Read the paper here

BioBO:

Biology-informed Bayesian Optimization for Perturbation Design

Yanke Li†, Tianyu Cui†, Tommaso Mansi, Mangal Prakash*, Rui Liao*

† *Equal Contribution as first authors * Equal Contribution as last authors*

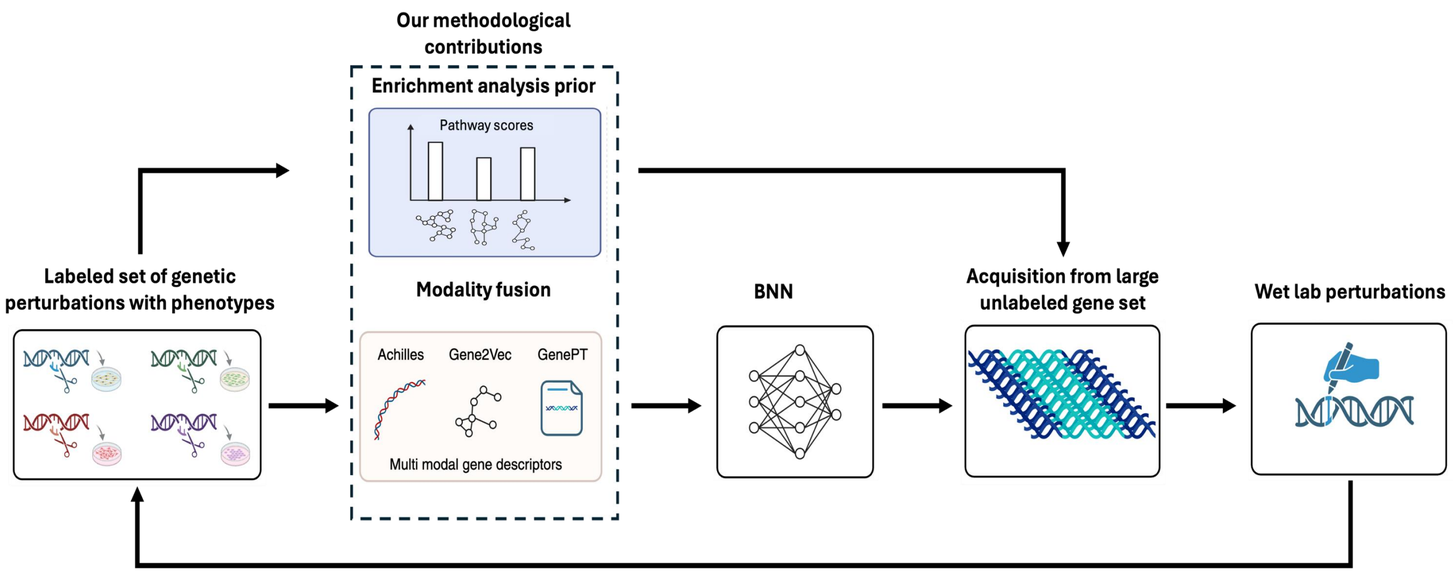
Motivation

- CRISPR screens enable gene perturbation studies, but exhaustive testing is infeasible due to large number of perturbations.
- Bayesian Optimization (BO) helps to select informative genes via a probabilistic surrogate and acquisition.
- **Limitation of existing works**: Current BO strategies do not generally leverage prior biological knowledge and multimodal gene representations.

Our contributions (BioBO)

- Using **rich multimodal gene embeddings** (Achilles, Gene2Vec, GenePT) for **better surrogate modelling.**
- **Biology inspired acquisition function**: novel Acquisition Functions (AFs) incorporating pathway enrichment analysis as a biological prior. Theoretically sound framework with no-harm guarantee under π-BO (Hvarfner et al., 2022).
- **BioBO improves** over conventional BO in **labelling efficiency by 25–40%** with **interpretable biologically coherent pathways.**

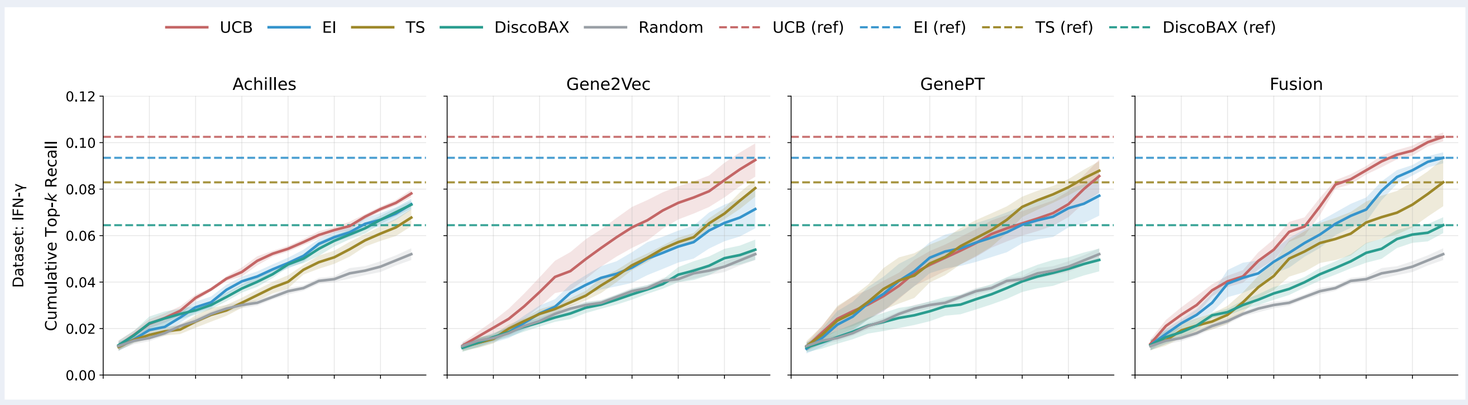
BioBO pipeline



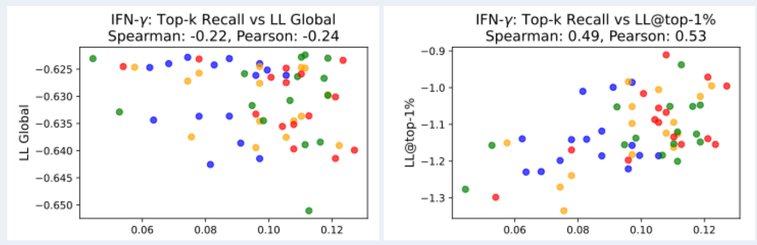
Method

- Surrogate: Two-layer Bayesian Neural Networks (BNN) over fused/single embeddings
- EA prior: Combine odds ratio & p-value into a pathway score; map to gene-wise probabilities
- AFs: Baselines: UCB / EI / TS / DiscoBax; Ours: BioUCB/ BioEI / BioTS (multiply UCB / EI / TS by EA-derived prior $\pi(x)$ with no-harm guarantee under π-BO (Hvarfner et al., 2022))
- Datasets from GeneDisco and DiscoBax benchmark (Lyle et al., 2023)

Multimodal BO improves perturbation design over any single modality by 4%-40%

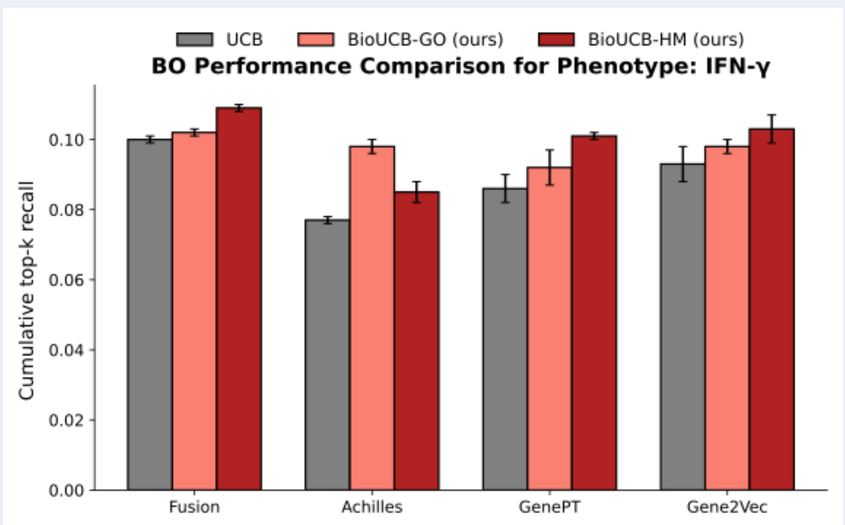


Multimodal BO improves design by improving the surrogate near optimum

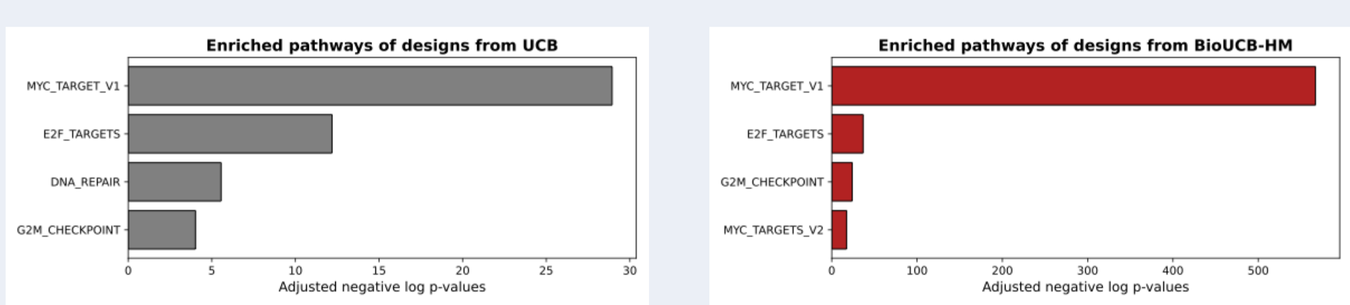


BO design accuracy (top-k recall) correlates more with the surrogate model's log-likelihood near the optimum (top-1%, right) than across the whole space (global, left).

Enrichment analysis priors with different pathway databases in BioBO improve perturbation design consistently



BioBO produces stronger enrichment signals in pathways closely tied to IFN-γ regulation in T cells than BO



Designs from BioUCB-HM (AF with enrichment based prior) are more significantly enriched (red) in pathways closely tied to IFN-γ regulation in T cells compared with vanilla UCB (grey).

1. Hvarfner, Carl, et al. "πBO: Augmenting Acquisition Functions with User Beliefs for Bayesian Optimization" ICLR 2022.
2. Mehrjou, Arash et al. "GeneDisco: A Benchmark for Experimental Design in Drug Discovery" ICLR 2022.
3. Lyle Clare, et al. "Discobax: Discovery of optimal intervention sets in genomic experiment design." ICML 2023.