

# Human Comprehensible Active Learning of Genome-Scale Metabolic Networks

Lun Ai<sup>1</sup>, Shishun Liang<sup>2</sup>, Wang-Zhou Dai<sup>3</sup>, Liam Hallett<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 <sup>3</sup>School of Intelligence Science and Technology, Nanjing University, China

<sup>□</sup>lun.ai15@imperial.ac.uk

## Outline

- Scientific problem
- Framework
- Novel matrix approach
- Results
- Summary

## Importance to Synthetic Biology

- Aim:
  - Synthesise <u>useful compounds</u>
- Method:
  - <u>Tuning</u> for correct precursors
  - Engineering of <u>metabolic networks</u>
  - Examination of <u>phenotypes</u>

## Exemplary metabolic network

m1 + m2↔m3 + m4



King et al. 2004

### Efficiently learn and navigate genome-scale metabolic networks?



iML1515 (Monk et al. 2017), 100 times increase in model complexity

### Reduce cost and design space?





<u>**1515**</u> genes + <u>**2719**</u> reactions

- Scientific problem
- Framework
  - Model-Comprehend
- Novel matrix approach
- Results
- Summary

### Model-Comprehend

### • Automate Design, Build, Test, Learn

- Rapid inferences
- Hypothesis space reduction
- Cost minimisation





### Model



- Scientific problem
- Framework

### Novel matrix approach

- 1. <u>Abduction</u>
- 2. Active learning
- Results
- Summary

## Abduction

### $p \leftarrow q \land r$

- Phenotypes in various nutrient media
- Simulation based on metabolite saturation
- Hypothesise <u>gene functions</u> to explain data

**Experiment data**:

Hypothesis:

phenotypic\_effect(Gene, Nutrients).

codes(Gene, Enzyme).

#### Enzymes/reactions



#### Metabolite saturation

% description of effect using metabolic network phenotypic\_effect(Gene, Medium):-% abduced fact codes(Gene, Enzyme), cant\_use\_enzyme(Enzyme), %metabolic pathways metabolic\_path(Medium, Metabolites), no\_essential\_molecule(Metabolites).



## Active learning



- a) Binary discrimination of hypotheses
- b) Approx. minimum cost binary decision tree



- Scientific problem
- Framework
- Novel matrix approach

### • <u>Results</u>

- Runtime improvement
- Cost reduction
- Summary

### > 4000 times better in runtime

Time per simulation	Robot Scientist (King et al. 2004)	MC
Without multi-threads	<u>≈250s</u>	≈0.6s
With multi-threads 20 cpus	≈27s	<u>≈0.06s</u>



10 times saving in cost

- Scientific problem
- Framework
- Novel matrix approach
- Results
- <u>Summary</u>

## Summary

- To automate metabolic network engineering
- Overcome network complexity
  - 4000+ times better runtime
- Experimental design
  - 10 times lower cost
  - Most informative trials

### Future work

- Generalisation of framework
  - Quantification
  - Multi-clause theories
- Optimisation of metabolic network
  - Validate hypotheses (CRISPRi)
  - Multiple gene loci
- Hypothesis comprehensibility (Ai et al. 2021)

Lun Ai

Email: <u>lun.ai15@imperial.ac.uk</u>

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

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

### MODEL



### Logical knowledge base



% description of effect using metabolic network phenotypic\_effect(Gene, Medium):-% abduced fact codes(Gene, Enzyme), cant\_use\_enzyme(Enzyme), %metabolic pathways metabolic\_path(Medium, Metabolites), no essential molecule(Metabolites).

% static knowledge codes(gene\_b, e\_b). codes(gene\_c, e\_c). codes(gene\_e, e\_e). metabolic\_step(n1, m3). metabolic\_step(n1, m4). enzyme(e\_a, n1, m3). enzyme(e\_b, n1, m4). essential molecule(m7). essential molecule(m8).

Four metabolites:	m1, m2, m3, m4
Two orfs:	g1, g2
Initial metabolic state:	m1, m2,m3
Representation:	[1, 1, 1, 0]
Logic encoding:	mstate(0, 14)



#### KO g1



#### KOg1+g2

