Engineering microbial consortia for distributed signal processing

Katherine E. Duncker[1,2], Ashwini R. Shende[1,2], Irida Shyti[1,2], Ashley Ruan[3,4], Ryan D'Cunha[1,3], Harshitha Venugopal-Lavanya[1], Helena R. Ma[1,2], Sizhe Liu[1], Neil Gottel[5], Deverick J. Anderson[6,7], Claudia Gunsch[5], Lingchong You[1,2,*]

1. Department of Biomedical Engineering, Duke University, Durham, NC, USA.
2. Center for Quantitative Biodesign, Duke University, Durham, NC, USA.
3. Department of Computer Science, Duke University, Durham, NC, USA.
4. Department of Mathematics, Duke University, Durham, NC, USA.
5. Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA.
6. Division of Infectious Diseases, Department of Medicine, Duke University School of Medicine, Durham, NC, USA.
7. Duke Center for Antimicrobial Stewardship and Infection Prevention, Duke University School of Medicine, Durham, NC, USA.

*Correspondence should be addressed to Lingchong You (you@duke.edu).

Supplementary Materials for this manuscript include the following:

Supplementary Notes 1-4

Supplementary Tables 1-3

Supplementary Figures 1-19

Supplementary Notes

1. VAE architecture and latent dimension optimization

To determine the optimal VAE architecture, we iteratively defined, trained, and tested the VAE with different architectures on the same dataset. Specifically, we tested architectures that varied in number of hidden layers and in the width of each hidden layer. We did this optimization for both the aTc-GFP and IPTG-mCherry community and the THS-CFP and TTR-YFP community datasets separately. Each architecture was trained in triplicate, and the resulting weights were saved. Architectures that yielded $R^2 > 0.989$ for reconstructions of the augmented training and validation time courses for that training iteration were subsequently used in the VAE-MLP framework. The latent dimensions from the architectures passing this threshold were used as inputs to train the MLP. The VAE architecture that yielded the highest average R^2 for predicting the experimental validation set concentrations for both datasets and that passed the VAE reconstruction threshold more than once in both analyses was selected for the final VAE-MLP pipeline used in this study.

Using the optimized VAE architecture, we systematically tested VAE latent space sizes ranging from 5 to 20 dimensions to determine the optimal number for subsequent MLP predictions of input concentrations. We performed this optimization on both the aTc-GFP and IPTG-mCherry community and the THS-CFP and TTR-YFP community datasets. The VAE was trained three times for each latent space size, and the model was saved after each iteration. Latent variables that yielded $R^2 > 0.97$ for VAE reconstructions of the augmented training and validation time courses for that training iteration were subsequently used in the MLP to predict inputs. The latent dimension size that yielded the highest average R^2 for predicting experimental validation set concentrations for both datasets was selected for the final pipeline used in this study.

2. Evaluating robustness of pipeline to model formulation and training procedure

To determine whether pipeline performance depends on specific modeling assumptions or fitting protocols, we conducted a series of robustness analyses examining (i) alternative kinetic model formulations, (ii) kinetic parameter non-identifiability, and (iii) training protocol (Supplementary Fig. 7).

Importantly, in our framework the kinetic ODE model is not used for prediction at inference time. Rather, it serves solely as a structured generator to augment sparse experimental datasets. Therefore, we hypothesized that inference accuracy should depend primarily on preserved trajectory-level input–output relationships, rather than on a specific kinetic formulation or parameterization.

(a) Robustness to kinetic model formulation

In addition to the baseline phenomenological kinetic ODE model used throughout this study (Methods), we implemented an alternative formulation in which reporter dynamics explicitly incorporate growth-mediated dilution in addition to baseline degradation. Specifically, the fluorescence dynamics were modified to:

$$\frac{dy_i}{dt} = \left(r_{0,i} + \sum_j \alpha_{i,j} \frac{k_i S_j^{h_{i,j}}}{K_{i,j}^{h_{i,j}} + S_j^{h_{i,j}}} \right) \mu_{eff,i} - (\mu_{eff,i} + d_{p0,i}) y_i \quad \text{(S1)}$$

where $\mu_{eff,i}$ is the instantaneous growth rate of strain i, and $d_{p0,i}$represents baseline protein degradation. All other variables and equations were identical to the base model.

Using this revised kinetic formulation, we repeated the entire computational pipeline:

1. Kinetic parameter estimation via nonlinear least-squares optimization
2. Generation of 10,000 synthetic trajectories
3. VAE training on augmented datasets
4. VAE–MLP training and evaluation

Despite differences in kinetic model structure, inference accuracy on experimental validation data remained qualitatively unchanged (Supplementary Fig. 7a). These results indicate that inference performance was preserved across the two kinetic formulations tested here, provided that each model captured the experimental time-course data sufficiently well.

(b) Robustness to kinetic parameter non-identifiability

Because the kinetic model parameters are not uniquely identifiable from the experimental time-course data, multiple distinct parameter sets can fit the observed trajectories comparably well. To assess whether this ambiguity affects downstream inference performance, we generated three independent parameter sets using different optimization strategies and parameter bound constraints:

- Trust Region Reflective (TRF) algorithm with restrictive bounds (dp0 ∈ [0.02, 0.2])
- Dogbox algorithm with moderate bounds (dp0 ∈ [0.03, 0.3])
- TRF algorithm with moderate bounds

All fits were performed using the nonlinear least-squares function from the SciPy optimize package. Each parameter set achieved comparable least-squares error on the experimental trajectories but yielded distinct kinetic parameter values. In the parameter list above dp0 represents the non-dilution protein decay rate (degradation independent of cell growth).

For each parameter set, we:

1. Generated 10,000 synthetic time-course trajectories
2. Combined these synthetic trajectories with experimental training data
3. Trained a separate VAE–MLP model
4. Evaluated performance on the same experimental validation partition

All three VAE–MLP models exhibited similarly high predictive accuracy and produced nearly identical predictions, as quantified by cross-model Pearson correlation (Supplementary Fig. 7b). These results demonstrate that uncertainty and non-identifiability in kinetic parameters did not compromise the effectiveness of using the ODE model for structured data augmentation. The VAE–MLP learns the input–output mapping that is statistically preserved across different kinetic parameterizations, rather than relying on a unique parameter set.

(c) Robustness to training protocol (split-before-fitting)

In the standard pipeline, the kinetic model is fit using all available experimental data before synthetic data generation. While the downstream ML model is trained and evaluated on explicit

experimental train/validation splits, this procedure allows the synthetic data distribution to be influenced by all experimental samples. To evaluate performance under a more conservative protocol, we implemented a strict split-before-fitting strategy:

1. Experimental data (n = 126 samples) were partitioned into 80% training (n = 101) and 20% validation (n = 25).
2. The kinetic model was fit exclusively to the experimental training subset.
3. Synthetic trajectories were generated using only parameters estimated from the training data.
4. The VAE–MLP was trained on synthetic data plus experimental training data.
5. Final performance was evaluated solely on the experimental validation set.

Under this protocol, predictive accuracy was similarly high as the base case (Supplementary Fig. 7c). These results indicate that the qualitative conclusions were preserved under the tested split-before-fitting protocol.

3. Probing the learned latent space

To explore how the VAE–MLP framework encodes time-resolved trajectories, we performed a series of post hoc analyses of the learned latent space (Supplementary Fig. 8). All analyses described here were conducted using the originally trained VAE–MLP model on the first example (Figure 2) from the main pipeline; no retraining or fine-tuning was performed for this analysis.

(a) Correlation between latent dimensions and inputs

For each latent variable z_0–z_9, we computed the Pearson correlation coefficient between latent values and the corresponding input concentrations (aTc and IPTG), separately for training and validation datasets.

Several features are notable (Supplementary Fig. 8a):

1. Multiple latent dimensions exhibit non-zero correlation with each input.
2. No single latent dimension uniquely corresponds to a single input.
3. Correlation patterns are qualitatively consistent between training and validation datasets.

These observations indicate that the learned representation distributes input-relevant information across multiple latent dimensions rather than encoding a one-to-one mapping between latents and chemicals.

(b) Global geometry of the latent space

To visualize the global organization of the learned representation, we performed principal component analysis (PCA) on the latent vectors computed from the training dataset and projected the validation dataset into the same PCA basis.

When colored by true input concentration, both training and validation samples exhibit smooth gradients and coherent geometric organization in the latent space (Supplementary Fig. 8b). This structure suggests that the VAE has learned a low-dimensional manifold reflecting systematic variations in experimental conditions, rather than memorizing individual samples.

(c) Functional sensitivity of predictions to individual latents

To assess how latent dimensions contribute to downstream predictions, we performed a sensitivity analysis using the trained MLP (Supplementary Fig. 8c).

For each latent dimension z_i:

1. The value of $z_i$was varied across its empirical range (±2 standard deviations).
2. All other latent dimensions were held fixed at their mean values.
3. The MLP was used to predict the corresponding input concentrations.

This analysis reveals that the VAE–MLP does not rely on a single latent variable for inference, but instead uses a distributed, nonlinear encoding of trajectory features.

4. Example of five-input inference from three time-resolved readouts

To examine one calibrated case in which the number of inferred inputs exceeds the number of measured readouts, we constructed a two-strain community with three time-resolved readouts (OD, mCherry, and YFP) and five chemical inputs (Supplementary Fig. 17). This configuration allowed us to evaluate whether shared, overlapping dynamical responses contained sufficient information to infer all five inputs. The overall computational workflow and VAE–MLP architecture were unchanged, whereas the kinetic ODE model was extended to represent the five-input system.

(a) Community configuration and kinetic model formulation

The system consists of two sensor strains:

- MG1655 IPTG–mCherry
- MG1655 ttr–YFP

The five chemical inputs are:

- IPTG
- Tetrathionate (ttr)
- Kanamycin (kan)
- Chloramphenicol (cm)
- Glucose

Growth (OD), mCherry, and YFP were measured for multiple combinations of input concentrations.

(b) Kinetic model formulation

To generate augmented training data, the kinetic ODE model was extended to incorporate five chemical inputs. The cell growth equations for each population are given as follows:

$$\frac{dn_i}{dt} = \mu_i \left(\frac{1}{1+\left(\frac{n_i}{N_m K_{s,i}}\right)^{\theta_i}} \right) \left(1 - \frac{n_i}{N_m}\right) \left(1 + \frac{S_{glucose}}{1+\sum_j d_{S,j} S_j}\right) n_i \quad \text{(S2)}$$

$$\sum_j d_{S,j} S_j = d_{s,1} S_1 + d_{s,2} S_2 + d_{s,3} S_3 + d_{s,4} S_4 + d_{s,5} S_5 \quad \text{(S3)}$$

Note that $j = 5$, i.e. the parameter $d_{S,j}$ is fitted for each of the five chemical inputs and the term $\sum_j d_{S,j} S_j$ is summed over these five inputs. The other change in the cell growth equations arises from the addition of glucose as a chemical input to be detected. We expect glucose to have a beneficial effect on growth in a dose-dependent manner. As such, we modify the final term multiplying cell density to reflect this. We find that this ODE formulation is sufficiently expressive to capture experimental data.

For the equations describing fluorescence, again note that $j = 5$ in the summation term.

$$\frac{dy_i}{dt} = r_{0,i} + \sum_j \alpha_{i,j} \frac{k_i S_j^{h_{i,j}}}{K_{i,j}^{h_{i,j}} + S_j^{h_{i,j}}} - d_{p,i} y_i \quad \text{(S4)}$$

In the pipeline, concentrations and K values are normalized to the K value corresponding to the fluorescence readout that the chemical is intended to activate. There is no intended dose response for glucose, kanamycin, and chloramphenicol in this system, so we use a similar approach as with the antibiotic-inhibitor datasets (see **Methods**) and identify the fluorescence

which is most responsive to changes in these chemical concentrations. As a result, we choose the K value associated with YFP for normalization for these inputs.

(c) Computational pipeline

The full VAE–MLP pipeline was structurally identical to other examples:

1. The kinetic model was fit to experimental time-course data.
2. Synthetic trajectories were generated across a range of five-dimensional input combinations.
3. Synthetic trajectories were combined with experimental training data.
4. The VAE was trained to encode concatenated OD and fluorescence time courses.
5. The MLP was trained to predict the five input concentrations from the learned latent variables.
6. Performance was evaluated on experimental validation samples.

As in previous analyses, evaluation was performed exclusively on experimental data.

Supplementary Tables

Supplementary Table 1: Addgene identifiers for plasmids used for this study

Plasmid	Addgene reference
pKD236-4b	Addgene plasmid # 90956; http://n2t.net/addgene:90956; RRID:Addgene_90956
pKD237-3a-2	Addgene plasmid # 90957; http://n2t.net/addgene:90957; RRID:Addgene_90957
pKD238-1a	Addgene plasmid # 90958; http://n2t.net/addgene:90958; RRID:Addgene_90958
pKD239-1g-2	Addgene plasmid # 90959; http://n2t.net/addgene:90959; RRID:Addgene_90959
pAJM.1642	Addgene plasmid # 108535; http://n2t.net/addgene:108535; RRID:Addgene_108535
pAJM.657	Addgene plasmid # 108525; http://n2t.net/addgene:108525; RRID:Addgene_108525
pAJM.011	Addgene plasmid # 108529; http://n2t.net/addgene:108529; RRID:Addgene_108529
pAJM.773	Addgene plasmid # 108527; http://n2t.net/addgene:108527; RRID:Addgene_108527
pAJM.847	Addgene plasmid # 108524; http://n2t.net/addgene:108524; RRID:Addgene_108524
pAJM.661	Addgene plasmid # 108532; http://n2t.net/addgene:108532; RRID:Addgene_108532

Supplementary Table 2: Kinetic model parameter values for initial estimate and bounds of the optimization

Parameter	Initial estimate	Bounds
μ_i	0.756	0.1 – 2
$K_{s,i}$	0.099	0.01 – 1
θ_i	3	0.6 – 6
$d_{s,j}$	0.5	0 – 1
$r_{0,i}$	0	0 – 1
k_i	1	0.0001 –1
$d_{p,i}$	0.1	0 – 1

Note: The carrying capacity was fixed at 1.5 to stabilize parameter estimation. This value is substantially larger than the maximum OD observed experimentally (~0.4–0.8) and therefore serves as a soft constraint rather than a fitted biological quantity.

Supplementary Table 3: Crosstalk statistics for all sensor systems

Entries report mean α, 95% bootstrap confidence interval (**CI_{low}**, **CI_{high}**), bootstrap standard deviation (sd), and n (# replicates per condition). Bootstrap resampling was performed over replicates (B = 2000). Negative values indicate input-induced suppression. See detailed interpretation below.

dataset	channel	inducer	mean	CI_{low}	CI_{high}	sd	n
aTc_IPTG	GFP	IPTG	-0.0228	-0.0298	-0.0170	0.0033	2
aTc_IPTG	GFP	aTc	1.0000	0.8258	1.2110	0.0954	2
aTc_IPTG	mCh	IPTG	1.0000	0.9680	1.0331	0.0159	2
aTc_IPTG	mCh	aTc	-0.0698	-0.0780	-0.0618	0.0050	2
TTR_THS	CFP	THS	1.0000	0.9561	1.0459	0.0219	2
TTR_THS	CFP	TTR	0.8466	0.8259	0.8683	0.0135	2
TTR_THS	YFP	THS	0.0796	0.0570	0.1027	0.0152	2

TTR_THS	YFP	TTR	1.0000	0.9825	1.0179	0.0088	2
cuma_ohc_atc	CFP	atc	0.0235	0.0189	0.0287	0.0027	2
cuma_ohc_atc	CFP	cuma	-0.0053	-0.0076	-0.0032	0.0014	2
cuma_ohc_atc	CFP	ohc14	1.0000	0.8970	1.1148	0.0532	2
cuma_ohc_atc	YFP	atc	0.0528	0.0386	0.0681	0.0090	2
cuma_ohc_atc	YFP	cuma	1.0000	0.9227	1.0838	0.0399	2
cuma_ohc_atc	YFP	ohc14	0.0729	0.0628	0.0838	0.0056	2
cuma_ohc_atc	mCh	atc	1.0000	1.0000	1.0000	0.0000	2
cuma_ohc_atc	mCh	cuma	0.0116	0.0000	0.0233	0.0083	2
cuma_ohc_atc	mCh	ohc14	-0.0058	-0.0233	0.0116	0.0121	2
van_dapg_nar	CFP	dapg	1.0000	0.9125	1.0959	0.0468	3
van_dapg_nar	CFP	nar	-0.0767	-0.0995	-0.0602	0.0109	3
van_dapg_nar	CFP	van	0.1007	0.0908	0.1115	0.0053	3
van_dapg_nar	YFP	dapg	0.0240	0.0229	0.0252	0.0006	3
van_dapg_nar	YFP	nar	0.0096	0.0065	0.0126	0.0015	3
van_dapg_nar	YFP	van	1.0000	0.9488	1.0540	0.0265	3
van_dapg_nar	mCh	dapg	0.1339	0.1190	0.1471	0.0069	3
van_dapg_nar	mCh	nar	1.0000	0.8667	1.1538	0.0739	3
van_dapg_nar	mCh	van	0.0536	0.0268	0.0804	0.0132	3
atc_van	YFP	atc	0.0012	0.0007	0.0018	0.0003	2
atc_van	YFP	van	1.0000	0.8343	1.1986	0.0888	2
atc_van	mCh	atc	1.0000	0.8784	1.1385	0.0644	2
atc_van	mCh	van	-0.0038	-0.0061	-0.0018	0.0013	2
Bla_TAZ	BFP	AMX	-1.0000	-1.0000	-0.6289	0.0919	3
Bla_TAZ	BFP	TAZ	0.6429	0.0714	1.0000	0.2657	3
Bla_TAZ	GFP	AMX	0.6793	0.5757	0.7945	0.0543	3
Bla_TAZ	GFP	TAZ	-1.0000	-1.0000	-1.0000	0.0000	3
Bla_SUL	BFP	AMX	-0.2642	-0.4048	-0.1719	0.0610	3
Bla_SUL	BFP	SUL	-1.0000	-1.0000	-1.0000	0.0000	3
Bla_SUL	GFP	AMX	1.0000	1.0000	1.0000	0.0000	3

Bla_SUL	GFP	SUL	-0.0532	-0.0873	-0.0057	0.0204	3
BlaM_TAZ	BFP	AMX	-1.0000	-1.0000	-0.8219	0.0521	3
BlaM_TAZ	BFP	TAZ	0.6407	0.2403	1.0000	0.2050	3
BlaM_TAZ	GFP	AMX	0.9986	0.7805	1.0000	0.0623	3
BlaM_TAZ	GFP	TAZ	-1.0000	-1.0000	-0.8093	0.0534	3
BlaM_SUL	BFP	AMX	-0.5779	-1.0000	-0.3690	0.1523	3
BlaM_SUL	BFP	SUL	-1.0000	-1.0000	-1.0000	0.0141	3
BlaM_SUL	GFP	AMX	0.3027	0.2511	0.3401	0.0216	3
BlaM_SUL	GFP	SUL	-1.0000	-1.0000	-1.0000	0.0000	3
HSGBla_TAZ	BFP	AMX	-1.0000	-1.0000	-1.0000	0.0000	3
HSGBla_TAZ	BFP	TAZ	0.3422	0.2274	0.4525	0.0561	3
HSGBla_TAZ	GFP	AMX	1.0000	1.0000	1.0000	0.0000	3
HSGBla_TAZ	GFP	TAZ	-0.1567	-0.1674	-0.1482	0.0049	3
HSGBla_SUL	BFP	AMX	-0.8413	-0.9736	-0.7278	0.0668	3
HSGBla_SUL	BFP	SUL	-1.0000	-1.0000	-1.0000	0.0053	3
HSGBla_SUL	GFP	AMX	1.0000	1.0000	1.0000	0.0000	3
HSGBla_SUL	GFP	SUL	-0.4070	-0.4355	-0.3806	0.0130	3
HSGBlaM_TAZ	BFP	AMX	-1.0000	-1.0000	-1.0000	0.0000	3
HSGBlaM_TAZ	BFP	TAZ	0.0000	-0.3034	0.4382	0.1835	3
HSGBlaM_TAZ	GFP	AMX	-0.1768	-0.3635	0.0182	0.0881	3
HSGBlaM_TAZ	GFP	TAZ	-1.0000	-1.0000	-1.0000	0.0000	3
HSGBlaM_SUL	BFP	AMX	-0.1604	-0.4133	0.0936	0.1264	3
HSGBlaM_SUL	BFP	SUL	-1.0000	-1.0000	-1.0000	0.0000	3
HSGBlaM_SUL	GFP	AMX	0.0057	-0.0229	0.0264	0.0121	3
HSGBlaM_SUL	GFP	SUL	-1.0000	-1.0000	-1.0000	0.0000	3

Notes:

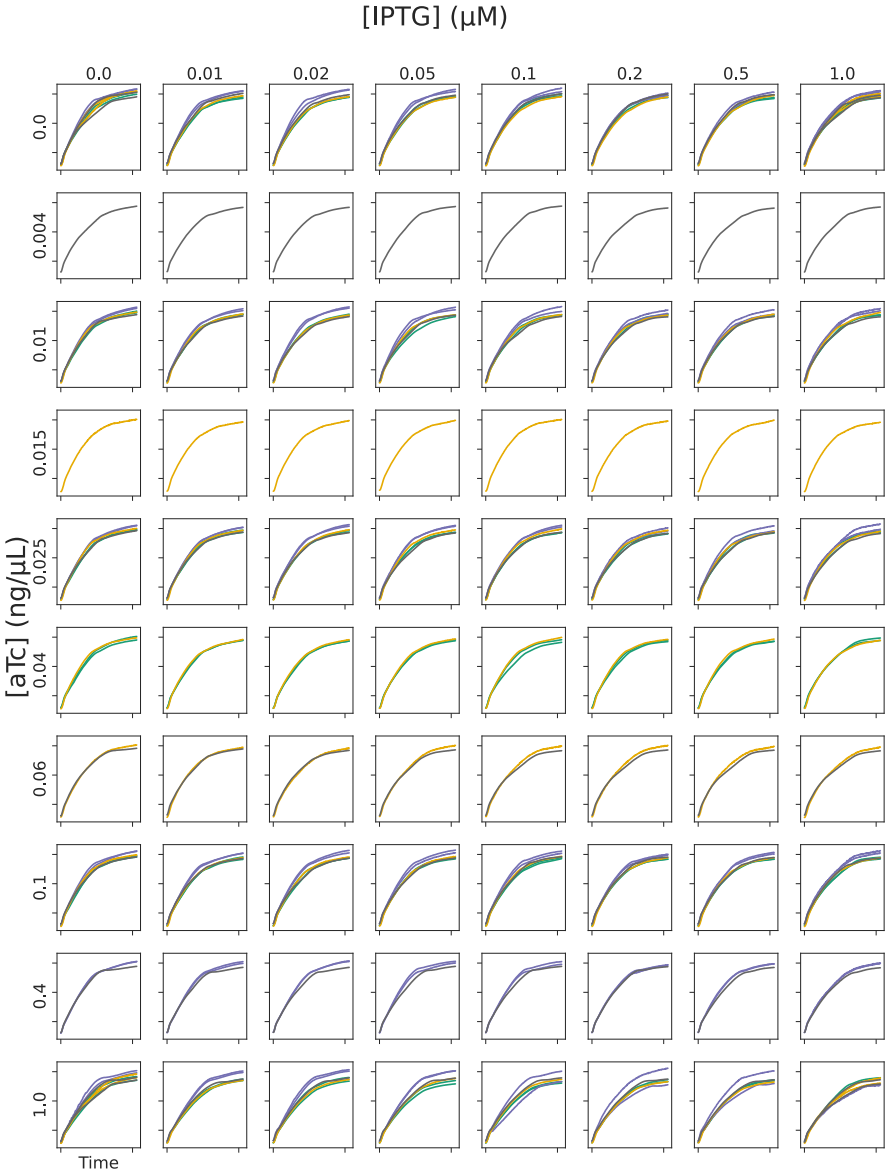
Crosstalk coefficient (α): Quantifies sensor response strength, normalized so that the maximum response for each sensor equals 1.0. For intended sensor-input pairs, α ≈ 1.0 indicates strong activation. For off-target pairs, α ≈ 0 indicates good specificity.

Negative α values: Indicate that the input suppresses fluorescence below baseline rather than activating it. This phenomenon occurs primarily in antibiotic sensor systems, where β-lactam

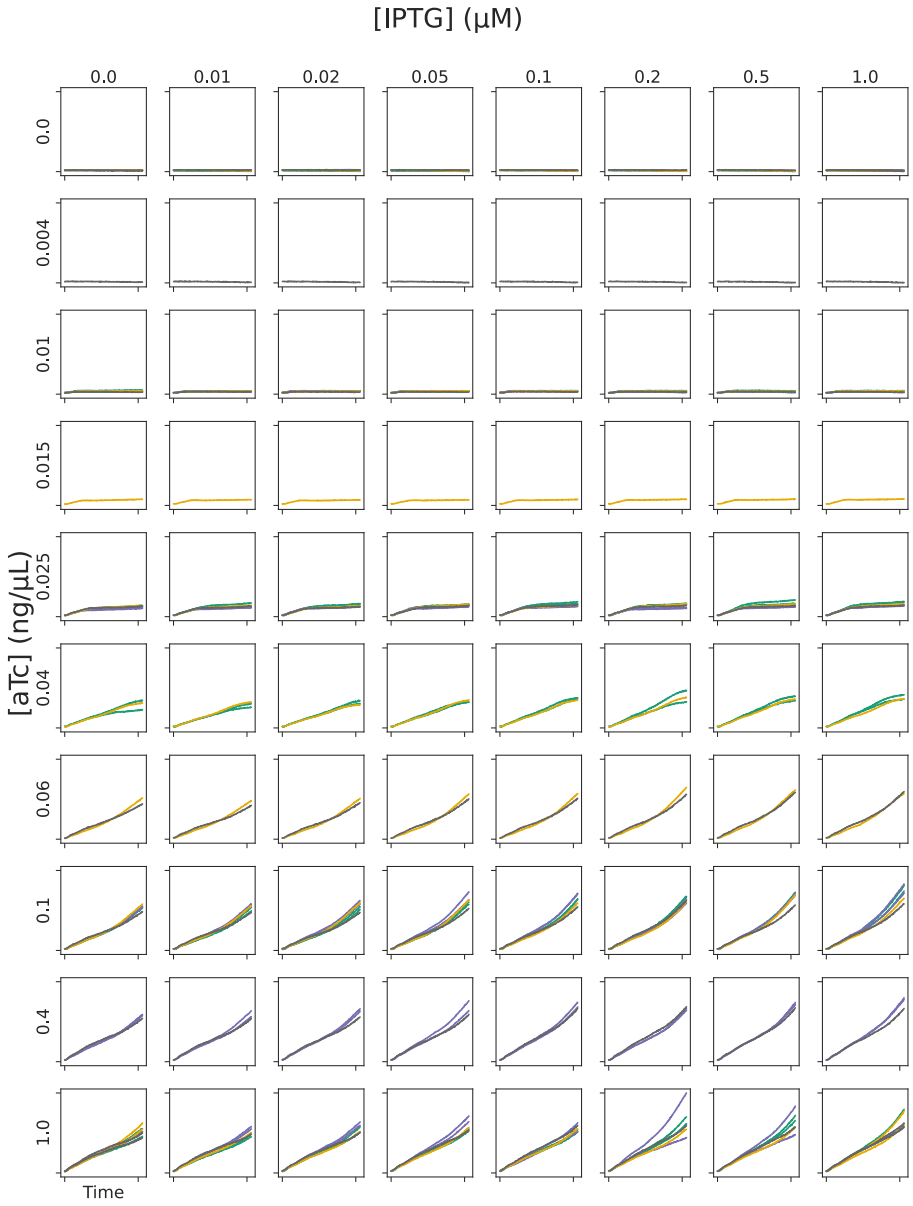
antibiotics stress bacterial cells, reducing overall protein expression including basal sensor output. An α value of -1.0 indicates complete suppression to the no-input baseline.

Confidence intervals (CI): Represent 95% bootstrap confidence intervals computed via resampling with replacement (B=2,000 iterations). Narrow CIs indicate precise, reproducible measurements.

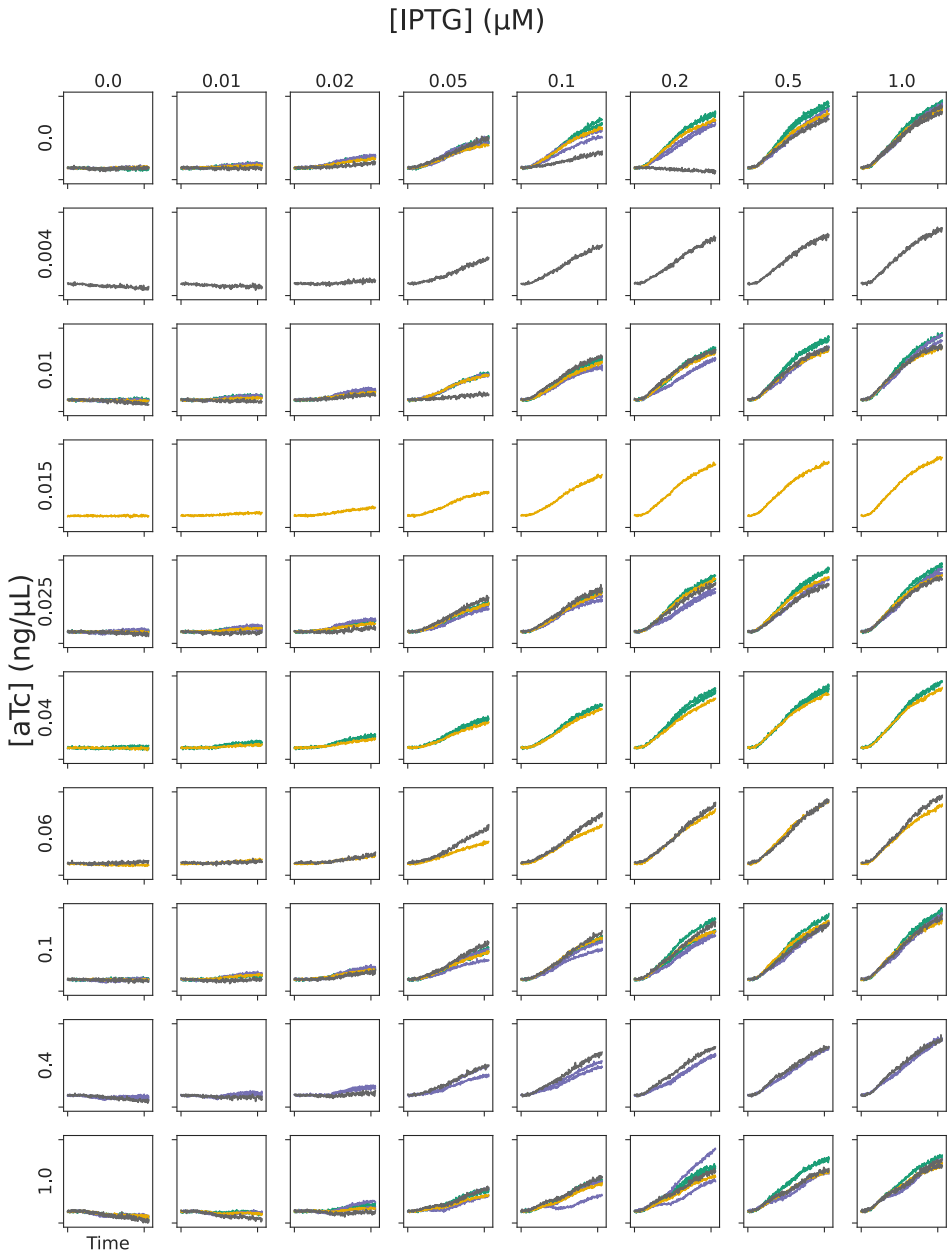
Supplementary Figures



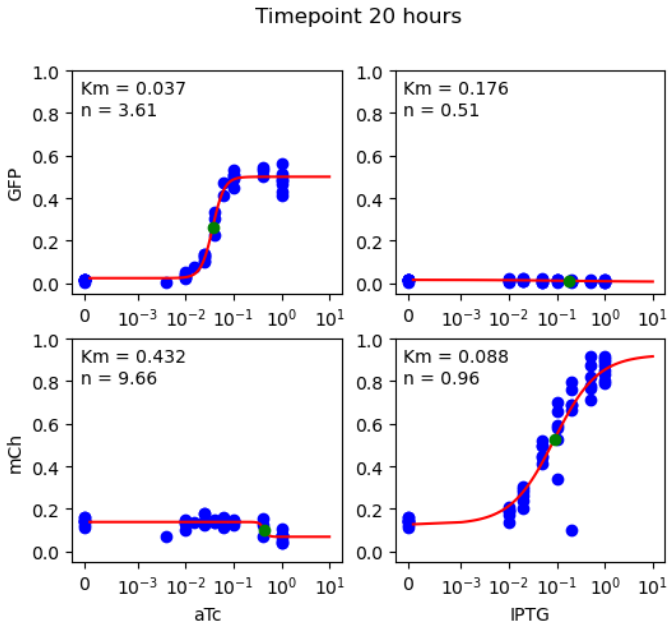
Supplementary Figure 1. OD time courses after data processing for all experimental conditions of the low-crosstalk aTc-GFP and IPTG-mCherry sensor set. Some samples here were excluded from full ODE model fitting and the VAE-MLP pipeline after determining the minimum and maximum threshold limits for the input concentrations based on the dose response curves (Methods). Rows and columns indicate aTc and IPTG concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



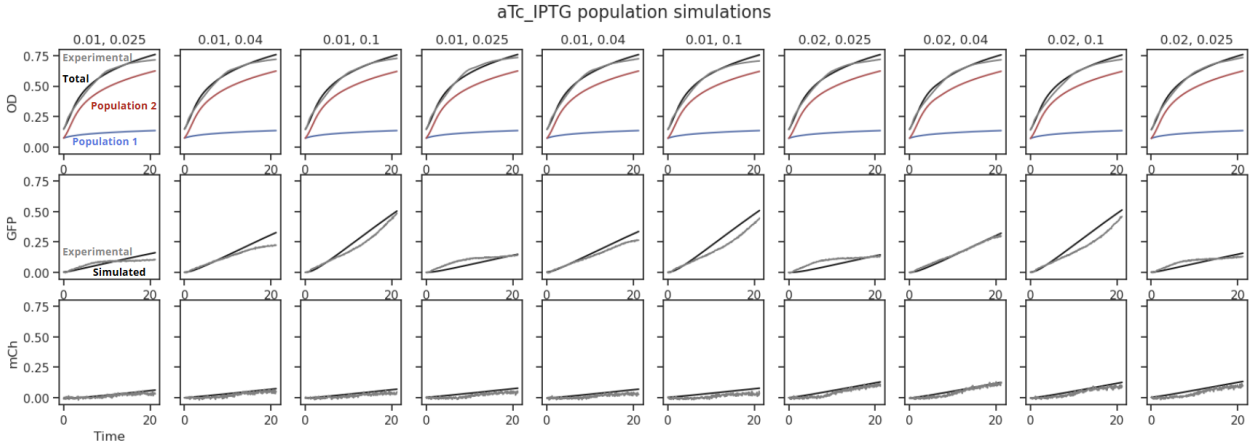
Supplementary Figure 2. GFP time courses after data processing for all experimental conditions of the low-crosstalk aTc-GFP and IPTG-mCherry sensor set. All samples here were used to estimate crosstalk and Hill equation parameters, but some were excluded from full ODE model fitting and the VAE-MLP pipeline after determining the minimum and maximum threshold limits for the input concentrations based on the dose response curves (Methods). Rows and columns indicate aTc and IPTG concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



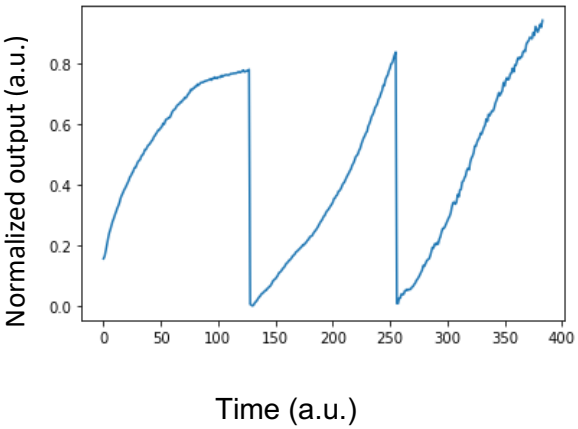
Supplementary Figure 3. mCherry time courses after data processing for all experimental conditions of the low-crosstalk aTc-GFP and IPTG-mCherry sensor set. All samples here were used to estimate crosstalk and Hill equation parameters, but some were excluded from full ODE model fitting and the VAE-MLP pipeline after determining the minimum and maximum threshold limits for the input concentrations based on the dose response curves (Methods). Rows and columns indicate aTc and IPTG concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



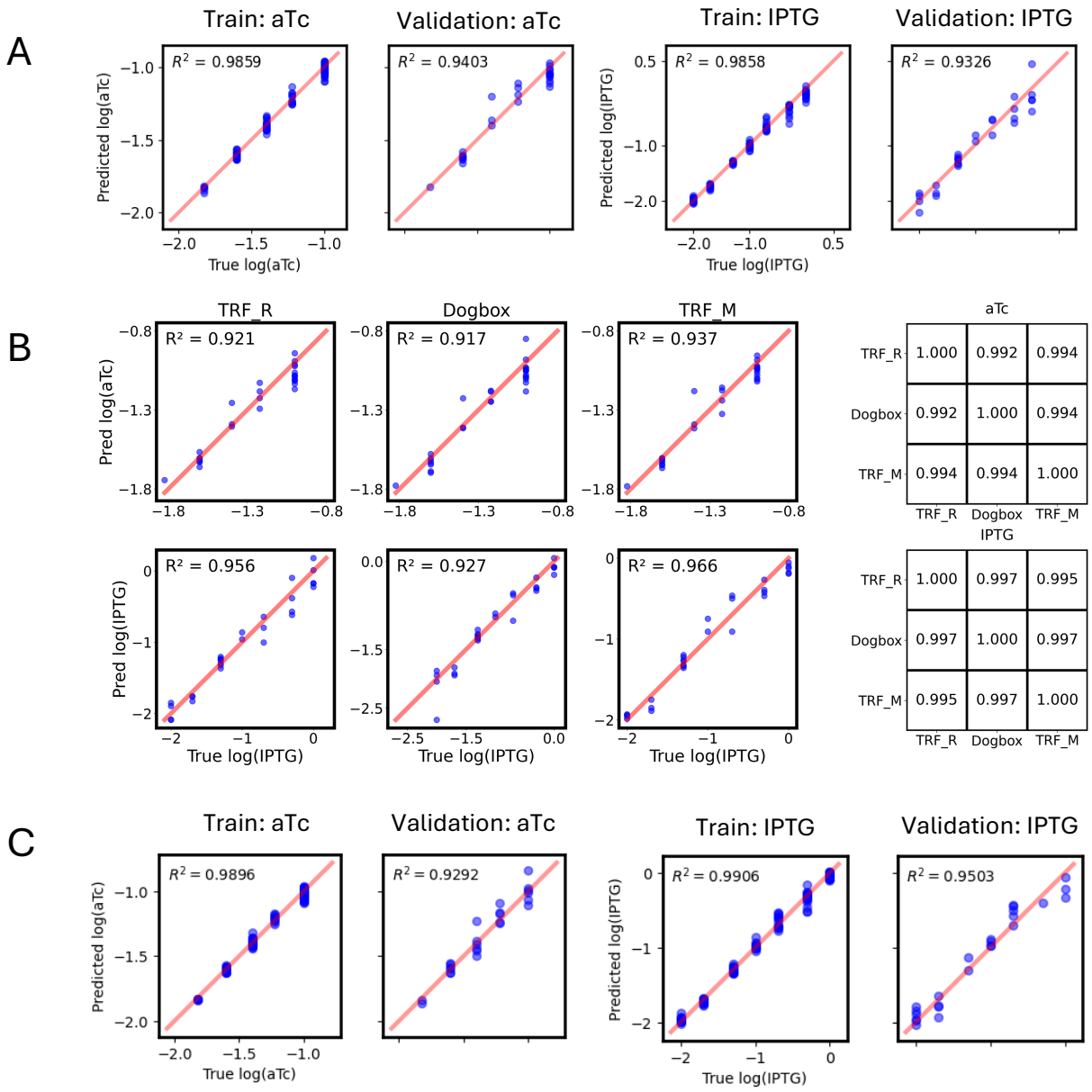
Supplementary Figure 4. Dose response data at 20 hours for aTc-GFP and IPTG-mCherry dataset. Blue points are experimental data, red line is the Hill equation with parameters optimized to fit the experimental data. Km and n parameter values of the hill equation are estimated using this optimization. Source data are provided as a Source Data file.



Supplementary Figure 5. Simulated growth curves of individual populations in the low-crosstalk microbial community for sensing aTc and IPTG, with different input concentrations. Population 1 (blue line) doesn't reach the same cell density as population 2 (red line). Total simulated OD (black line) and experimentally measured OD (gray line) are also plotted (top row). Simulated and experimental fluorescence curves are also plotted (middle and bottom row). Source data are provided as a Source Data file.



Supplementary Figure 6. Example of concatenated OD, GFP, and mCherry time courses, which form a one-dimensional data series. The blue trace shows the concatenated normalized output series. Source data are provided as a Source Data file.



Supplementary Figure 7. Robustness of pipeline performance to modeling and training assumptions

a. **Robustness to kinetic model formulation**. Pipeline performance is shown for synthetic datasets generated using an alternative formulation that explicitly incorporates growth-dependent expression and dilution of each reporter.

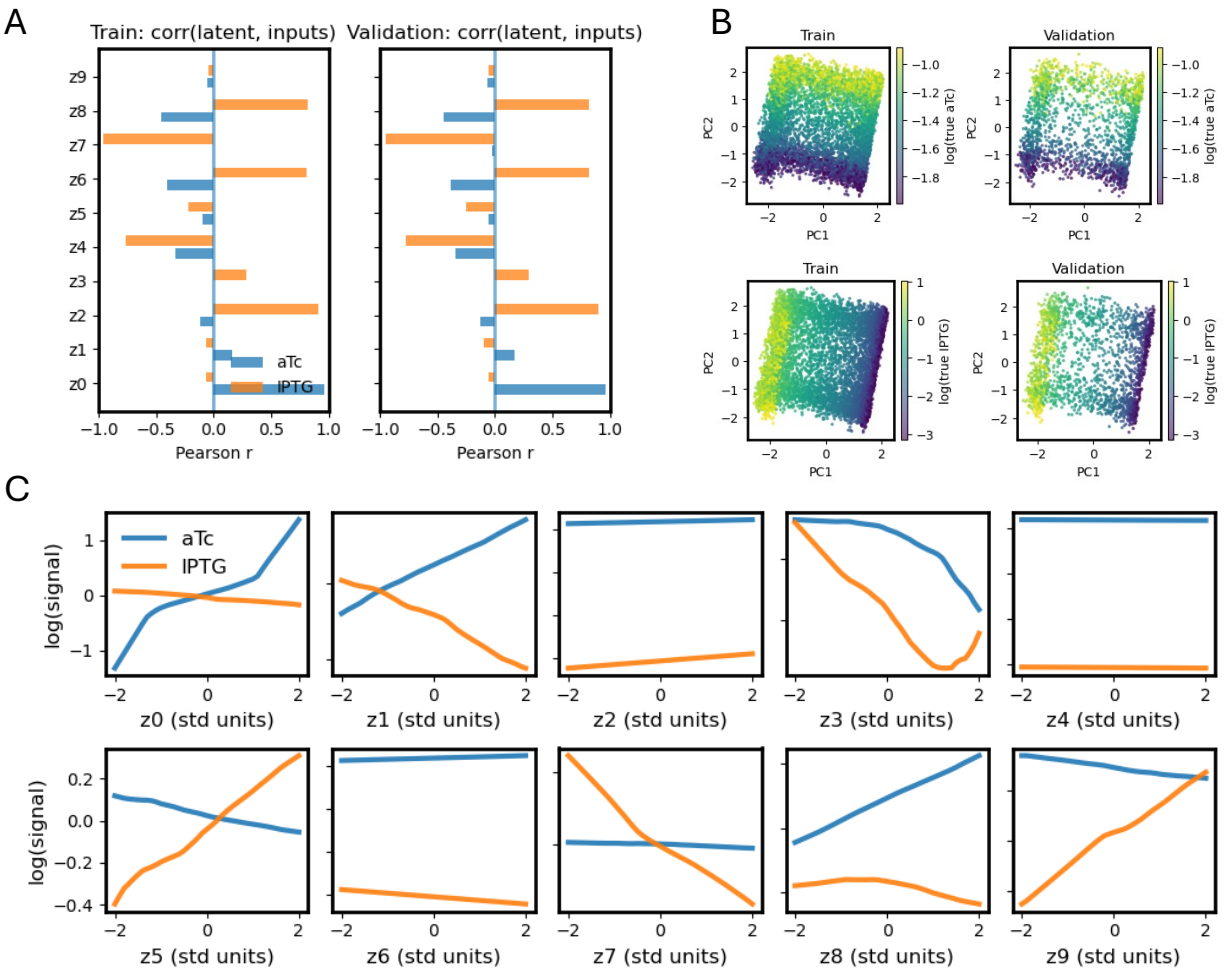
$$\frac{dy_i}{dt} = \left(r_{0,i} + \sum_j \alpha_{i,j} \frac{k_i S_j^{h_{i,j}}}{K_{i,j}^{h_{i,j}} + S_j^{h_{i,j}}} \right) u_{eff} - (u_{eff} + d_{p,i0}) y_i \quad \text{(S1)}$$

All the variables and the other equations are identical to the base model (Methods). The subsequent analysis (kinetic parameter estimation → synthetic data generation → VAE training → VAE-MLP training) remains the same as the use of the base model. Despite differences in

kinetic model structure, inference accuracy on held-out experimental data remains qualitatively unchanged.

b. **Robustness to kinetic model parameter non-identifiability.** The ODE model was fit to the same experimental data using multiple optimization strategies and parameter bound constraints, yielding distinct parameter sets that fit the observed trajectories comparably well. Synthetic training datasets were generated independently from each parameter set and used, together with experimental training data, to train separate VAE–MLP models. When evaluated on held-out experimental samples, all models achieved similarly high prediction accuracy (left panels) and produced nearly identical predictions, as quantified by cross-model correlation (right panels). These results demonstrate that uncertainty and non-identifiability in kinetic parameters do not compromise the effectiveness of using the ODE model for data augmentation, and that inference depends on learned input–output structure rather than on a unique parameterization.

c. **Robustness to training protocol**. Pipeline performance under the standard training procedure is compared to a conservative split-before-fitting protocol, in which the base kinetic model is fit exclusively to the experimental training subset prior to synthetic data generation. Although this protocol modestly reduces accuracy, the qualitative conclusions and overall inference performance are preserved, indicating that results do not rely on information reuse across experimental partitions. The split-before-fitting analysis used n = 126 experimental samples (n = 101 training and n = 25 validation). Training and validation data are identified within the panels. Source data are provided as a Source Data file.

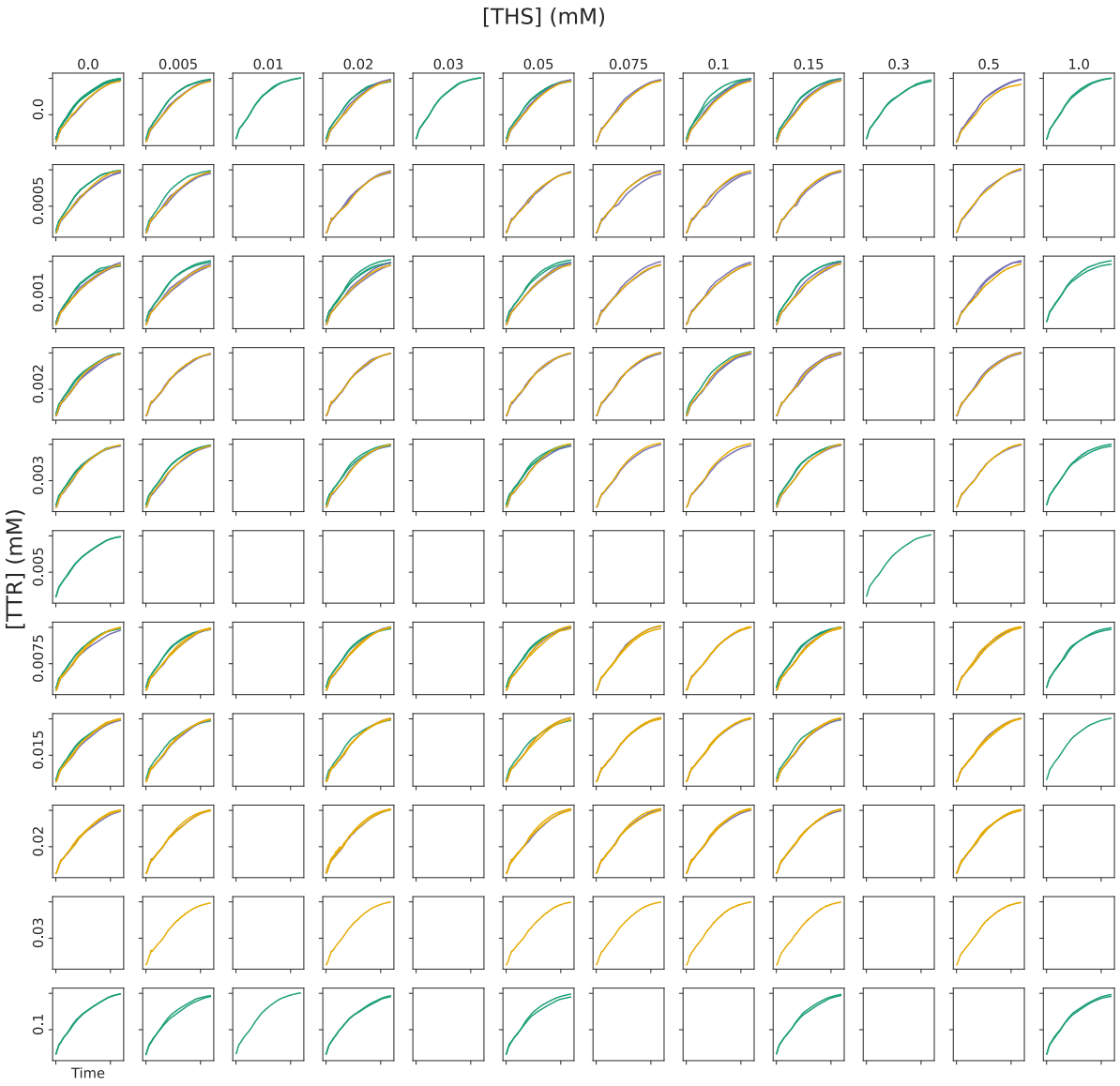


Supplementary Figure 8. Probing of the VAE latent space

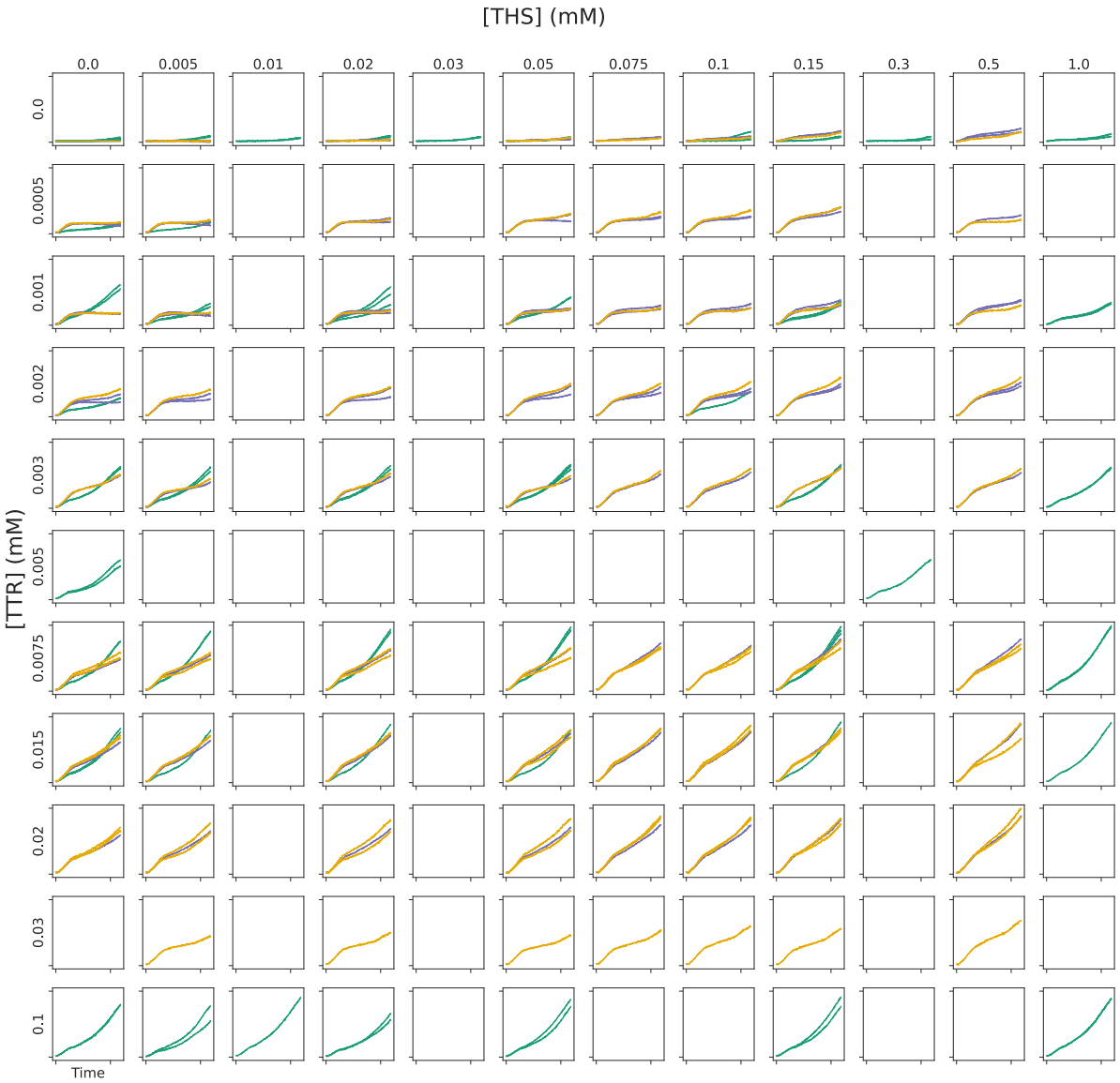
(a) Correlation between individual latents (z0–z9) and experimental input signals (aTc and IPTG), shown separately for training (left) and validation (right) datasets. Each latent contributes to the prediction of both inputs, with similar correlation patterns across train and validation, indicating a stable, distributed encoding rather than one-to-one correspondence between latents and inputs.

(b) Two-dimensional PCA projection of the latent space, computed from the training set and applied to the validation set, with points colored by the true input levels of aTc (top) or IPTG (bottom). Smooth gradients and consistent geometric organization across train and validation demonstrate that the latent space captures systematic variations in experimental conditions rather than noise.

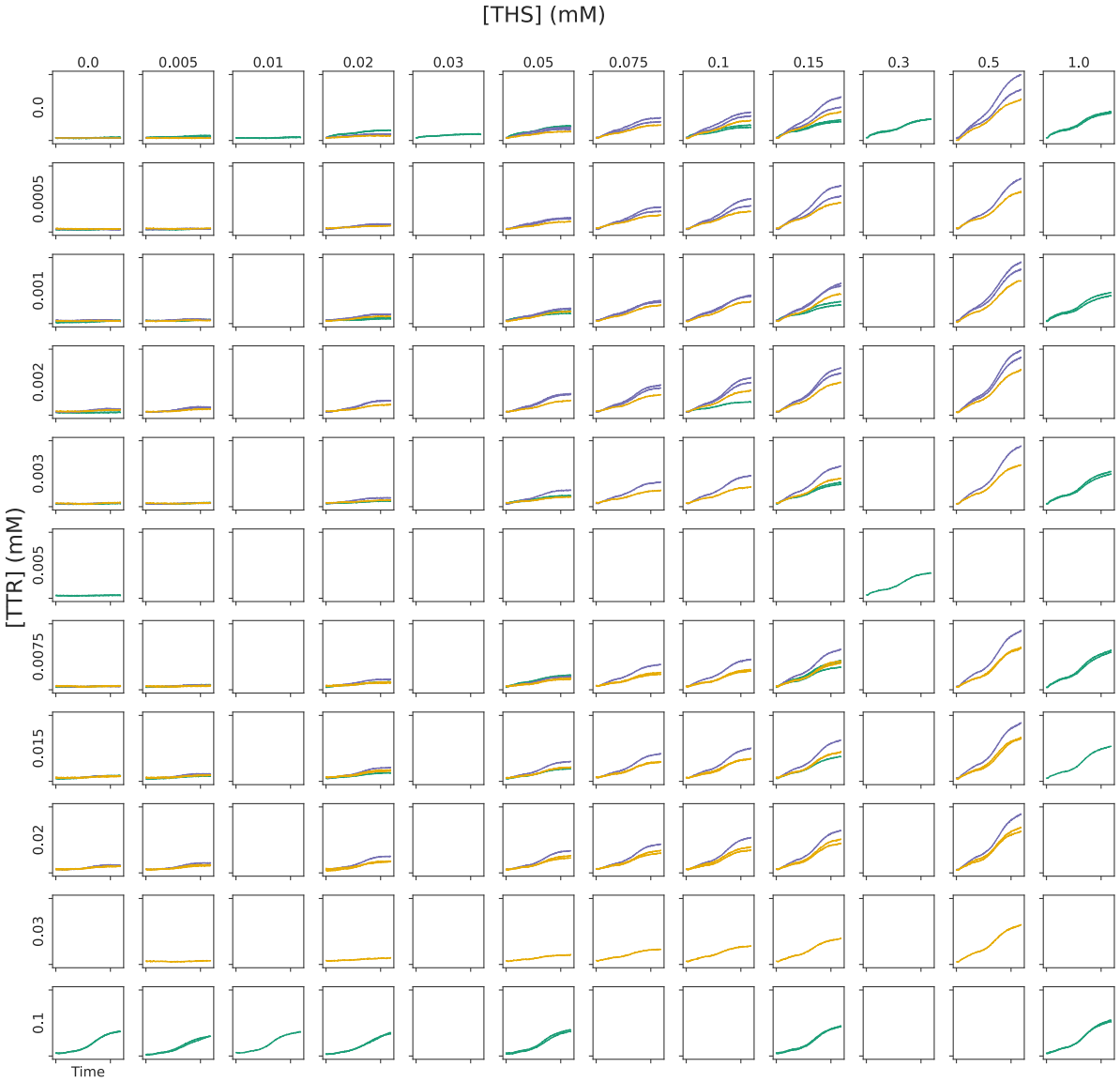
(c) Sensitivity of each signal prediction on each latent: individual latents are varied across their empirical range (±2 standard deviations) while all other dimensions are held fixed; the trained MLP is used to predict the corresponding input signals. Different latents exhibit distinct and sometimes nonlinear influences on aTc and IPTG, revealing how the learned representation functionally encodes input-relevant features of the timecourse trajectories. Training and validation data and the plotted color scales are identified within the panels. Source data are provided as a Source Data file.



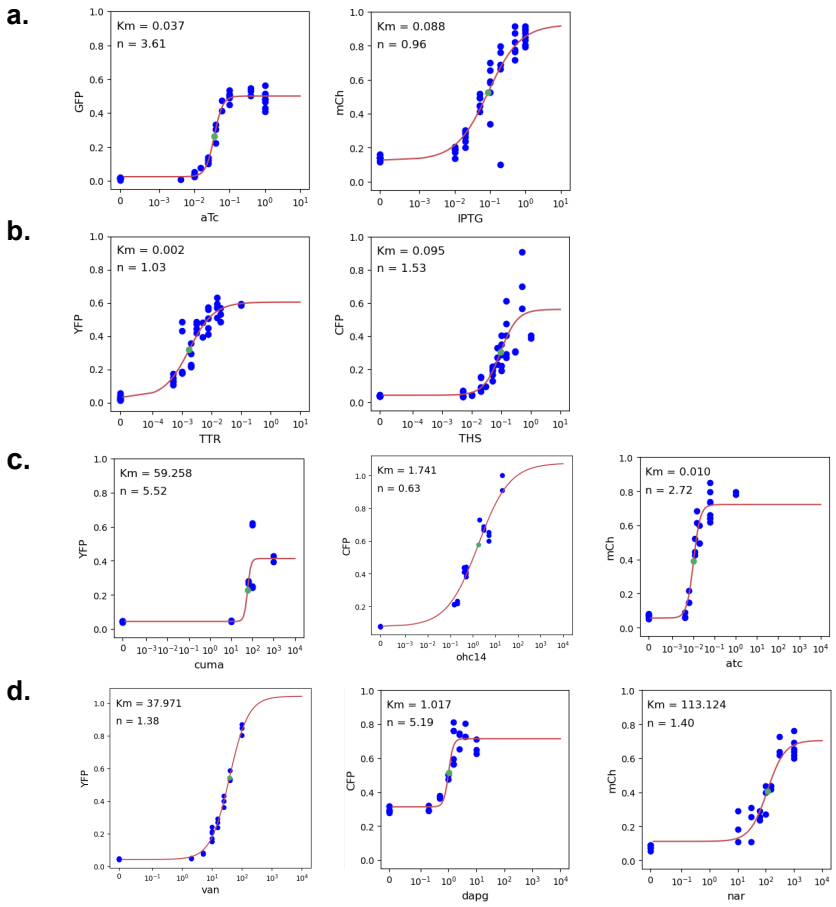
Supplementary Figure 9. OD time courses for all conditions from the high crosstalk TTR-THS sensor set after data processing, which were used as inputs into the VAE-MLP as well as for fitting the kinetic model parameters to. Rows and columns indicate TTR and THS concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



Supplementary Figure 10. YFP time courses for all conditions from the TTR-THS sensor set after data processing, which were used as inputs into the VAE-MLP as well as for fitting the kinetic model parameters to. Rows and columns indicate TTR and THS concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



Supplementary Figure 11. CFP time courses for all conditions from the TTR-THS sensor set after data processing, which were used as inputs into the VAE-MLP as well as for fitting the kinetic model parameters to. Rows and columns indicate TTR and THS concentrations, respectively, as labeled in the figure; colored traces show replicate time courses. Source data are provided as a Source Data file.



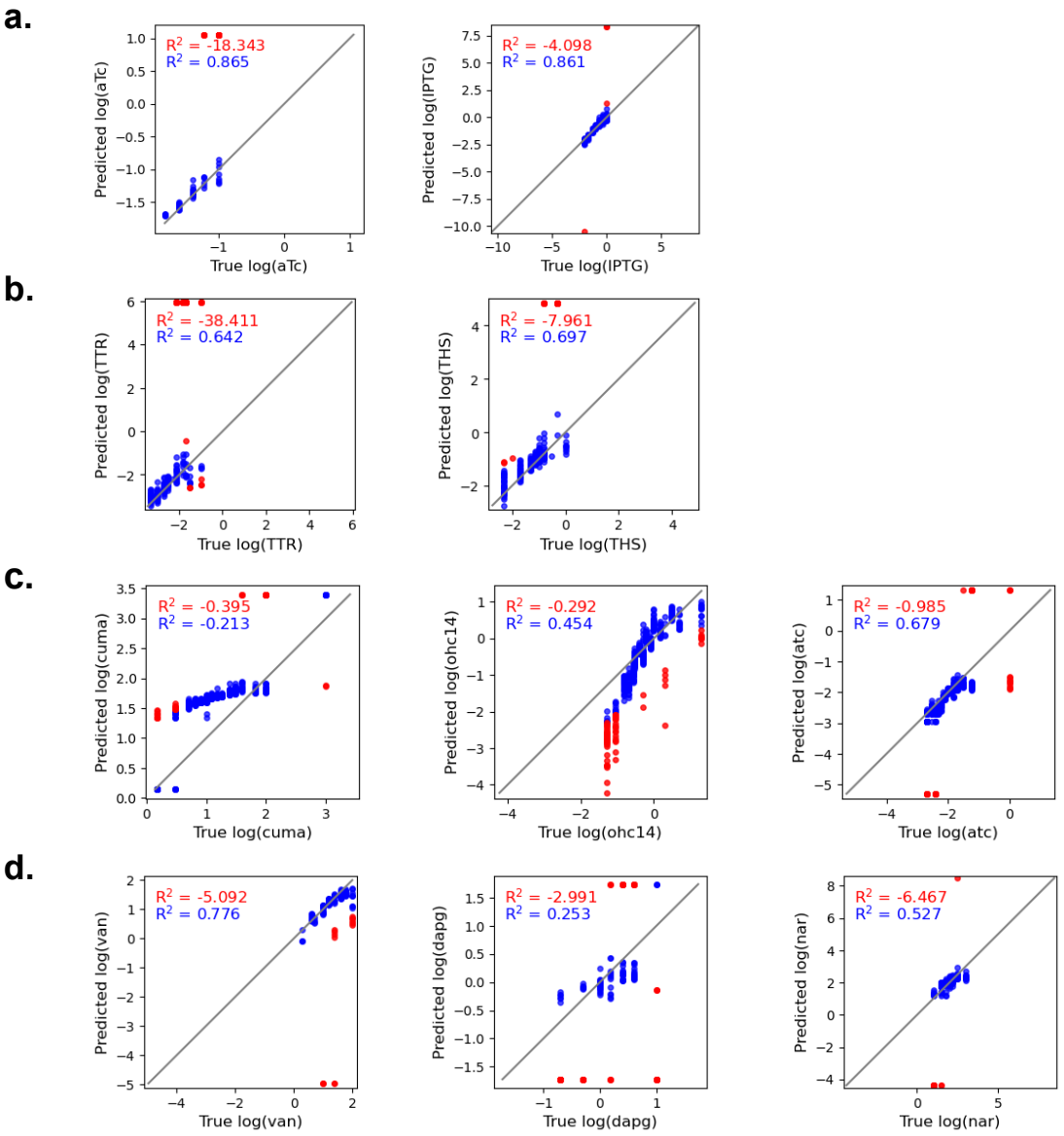
Supplementary Figure 12. Fitted dose responses for various sensors. In each case, data (at a fixed timepoint of 20 hours) were fitted to a Hill equation, with estimated half-maximum activation concentration (K_m) and Hill coefficient (n) being listed in the panel. For compactness, here we show the responses to their intended signals only.

(a) Dose response data for aTc-GFP and IPTG-mCherry dataset, corresponding to Figure 2 in the main text. These two panels are reproduced from Supplementary Figure 4 for easy comparison.

(b) Dose response data for TTR-YFP and THS-CFP dataset, corresponding to Figure 3 in the main text.

(c) Dose response data for cuma-YFP, ohc14-CFP, and aTc-mCherry dataset, corresponding to Figure 4a,b in the main text.

(d) Dose response data for van-YFP, dapg-CFP, and nar-mCherry dataset, corresponding to Figure 4c,d in the main text. Blue points are experimental measurements, and red curves are Hill-function fits. Source data are provided as a Source Data file.



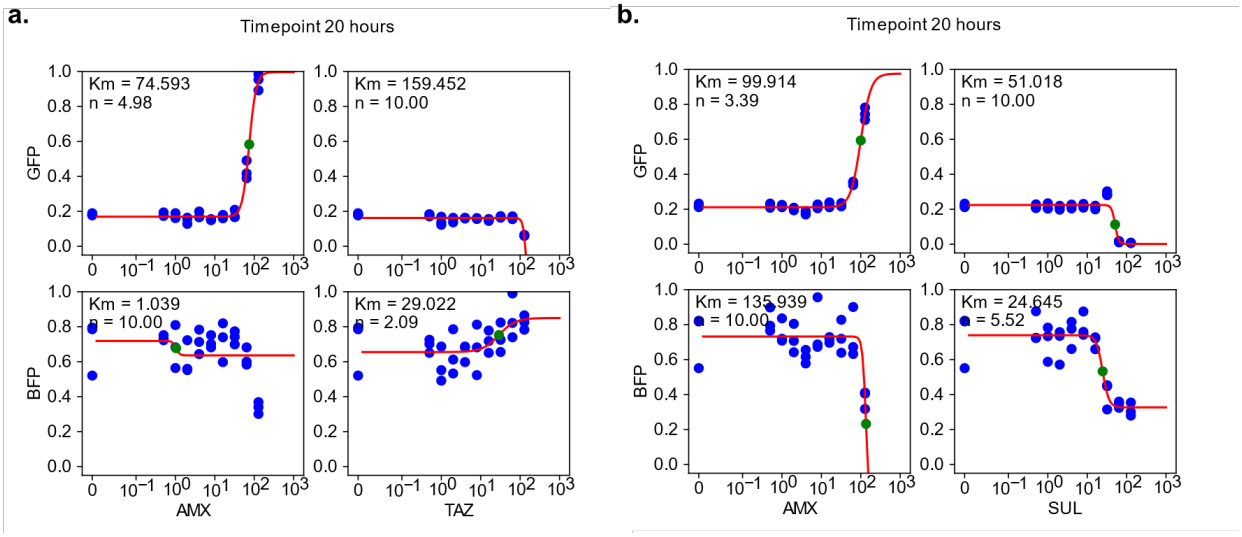
Supplementary Figure 13. Input chemical concentrations predicted from single-timepoint Hill function fitting. A fixed timepoint of 20 hours was used for the entire analysis, as dose-response curves showed maximal dynamic range at this timepoint. First, nonlinear optimization was used to fit the dose-response data to the Hill function. Using the fitted equation, input chemical concentrations were then back-calculated from corresponding fluorescence values. After back-calculations, predicted concentrations were plotted against the true values. In all plots, the R^2 values for the entire dataset (includes red points) are shown in red and the R^2 values for the dataset excluding edge cases (blue points only) are shown in blue. Points were classified as edge cases if the absolute value of the difference between the predicted and true value was at least 10-fold. We can compare the red R^2 scores from the single-timepoint fitting to the original pipeline performance, as they both correspond to the same filtered datasets. We find that across all datasets, the machine learning pipeline shows higher performance in predicting input chemical concentrations.

a. Predictions of aTc and IPTG concentrations from transfer function inverse calculation, compared to true values. To compare the VAE-MLP pipeline predictive performance on the same dataset, see Figure 2f in the main text.

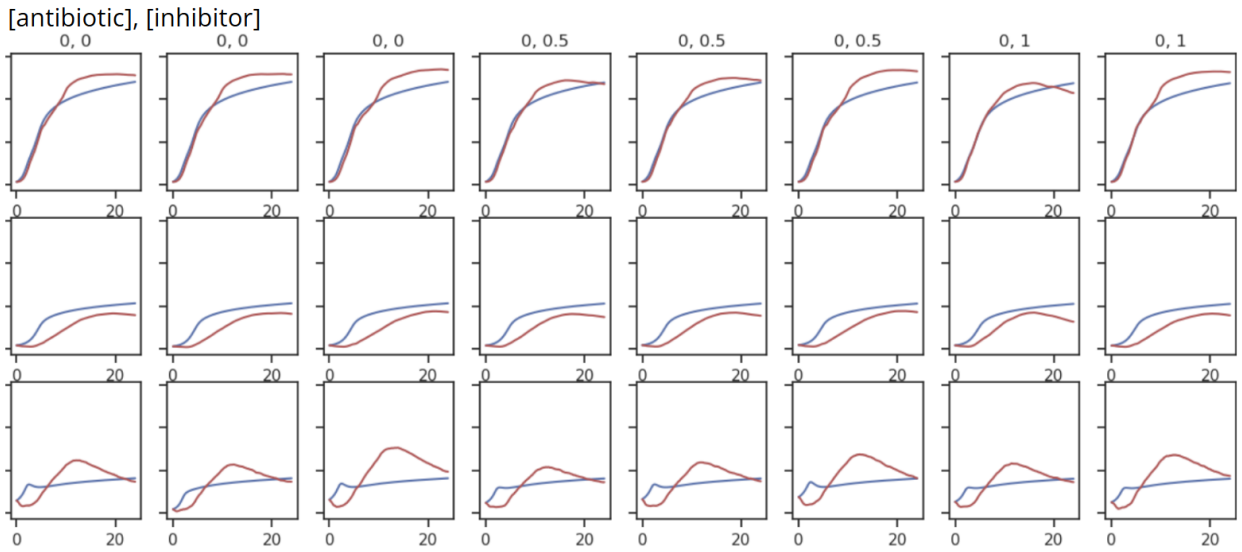
b. Predictions of TTR and THS concentrations from transfer function inverse calculation, compared to true values. To compare the VAE-MLP pipeline predictive performance on the same dataset, see Figure 3d in the main text.

c. Predictions of cuma, ohc14, and aTc concentrations from transfer function inverse calculation, compared to true values. To compare the VAE-MLP pipeline predictive performance on the same dataset, see Figure 4b in the main text.

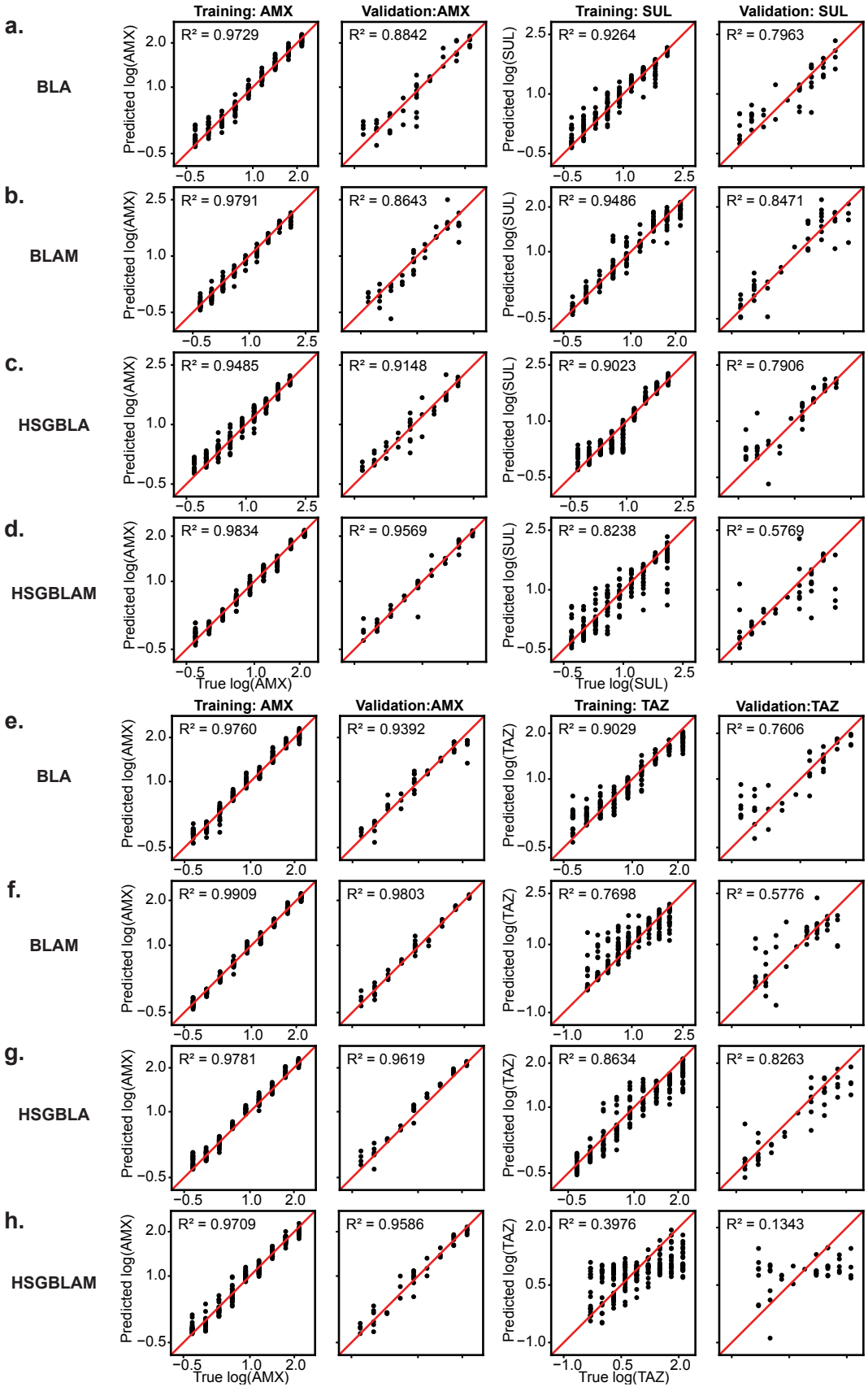
d. Predictions of van, dapg, and nar concentrations from transfer function inverse calculation, compared to true values. To compare the VAE-MLP pipeline predictive performance on the same dataset, see Figure 4d in the main text. Source data are provided as a Source Data file.



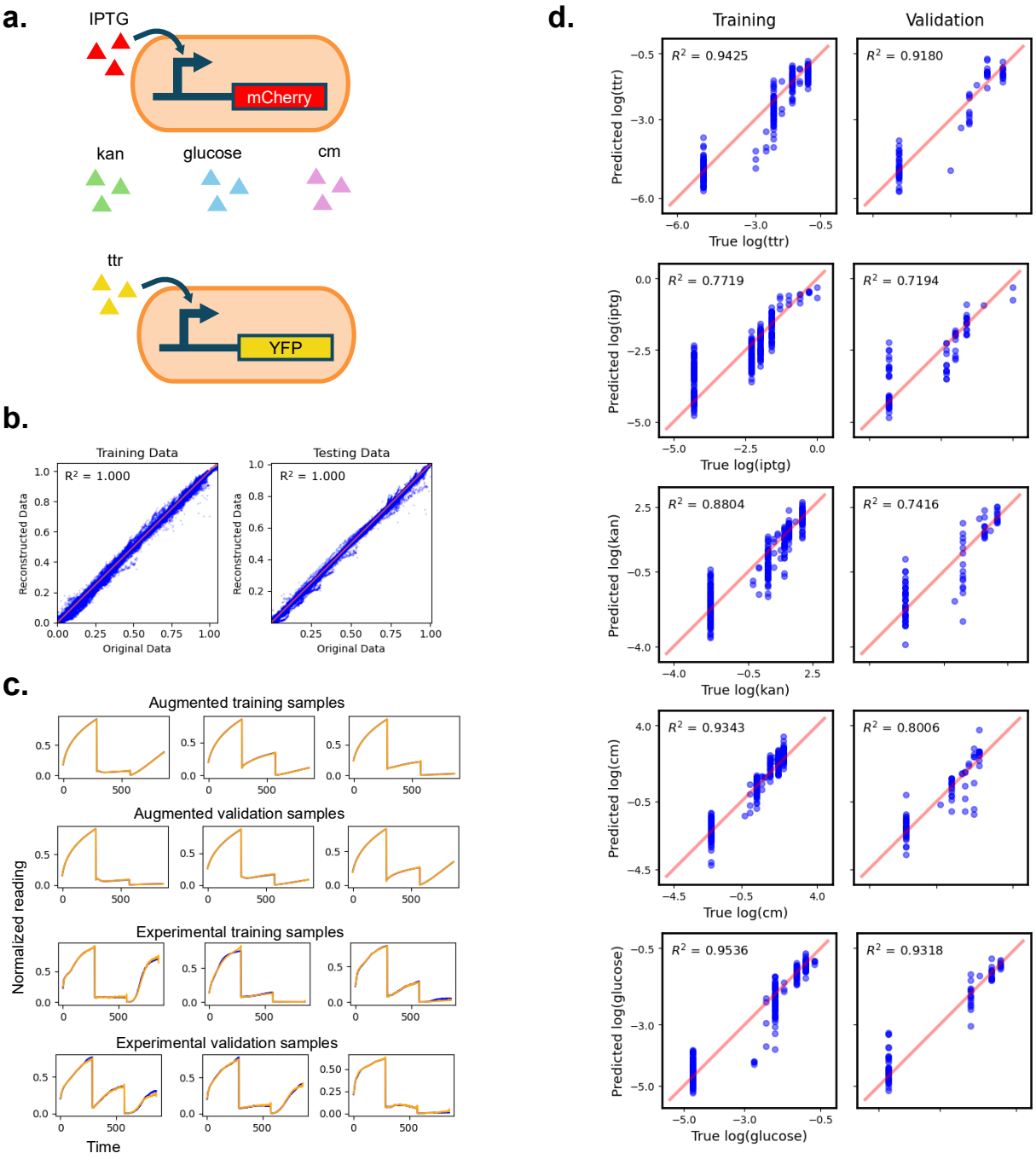
Supplementary Figure 14. Fluorescence dose responses of the antibiotic-treated microbial community when either antibiotic or inhibitor was added. Color bar and labeled value represent the fluorescence fold-change with respect to the fluorescence when neither antibiotic or inhibitor were added. For the corresponding crosstalk statistics, n = 3 replicate wells per condition (Supplementary Table 3). Source data are provided as a Source Data file.



Supplementary Figure 15. Sampling of time course fitting results for antibiotic treatment data. Blue line is simulated data and red line is experimental data. OD (top row), GFP (middle row), and BFP (bottom row) are plotted. Source data are provided as a Source Data file.


a.
BLA
b.
BLAM
c.
HSGBLA
d.
HSGBLAM
e.
BLA
f.
BLAM
g.
HSGBLA
h.
HSGBLAM
Training: AMX
Validation:AMX
Training: SUL
Validation: SUL
Training: TAZ
Validation:TAZ
Predicted log(AMX)
Predicted log(SUL)
Predicted log(TAZ)
True log(AMX)
True log(SUL)
True log(TAZ)
R² = 0.9729
R² = 0.8842
R² = 0.9264
R² = 0.7963
R² = 0.9791
R² = 0.8643
R² = 0.9486
R² = 0.8471
R² = 0.9485
R² = 0.9148
R² = 0.9023
R² = 0.7906
R² = 0.9834
R² = 0.9569
R² = 0.8238
R² = 0.5769
R² = 0.9760
R² = 0.9392
R² = 0.9029
R² = 0.7606
R² = 0.9909
R² = 0.9803
R² = 0.7698
R² = 0.5776
R² = 0.9781
R² = 0.9619
R² = 0.8634
R² = 0.8263
R² = 0.9709
R² = 0.9586
R² = 0.3976
R² = 0.1343

Supplementary Figure 16. VAE-MLP results of experimental antibiotic and inhibitor concentration predictions for all antibiotic/inhibitor datasets. a-d. Predictions of the amoxicillin and sulbactam concentrations that were used to treat communities with different plasmids conferring antibiotic resistance in the resistant strain. **e-h.** Predictions of the amoxicillin and tazobactam concentrations that were used to treat communities with different plasmids conferring antibiotic resistance in the resistant strain. Resistance plasmids were Bla (**a,e**), BlaM (**b,f**), HSGBla (**c,g**), and HSGBlaM (**d,h**). Training and validation samples are distinguished as labeled in each panel; diagonal reference lines indicate y = x. Source data are provided as a Source Data file.



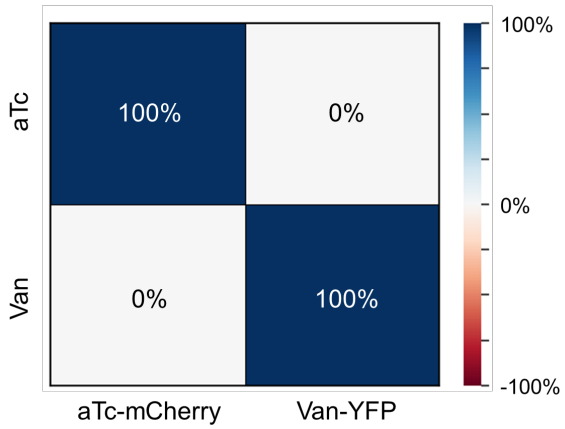
Supplementary Figure 17. Pipeline predicts concentrations of five input chemicals in a two-strain community with two fluorescence readouts.

a. A community of two sensor strains, MG1655 IPTG-mCherry and MG1655 ttr-YFP, was used to detect the concentrations of five input chemicals: tetrathionate (ttr), IPTG, kanamycin (kan), chloramphenicol (cm), and glucose. Temporal growth (OD), mCherry, and YFP readouts were measured for various combinations of input chemical concentrations. 176 unique input combinations were tested, yielding 453 time-course samples in total.

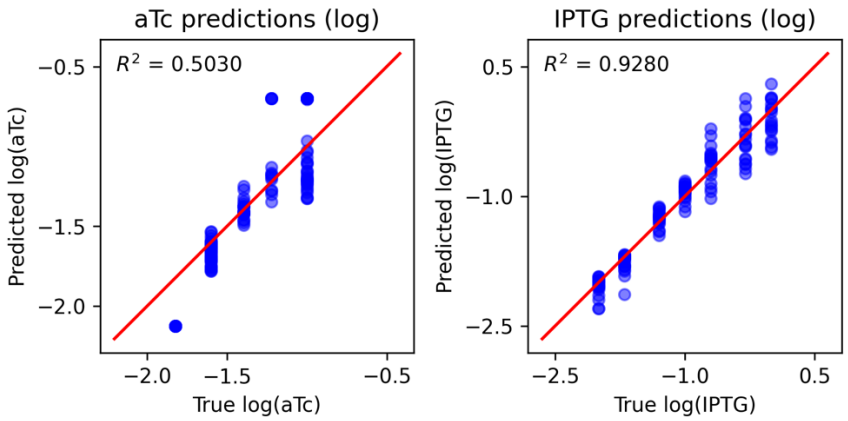
b. VAE reconstructed vs. true time course data with R^2. Using the latent dimensions the VAE generated from time courses, the VAE decoder accurately reconstructs the features of the full time-course data.

c. Examples of VAE reconstructed time courses (orange) and their respective true time courses (blue). The VAE reconstructs the shape of the time courses for various training and validation datasets with high fidelity.

d. MLP predictions of ttr, IPTG, kan, cm, and glucose concentrations vs. true concentrations for experimental training and validation datasets. Training and validation samples are distinguished as labeled in each panel; diagonal reference lines in the prediction panels indicate y = x. Source data are provided as a Source Data file.



Supplementary Figure 18. Crosstalk heatmap for aTc-mCherry and Van-YFP sensor community responses to added Van and aTc in hospital sink wastewater. Under the tested condition, neither sensor showed measurable activation by the noncognate input, but the complex background environment of the sink wastewater can still introduce crosstalk. See Supplementary Table 3 for details of statistical analysis. The color scale reports the normalized crosstalk coefficient α. Crosstalk statistics used n = 2 replicate wells per condition (Supplementary Table 3). Source data are provided as a Source Data file.



Supplementary Figure 19. IPTG and aTc concentration predictions using non-linear optimization with the kinetic ODE model. Using the optimized kinetic model parameters for the IPTG-mCherry and aTc-GFP experimental dataset, we used a least squares optimization to estimate the IPTG and aTc concentrations of each sample. To fit the equations to the experimental curves, we only optimized IPTG and aTc concentrations to minimize residuals, keeping the parameters fixed. Initial guesses and bounds for both chemicals were the K value of the Hill equation and [0, 15] respectively. The aTc predictions were significantly lower fidelity than the predictions from the MLP, particularly for the lowest concentration. The kinetic model also predicted IPTG concentrations with less accuracy than the MLP. R^2 values were calculated from the $\log_{10}$ transformed input values. Red line is Y=x for reference. Blue points represent individual samples. Source data are provided as a Source Data file.