|  |  |
| --- | --- |
| ­­Agreement #: 25416  FY18  WBS #: 2.5.3.104 | Completion Date: September 30, 2021  Scheduled Completion: September 30, 2021  Technology Area: Biochemical Conversion |
| Project Title: | Agile BioFoundry |
| Principal Investigator: | Nathan Hillson (LBNL) |
| Milestone Title: | Demonstrate a cross-validated 20% improvement in predictive power (e.g., improved accuracy/precision/correlation between predicted and observed) for two or more ABF Learn methodologies as a function of data volume, velocity, and/or veracity, for multiple vs. single data modalities. |
| Authors: | Jeremy Zucker, Peter St. John |
| Participating Researchers: | Jeremy Zucker, Peter St. John, Shant Mahserejian, Joseph Cottam |
| Key Words: | Bayesian metabolic control analysis; Learn algorithm; Aspergillus niger; multi-omics experiment; DBTL |
| Reviewed By: |  |

Milestone Completion Report



**Executive Summary**

**Abstract:** To improve titers, rates and yields for 3-hydroxypropionate production in ABF host organism *Aspergillus niger,* we employed Bayesian metabolic control analysis (BMCA) to multiomics data generated for 17 engineered strains. Top overexpression candidates included nitrate transporter (r1086:1189116), 5-formyltetrahydrofolate deformylase (r95:1182700), malate dehydrogenase (r44:1144118), pyruvate carboxylase (r19:1031996), and mitochondrial aspartate transaminase (r258m: 1184650). ALD6 was found to be the top repression candidate, which was expected, because it was knocked out in the reference strain. These results were compared against single modal correlation analyses between enzymes and 3HP flux.

**Milestone text:**

**Introduction**

1. **Was the milestone met or not met?**

Milestone was met

1. **How the performers did it.**

**Results**

***Bayesian metabolic control analysis***

In this application, we employed the Bayesian metabolic control analysis (BMCA) methodology previously developed under the Agile BioFoundry.1In BMCA, a low-fidelity kinetic model of microbial metabolism is constructed leveraging linear-logarithmic kinetics.2 With known kinetic parameters, a kinetic model enables the expected steady-state internal metabolite concentrations and metabolic fluxes to be estimated as a function of enzyme expression and media conditions. With measurements of both the input variables (extracellular metabolite concentrations and enzyme expression) and the output variables (steady-state fluxes and internal metabolite concentrations), posterior distributions in the kinetic parameters that are consistent with the observed data can then be estimated.

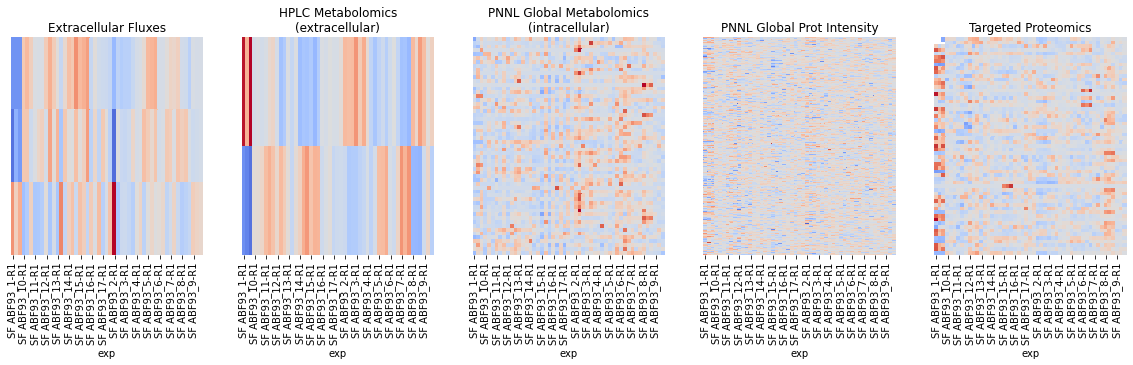
In this application, we used a reduced model of *A. niger* (172 reactions, 171 metabolites) adapted from a recently published model3 and experimental data for 17 strains under glucose media conditions producing 3-hydroxypropionate (3HP) from a recent large omics campaign to demonstrate the ability of the method to generate actionable metabolic engineering predictions (see Table 1).

Table : Engineered strains designed to produce 3HP in A. niger

Text, table

Description automatically generated

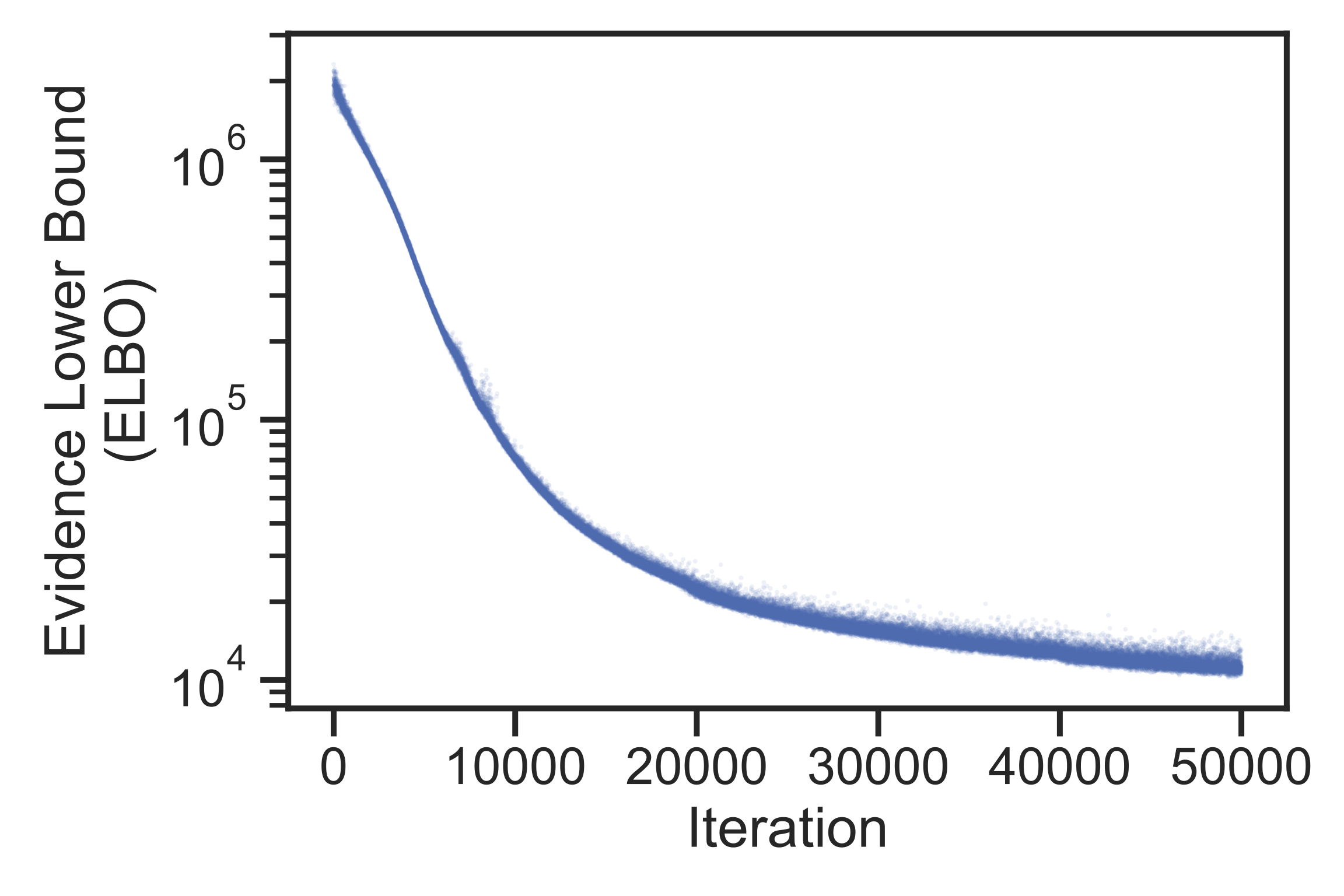
The experimental data consists of proteomics and metabolomics measurements along with quantifications of the spent media and is depicted graphically in Fig. 1. The spent media and time and OD at collection were used to construct a simple exponential growth model of the organism and estimate strain-specific update and excretion rates for key measured extracellular metabolites. Metabolomics and proteomics measurements indicate relative changes in abundance of key intermediate species and were mapped back to their appropriate compartment and identifier in the core-carbon metabolic model.



**Fig. 1. Depiction of the data collected in the multi-modal *A. niger* experimental campaign, including extracellular fluxes, metabolomics and proteomics.** Rows represent measured quantities, while columns indicate different experimental conditions. Red values indicate expression levels higher than the mean for each measured quantity, while blue indicates a decrease relative to the mean.

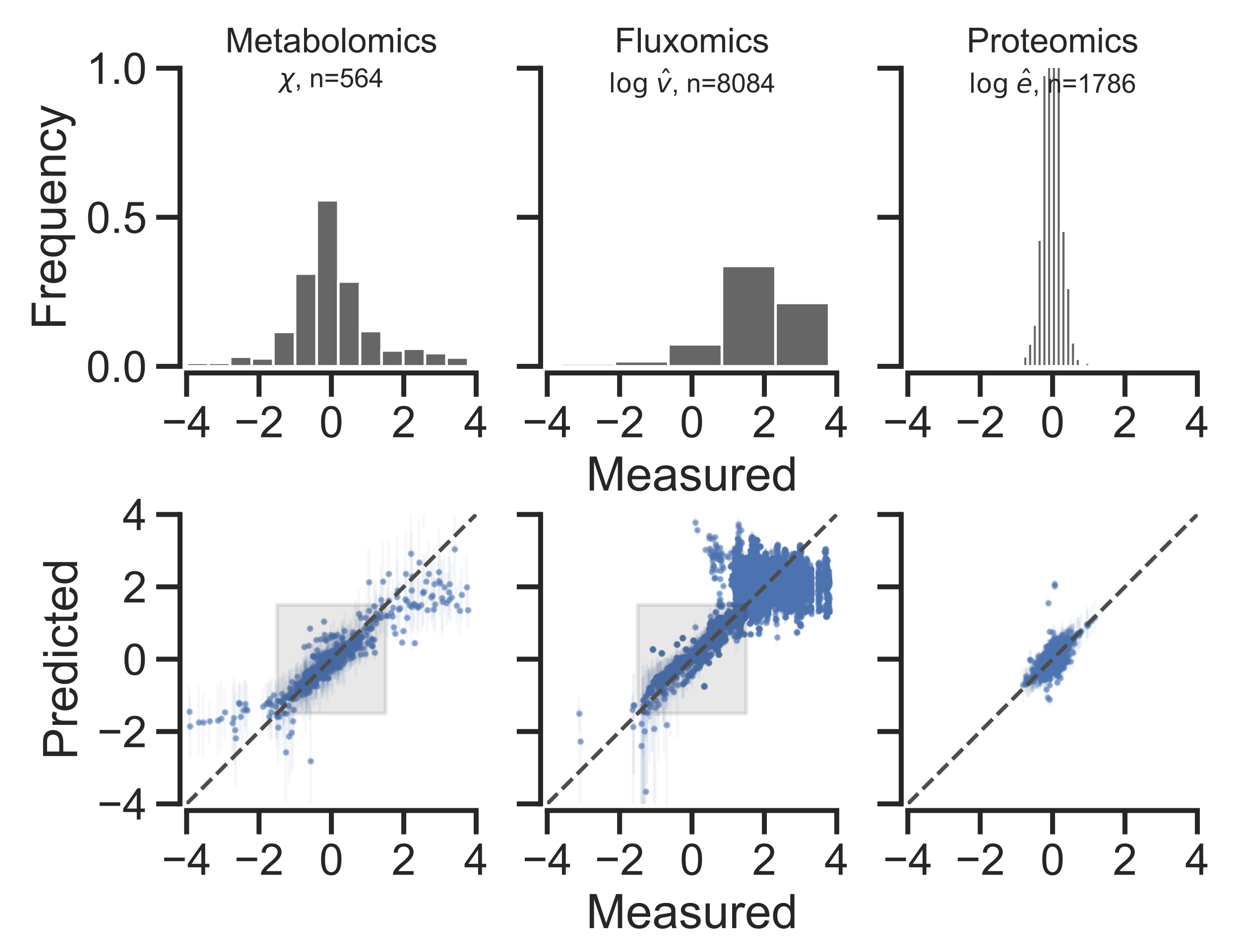
In addition, we calculated internal fluxes using eflux2 [6-7] with the extracellular fluxes and global proteomics as constraints for input to Bayesisan MCA (Appendix A)

Due to the size of the kinetic model considered, posterior distributions in kinetic parameters as a function of the observed data was estimated using automatic differentiation variational inference as implemented in the PyMC3 Python library. The model was optimized until convergence of the evidence lower bound score using the Adagrad optimizer (Fig. 2).

****

**Fig. 2. Optimization of the ELBO score**

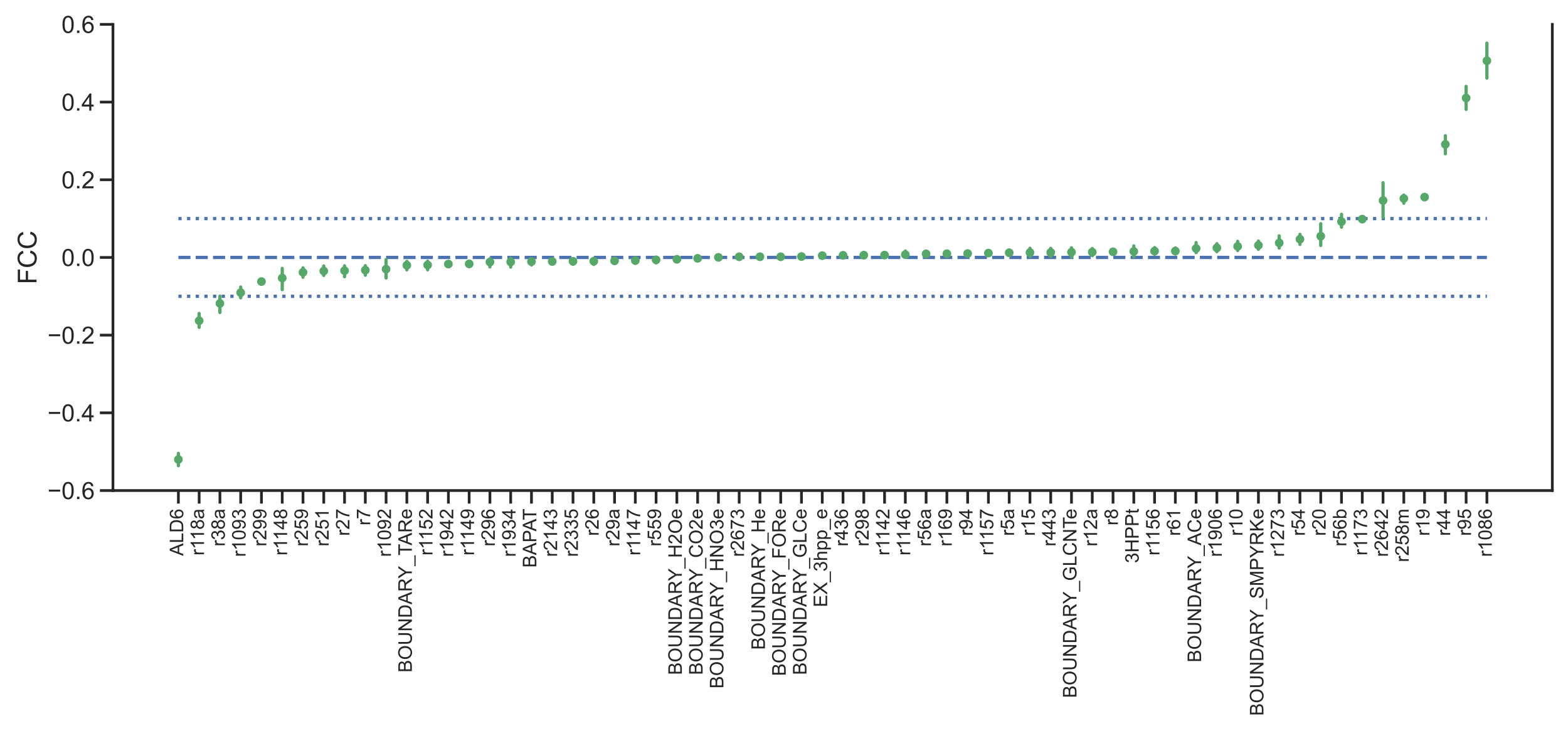
The posterior predictive distribution (PPD) of the model shows the ability of the model to reproduce the variability found in the experimental dataset. The PPD of the fitted model closely reproduces the measured steady-state flux and metabolite concentration data within the unclipped shaded region (Fig. 3). Outside this region, predicted metabolomics and the protein measurements fluxes were based on clipped measurements (hence the horizontal cluster of sample points). The proteomics measurements were accidentally withheld from the analysis which is why the measured distribution for all strains is 0. This figure will be updated with the measured proteomics data.

****

**Fig. 3. Posterior predictive distribution of the fitted model.** The metabolomics (left), intracellular and extracellular fluxes (center) and proteomics (right) closely match the experimentally measured values. Fluxes are in units of mmol/gDCW\*hr, while metabolomics and proteomics data are in log-transformed, dimensionless units relative to the reference strain (**[panD+,bapat+, hpdh+,pyc+,Δald6](https://registry.agilebiofoundry.org/entry/8348" \t "_blank)** ). Measured fluxes were calculated using strain-specific glucose uptake and 3HP secretion rates, under stationary state growth conditions due to phosphate starvation, with a proton export objective, using Eflux2 to constrain strain-specific intracellular fluxes with global proteomics data. Shaded boxes indicate the unclipped region where measured data is near the reference state.

With a kinetic model and estimated probability distributions in kinetic parameters, we can next conduct the Metabolic Control Analysis (MCA) portion of the BMCA framework. Here, we propagate the uncertainty in the estimated kinetic parameters to the metabolic design strategies suggested by MCA. In Fig. 4, we show the 95% highest posterior density regions of flux control coefficients (FCCs) on 3HP export calculated from the posterior distribution. FCCs capture the systems-level regulation of changing enzyme concentration on steady-state metabolic flux. The largest flux control coefficient was a nitrate transporter (r1086:1189116), which is rather surprising. Other overexpression candidates ordered by decreasing control over 3HP flux include 5-formyltetrahydrofolate deformylase (r95:1182700), malate dehydrogenase (r44:1144118), pyruvate carboxylase (r19:1031996), and mitochondrial aspartate transaminase (r258m: 1184650).

We also look at the most negative FCC’s, which represent targets for knockout or repression. In order of decreasing control over 3HP flux is ALD6, which converts 3-oxoproprionate to Acetyl-CoA, mitochondrial alcohol dehydrogenase (r118a), but this has 18 isoenzymes, so this is not a good candidate, and succinate CoA ligase (r38a:1145655 and 1141712). Of these, ALD6 is likely the only metabolic engineering target for knockout or repression, and this candidate was, in fact, knocked out in the fastest growing reference strain, further confirming the predictive power of Bayesian MCA.

****

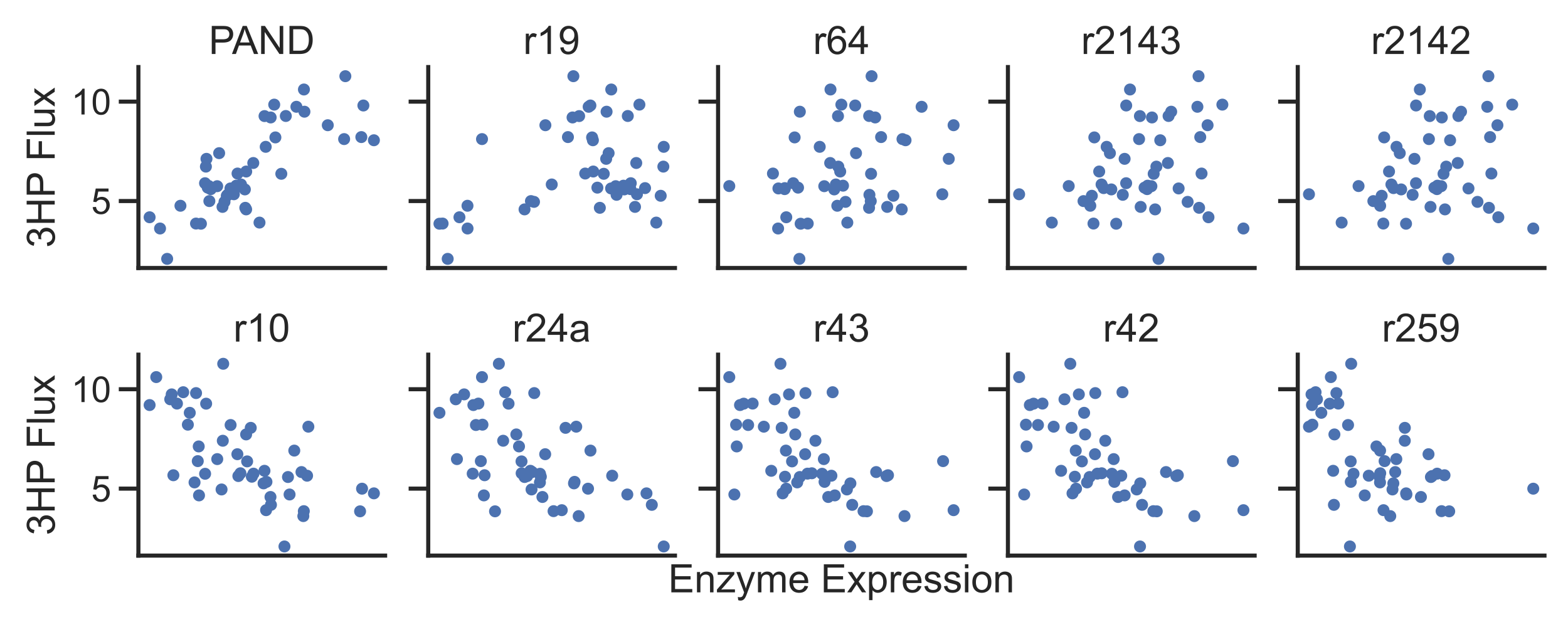
**Fig. 4. Posterior distributions in 3HP flux control coefficients.** A positive FCC indicates that an increase in the corresponding enzyme concentration will increase 3HP flux, while a negative FCC indicates that a decrease in enzyme concentration will increase 3HP flux. The dotted lines are provided as a qualitative aid for selecting candidate targets. FCC’s whose credible interval crosses the dashed line are not considered a target.

Figure 5 shows the engineered 3HP pathway that was inserted into *A. niger*.



**Fig. 5. Engineered pathways for 3-hydroxypropionate (3HP) flux in** [**panD+,bapat+, hpdh+,pyc+,Δald6**](https://registry.agilebiofoundry.org/entry/8348)  **strain under glucose media conditions designed to induce phosphate starvation.**

The performance of the BMCA methodology was then compared to approaches using only a single data modality. A simple approach for determining enzyme targets that does not require additional experimental data is by searching for proteins whose expression is correlated with higher 3HP flux. A list of the five more positively correlated and most negatively correlated genes is shown in Fig. 6. Many of these proteins are distantly or unrelated to 3HP production, underscoring that correlation does not always imply causation. Interestingly, aspartate alpha-decarboxylase (PAND), is a non-native enzyme that is part of the engineered 3HP pathway, so it is perhaps worth considering as a target for overexpression.



**Fig. 6. An analysis using only a single data modality to generate metabolic engineering predictions.** Proteins with the highest correlation with 3HP flux are shown in the top row, while proteins with the most negative correlation are shown in the bottom row. While comparison of hit rates for these methods against BMCA would require experimental confirmation, these sets likely include more false positive targets.

**References**

1. St. John, P. C., Strutz, J., Broadbelt, L. J., Tyo, K. E. J. & Bomble, Y. J. Bayesian inference of metabolic kinetics from genome-scale multiomics data. *PLOS Comput. Biol.* **15**, e1007424 (2019).

2. Visser, D. & Heijnen, J. J. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab. Eng.* **5**, 164–176 (2003).]

3. Brandl, J., Aguilar-Pontes, M.V., Schäpe, P. *et al.* A community-driven reconstruction of the *Aspergillus niger* metabolic network. *Fungal Biol Biotechnol* **5,**16 (2018). https://doi.org/10.1186/s40694-018-0060-7

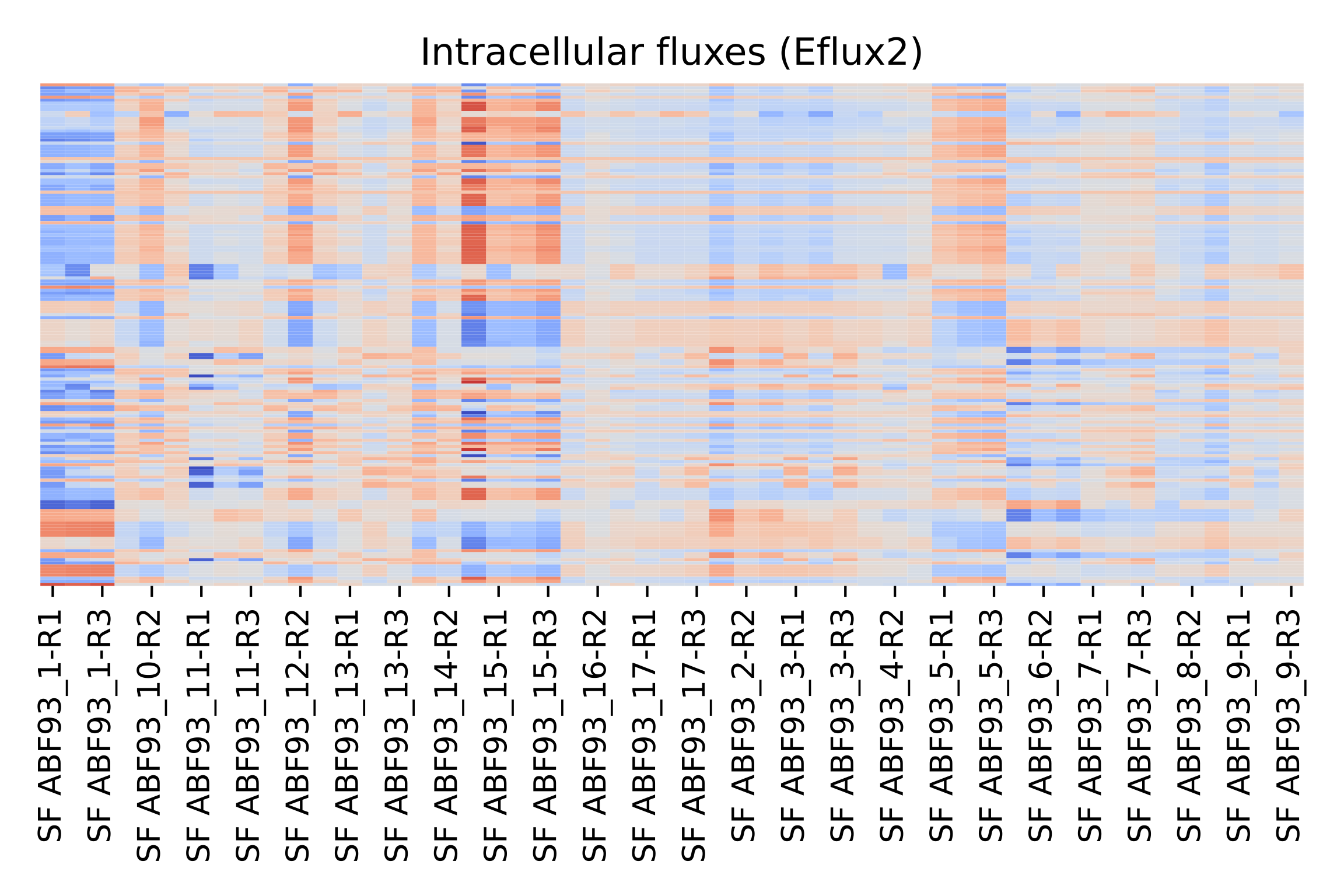
4. Lu, H., Cao, W., Liu, X., Sui, Y., Ouyang, L., Xia, J., Huang, M., Zhuang, Y., Zhang, S., Noorman, H. & Chu, J. Multi-omics integrative analysis with genome-scale metabolic model simulation reveals global cellular adaptation of Aspergillus niger under industrial enzyme production condition. *Sci. Rep.* **8,** 14404 (2018).

5. Pomraning KR, Dai Z, Munoz N, Kim YM, Gao Y, Deng S, Kim J, Hofstad BA, Swita MS, Lemmon T, Collett JR, Panisko EA, Webb-Robertson BM, Zucker JD, Nicora CD, De Paoli H, Baker SE, Burnum-Johnson KE, Hillson NJ, Magnuson JK. Integration of Proteomics and Metabolomics Into the Design, Build, Test, Learn Cycle to Improve 3-Hydroxypropionic Acid Production in *Aspergillus pseudoterreus*. Front Bioeng Biotechnol. 2021 Apr 7;9:603832. doi: 10.3389/fbioe.2021.603832. PMID: 33898398; PMCID: PMC8058442.

6. Kim MK, Lane A, Kelley JJ, Lun DS. E-Flux2 and SPOT: Validated Methods for Inferring Intracellular Metabolic Flux Distributions from Transcriptomic Data. PLoS ONE 2016;11:e0157101.

7. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput Biol 2009;5:e1000489.

**Appendix A**



**Supplemental figure 1: Eflux2 predictions** Measured fluxes for a reduced model with 172 reactions and 171 metabolites based on iJB1325 were calculated using strain-specific glucose uptake and 3HP secretion rates, under stationary state growth conditions due to phosphate starvation, with a proton export objective, using Eflux2 to constrain strain-specific intracellular fluxes with global proteomics data.

**Appendix A**