Active Abductive Learning of Multigenic Interactions from High-Throughput Repression Screening Data

Lun Ai[1], **Shi-Shun Liang**[1], **Stephen H. Muggleton**[2], **Geoff S. Baldwin**[1]
[1]Department of Life Sciences, Imperial College London, UK
[2]Department of Computing, Imperial College London, UK
{lun.ai15 ✉,shishun.liang20,s.muggleton, g.baldwin}@imperial.ac.uk

Abstract

Our objective is to automate genome-scale function discovery. A key application of Synthetic Biology is to genetically engineer reliable microbial systems for maximising the production of valuable compounds. However, even in simple systems, genes frequently interact with other genes, making it difficult to interpret their effects comprehensively. In the quest for more robust Synthetic Biology, it is crucial to accelerate our understanding of gene functions across the entire genome systematically.

To this goal, we will present a novel workflow (Figure 1) of our abductive learning system $BMLP_{active}$ [1] and a new pair-wise repression library screening procedure. $BMLP_{active}$ leverages boolean matrices and is the first symbolic AI system applied to genome-scale metabolic network models (GEMs) such as iML1515 [3]. Notably, this GEM has 1515 genes and 2719 reactions, significantly larger than the model in the Robot Scientist [2]. $BMLP_{active}$ can actively learn single-gene and digenic functions with only 20 synthetic gene-knockout phenotype data [1]. In comparison, sub-symbolic methods rely on opaque models and require extensive training data. Our workflow is a realistic approach for creating a self-driving lab to reliably engineer biological systems.

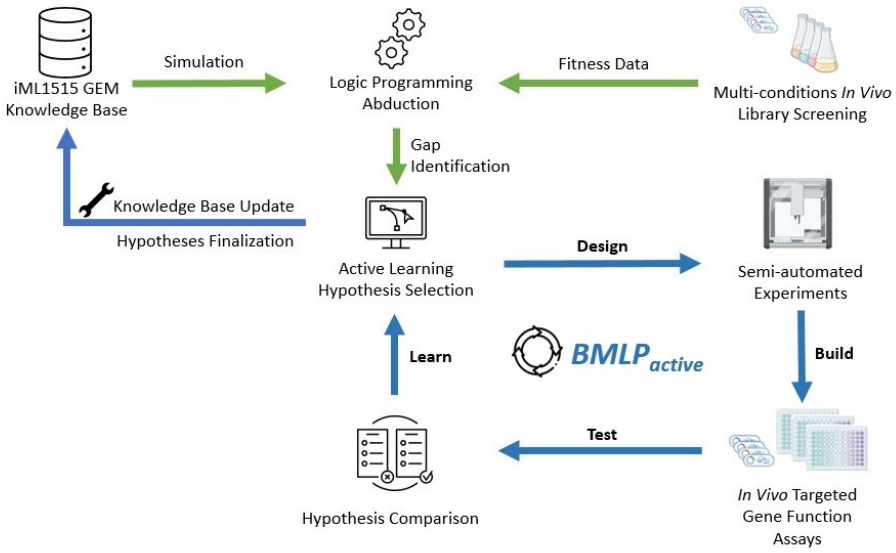

Figure 1: Implementation of Design, Build, Test and Learn using actively abductive learning, targeted multi-gene repressions and pair-wise library screening. Gene repression library screening [4] supplies a diverse pool of genotypes for generating multigenic interaction hypotheses. $BMLP_{active}$ efficiently navigates exponentially growing spaces of hypotheses and experimental designs by actively proposing multi-gene repression trials.

References

[1] L. Ai, S. H. Muggleton, S.-S. Liang, and G. S. Baldwin. Boolean matrix logic programming for active learning of gene functions in genome-scale metabolic network models. *arXiv*, 2024.

[2] R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare. The Automation of Science. *Science*, 324(5923):85–89, 2009.

[3] J. M. Monk, C. J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, A. M. Feist, and B. O. Palsson. iML1515, a knowledgebase that computes escherichia coli traits. *Nature Biotechnology*, 35(10):904–908, 2017.

[4] T. Wang, C. Guan, J. Guo, B. Liu, Y. Wu, Z. Xie, C. Zhang, and X.-H. Xing. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nature Communications*, 9(1):2475, 2018.