Efficient prediction of functional gene modules in *Pseudomonas putida* with machine learning

Background

- A deep understanding of gene function in *P. putida*, especially for metabolism and stress tolerance, is essential for its deployment as an industrial biocatalyst.
- Transposon sequencing (RB-TnSeq) data can illuminate gene function, but these experiments often focus on single conditions, making identification of functionally related gene groups difficult.

Approach

- Existing public gene fitness values on 4732 *P. putida* genes in RB-TnSeq libraries tested under 332 experimental conditions were gathered.
- The dataset underwent matrix decomposition by independent component analysis (ICA), an unsupervised, multivariate signal separation algorithm.

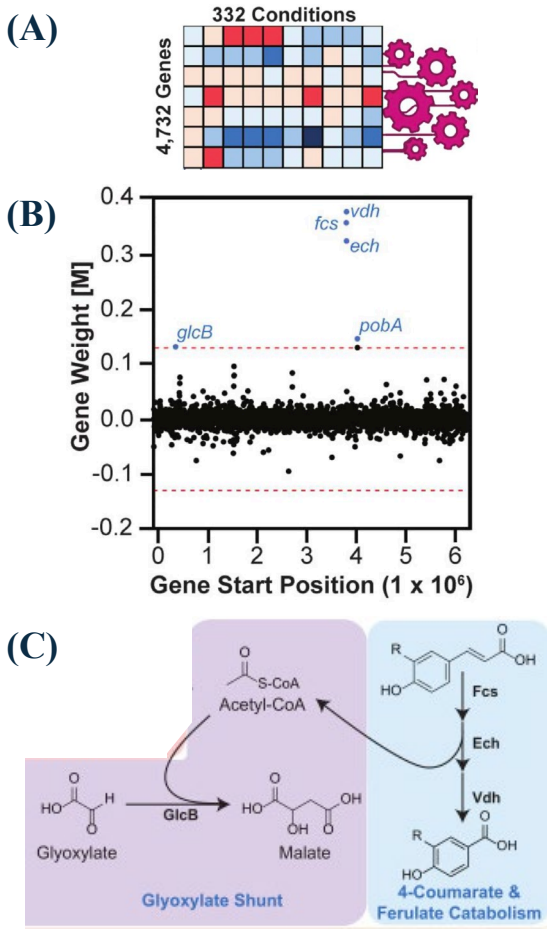
Results

- ICA identified 84 groups of genes unified by shared functional influence upon specific cellular processes, hereafter termed **fModules** (functional modules).
- Selected engineered mutants of *P. putida* validated gene members from fModules for hydroxycinnamate catabolism, acetyl-CoA assimilation, and nitrogen metabolism.
- Comparing functional gene clusters from ICA of RB-TnSeq data to the regulatory gene clusters from ICA of RNAseq data identified relationships between gene regulation and function.

Significance

- Obtaining functional modules through ICA can drastically reduce the amount of time required to annotate gene function, thereby accelerating efforts in metabolic engineering and metabolic model reconstruction.
- Comparison of data obtained from independent component analysis of transcriptomics and gene fitness datasets can elucidate regulatory-functional relationships between genes

Borchert, A.J., *et al*. *mSystems* (2024) 9:e00942-23. doi.org/10.1128/msystems.00942-23



Identification of an fModule for hydroxycinnamate metabolism. (A) ICA matrix decomposition of RB-TnSeq fitness data. **(B)** Genes with weights above an outlier threshold were grouped into a single fModule. **(C)** Each gene in one identified fModule plays a role (directly or indirectly) in hydroxycinnamate metabolism.