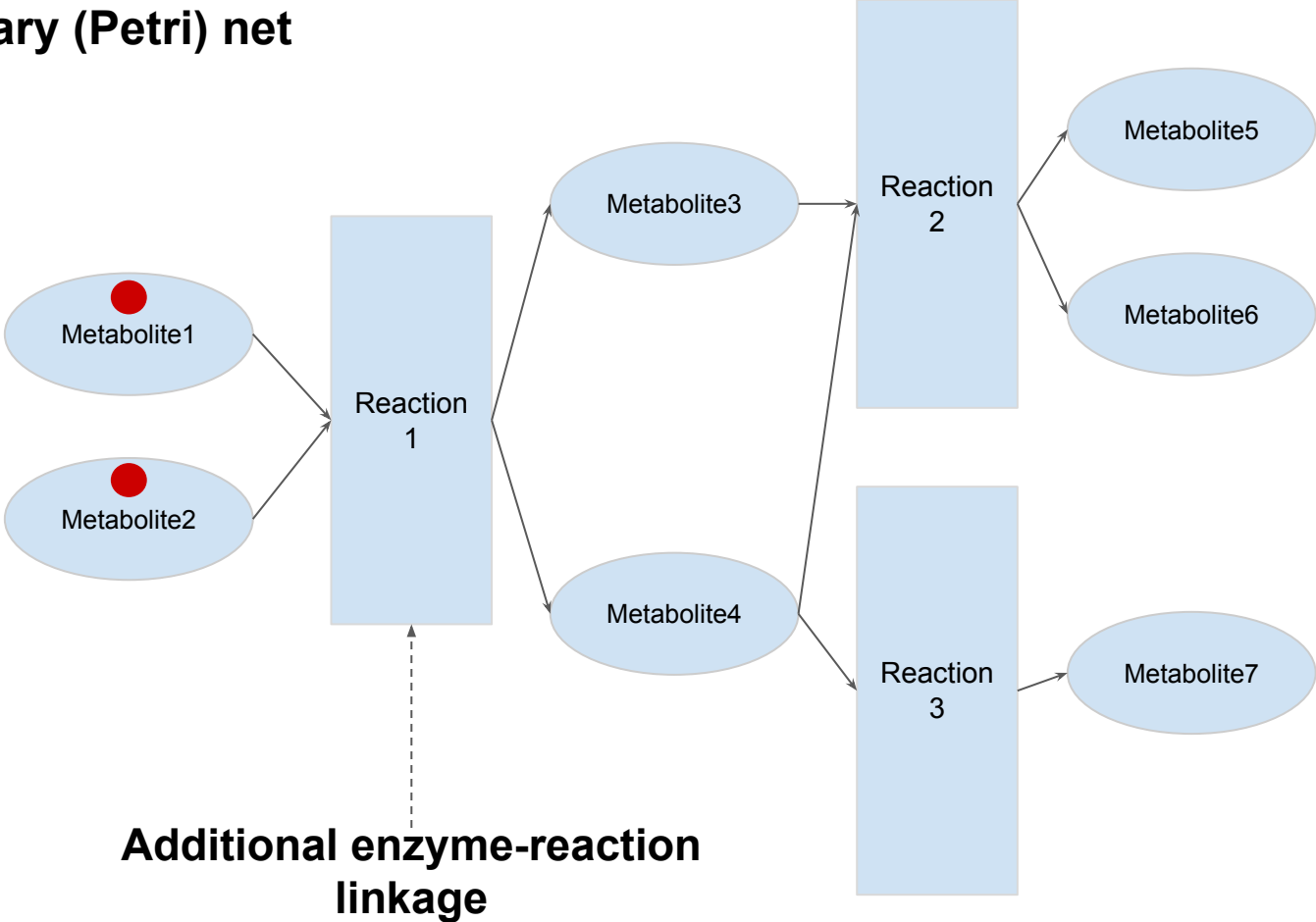


# A Symbolic AI platform for **discovery** (ILP-iML1515)



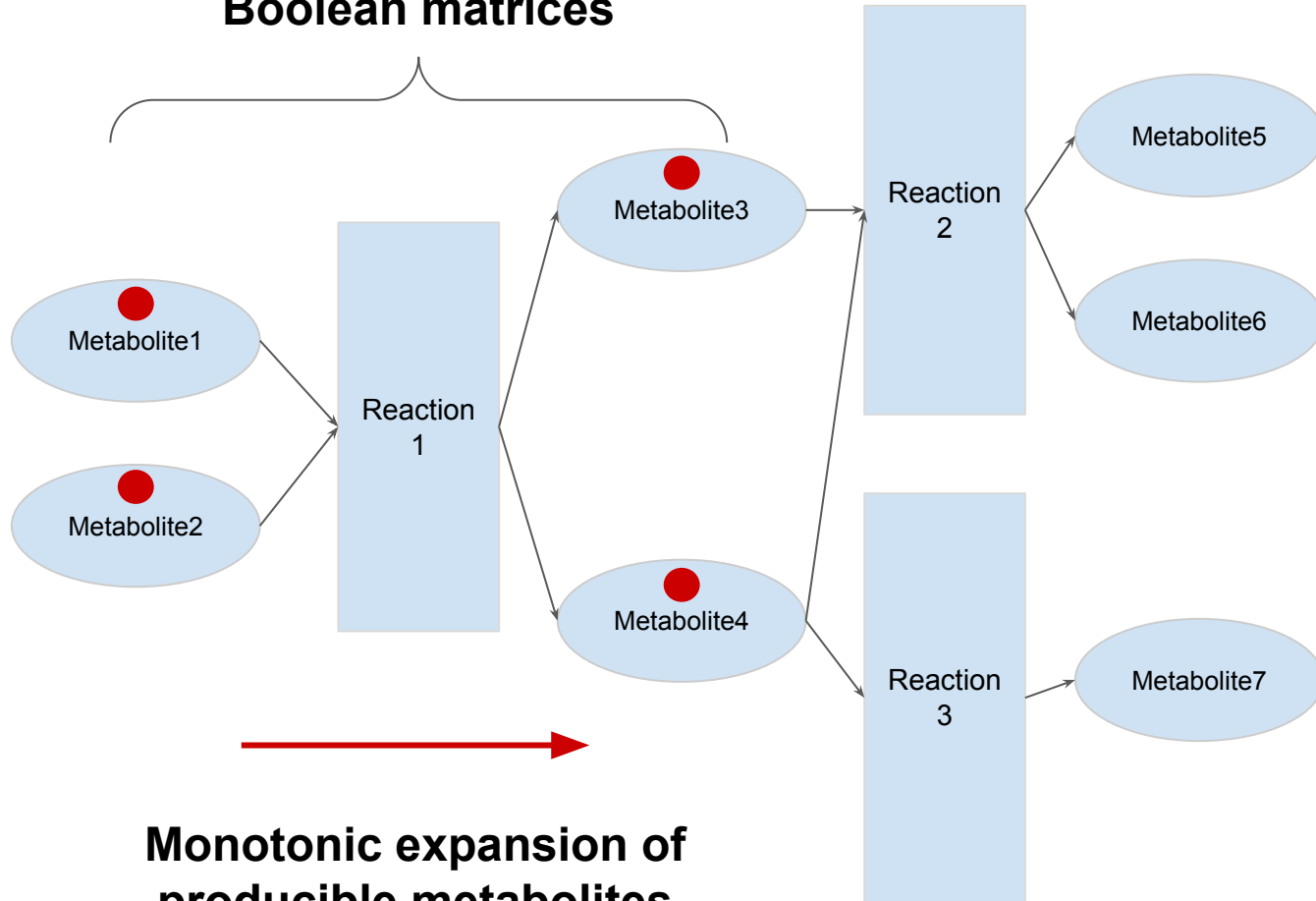
- Logical reasoning engine
- Active selection of experiments
- Recovering gene functions
- Remarks

# Elementary (Petri) net

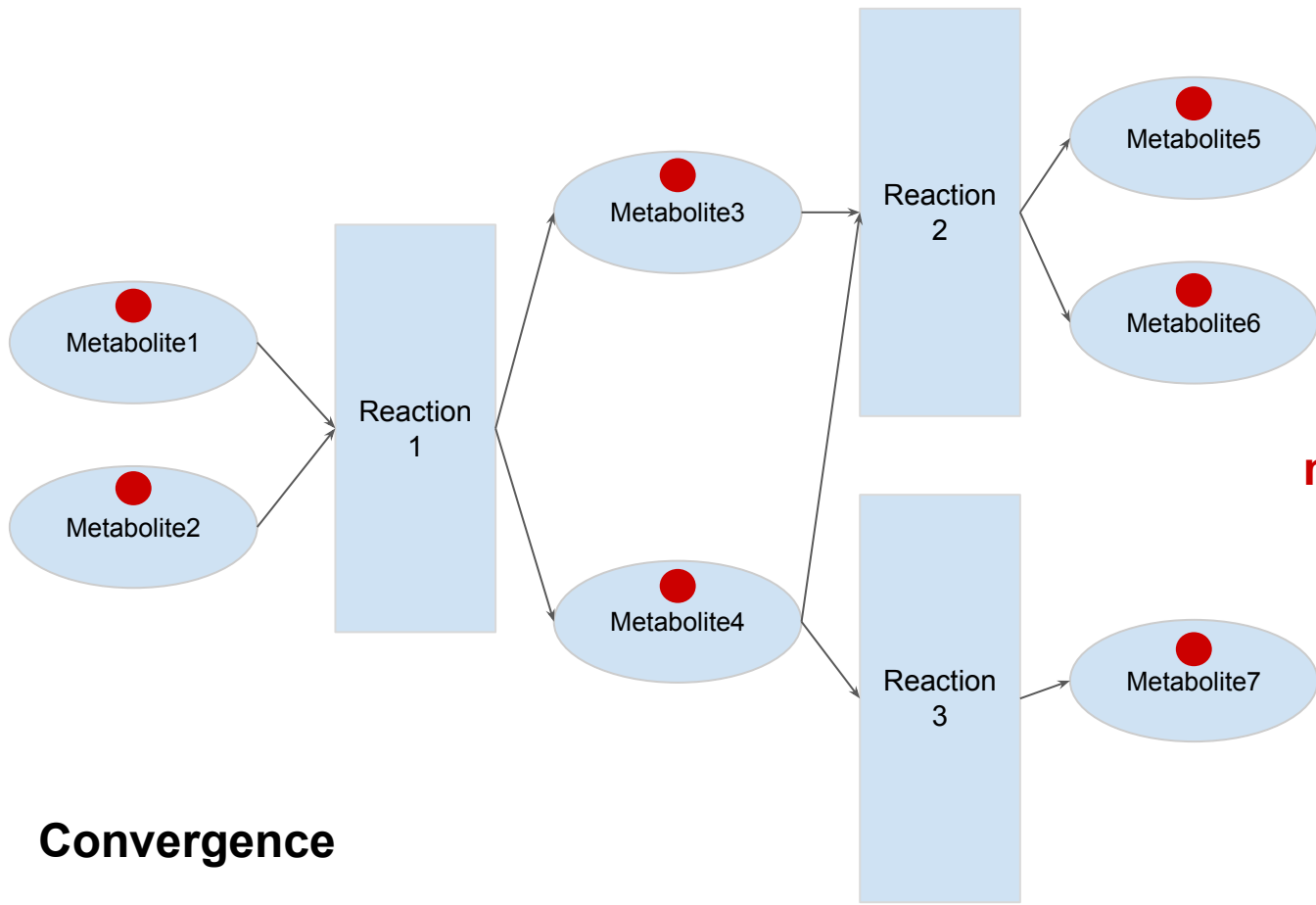




# Boolean matrices



**Source**

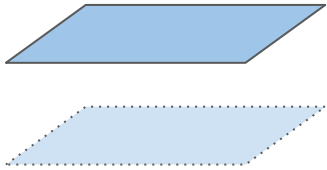


**Convergence**

**Essential  
metabolites  
reachable**

# Use Inductive Logic Programming (ILP) for navigation and learning

*Logic programs*



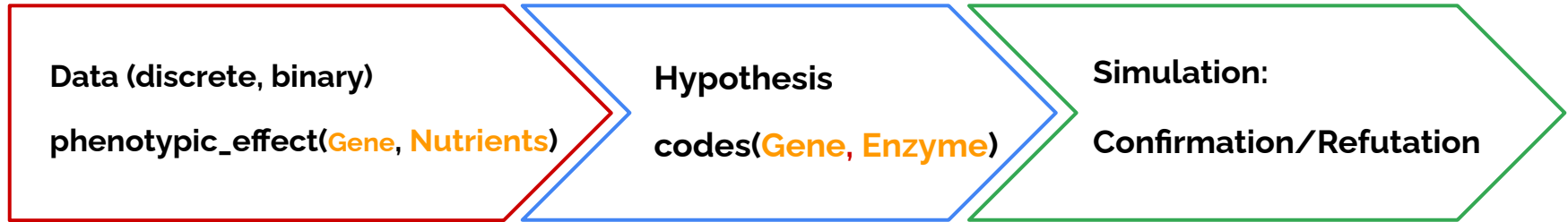
```
% static knowledge
codes(gene_b, e_b).
codes(gene_c, e_c).
codes(gene_e, e_e).
...
metabolic_step(m1, m3).
metabolic_step(m1, m4).
...
enzyme(e_a, m1, m3).
enzyme(e_b, m1, m4).
essential_molecule(m7).
essential_molecule(m8).
```

```
% description of effect using metabolic network
phenotypic_effect(Gene, Medium):-
    % abduced fact
    codes(Gene, Enzyme),
    cant_use_enzyme(Enzyme),
    %metabolic pathways
    metabolic_pathway(Medium, Metabolites),
    no_essential_molecule(Metabolites).
```

**Human-comprehensible**

# Abduction

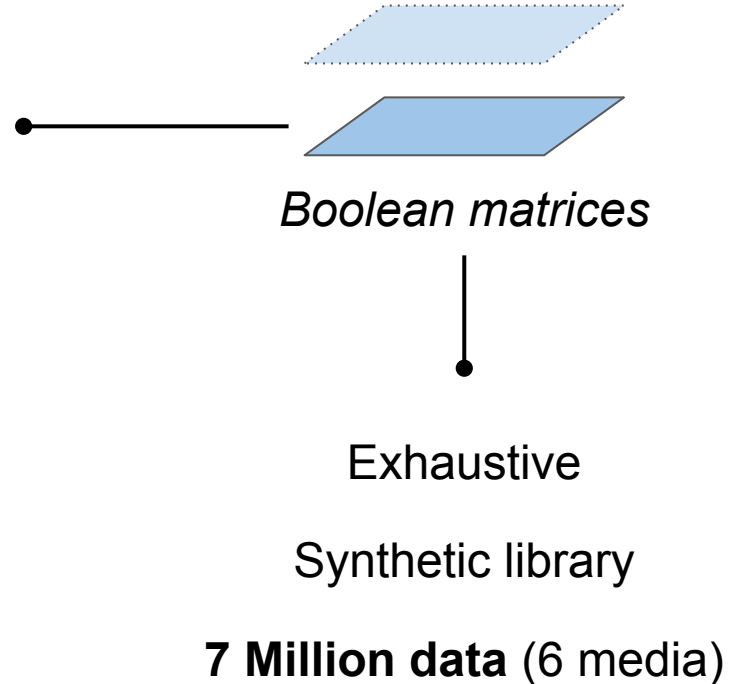
*“Logical inference that seeks the best conclusion that explains observations”*



- **Phenotypes:** gene knockouts in various nutrient media
- **Hypothesis:** of gene functions to explain data
- **Simulation:** hypothetical phenotypic observations

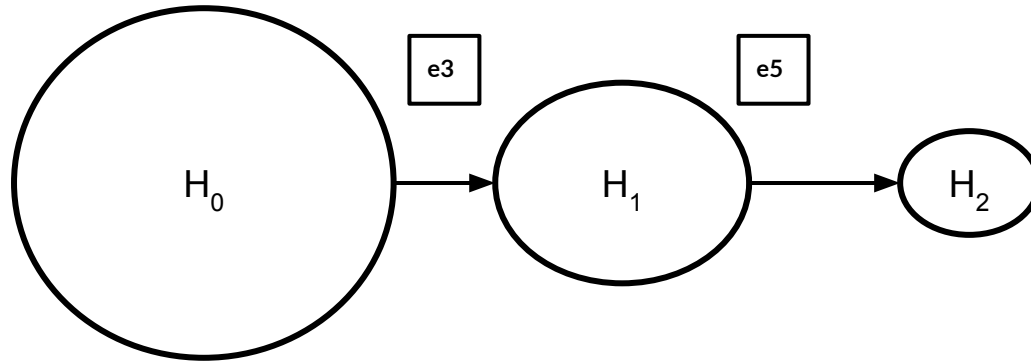
# Accelerated simulation (> 4000 times speed up)

Time per simulation	Without Binary Matrices	With Binary Matrices
Without multi-threads	<b><u>≈250s</u></b>	≈0.6s
With multi-threads 20 cpus	≈27s	<b><u>≈0.06s</u></b>
HTCondor HPC 10 nodes	-	<b><u>≈0.016s</u></b>



- Logical reasoning engine
- Active selection of experiments
- Recovering gene functions
- Remarks

# Active learning



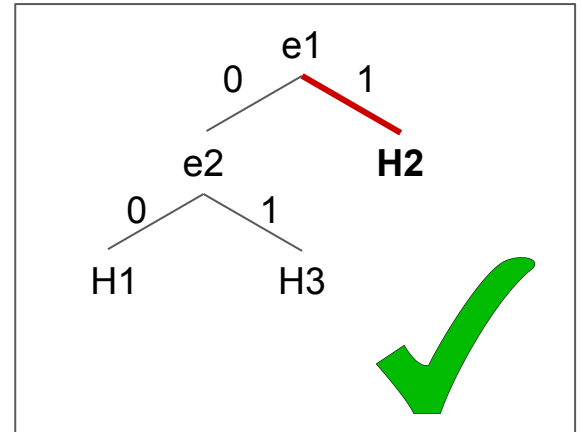
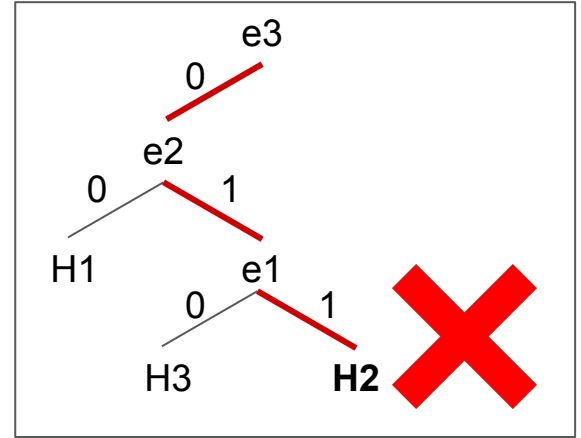
**Binary discrimination** of the candidate hypotheses

Experiment selection = Playing “**Yes/No**” game

# Experiment sample and hypothetical outcomes

<i>Phenotypic effects when we assume <math>H_i</math> is true</i>	e1	e2	e3
H1	0	0	0
H2 (target hypothesis)	1	1	0
H3	0	1	0

**Logarithmic reduction** of candidate hypotheses  
by identifying a (near-)optimal **binary tree**





- Logical reasoning engine
- Active selection of experiments
- Recovering gene functions
- Remarks



**Synthetic  
Data**

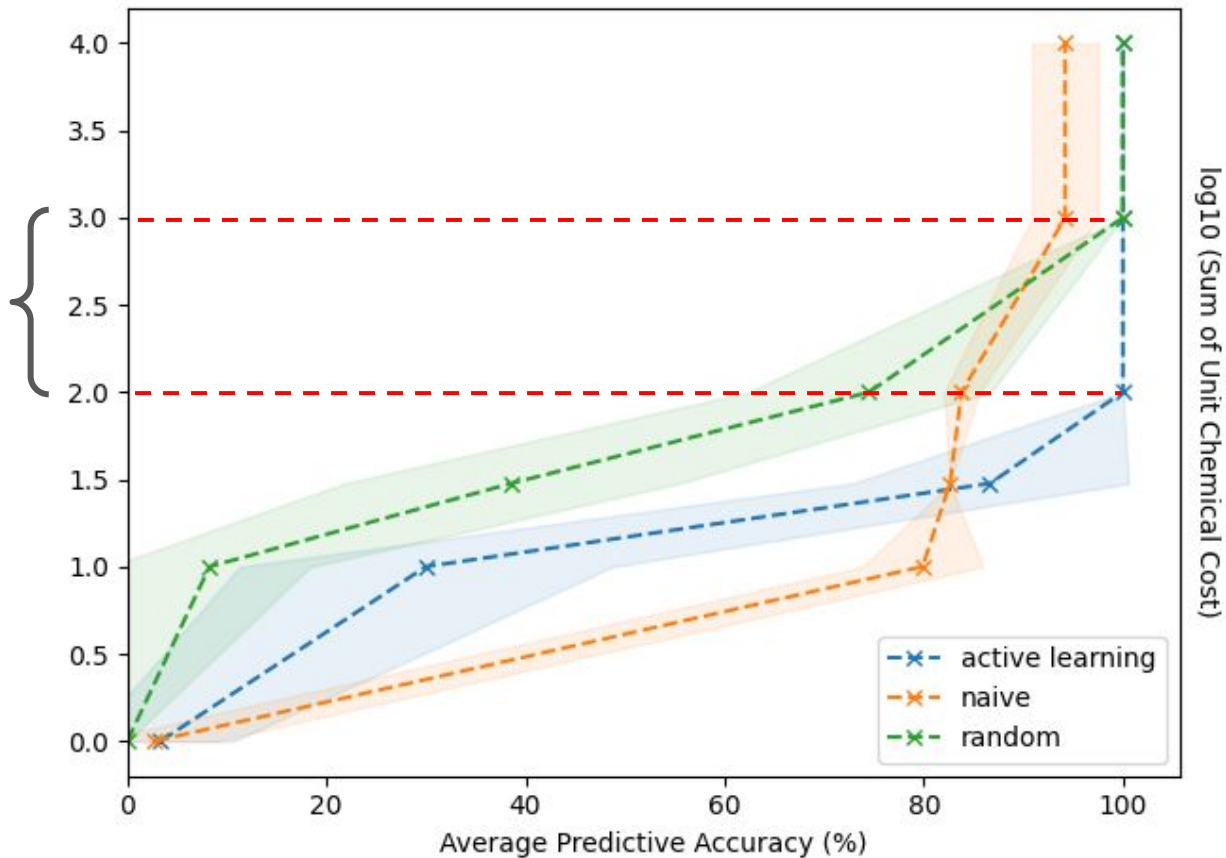
**Actively learn  
comprehensible  
gene functions**



**Compare with  
original model**

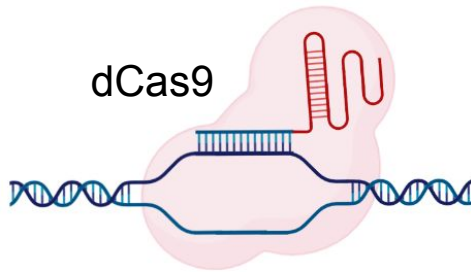
# Active learning known gene functions: **lower cost**

**10 times saving in cost**



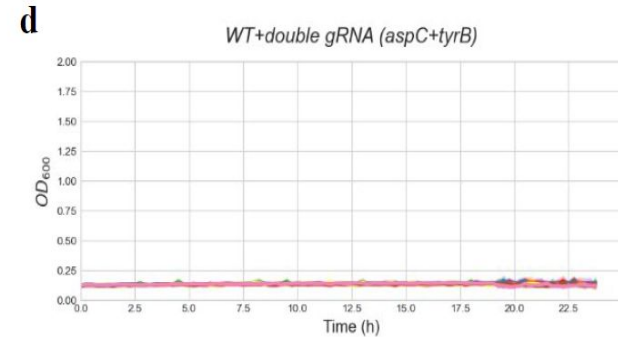
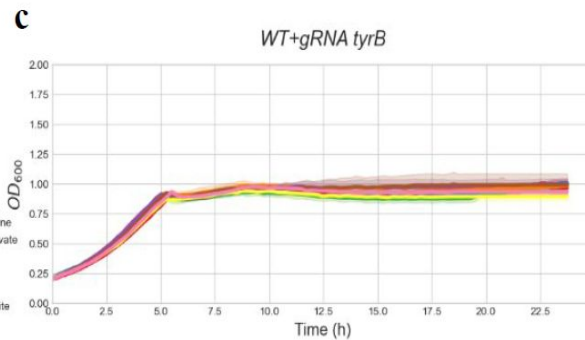
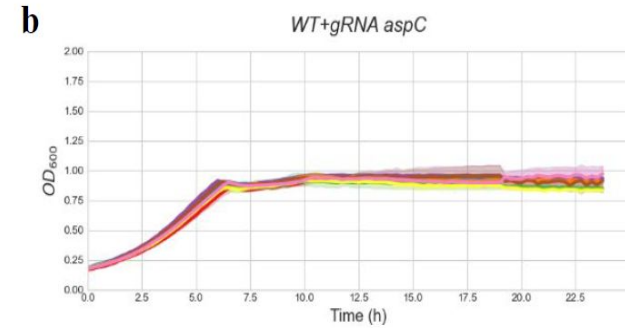
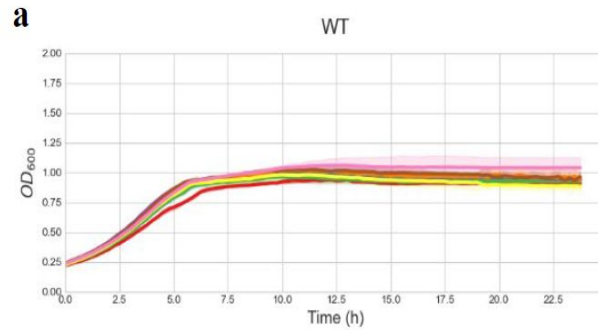
# Isoenzyme identification

## CRISPRi



gRNA targets dCas9 to silence target genes in *E. coli* genome

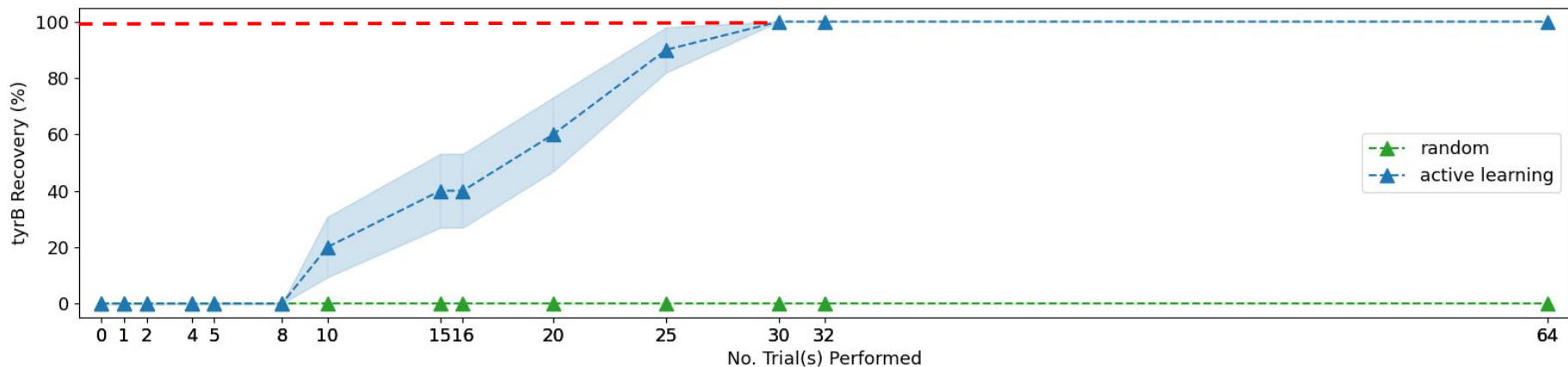
## Synergistic deactivation of TyrB & AspC



# Isoenzyme recovery

- Remove isoenzyme association of tyrB with aspC
- Hypotheses (1089)
  - 33 distinct enzyme functions and **a library of 33 genes**
- Training dataset
  - Double-knockout (synthetic data, 3696)
  - Single-knockout (online experimental data, 231)

# Active learning isoenzyme (tyrB): fewer data usage



- Logical reasoning engine
- Active selection of experiments
- Recovering gene functions
- Remarks

# Future work

- Generalisation of ILP-iML1515
  - Quantification of metabolite concentration
  - Introduction of regulatory interactions
- Knowledge transfer
  - Helping people comprehend biological concepts  
[Muggleton et al., 2018]



# Enhance automation of discovery

- Overcome network **complexity**
  - > 4000+ times better runtime
- **Active** selection of experiments (ASE)
  - 10-fold saving in cost
  - Use fewer training data
- Can target **errors** in SOTA metabolic networks
  - Remain efficient for multi-knockouts

What people think I do ...



This is what I actually do ...



This is what I actually do ...



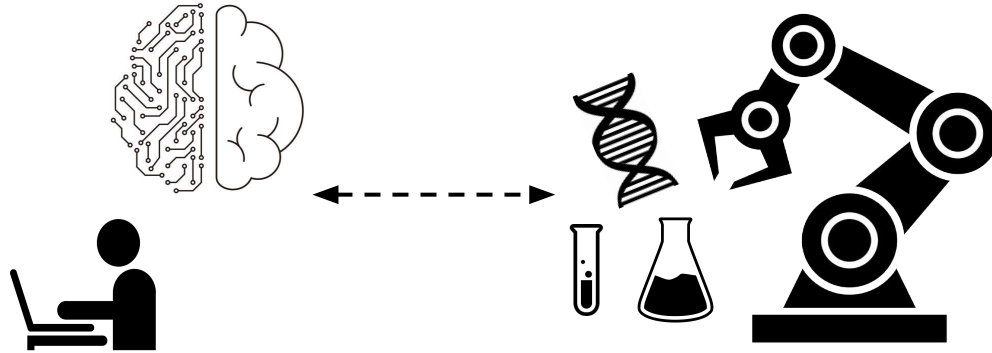
What are biologists **actually** doing?

Pathway analysis

Experimentations

Understanding AI models

Our vision is to bridge ...



**Human-Comprehensible AI and Synthetic Biology**

## Lun Ai

Email: [lun.ai15@imperial.ac.uk](mailto:lun.ai15@imperial.ac.uk)

Website: <https://lai1997.github.io/>

Linkedin: <https://www.linkedin.com/in/lun-ai-46481a128/>