

TorchCell Summary

TorchCell Project Summary

TorchCell is a multi-modal machine learning framework for predicting cellular phenotypes in single-cell organisms used in biotechnology applications, with explicit metabolic modeling designed to solve metabolic engineering strain design problems. Initial development focuses on model organism *S. cerevisiae*. The project addresses a fundamental challenge in systems biology and metabolic engineering: the fragmentation of phenotypic data across diverse experimental datasets, where individual studies are typically too small to support deep learning approaches. The solution to this problem is a comprehensive graph database that standardizes and aggregates heterogeneous data sources. TorchCell makes deep learning applications more feasible, and helps enable multimodal phenotype prediction under multiplexed genetic perturbations.

The TorchCell *S. cerevisiae* virtual-cell-model (VCM) helps map cellular graphs, environmental conditions, and genetic perturbations to help predict phenotypic outcomes like fitness, gene interactions, morphology, and expression. While traditional tabular machine learning models are sufficient for some phenotype prediction, like fitness, they fail to capture complex gene interactions, motivating development of deep learning models that model biological network topology. The TorchCell *S. cerevisiae* VCM employs a graph-regularized equivariant transformer that encodes genetic perturbations through cross-attention mechanisms, allowing the model to learn how perturbations interact with cellular networks. Recent results have demonstrated state-of-the-art performance in predicting trigenic deletion interactions in yeast, showing the model can identify high-order functional complexes in yeast. Future work will focus more on predicting metabolically relevant phenotypes for industrially applications.