

Engineered bacterial swarm patterns as spatial records of environmental inputs

Received: 17 May 2022

Accepted: 6 April 2023

Published online: 04 May 2023

 Check for updates

Anjali Doshi¹, Marian Shaw¹, Ruxandra Tonea¹, Soonhee Moon¹,
Rosalia Minyety¹, Anish Doshi², Andrew Laine¹, Jia Guo^{3,4} & Tal Danino^{1,5,6}✉

A diverse array of bacteria species naturally self-organize into durable macroscale patterns on solid surfaces via swarming motility—a highly coordinated and rapid movement of bacteria powered by flagella. Engineering swarming is an untapped opportunity to increase the scale and robustness of coordinated synthetic microbial systems. Here we engineer *Proteus mirabilis*, which natively forms centimeter-scale bullseye swarm patterns, to ‘write’ external inputs into visible spatial records. Specifically, we engineer tunable expression of swarming-related genes that modify pattern features, and we develop quantitative approaches to decoding. Next, we develop a dual-input system that modulates two swarming-related genes simultaneously, and we separately show that growing colonies can record dynamic environmental changes. We decode the resulting multicondition patterns with deep classification and segmentation models. Finally, we engineer a strain that records the presence of aqueous copper. This work creates an approach for building macroscale bacterial recorders, expanding the framework for engineering emergent microbial behaviors.

Swarming behaviors are ubiquitously found in natural systems, ranging from bird flocks to microbial communities, and have inspired the creation of artificial systems such as robot swarms^{1–3}. A collective movement stemming from individual interactions, swarming can greatly increase a community’s scale as well as robustness to noisy individuals and environments. The swarming of many microbial species creates complex emergent patterns at the centimeter scale on solid surfaces^{3–7}. Although a long-standing goal of synthetic biology has been to program self-organization in such a fashion, swarming motility has yet to be engineered or used for biotechnological applications^{8–10}. Previous approaches have focused on prototypical microbes such as *Escherichia coli*, which forms homogenous colonies, and have engineered swimming and quorum-sensing systems in liquid-agar environments, or used external prepatterning to generate coordinated behavior^{11–14}. One promising application of engineering natural swarming is the creation of a durable spatial recording system, using the sensing capabilities of millions of individual bacteria within a swarm to visibly

‘write’ information onto a solid surface. Thus far, synthetic cellular information recording efforts have achieved the recording of multiple inputs, cellular lineage and transient signals, primarily within DNA, but rely on sequencing and other technologies for decoding, or are in silico^{15–19}.

We focused on engineering the unique swarming of *Proteus mirabilis*—a commensal gut bacterium also commonly found in soil and water, which produces a bullseye pattern on solid agar defined by concentric rings of high bacteria density that are visible to the naked eye²⁰ (Fig. 1a). The inherent clock-like timing and internal consistency of the colony ring formation naturally suggest application as a recording system, similar to the way a growing tree records information in the rings in its trunk²¹. Although the ability of *P. mirabilis* to produce rings has been known for over 100 years, it has not been developed as a synthetic biology platform and quantification of its macroscale patterns has been limited²². Beyond the large-scale features of *P. mirabilis* that enable simple decoding visually, applying methods of deep learning and

¹Department of Biomedical Engineering, Columbia University, New York City, NY, USA. ²Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. ³Department of Psychiatry, Columbia University, New York City, NY, USA. ⁴Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, NY, USA. ⁵Herbert Irving Comprehensive Cancer Center, Columbia University, New York City, NY, USA. ⁶Data Science Institute, Columbia University, New York City, NY, USA. ✉e-mail: tal.danino@columbia.edu

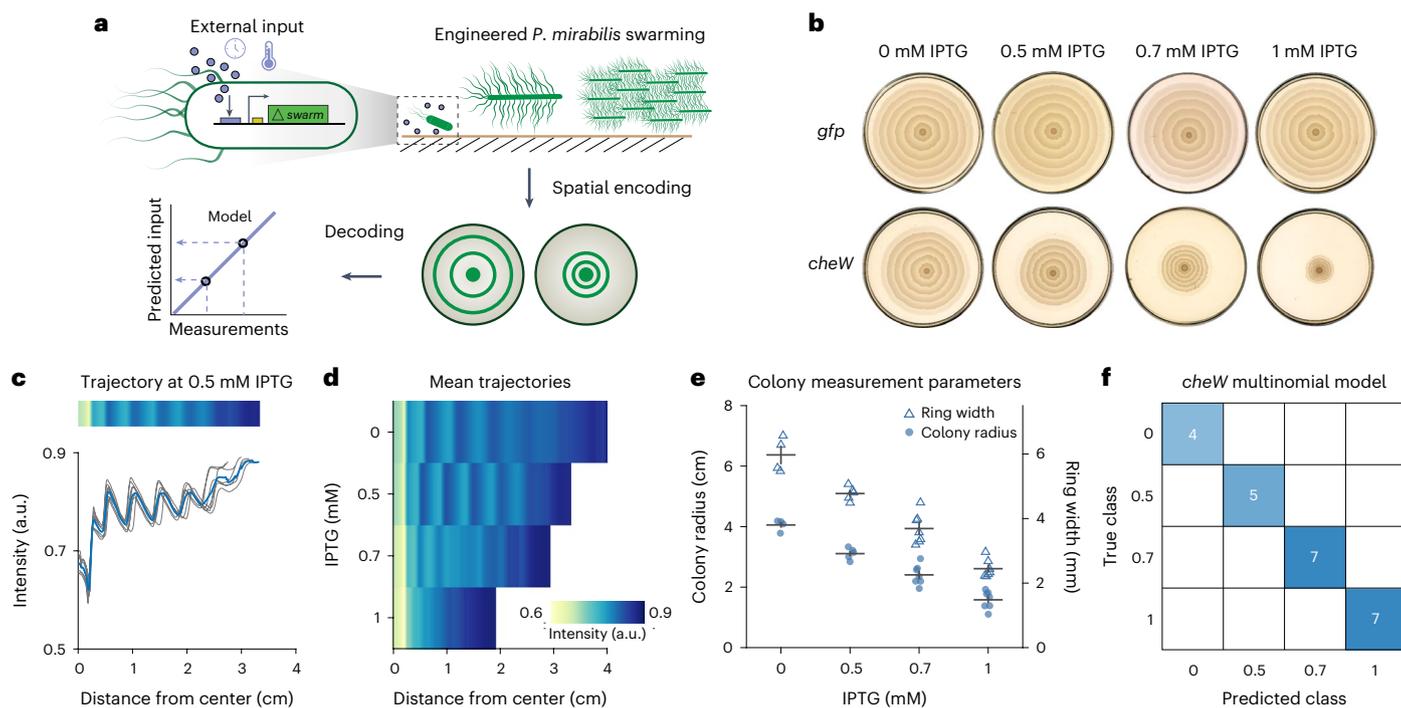


Fig. 1 | Engineered *P. mirabilis* swarm patterns as spatiotemporal records.

a, Wild-type (WT) *P. mirabilis* cells undergo oscillatory swarming on solid agar to grow into a characteristic bullseye colony via elongation, hyperflagellation and raft formation. *P. mirabilis* is engineered with an externally inducible genetic circuit driving swarming-related genes to modify the macroscale pattern output, which can then be decoded using quantitative methods to predict the input conditions. **b**, Representative images of colony patterns formed by a strain containing a control circuit with green fluorescent protein (*gfp*) (top) compared to one containing a circuit with the chemotaxis gene *cheW* (bottom), grown for 24 h on agar supplemented with various IPTG concentrations. **c**, The *cheW* colony pattern is distilled into radially averaged pixel intensity profiles, with distinct peaks matching low-density ring boundaries when plotted as a heatmap

or line plot. The blue line denotes the mean profile of the individual plates (each gray line represents one plate). **d**, Heatmaps of average *cheW* profiles at varying IPTG concentration ($n = 4, 5, 7$ and 7 plates at IPTG = 0, 0.5, 0.7 and 1 mM, respectively). Colormap is on the same scale for **c** and **d**. **e**, Radii of the colonies plotted by IPTG concentration after 24 h (filled circles) and calculated ring width (empty triangles), derived from Fourier analysis of the radially averaged profiles of individual images. The mean and standard error of the mean (s.e.m) are shown in black. $n = 4, 5, 7$ and 7 plates for IPTG = 0, 0.5, 0.7 and 1, respectively. **f**, A multinomial model was fit to the measurements in **e**, with predicted IPTG concentration as the output variable. The model's predictions for each plate shown in **e** are shown as a confusion matrix. Color reflects n per square (same as listed in **d**; white squares represent 0).

image segmentation could further decode multiple external inputs and potentially dynamic conditions from more complex pattern features.

The bullseye pattern of *P. mirabilis* is created from a sequence of phases starting with initial colony growth (lag), followed by oscillatory cycles of synchronized colony expansion (swarming) and stationary periods of cell division (consolidation). The synchronicity of its swarming is achieved by a complex coordination of cell elongation, secretion of surfactant to aid movement, intercellular communication and alignment of swarmer cells into rafts by intercellular bundling of overexpressed flagella^{23,24}. While the investigation of the mechanisms governing these behaviors is ongoing, studies have identified an array of genes upregulated during consolidation phases, including those responsible for the synthesis of flagella, metabolism and cell division, and during swarming phases, such as the master regulator *flhDC*^{24–27}. These works have shown that modification of expression of these genes and choice of growth conditions can lead to different variations of the ring pattern^{28,29}. In this work, we engineer *P. mirabilis* by tuning the expression of a set of swarming genes so that the presence of various inputs is encoded in the output ring patterns of the engineered strains, which can then be decoded by analysis of the resulting pattern images.

Results

Preliminary engineering of *P. mirabilis* patterns

We chose the strain PM7002 with baseline conditions that would create a pattern with distinct ring boundaries, such that modifications

to the pattern would be easily visible and quantifiable (Extended Data Figs. 1 and 2). After establishing these conditions, we expressed swarming-related genes to controllably modify specific colony pattern features, which could be subsequently analyzed and decoded to report on conditions during colony growth (Fig. 1a).

To initially demonstrate swarm pattern modulation, we engineered *P. mirabilis* with a high-copy plasmid carrying an isopropyl β -D-1-thiogalactopyranoside (IPTG) inducible promoter, pLac, expressing *cheW*, a chemotaxis-related gene upregulated in the swarming process^{26,30}. In *E. coli* chemotaxis, CheW is a membrane-bound coupling protein part of a signaling complex in which it bridges the kinase CheA to chemoreceptors, allowing phosphotransfer to CheY and CheB, where CheY is involved in the control of flagellar motor rotation³¹. Although the exact role of *cheW* in swarming is not fully known, a *cheW* mutant of *P. mirabilis* was previously found to be unable to swarm³⁰. Here increasing concentrations of IPTG in the agar were used to induce constitutive *cheW* expression, which resulted in colonies with decreasing ring width and size at 24 h, as compared to a control green fluorescent protein (*gfp*)-expressing strain, which showed no change in pattern in response to IPTG (Fig. 1b). To quantify the patterns, we examined the radially averaged pixel intensity as a proxy for colony density in each of the conditions, where high pixel intensity (lighter regions in the colony) represents lower density (Fig. 1c and Extended Data Fig. 3). All colonies had a characteristically dense boundary around the central inoculum, seen as a dip in the intensity plot around 0.25 cm from the

center of the colony ($x = 0$ in the plot), and showed periodic changes in density across the colony. As expected, the radially averaged intensity profiles showed peaks of intensity corresponding to the periodic ring boundaries. With increasing *cheW* expression, the profiles showed greater density near the inoculum and at the ring boundaries, which can be seen as lighter areas on heatmaps of radially averaged intensities (Fig. 1d). We constructed a small dataset and measured colony radius manually using image processing tools and ring widths using a custom algorithm (Supplementary Note on Computational Methods and Fig. 1e). The colony radius and ring width correlated well with IPTG concentrations ($R^2 = 0.90$ for each). To more accurately decode input conditions from the pattern, we fit a multinomial regression model on these measurements and found that the model correctly predicted each colony's input IPTG from the combination of its radius and ring width in all cases (Fig. 1f). Therefore, we reasoned that colony features could potentially encode information about external inputs received by the bacteria and feature measurements could subsequently be used to decode the information.

Manipulation of multiple swarming-pathway genes

Given the variety of *P. mirabilis* pattern features observed in the literature beyond ring widths and overall colony radii, we explored the potential for multiplexed information encoding. Here we sought to identify additional genes that could distinctly modulate colony pattern features (Fig. 2a). We chose genes that were previously implicated in a range of points in the swarming process, including *umoD*, which controls the master regulator of swarming, *flhDC*; the signaling factors *fliA* and *flgM*, which are involved in flagellar gene transcription and *lrp*, which affects general cellular processes in response to the presence of leucine^{26,32–36}. Induced expression of these genes via IPTG generated a variety of patterns, ranging from dense ruffled textures, to 'spikes,' to indistinct ring boundaries (Fig. 2b). Scanning a range of IPTG concentrations showed graded changes in patterns (Extended Data Fig. 4). For example, with minimal IPTG-induced expression, the *lrp* strain formed spikes in the inner colony rings, and at maximal induction, each ring boundary was spiky. Increased expression of *flgM* caused the colony radius at 24 h to shrink, while *fliA* caused the formation of visible light-colored dots just within the boundaries of each ring. As *umoD* expression increased, colonies became more symmetric and ring boundaries and the inoculum edge became fainter. Expression of another flagellar gene, *flgA*, resulted in no visible effect on the pattern, confirming that the graded changes seen were due to the specific swarm gene overexpressed in each of the above strains. Taken together, these various qualitative characteristics suggested that the induced expression of certain swarm genes could indeed affect several pattern features that could be measured and quantified.

We next examined the radially averaged profiles of each pattern, which revealed distinct characteristics for each strain (Fig. 2c). For example, overall colony density was higher with induced *cheW* expression than with *umoD*. The spikes visible in the *lrp* pattern, which caused ring boundaries to spread over greater widths, reduced the sharpness of the ring boundaries in the radially averaged profiles. Given the repeating nature of features in the patterns, we also explored visualization of the Fourier spectra of the polar transforms of the images, which highlight the presence of frequency information, to see if the spectra varied between strains (Fig. 2d). The periodic features in the patterns resulted in high visible intensities in certain regions of the Fourier transform images. For example, the *umoD* strain displayed a higher magnitude in the outer regions of the transformed images, representing high-frequency (that is, short distances between repeated features) information, while the other strains showed greater magnitude in the central regions, which represents lower-frequency features. In summary, we saw that the visible differences in patterns between the engineered strains were reflected in, and thus could be analyzed from, quantitative representations of the images.

We then sought to identify features of each strain's pattern that could allow for the determination of the input IPTG concentration. We generated a dataset of images for each strain grown at a range of inducer concentrations and measured a range of features for each (Extended Data Fig. 5a). The low frequencies of the Fourier spectra were found to increase with IPTG induction for the *flgM* and *cheW* strains, reflecting the visual observation of thinner, fewer rings of increased density at higher IPTG (Fig. 2e and Extended Data Fig. 5a). A second measure, the local coefficient of variation (CV), increased with increasing IPTG for the *lrp* strain, which could be observed visually in the spiked rings (Fig. 2f). Finally, the distinctness of the inoculum border, measured by the change in intensity over the border, decreased with increasing IPTG for the *umoD* strain, particularly from 0.1 mM to 1 mM IPTG (Fig. 2g). These measurements showed that induced expression of these genes could quantifiably affect the pattern in response to changes in IPTG.

As an approach for decoding information from the patterns, we explored fitting regression models on these measurements. The samples were binned into three classes (0–1, 1–5 and 5–10 mM IPTG), and then each feature individually, and all possible combinations of the measured features, were used to fit multinomial regression models, to identify which combination would best decode a given strain's pattern. The performance of such models can be evaluated using a multiclass area under the receiver-operating curve (ROC) measurement, known as the AUC metric, where the more accurately a model predicts true positives compared to false positives for each class, the closer the AUC will be to 1. The AUC of each strain's fitted model was evaluated on the input data (Fig. 2h and Supplementary Table 1). For each strain, the combination of parameters that gave the highest AUC varied, confirming that each strain was encoding information in a characteristic combination of pattern features. The best models for the experimental strains with *cheW*, *umoD*, *fliA* and *lrp* generally had AUC > 0.9, showing how well the models were able to differentiate true positives in each IPTG class from false positives. The AUCs were 0.6 for the *gfp* control strain, just slightly above a random classifier (AUC = 0.5), suggesting that pattern parameters were not strongly affected by increasing IPTG for control strains. The confusion matrices showed that the fitted models correctly classified a majority of the plates for each strain (Extended Data Fig. 5b). Thus, information about the environment encoded within the engineered strains' patterns can be decoded using combinations of relevant pattern features.

Dynamics of engineered *P. mirabilis* strains

P. mirabilis swarming creates patterns not only in space but also in time; this temporal regularity suggests the possibility of encoding information in both the endpoint patterns and their dynamic growth phases. We aimed to gain an understanding of the dynamics of the engineered strains by time-lapse imaging of colony growth (Fig. 3a). To capture high-resolution images of swarming, we developed a time-lapse setup using a commercial flatbed scanner. For each strain, a time-lapse was captured with maximal IPTG concentration at 25 °C; images were taken every 10 min over the course of the time-lapse (Fig. 3a and Supplementary Video). The individual images were then radially averaged, and full time-lapses were visualized via heatmaps (Fig. 3b). Using a custom semi-automated algorithm (Supplementary Note on Computational Methods), we identified the location of the colony front at each timepoint and obtained trajectories with high spatiotemporal resolution (Fig. 3c). The colony growth trajectories showed that each of the engineered strains maintained the classic alternation in phases, but with changes in aspects such as initial lag time and length of the phases compared to the control *gfp* strain. We then measured the mean length of time of each phase from each of these trajectories (Fig. 3d), which, together with the distance swarmed during each swarm phase, enabled the calculation of swarm speed (Fig. 3e).

To explore whether certain dynamic parameters would show a trend with increasing IPTG for each strain, we generated individual

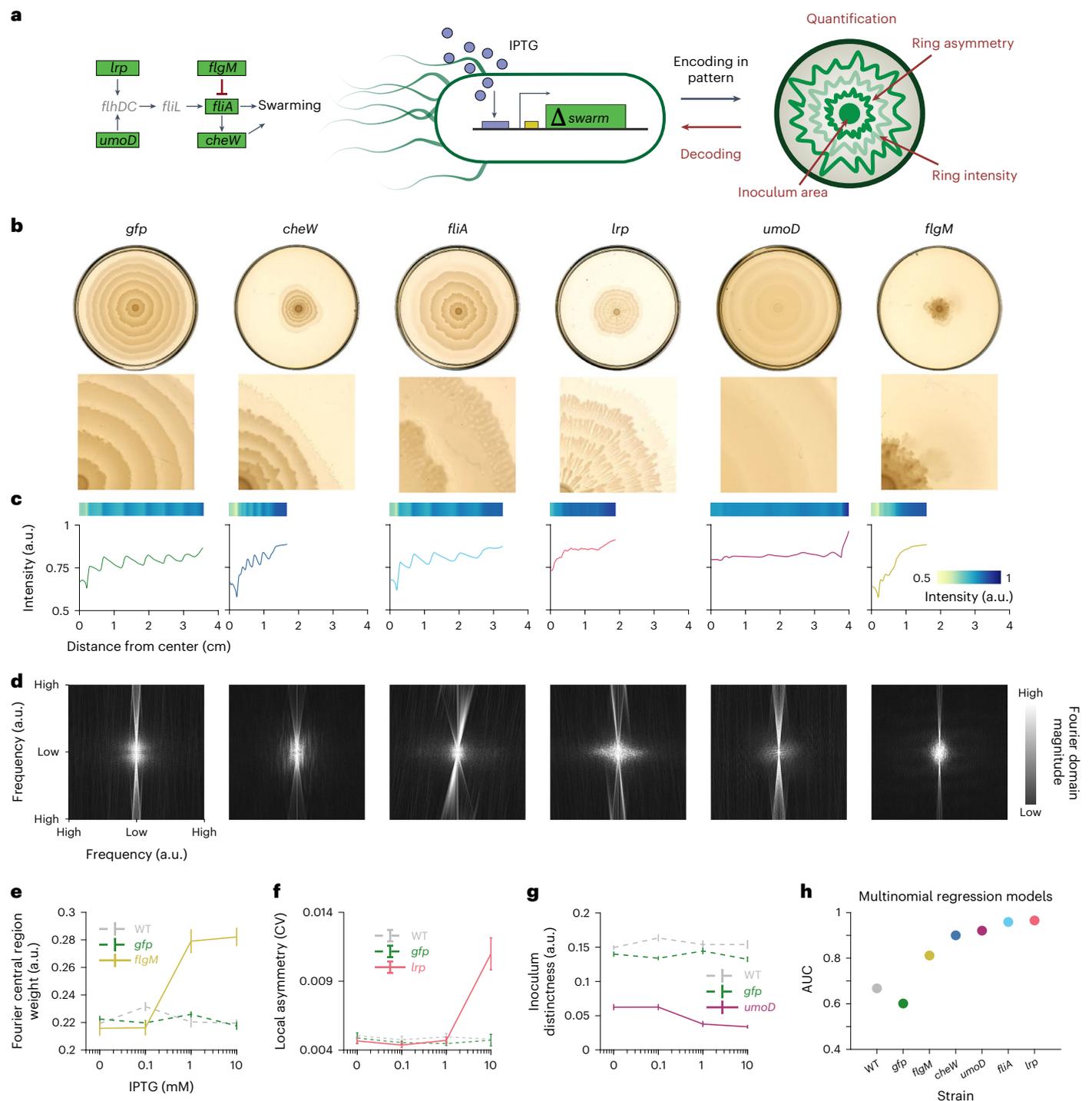


Fig. 2 | Modulation of swarm genes in engineered *P. mirabilis* results in quantifiable changes to distinct spatial pattern features. **a**, Candidate genes involved in the *P. mirabilis* swarming pathway were chosen for the construction of inducible strains. The patterns in the presence of the inducer were characterized by growth on IPTG-supplemented solid agar, then by specific feature measurements used to recover the inducer concentration. **b**, Characteristic patterns of engineered strains in the presence of IPTG and closeups of pattern features. **c**, For each induced strain, heatmaps and plots of radially averaged intensity profiles across the colony for the representative images in **b**. **d**, Fourier transforms of the polar images visualize the magnitudes of the intensity frequencies of each induced strain. **e–g**, Quantification of aspects

of colony patterns of engineered strains at increasing IPTG concentrations. All strains had at least $n = 3$ plates measured at each IPTG concentration. Error bars represent s.e.m. Details can be found in Supplementary Note on Computational Methods. **e**, Intensity of the central region compared to the total intensity of the Fourier transform of the polar image. **f**, Local radial coefficient of variation (CV), which increases with colony asymmetry. **g**, Change in intensity from the densest edge of the inoculum (innermost circular region of colony) to the low-density region immediately surrounding it, that is, distinctness of the inoculum edge, where low values correspond to less distinct edges. **h**, Area under the curve (AUC) of multinomial regression models for predicting IPTG concentration, fit with specific pattern measurements for each strain.

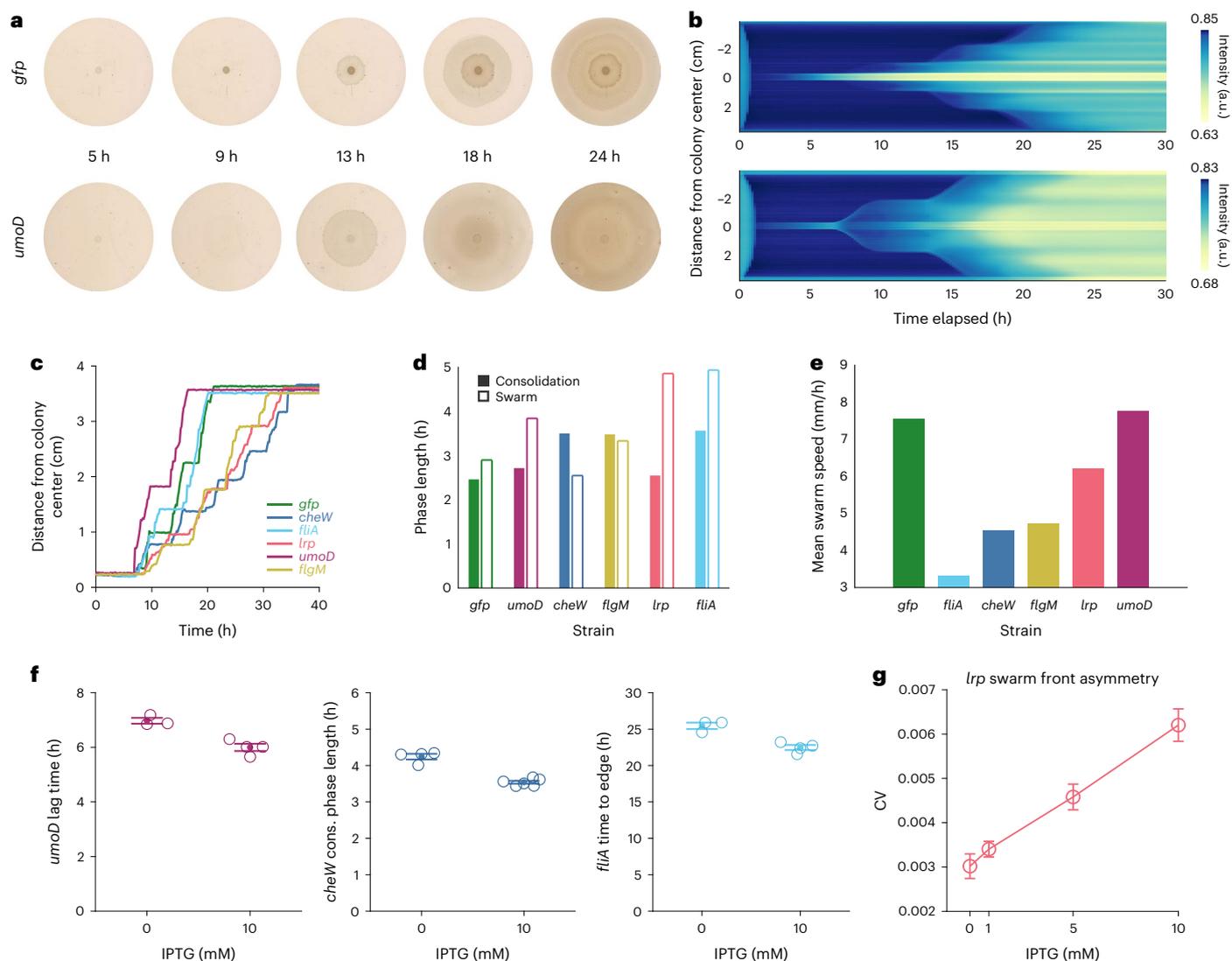


Fig. 3 | Dynamics of engineered *P. mirabilis* pattern formation. **a**, Time-lapse of *P. mirabilis* with inducible *umoD* expression or inducible *gfp* expression (control). Plates contained 20 ml (1.3%) agar with 10 mM IPTG. **b**, Heatmap visualizations of swarming pattern development from the center of plate (0 cm on the left axis) to the edge (top and bottom of heatmap) for each image in the time-lapses in **a**. Radially averaged pixel intensity, a proxy for local colony density, at each location on the plate is represented by heatmap color, with blue indicating the least dense and yellow indicating the most dense regions. Active regions and time periods of colony expansion via swarming are visible as faint blue diagonal edges. Consolidation phases appear as horizontal edges corresponding with increasing density (lighter colors) within the colony. **c**, Colony front distance from center plotted as a function of time for a single time-lapse of six plates.

All plates contained 10 mM IPTG. **d**, Mean consolidation (filled bars, left) and swarm (outlined bars, right) phase lengths calculated from the trajectories in **c**. **e**, Mean of the swarm speeds for each strain in the same time-lapse. **f**, Measurements of dynamic features at 0 mM versus 10 mM IPTG for the indicated strains. Each condition and strain was tested on at least $n = 3$ separate plates. All plots represent a significant difference between induced and uninduced conditions (P values from two-tailed two-sample t -tests were 0.003, 2×10^{-5} and 0.003 for the plots of *umoD*, *cheW* and *fliA*, respectively). **g**, The local CV of the swarm front for an *lrp* colony at each given IPTG (averaged over all swarm phases the colony underwent, $n = 3$ phases at 0 mM and 1 mM IPTG and 4 phases at 5 mM and 10 mM IPTG). Error bars in **f** and **g** represent s.e.m.

time-lapses of each strain grown at a range of IPTG concentrations (Supplementary Fig. 1 and Extended Data Fig. 6). When comparing uninduced to induced conditions, we observed distinct measurements that changed for each strain, such as the lag time for *umoD*, the length of the middle consolidation phases for *cheW* and the time for the colony to cover the plate for *fliA* (Fig. 3f). More complex dynamic parameters also encoded information; for example, the asymmetry of the colony front during swarming phases increased with IPTG for the *lrp* strain (Fig. 3g). These results suggest that dynamic parameters can potentially also be used to encode and decode information from these spatiotemporal patterns, and that in the future, strains can be chosen for a given application depending on the desired timescale of recording.

Multiplexed recording using a dual-input strain

To build a strain that could provide information about multiple inputs simultaneously, we induced a second swarming-related gene with the pBAD operon and promoter, transcribed in the presence of arabinose (Supplementary Fig. 2). Because swarming-related genes have interdependent effects, we sought to try two genes that robustly changed distinct pattern features on their own. Thus, we built a dual-input strain with *cheW* expression induced by the pLac promoter and *umoD* expression induced by the pBAD promoter (Fig. 4a and Supplementary Fig. 2). Initial characterization of this strain demonstrated that its swarm patterns indeed distinctly reflected the presence or absence of each input, unlike a combination of *lrp* and *cheW* (Fig. 4b and Supplementary Fig. 3).

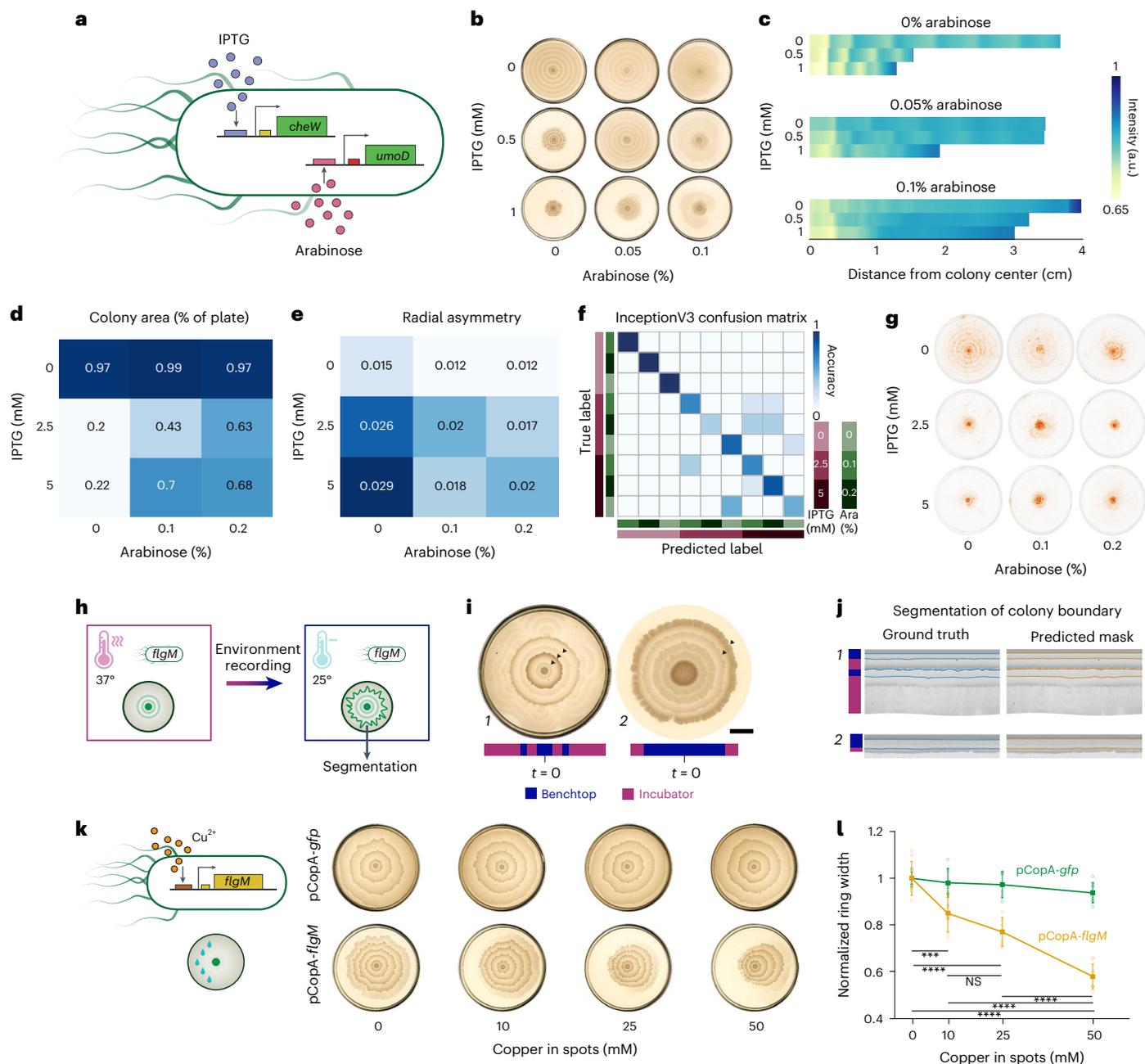


Fig. 4 | Multicondition pattern encoding and deep-learning models for decoding. **a**, Dual-input swarming strain with IPTG-inducible expression of *cheW* and arabinose-inducible expression of *umoD*. **b**, Representative images of colony patterns produced by the dual *cheW/umoD* strain in response to combinations of IPTG and arabinose. **c**, Heatmaps of radially averaged profiles of the patterns in **b**. **d–e**, Mean colony area (calculated as percent of agar area in the flattened image) and coefficient of variation for all plates at each combination of IPTG and arabinose. From left to right, top to bottom, $n = 41, 23, 27, 35, 22, 14, 31, 23$ and 26 plate images. **f**, Confusion matrix for the InceptionV3 model's accuracy of predicting combinations of IPTG and arabinose concentrations from endpoint patterns unseen during training. Total available images per class were 26, 27, 45, 22, 13, 26, 24, 27 and 31 by column from left to right in the matrix; an 80/20 train/test split was used. Accuracy colors on the matrix represent a fraction of predicted test images per true class; clear squares represent 0. **g**, Visualization of the pixel attributions from the InceptionV3 model for representative, correctly predicted images of each class. Darker orange represents higher weight of that pixel on the final prediction. **h**, Schematic representation of encoding of environmental changes within developing *flgM* pattern. **i**, Example patterns of

the *flgM* strain grown with 10 mM IPTG and moved between the benchtop and incubator. Arrows mark boundaries between regions of the pattern formed in different conditions, that is, the location of the colony edge at the time of a switch in conditions. Scale bar for the righthand image represents 0.5 cm, while the lefthand image is an uncropped standard-size Petri dish. **j**, Examples of the predicted boundary masks generated by the trained U-Net compared with ground truth annotations for the pattern images shown in **i**, which were unseen during training. **k**, Schematic representation of copper-sensing strain pCOP-A-*flgM* and location of spots in plate images. Representative images of colonies of indicated strains with added spots of copper at the indicated concentrations. **l**, Mean middle ring widths on the sides of the plates with added copper, normalized to the mean middle ring width for the same day 0 mM copper plates of the same strain. Individual points represent separate plates. $n = 9, 9, 6$ and 9 plates for each of the two strains at 0, 10, 25 and 50 mM copper, respectively. Squares represent the means of the individual points, and error bars represent the standard deviation. From top to bottom: $*P = 2.5 \times 10^{-5}$, $**P = 1.3 \times 10^{-6}$, 0.1 (NS), $***P = 3.9 \times 10^{-5}$, $****P = 6.3 \times 10^{-9}$ and $*****P = 2.3 \times 10^{-13}$. P values were calculated by one-way ANOVA.

Representative radially averaged profiles were visualized as heatmaps for comparison (Fig. 4c). The plates imaged followed a characteristic pattern at most of the conditions. Increasing IPTG from 0 mM to 1 mM, inducing *cheW* expression, resulted in a visible decrease in 24-h colony radius, ring width and colony symmetry, as seen previously in the single-input strain. Meanwhile, increasing arabinose from 0% to 0.1% resulted in a highly symmetric pattern with initially semi-distinct, narrow rings giving way to the indistinct wide rings more characteristic of the single-input *umoD* pattern. The combination of IPTG and arabinose presence resulted in a similar pattern, with narrower inner rings giving way to wider outer rings, but with smaller colonies at 24 h and asymmetric ring boundaries compared to those formed with arabinose alone.

To characterize the *cheW* and *umoD* combination patterns in more detail, a dataset of plate images at IPTG concentrations of 0, 2.5 and 5 mM combined with arabinose at 0%, 0.1% and 0.2% was created. The average percent of the plate covered by the colony at each condition decreased with increasing IPTG and increased with the addition of arabinose (Fig. 4d). However, the increase of arabinose from 0.1% to 0.2% had little effect on the colony area except at 2.5 mM IPTG (Supplementary Fig. 2). Similarly, average radial CV as a measure of colony asymmetry increased with the induced expression of *cheW* but decreased with the addition of arabinose inducing *umoD* expression (Fig. 4e and Supplementary Fig. 2).

As done previously for the single-input strains, a set of standard measurements was then taken on each image in the dataset, and a nine-class multinomial regression model was fit on the output (Supplementary Fig. 4 and Supplementary Table 2). The model performed poorly, predicting almost all images as 0% arabinose, and the maximum AUC achieved was only 0.72. This result suggested that the dual-input strain's patterns, involving interdependent swarm genes, were too complex for the previous feature measurement-based decoding method. However, the ease of distinguishing the patterns by the human eye suggested that deep-learning methods for image classification could prove useful for decoding the patterns. In particular, convolutional neural networks (CNNs) have clear applicability and have not yet been used to characterize macroscale bacterial swarm colony patterns. We anticipated that CNNs could learn to extract salient features from bacterial images and classify patterns to predict the image class.

We fine-tuned models, including the ResNet and Google InceptionV3 networks, to classify images in the dataset into one of the nine classes (details in Supplementary Note on Computational Methods; Extended Data Fig. 7a,b and Supplementary Table 3). The models had been pretrained on ImageNet data, a common strategy in deep learning (Extended Data Fig. 7b)³⁷. Here the fine-tuned Google InceptionV3 model was able to successfully classify the majority of the images (Fig. 4f). An ROC curve was calculated (Methods) and the AUC was 0.96, a noticeable improvement from the multinomial regression model. Such models can also be characterized by 'top-3' accuracy, that is, when used to predict the three most likely classes of an image, whether one of the three is the correct class; the fine-tuned model achieved a top-3 accuracy of 0.98. We observed that intermediate concentrations of IPTG and arabinose reduced the model's accuracy due to two types of patterns forming at that condition, small dense colonies on some plates and colonies with larger asymmetric rings on other plates (Fig. 4f). Visualizing the pixel attributions of the model indicated the inoculum and inner rings had a large impact on the predictions, suggesting that these areas of the pattern were most affected by the induced expression of the different swarm genes (Fig. 4g). As the innermost portion of the colony was most critical to pattern prediction, pattern decoding may be possible after just a few hours of growth, rather than needing to wait for 24 h until the full plate is covered.

As the classification approach requires all potential concentrations to have been acquired in the training dataset, we also explored a deep regression-output approach to predict intermediate values that would provide more utility for future applications (Extended Data Fig. 7c,d).

The previous dataset was combined with images acquired at additional concentrations of IPTG and arabinose. When the EfficientNetB2 architecture, pretrained on ImageNet and directed to produce a numerical output, was trained on a subset of this new dataset, it achieved mean squared errors of less than 0.1 on held-out validation and test subsets (Extended Data Fig. 7c) and was able to relatively accurately predict IPTG and arabinose over the range of concentrations seen (Extended Data Fig. 7d). Overall, these results suggest that our system can be used to encode and decode multiple inputs and that the use of deep networks along with transfer learning will enable decoding of complex pattern feature changes.

Decoding multicondition patterns with deep learning

After determining that our system could be used to encode more complex information, we sought to explore its robustness, which would be key to its application in a real-world or field setting. We observed that our strains' patterns were robust to environmental water and certain strains' IPTG readouts were robust to changes in temperature (Extended Data Figs. 8 and 9 and Supplementary Table 4). On the other hand, we noted certain strains showed a distinguishable pattern change in response to changing temperature; we predicted that such pattern changes could be used to record a dynamically fluctuating environment during pattern formation. Therefore, we next aimed to determine whether an engineered strain could record such changing conditions and how these changes could be decoded from the endpoint pattern, similar to the analysis of rings in a tree²¹. We used the *flgM* strain, which we had observed to form two strikingly different patterns in the presence of 10 mM IPTG in the incubator versus on the benchtop—a swarming-inhibited, ruffled and dense pattern in the incubator at 37 °C and a wide-ring, symmetric and less dense pattern on the benchtop at approximately 25 °C (Supplementary Fig. 5a). After inoculation, plates were first placed in one condition (that is, either the benchtop or the incubator); after some time, plates were switched to the second condition, and certain plates were switched a third time before the endpoint scans were captured (Fig. 4h). Plates were scanned before each switch, creating a dataset of 21 images. Representative pattern images are shown in Fig. 4i. This shift in environmental conditions resulted in the formation of rings alternating between indistinct, radially symmetric, wider rings and dense, asymmetrical, narrow rings, visible as bands on the polar-transformed images (Fig. 4j). In general, denser regions corresponded to incubator growth, while fainter regions with wider rings corresponded to benchtop growth.

To decode these alternating ring patterns, we first manually annotated the dataset, creating ground truth masks of the boundaries marking the shift in the pattern corresponding to a shift in the environment. We then trained a U-Net model, a type of network frequently used for segmentation problems, pretrained on ImageNet to predict these boundaries given an input pattern image (details in Supplementary Note on Computational Methods)³⁸. The model achieved above 95% training and validation accuracy and above 90% recall within the first 25 epochs of training, showing that it could learn the features within the dataset (Supplementary Fig. 5b). Application of the trained model to previously unseen images resulted in the specific prediction of boundaries matching the ground truth and noticeably did not simply highlight all ring boundaries (Fig. 4j). In the future, these predicted boundaries could potentially be used to back-calculate the time at which a given perturbation was experienced, by generating prior control measurements of the time of formation of rings at different conditions. Taken together, these results demonstrated that our approach could be used to decode information about changing environments from the engineered strains' patterns.

Engineering *P. mirabilis* to detect copper

We finally explored whether the engineered strains could be used to detect inputs beyond IPTG, arabinose, and environmental temperature

and humidity. We focused on the detection of heavy metals, given the innate resistance of *P. mirabilis* to certain heavy metals and previous microbial biosensor efforts³⁹. We cloned the intergenic regions, containing putative promoters, before the native *P. mirabilis* 7002 *copA* (copper-exporting P-type ATPase) and *cadA* (heavy metal-translocating P-type ATPase, for zinc, cadmium and mercury) onto the previously used pLac-*gfp* plasmid, replacing the pLac promoter region. A plate-reader experiment for GFP expression demonstrated that the *copA* promoter responded to the presence of increasing concentrations of copper from 0.001 mM to 1 mM and the *cadA* promoter responded to zinc (0.001–1 mM) and mercury (0.0001–0.1 mM), by increasing GFP expression during log phase growth (Extended Data Fig. 10a,b). We then replaced the *gfp* gene with *flgM* and explored the strain's ability to read out a copper presence in spots of water, using a modified protocol involving spotting samples on top of the agar for greater applicability (Fig. 4k). The pCopA-*flgM* strain visually recorded the presence of up to 50 mM copper in the water spots, with a graded change from 0 mM to 50 mM in the ring widths and colony radii at the spot locations (Fig. 4k,l and Extended Data Fig. 10c,d).

Discussion

We have developed a proof-of-concept approach to engineering the spatial patterns of *P. mirabilis* for information encoding and decoding. To advance toward real-world applications, one important consideration is the robustness of the system to environmental conditions. We observed that the IPTG readout of the current system was robust to varied temperatures (Extended Data Fig. 9) and background samples of natural water sources (Extended Data Fig. 8). To further control for noise as well as differences among sample types, we propose the use of a 2 × 2 design consisting of controls with a nonengineered strain alongside the engineered strain, each with and without the sample of interest (Extended Data Fig. 8). Such controls would help determine relative changes in the pattern from conditions specific to the environment, as we observed in copper induction experiments (Fig. 4k,l and Extended Data Figs. 10c,d). For decoding, control images side-by-side with experimental images could be fed to a classifier for use with deep models. Alternatively, an induction curve for the molecule to be sensed could be pregenerated to improve decoding from feature measurements.

Further engineering of strains, such as reducing promoter leakiness, chromosomal integration and tuning of promoter strength with RBS modifications, could achieve tighter control over the final patterns for greater robustness to the environment and improvement of the limit of detection (LOD) for the molecule of interest⁴⁰. For example, the LOD of the pCopA-*flgM* strain on the plate could potentially be decreased to the levels of surface water contamination, which have been measured up to the millimolar range, by using promoters tuned in this range, or approaches to modulate promoter sensitivities^{41,42}. To improve decoding, we anticipate advances in segmentation and classification models will allow for the use of more complex spatiotemporal patterns⁴³. Transition from classification to regression approaches can enable the prediction of intermediate concentrations, allowing more freedom when generating training data. Although we aimed to standardize the data acquisition method in this work, aspects such as lighting and optics can be intentionally varied experimentally to capture a more diverse dataset, applicable to a broader range of settings.

A notable feature of the *P. mirabilis* system is that pattern changes can easily be decoded by eye in comparison to other biosensor approaches that may require additional laboratory equipment⁴⁴. Furthermore, due to their periodic nature, swarming rings provide repeated measurements that can improve accuracy. The use of deep learning can increase granularity, while also being kept to lightweight formats suitable for mobile and low-cost devices. One limitation of the system is that pattern changes are controlled by the timescale of ring and colony formation, which require several hours to develop.

However, slower colony development can indeed be useful to report on longer-term changes, and strains can be accordingly chosen for application-specific timing.

Beyond the application of heavy metal detection in water, the approach can be used to explore sensing other inputs, such as light, radiation or gaseous molecules, or to develop a longer-running recorder for changes in temperature or air quality. For example, previously characterized promoter systems in *E. coli* for the sensing of radiation and airborne nitric oxide can potentially be adapted for use in *P. mirabilis*^{45,46}. Other species with natural swarming properties could be manipulated, such as *Pseudomonas aeruginosa*, *Paenibacillus vortex* or *Bacillus subtilis*, and strains with slower swarming can be exploited to record over several days or weeks^{4,47,48}. Beyond biosensing and recording, engineering bacterial swarming behaviors could be useful in applications ranging from living material assembly to in situ bacteria drug delivery^{49,50}. The approaches developed here can, in turn, shed light on the growth dynamics and virulence of *P. mirabilis* and can be applied to understanding the coordinated and emergent behaviors of microbes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests and statements of data and code availability are available at <https://doi.org/10.1038/s41589-023-01325-2>.

References

- Schuerle, S. et al. Synthetic and living micropropellers for convection-enhanced nanoparticle transport. *Sci. Adv.* **5**, eaav4803 (2019).
- Rubenstein, M., Cornejo, A. & Nagpal, R. Programmable self-assembly in a thousand-robot swarm. *Science* **345**, 795–799 (2014).
- Kearns, D. B. A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* **8**, 634–644 (2010).
- Ingham, C. J. & Jacob, E. B. Swarming and complex pattern formation in *Paenibacillus vortex* studied by imaging and tracking cells. *BMC Microbiol.* **8**, 36–36 (2008).
- Rauprich, O. et al. Periodic phenomena in *Proteus mirabilis* swarm colony development. *J. Bacteriol.* **178**, 6525 (1996).
- Kohler, T., Curty, L. K., Barja, F., van Delden, C. & Pechere, J. C. Swarming of *Pseudomonas aeruginosa* is dependent on cell-to-cell signaling and requires flagella and pili. *J. Bacteriol.* **182**, 5990–5996 (2000).
- Fujikawa, H. & Matsushita, M. Fractal growth of *Bacillus subtilis* on agar plates. *J. Phys. Soc. Jpn.* **58**, 3875–3878 (1989).
- Prindle, A. et al. A sensing array of radically coupled genetic 'biopixels'. *Nature* **481**, 39–44 (2012).
- Barbier, I., Kusumawardhani, H. & Schaerli, Y. Engineering synthetic spatial patterns in microbial populations and communities. *Curr. Opin. Microbiol.* **67**, 102149 (2022).
- Lu, J., Şimşek, E., Silver, A. & You, L. Advances and challenges in programming pattern formation using living cells. *Curr. Opin. Chem. Biol.* **68**, 102147 (2022).
- Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
- Liu, C. et al. Sequential establishment of stripe patterns in an expanding cell population. *Science* **334**, 238–241 (2011).
- Curatolo, A. et al. Cooperative pattern formation in multi-component bacterial systems through reciprocal motility regulation. *Nat. Phys.* **16**, 1152–1157 (2020).
- Kim, H. et al. 4-bit adhesion logic enables universal multicellular interface patterning. *Nature* **608**, 324–329 (2022).

15. Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).
16. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
17. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345–349 (2017).
18. Riglar, D. T. et al. Bacterial variability in the mammalian gut captured by a single-cell synthetic oscillator. *Nat. Commun.* **10**, 4665 (2019).
19. Lu, J. et al. Distributed information encoding and decoding using self-organized spatial patterns. *Patterns* **3**, 100590 (2022).
20. Schaffer, J. N. & Pearson, M. M. *Proteus mirabilis* and urinary tract infections. *Microbiol. Spectr.* **3**, (2015).
21. Cook, E. R. & Pederson, N. *Dendroclimatology*, pp. 77–112 (Springer, 2011).
22. Hauser, G., *Über Faulnisbakterien und deren Beziehung zur Septicämie* (FCW Vogel, 1885).
23. Saak, C. C., Gibbs, K. A. & DiRita, V. J. The self-identity protein IdsD is communicated between cells in swarming *Proteus mirabilis* colonies. *J. Bacteriol.* **198**, 3278–3286 (2016).
24. Fraser, G. M. & Hughes, C. Swarming motility. *Curr. Opin. Microbiol.* **2**, 630–635 (1999).
25. Howery, K. E., Simsek, E., Kim, M. & Rather, P. N. Positive autoregulation of the flhDC operon in *Proteus mirabilis*. *Res. Microbiol.* **169**, 199–204 (2018).
26. Pearson, M. M., Rasko, D. A., Smith, S. N. & Mobley, H. L. Transcriptome of swarming *Proteus mirabilis*. *Infect. Immun.* **78**, 2834–2845 (2010).
27. Şimşek, E., Dawson, E., Rather, P. N. & Kim, M. Spatial regulation of cell motility and its fitness effect in a surface-attached bacterial community. *ISME J.* **16**, 1004–1011 (2022).
28. Armbruster, C. E., Hodges, S. A. & Mobley, H. L. Initiation of swarming motility by *Proteus mirabilis* occurs in response to specific cues present in urine and requires excess L-glutamine. *J. Bacteriol.* **195**, 1305–1319 (2013).
29. Little, K., Austerman, J., Zheng, J. & Gibbs, K. A. Cell shape and population migration are distinct steps of *Proteus mirabilis* swarming that are decoupled on high-percentage agar. *J. Bacteriol.* **201**, e00726-18 (2019).
30. Burall, L. S. et al. *Proteus mirabilis* genes that contribute to pathogenesis of urinary tract infection: identification of 25 signature-tagged mutants attenuated at least 100-fold. *Infect. Immun.* **72**, 2922–2938 (2004).
31. Huang, Z., Pan, X., Xu, N. & Guo, M. Bacterial chemotaxis coupling protein: structure, function and diversity. *Microbiol. Res.* **219**, 40–48 (2019).
32. Dufour, A., Furness, R. B. & Hughes, C. Novel genes that upregulate the *Proteus mirabilis* flhDC master operon controlling flagellar biogenesis and swarming. *Mol. Microbiol.* **29**, 741–751 (1998).
33. Hay, N. A., Tipper, D. J., Gygi, D. & Hughes, C. A nonswarming mutant of *Proteus mirabilis* lacks the Lrp global transcriptional regulator. *J. Bacteriol.* **179**, 4741–4746 (1997).
34. Clemmer, K. M. & Rather, P. N. The Lon protease regulates swarming motility and virulence gene expression in *Proteus mirabilis*. *J. Med. Microbiol.* **57**, 931–937 (2008).
35. Gygi, D., Fraser, G., Dufour, A. & Hughes, C. A motile but non-swarming mutant of *Proteus mirabilis* lacks FlgN, a facilitator of flagella filament assembly. *Mol. Microbiol.* **25**, 597–604 (1997).
36. Morgenstein, R. M. & Rather, P. N. Role of the Umo proteins and the Rcs phosphorelay in the swarming motility of the wild type and an O-antigen (waaL) mutant of *Proteus mirabilis*. *J. Bacteriol.* **194**, 669–676 (2012).
37. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
38. Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds). U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Vol. 9351 (Springer, 2015).
39. Liu, C. et al. Engineering whole-cell microbial biosensors: design principles and applications in monitoring and treatment of heavy metals and organic pollutants. *Biotechnol. Adv.* **60**, 108019 (2022).
40. Hicks, M., Bachmann, T. T. & Wang, B. Synthetic biology enables programmable cell-based biosensors. *ChemPhysChem* **21**, 132–144 (2020).
41. EPA. National primary drinking water regulations. <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations#inorganic> (2023).
42. Kang, Y. et al. Enhancing the copper-sensing capability of *Escherichia coli*-based whole-cell bioreporters by genetic engineering. *Appl. Microbiol. Biotechnol.* **102**, 1513–1521 (2018).
43. Casado-García, Á. et al. MotilityJ: an open-source tool for the classification and segmentation of bacteria on motility images. *Comput. Biol. Med.* **136**, 104673 (2021).
44. Kim, H. J., Jeong, H. & Lee, S. J. Synthetic biology for microbial heavy metal biosensors. *Anal. Bioanal. Chem.* **410**, 1191–1203 (2018).
45. Chen, J. X. et al. Redesign of ultrasensitive and robust RecA gene circuit to sense DNA damage. *Microb. Biotechnol.* **14**, 2481–2496 (2021).
46. Chen, X. J., Wang, B., Thompson, I. P. & Huang, W. E. Rational design and characterization of nitric oxide biosensors in *E. coli* Nissle 1917 and Mini SimCells. *ACS Synth. Biol.* **10**, 2566–2578 (2021).
47. Kearns, D. B. & Losick, R. Swarming motility in undomesticated *Bacillus subtilis*. *Mol. Microbiol.* **49**, 581–590 (2003).
48. Dietrich, L. E., Teal, T. K., Price-Whelan, A. & Newman, D. K. Redox-active antibiotics control gene expression and community behavior in divergent bacteria. *Science* **321**, 1203–1206 (2008).
49. Rodrigo-Navarro, A., Sankaran, S., Dalby, M. J., del Campo, A. & Salmeron-Sanchez, M. Engineered living biomaterials. *Nat. Rev. Mater.* **6**, 1175–1190 (2021).
50. Cubillos-Ruiz, A. et al. Engineering living therapeutics with synthetic biology. *Nat. Rev. Drug Discov.* **20**, 941–960 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Bacterial strains and growth conditions

P. mirabilis ATCC 7002 was kindly provided by P. Rather (Emory University). *E. coli* Mach1 for cloning was purchased from Thermo Fisher Scientific. *P. mirabilis* 29906 was purchased from ATCC. *P. mirabilis* and *E. coli* were cultured in Luria-Bertani (LB) media (Sigma-Aldrich) supplemented with 50 $\mu\text{g ml}^{-1}$ kanamycin as needed. *P. mirabilis* was grown on either 3% or 1.5% agar to suppress or allow for swarming, except for time-lapse assays as indicated. Plate-reader experiments were conducted with a Tecan Infinite 200 Pro using the i-control software version 2.0.10.0. For the experiment presented in Extended Data Fig. 10a, overnight cultures were subcultured for 2 h, then diluted 1:2 and transferred into the 96-well plate and supplemented with metals at the indicated concentrations (CuSO_4 , ZnCl_2 and HgCl_2).

Competent cell preparation

P. mirabilis (PM7002) cells and *E. coli* (Mach1) were made electrocompetent as follows. A fresh 2-ml overnight culture was subcultured 1:100 in 50 ml LB media, then grown at 30 °C with shaking until the logarithmic growth phase was reached, indicated when the optical density at 600 nm (OD_{600}) was 0.4–0.6. Growth was stopped by incubation of the culture on ice for 15 min. Cells were then pelleted by centrifuging for 10 min at 4 °C and 3,000g. After decanting, the pellet was washed three times in either 50 ml ice-cold filter-sterilized 10% glycerol (*P. mirabilis*) or 50 ml ice-cold filter-sterilized water (*E. coli*), then resuspended in 220 μl 10% glycerol. Finally, 50 μl aliquots were stored at –80 °C.

Strain construction

A previously constructed pZE24 (pLac-*gfp* pConst-*lacIQ*) plasmid, containing the ColE1 origin of replication and a kanamycin resistance cassette, was used as the backbone for the inducible swarming plasmids. Plasmids and chromosomal *P. mirabilis* DNA were prepared using standard procedures (Qiagen). Swarming gene and *copA/cadA* promoter sequences were obtained from GenBank (JOVJ000000000.1). Gibson primers were designed (Eton) to amplify the genes and promoter regions from the PM2006 chromosomal DNA via PCR (Phusion)⁵¹. A set of swarming plasmids was constructed using Gibson Assembly and standard restriction digest and ligation cloning to replace the *gfp* gene with the appropriate swarming gene. For plasmids that additionally contained pBAD pConst-*araC*, the operon was obtained from the pBAD-*mCherry* pConst-*araC* plasmid (ATCC54630). The *copA* and *cadA* promoter regions were amplified from the chromosomal DNA preparation. After cloning plasmids into Mach1 *E. coli*, clones were verified via colony PCR (Phusion) and sequencing (Eton, Plasmidsaurus). Clones were then grown at 37 °C with shaking overnight before being stored in 50% glycerol at –80 °C. All strains, plasmids and primer sequences are listed in Supplementary Tables 5 and 6; plasmid maps are shown in Supplementary Fig. 6.

P. mirabilis transformation

Plasmid DNA was introduced into *P. mirabilis* competent cells as follows. Fifty microliter aliquots of competent cells were thawed on ice for 10 min. DNA was added to the cells (100–400 ng DNA in a volume of 1–5 μl per aliquot). The mixture was then incubated on ice for 1 h. Cells were electroporated in prechilled electroporation 0.1 cm electrode gap cuvettes using a Bio-Rad MicroPulser set to E1 setting (1.8 kV) for bacterial electroporation. Cells were recovered by adding 1 ml prewarmed SOC (super optimal broth with catabolite repression, Thermo Fisher) media and incubated with shaking at 37 °C for 3 h. The cells were pelleted by centrifugation for 10 min at 4 °C and 3,000g, and 700 μl of the supernatant was decanted before resuspension in the remaining 300 μl . The cells were then plated on prewarmed 3% LB agar plates with antibiotics as necessary and incubated at 37 °C for 22–24 h. Single colonies were inoculated into liquid culture (2.5 mL LB

media with antibiotics as necessary), grown overnight with shaking at 37 °C, and fresh overnight cultures were stored in 50% glycerol at –80 °C.

Bacterial growth and swarm assay

Overnight liquid bacterial cultures were prepared by inoculating LB media with cells from the –80 °C glycerol stocks and supplementing with 50 $\mu\text{g ml}^{-1}$ kanamycin as appropriate. Cultures were incubated at 37 °C with shaking for 12–16 h. The OD_{600} of each culture was measured and normalized to 1.0 by dilution with LB media. Swarm assays were optimized from a protocol adapted from the literature. A study to develop standard conditions is shown in Extended Data Fig. 1. Precise maintenance of the selected conditions was necessary to achieve consistent results, as noted in previous work with *P. mirabilis*⁵². A freshly prepared solution of 1.5% agar was autoclaved, then cooled to 50–55 °C with stirring. As necessary, 5 $\mu\text{g ml}^{-1}$ kanamycin, IPTG and/or arabinose were then added. Fifteen milliliters of agar was poured into each 100 × 15 mm Petri dish and left to solidify partially uncovered under an open flame or in a BSL-2 hood for 30 min. Two microliters of the previously diluted liquid culture was inoculated on the center of each Petri dish and dried for 15 min partially uncovered under an open flame or in a BSL-2 hood. For the assays shown or analyzed in Fig. 4k and Extended Data Figs. 8 (environmental water) and 10c,d (copper sensing), the indicated inducer was spotted on the plates in 2 μl spots in a semi-circle around the inoculum, then left to dry concurrently with the inoculum (rather than being mixed into the molten agar). Where 0 inducer is specified, the same water was used but without added IPTG or metal. The environmental water samples (river water from the Hudson River in New York and ocean water from Venice Beach, California) in Extended Data Fig. 8 were supplemented with 1 M IPTG and then filtered with a 0.22 μm syringe filter (BD). The plates were incubated at 37 °C for 24 h, except for time-lapses and temperature-varying experiments described in Fig. 4h–j and Extended Data Fig. 9. Incubator humidity typically varied between 50% and 80% during the course of experiments.

Plates were individually imaged using a scanner (Epson Perfection V800 or V850 Pro Photo Scanner) set to 24-bit Color and 400 dpi, with the lid off and the colony side facing up. The scanner was kept on the benchtop, and ambient lighting was similar during all experiments; other scanner settings were also kept constant between experiments. For greater visibility, a consistent contrast increase and –20% saturation were applied to all colony scans shown in the figures, but the analysis was done on raw images (Extended Data Fig. 3b).

Time-lapses

For time-lapses on the benchtop (room temperature), up to six plates with 20 ml (1.3%) LB agar were inoculated and placed on the flatbed scanner using the previously described settings. The plates were kept upside down to prevent condensation and with lids on to prevent contamination. A custom AppleScript was written to scan plates every 10 min for a preset length of time (typically 48–72 h). Typical benchtop conditions were 25 °C and 40–50% humidity.

Computational methods overview

Image processing and statistical analysis were done in MATLAB R2022B (Mathworks), with a combination of built-in and custom-written functions. In general, most image analyses were done on polar transformations of the plate images (Extended Data Fig. 3). Deep-learning models were implemented in Tensorflow (v2.8) and Pytorch (v1.7), with preprocessing in MATLAB and Python; the ring-boundary segmentation model used for measurements of the pCopA-*flgM* strain's rings, as well as those for the pLac-*flgM* strain grown at 34, 36 and 37 °C (Extended Data Fig. 9b), was trained previously⁵³. Full details are given in Supplementary Note on Computational Methods.

Statistical analysis and regression

Statistical tests were calculated, and data were plotted either in MATLAB or in Python. R^2 values of pLac-*cheW* strain colony radii and ring widths were calculated with MATLAB's *fitlm* function, which fits a linear regression model. P values shown in Fig. 4I were generated using MATLAB's *multcompare*, and those shown in Fig. 3F, Supplementary Fig. 2b and Extended Data Fig. 9b were generated with the *ttest2* function (two-tailed). The exact statistical powers of two-sample, two-tailed t -tests (for pairwise comparisons of 0 mM and 10 mM IPTG at 34, 36 and 37 °C, as shown in Supplementary Table 4) were calculated using MATLAB's *extpowerStudent* function on MathWorks FileExchange, which employs the noncentral cumulative distribution function. Latex tables of results were generated using Overleaf. Multinomial regression models were fit to the measurements using the *mnrfit* function in MATLAB, returning the coefficients and P values in Supplementary Table 1. The functional form of *mnrfit* is a 'proportional odds model' (log of the change in the response variable = sum of each coefficient \times predictor variable + the intercept). For the single-input strain data used in Fig. 2h, each flattened image was divided into four sectors (each 250 pixels wide), and measurements were taken on each sector to increase the number of measurements available so that model fitting could converge. The models were evaluated using the *multiClassAUC* function on MathWorks FileExchange, which implements the Hand and Till function for the AUC for multiclass problems⁵⁴. The AUC for the InceptionV3 model was calculated using a function from scikit-learn, `metrics.roc_auc_score`.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Compressed folders of lower-resolution versions of the images generated and analyzed during this study are uploaded on GitHub and publicly available at <https://github.com/daninolab/teus-mirabilis-engineered>, DOI: 10.5281/zenodo.7637609; the time-lapses are uploaded as videos. The full, several-hundred-gigabyte dataset of the original high-resolution images is not publicly available due to large file sizes preventing them from being stored on GitHub. The high-resolution images are available for sharing upon request from the corresponding author (T.D.).

Code availability

The codes used in this study are deposited at GitHub at <https://github.com/daninolab/teus-mirabilis-engineered>, <https://doi.org/10.5281/zenodo.7637609>.

References

- Minogue, T. et al. Draft genome assemblies of *Proteus mirabilis* ATCC 7002 and *Proteus vulgaris* ATCC 49132. *Genome Announc.* **2**, e01064-14 (2014).
- Pearson, M. M. (ed.), *Proteus mirabilis: Methods and Protocols*, pp. 15–25 (Springer, 2019).
- Doshi, A. et al. A deep learning pipeline for segmentation of *Proteus mirabilis* colony patterns. *Proceedings of the IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5 (IEEE, 2022).
- Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001).
- Baumberg, S. & Roberts, M. Anomalous expression of the *E. coli* lac operon in *Proteus mirabilis*. *Mol. Gen. Genet.* **198**, 166–171 (1984).

Acknowledgements

We thank M. Pavlicova (Columbia University) for the helpful discussion of the statistical methods used. We thank the members of the Danino laboratory for review of the manuscript. We thank S. Li (Columbia University) for assisting in metal-sensing experiments and image processing. This work was supported by an NSF CAREER Award (1847356 to T.D.), Blavatnik Fund for Innovations in Health (T.D.), and NSF Graduate Research Fellowship (DGE-2036197 to A.D., Fellow ID 2018264757).

Author contributions

A.D. and T.D. conceived and designed the study. A.D., M.S., R.T., R.M. and S.M. conducted the experiments as follows: A.D. constructed the engineered strains with assistance from M.S. and R.T.; A.D., M.S. and S.M. established the initial swarm assay protocols; A.D., M.S., R.T., R.M. and S.M. performed swarm assays and time-lapses with single- and dual-input strains; A.D., M.S. and R.T. performed the condition-switching experiments and A.D. and M.S. performed the metal-sensing experiments. A.D. and M.S. performed the image-processing-based computational analysis, with assistance in preprocessing (image flattening) from R.T. and R.M. M.S. performed the deep segmentation work with U-Nets and temperature experiment classification work with Transformer models along with the power analysis, A.D. performed the dual-input strain regression work and AUC work and A.D. (Berkeley) performed the dual-input strain classification and attribution visualization, all with input from J.G. and A.L. A.D. and T.D. wrote the original manuscript draft, and A.D., M.S. and T.D. edited the manuscript with input from all authors.

Competing interests

A.D., M.S., J.G., A.L. and T.D. are named as inventors on a provisional patent application that has been filed by Columbia University with the US Patent and Trademark Office related to all aspects of this work. The remaining authors declare no competing interests.

Additional information

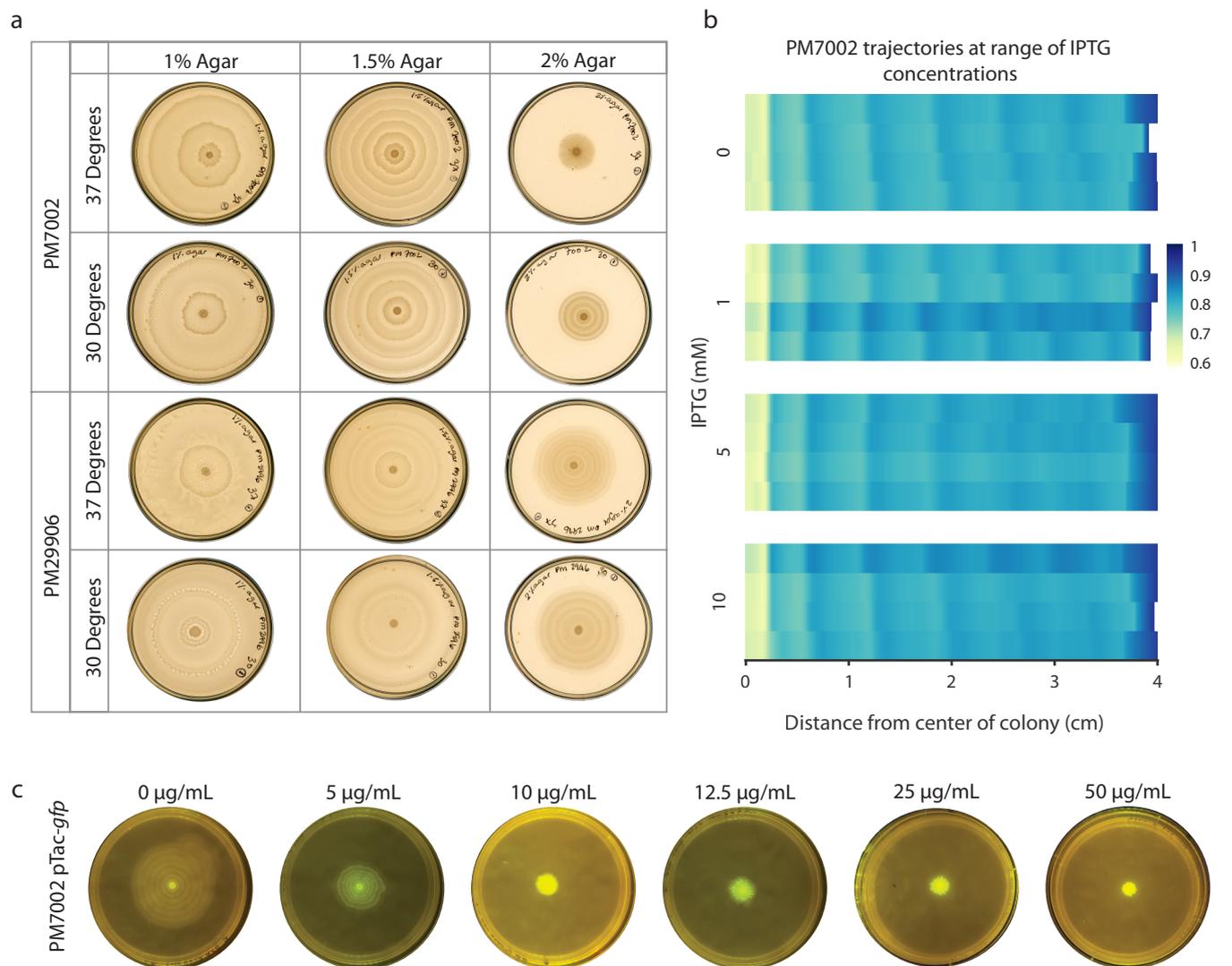
Extended data is available for this paper at <https://doi.org/10.1038/s41589-023-01325-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-023-01325-2>.

Correspondence and requests for materials should be addressed to Tal Danino.

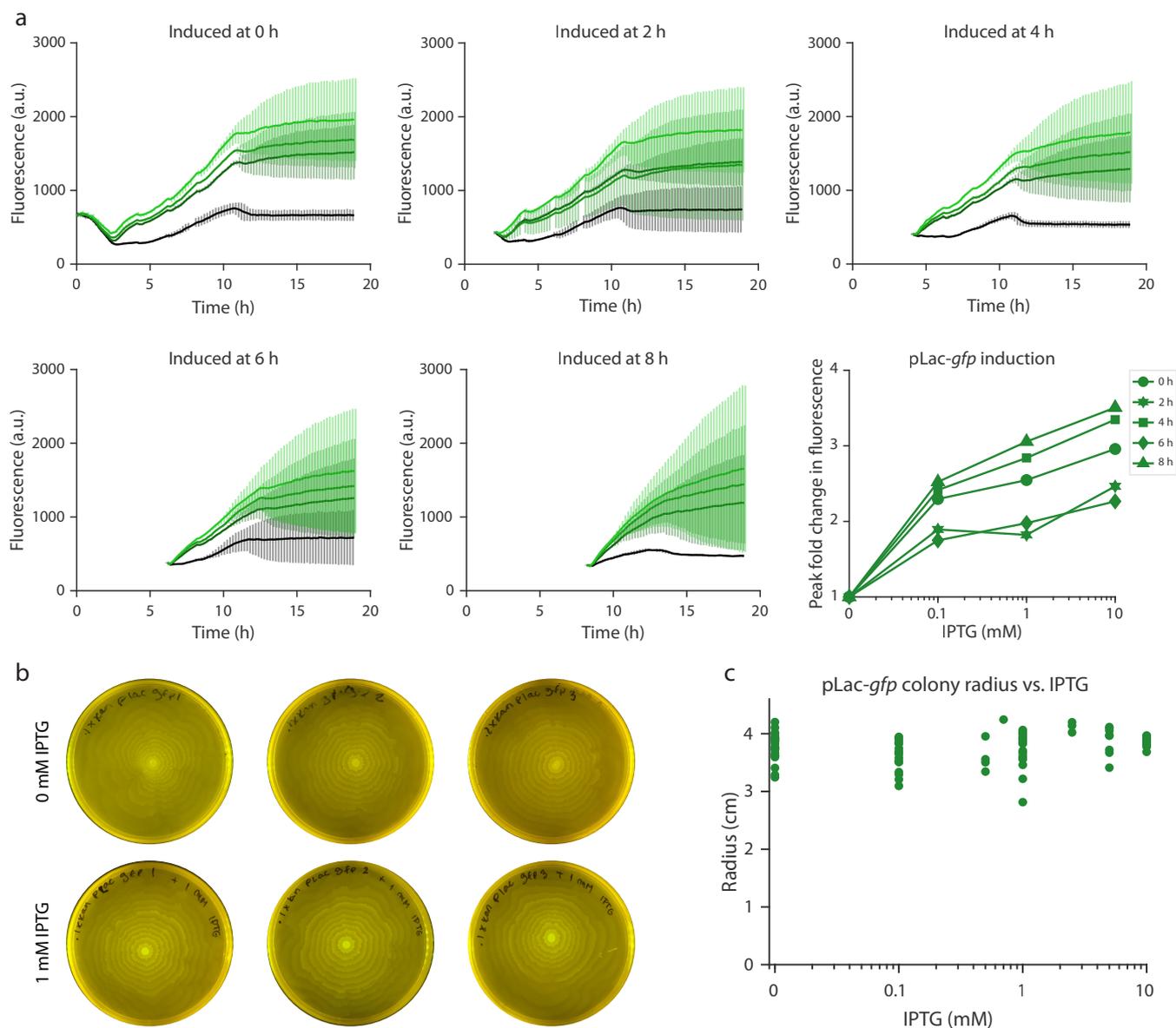
Peer review information *Nature Chemical Biology* thanks Mark Isalan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



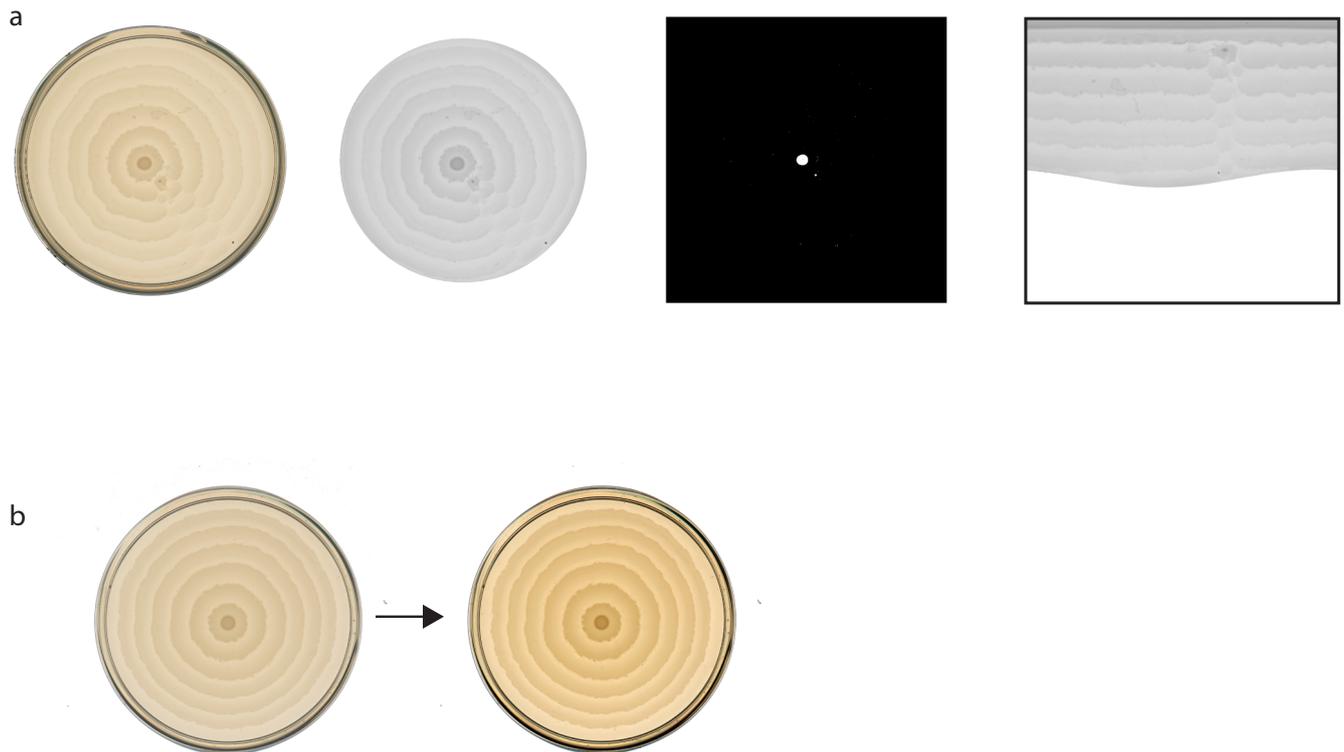
Extended Data Fig. 1 | Selection of conditions. **a.** Representative colony images for different *P. mirabilis* strains, temperatures, and agar hardness after 24 hours of growth. **b.** Heatmaps of radially averaged profiles of PM7002 colonies at a range of IPTG concentrations (each profile represents an individual colony).

Light green represents lowest pixel intensity (highest colony density), arbitrary units. **c.** PM7002 with pTac-*gfp* plasmid was grown at a range of kanamycin concentrations for 24 hours. Representative images of fluorescence taken under blue-light transilluminator are shown.



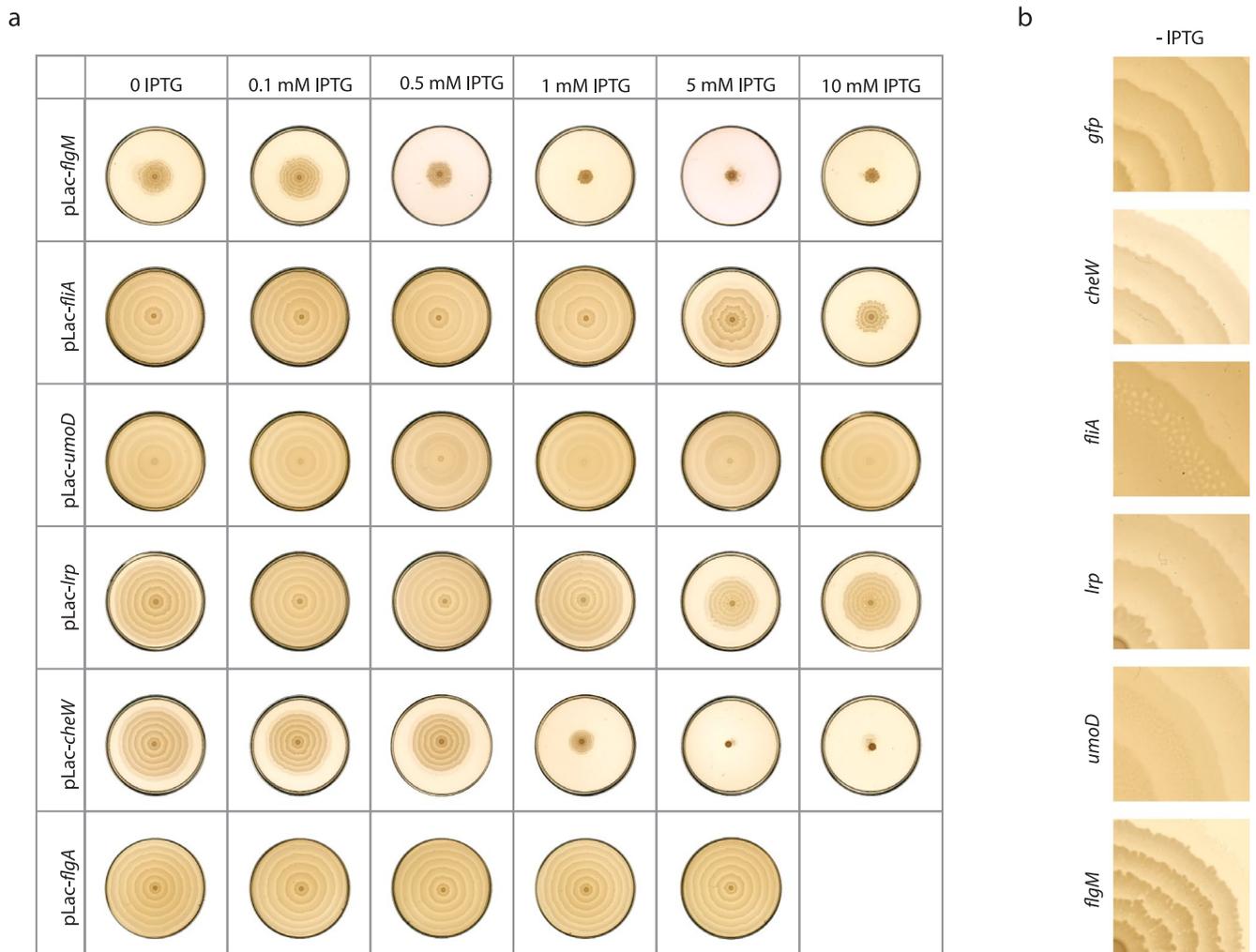
Extended Data Fig. 2 | Induction of pLac promoter in *P. mirabilis*. **a.** Fresh overnight cultures of PM7002 with pLac-*gfp* plasmid were subcultured with 50 $\mu\text{g}/\text{ml}$ kanamycin, grown for 0–8 hours, and induced with a range of IPTG concentrations. The mean GFP intensity (arbitrary units, shown as the line plot) after subtracting background fluorescence and normalizing to OD_{600} is plotted for each group; error bars indicate standard deviation ($n = 3$ for each). Bottom right panel: Peak fold change in fluorescence from uninduced (0 mM IPTG)

groups is plotted for each group induced at a different time after subculturing (symbols represent mean of three replicate wells, individual data shown in **(a)**). Maximal expression of around 3-fold was consistent with literature⁵⁵. **b.** Images of PM7002 pLac-*gfp* strain plates grown with either 0 or 1 mM IPTG, imaged after 24 hours under blue-light transilluminator. **c.** 24-hour colony radii of pLac-*gfp* strain at various concentrations of IPTG; each dot represents a plate. Plate images are from different days.

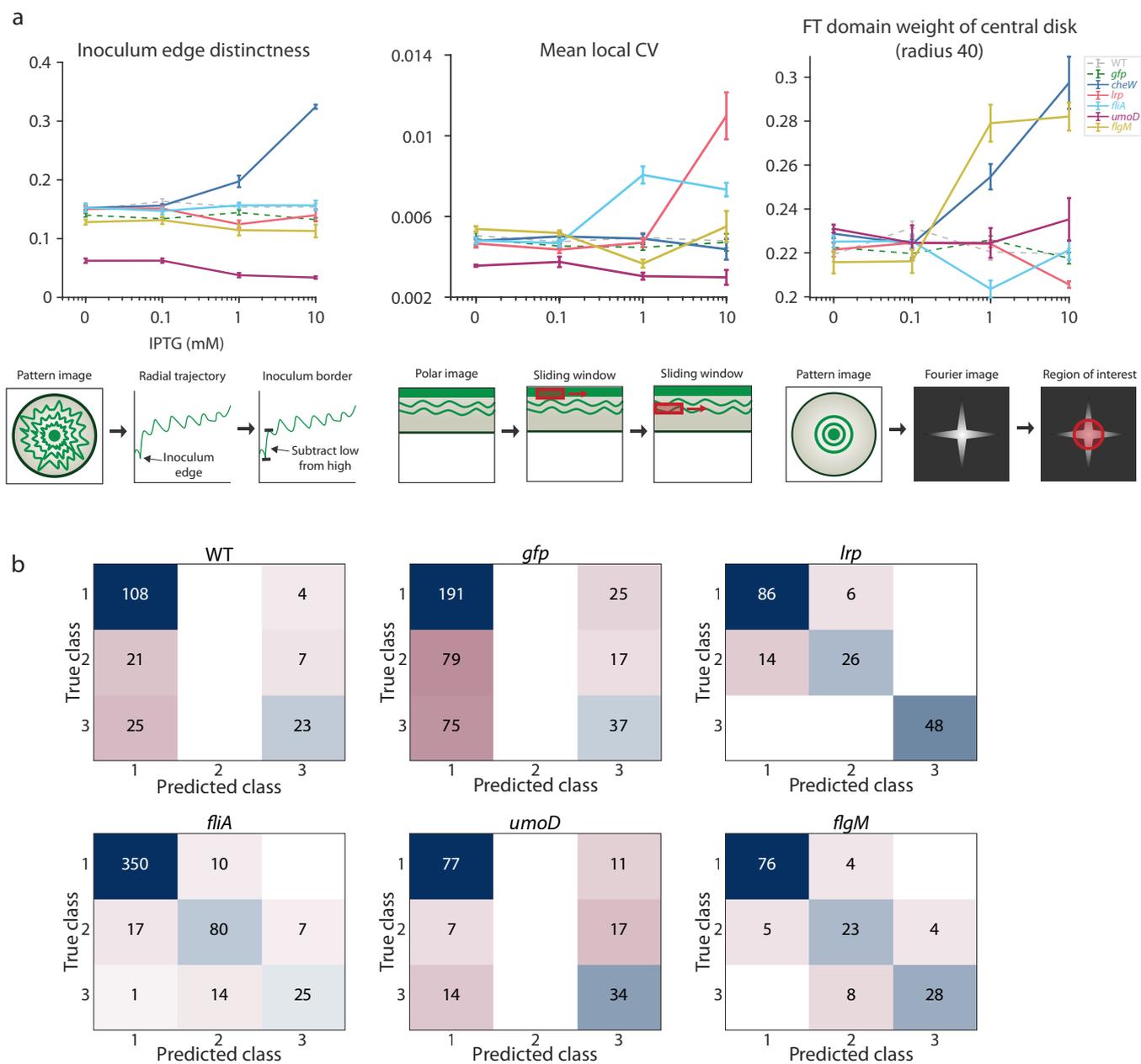


Extended Data Fig. 3 | Image processing. a. Petri dishes were scanned at high resolution. The Petri rim was identified and cropped out using MATLAB functions. Images were thresholded to show only the colony inoculum, and the center point was identified using MATLAB functions. Images were converted

from Cartesian to polar coordinates with interpolation, and the flattened images were used for subsequent analysis. See Supplementary Note on Computational Methods for details. **b.** A representative raw image and enhanced-contrast version showing the adjustment which was made for images shown in figures.

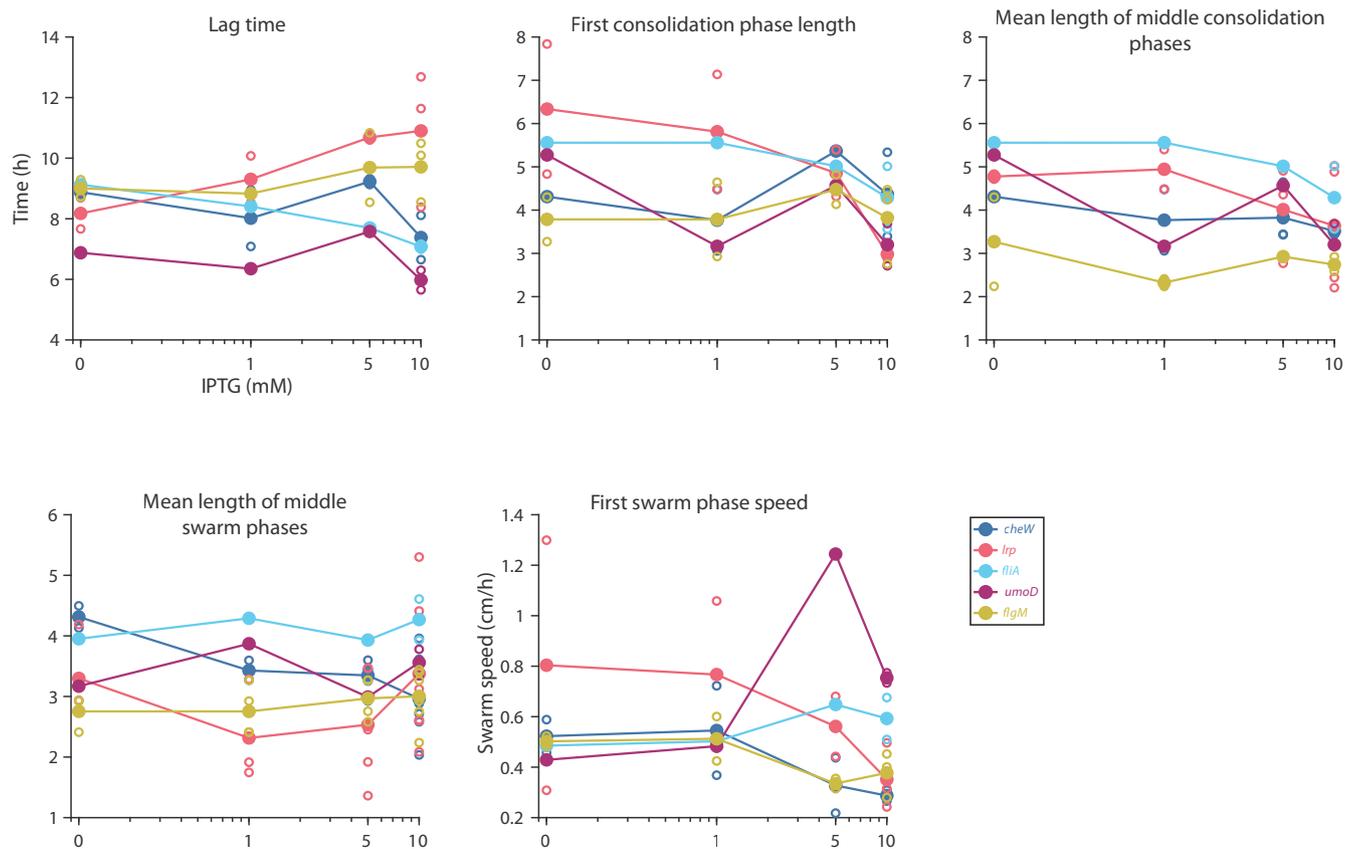


Extended Data Fig. 4 | Range of patterns formed. **a.** PM7002 strains with indicated inducible swarm plasmids were grown for 24 hours at a range of IPTG concentrations. Representative images of three replicates are shown. **b.** Closeups of patterns formed at 0 mM IPTG.

**Extended Data Fig. 5 | Quantification of engineered colony patterns.**

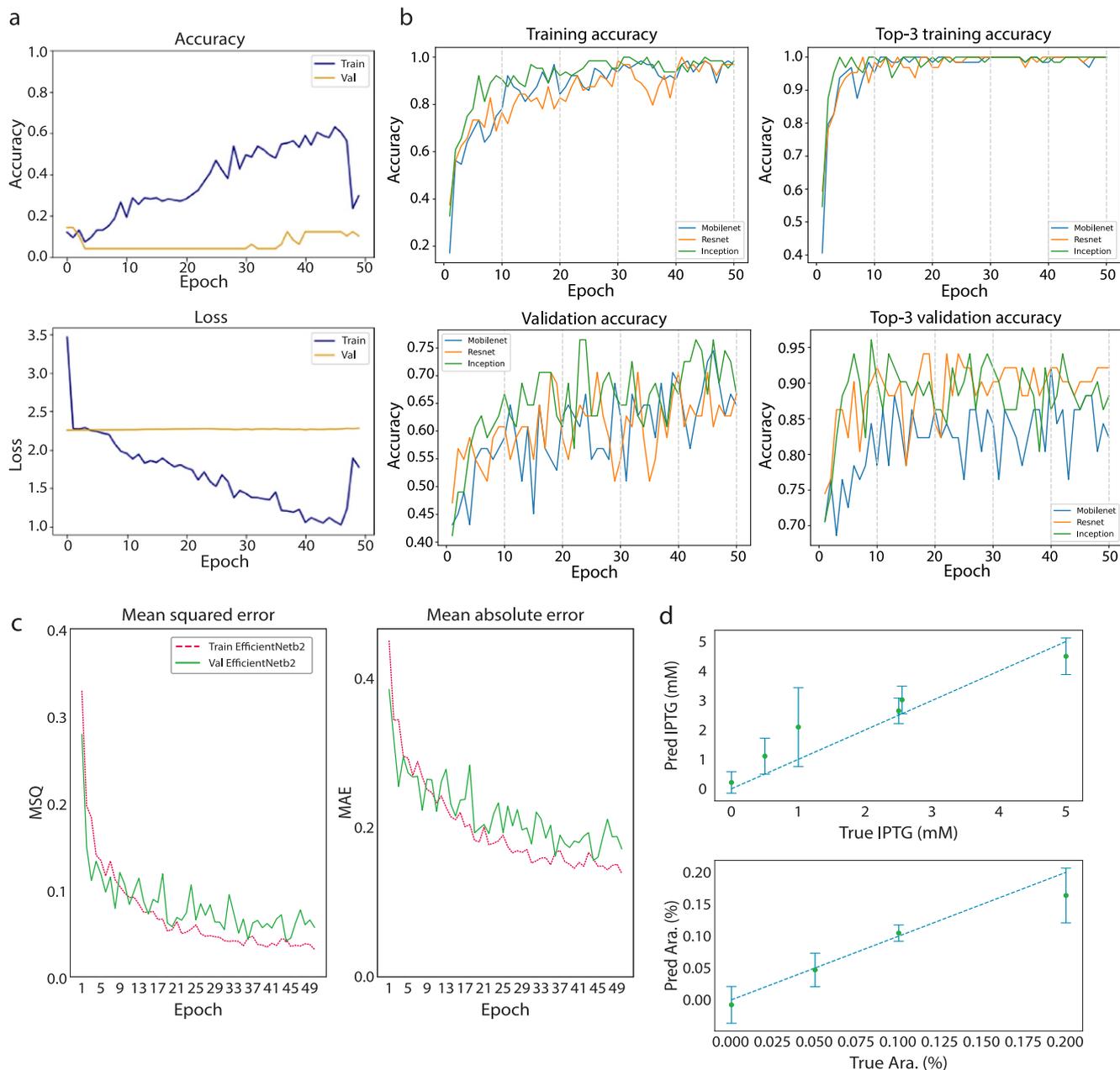
a. Quantification of aspects of colony patterns of engineered strains at increasing IPTG concentrations, measured as in Fig. 2e–g and as shown in the schematics below each graph. All strains had at least $n = 3$ plates measured at each IPTG concentration. Error bars represent standard error of the mean (s.e.m).

b. Confusion matrices of the fitted models' performance on each dataset; matrices are labeled with numbers of plate sectors (see Supplementary Note on Methods for details). Classes 1–3 represent 0 up to but not including 1 mM IPTG, 1 up to but not including 5 mM IPTG, and 5 up to and including 10 mM IPTG.



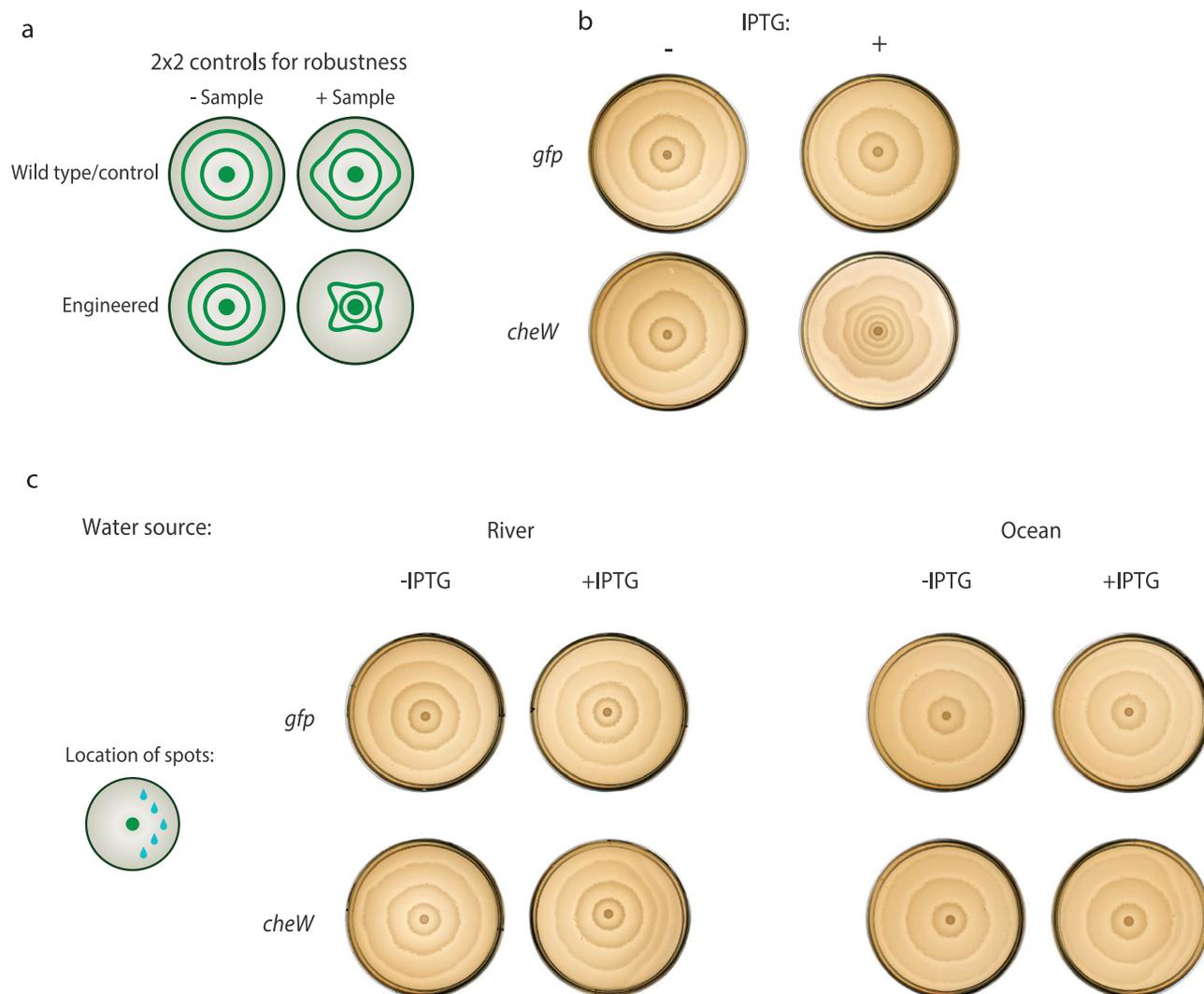
Extended Data Fig. 6 | Dynamics of pattern formation. Plots of dynamic characteristics of engineered strains vs IPTG concentration, measured from one to two time-lapses per condition, per strain. Individual dots represent individual time-lapses (lag times) or individual phases in all time-lapses (consolidation and swarm phase lengths, swarm speeds). Lines represent the mean of these

individual measurements. Middle swarm and consolidation phase lengths were determined by discarding the measurements of the first and last of these phases for each time-lapse, or discarding the last phase if the time-lapse had only two of the given phase.



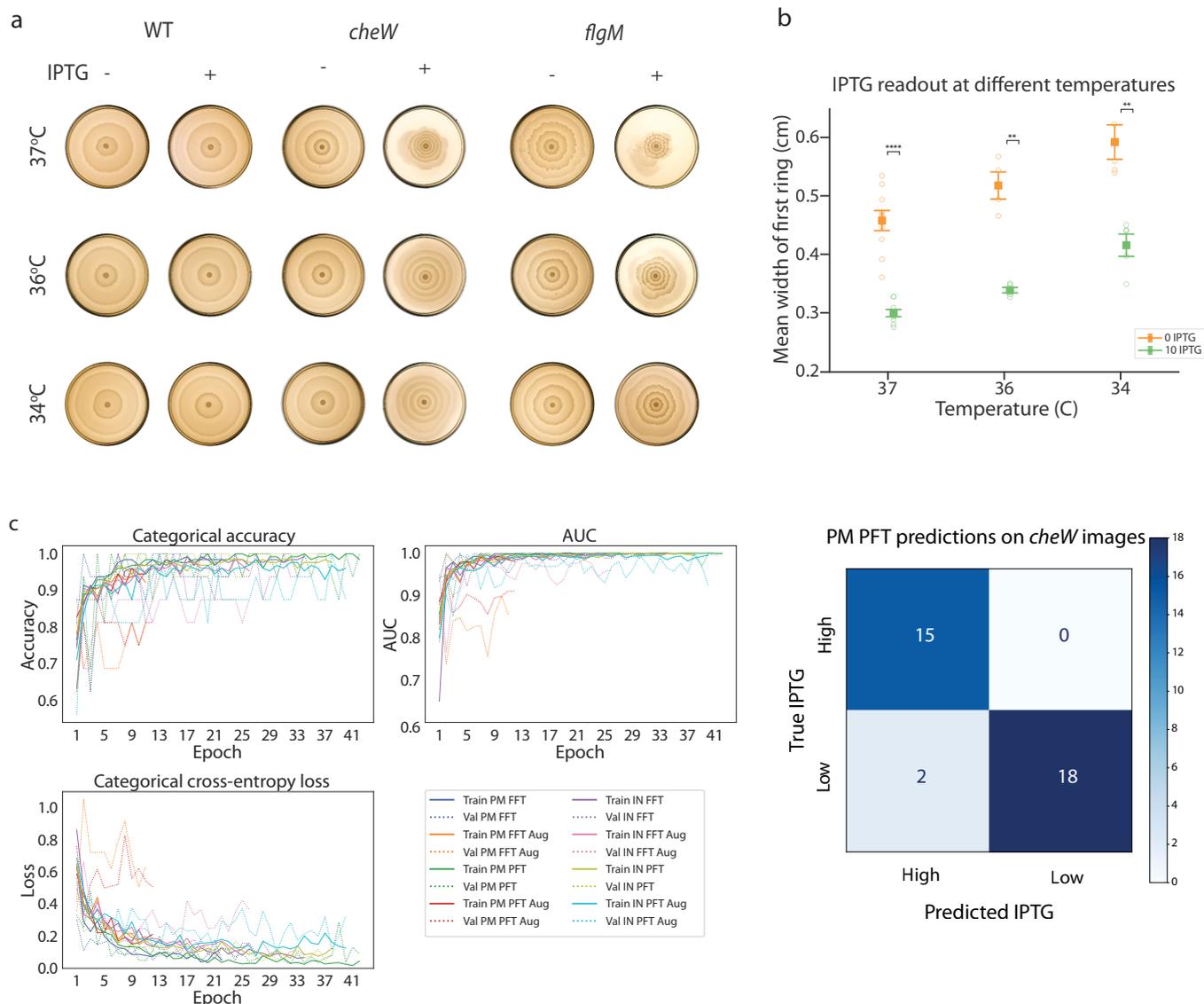
Extended Data Fig. 7 | Performance of CNNs on classification and regression tasks. **a.** Training and validation accuracy and loss for a CNN model which had three convolutional/max pooling blocks, trained on the dataset of images of the dual-input strain at various IPTG and arabinose conditions (same dataset as in Fig. 4f). **b.** Fine-tuning of three architectures pretrained on ImageNet; righthand panels represent models' ability to identify the correct image class within its top three predicted classes. **c.** Learning curves of EfficientNet model

trained on the dual-input pattern images with regression output. Loss = mean squared error; mean absolute error shown for further detail. **d.** Predictions of the trained model evaluated on images not seen in the original training dataset. Dotted lines represent location where predictions would match the true values. Error bars represent root mean squared error on the predictions for each given concentration.



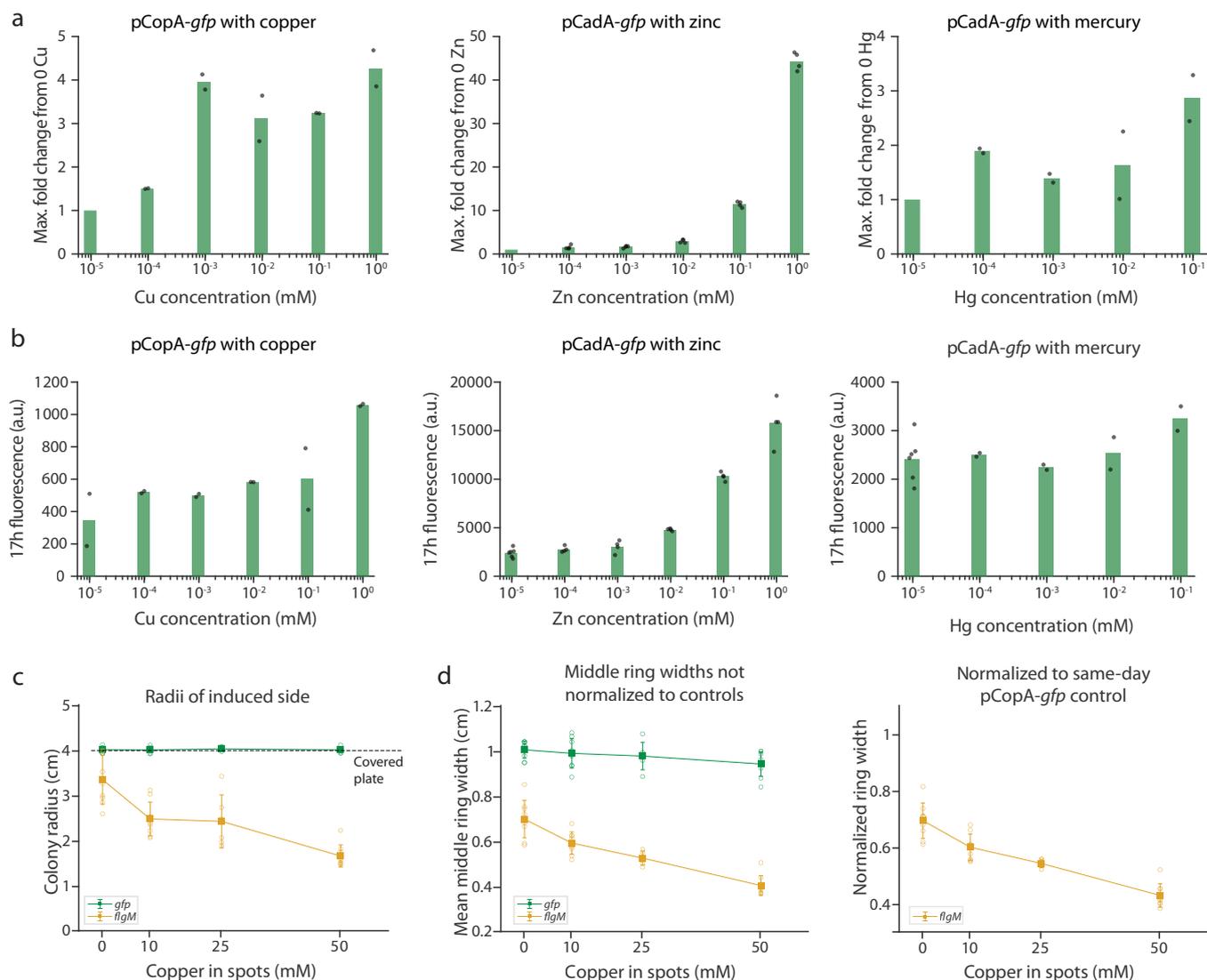
Extended Data Fig. 8 | Visual IPTG readout is robust to natural water samples.
a. Proposed controls for natural samples. **b.** Representative images of colonies grown with either river water only or river water with 5 mM IPTG throughout

the plate. **c.** Representative images of colonies grown with natural water sample spots with and without IPTG (1 M) as indicated. Schematic indicates location of natural water droplets on plate relative to colony inoculum.



Extended Data Fig. 9 | Qualitative and quantitative IPTG readouts are robust to growth at different temperatures. **a.** Representative images of colonies grown at the indicated temperature. Each pair of 0 and 10 mM IPTG colonies of a given strain were collected on the same day. **b.** Width of the first colony ring of the pLac-*flgM* strain grown at the indicated conditions. $n = 10$ plates (37 °C, 0 IPTG); 9 plates (37 °C, 10 mM IPTG); 4 plates (36 °C, 0 IPTG); and 5 plates each at (36 °C, 10 mM IPTG), (34 °C, 0 mM IPTG), and (34 °C, 10 mM IPTG). Error bars represent s.e.m. $P = 2.42 \times 10^{-6}$ (37 °C), 3.59×10^{-3} (36 °C), and 0.001 (34 °C) for the comparison

between 0 vs 10 mM IPTG, calculated with a two-tailed t -test. **c.** Learning processes of the SwinTransformer model variants for classifying pLac-*cheW* images acquired at 37 °C into low vs high IPTG (curves legend: 'PM' = *P. mirabilis*-pretrained, 'IM' = ImageNet-pretrained; 'FFT' = fully fine-tuned, 'PFT' = partially fine-tuned; 'Aug' = trained with augmentations). Confusion matrix shows absolute, pooled accuracy of the highest-performing model (PM PFT) evaluated on the held-out test set of 37 °C, 36 °C, and 34 °C pLac-*cheW* images. Numbers on the matrix indicate number of images in the given category.

**Extended Data Fig. 10 | Engineering metal-sensing strains of *P. mirabilis*.**

a. Maximum fold change of GFP, expressed from either pCopA or pCadA, achieved over the course of 17 hours at the indicated concentration normalized to uninduced wells, in plate reader. **b.** GFP fluorescence at 17 hours (end of experiment) for each concentration of each metal by strain. GFP fluorescence was calculated by subtracting background fluorescence value and dividing the raw fluorescence value by the media-background-subtracted OD600 value for the same well. In **(a)** and **(b)** dots represent individual wells. **c.** Colony radius

of the copper-induced side of the plates, as shown in Fig. 4k. **d.** Lefthand plot: mean of the middle ring widths (that is, neither innermost nor outermost) of the colonies on the sides with copper. Each open circle represents a separate plate. Righthand plot: the same measurements normalized to the same day *gfp* strain's measured mean widths at the given concentration. **(c)** and **(d)** show alternative representations of the data plotted in Fig. 4l; as in that figure, $n = 9, 9, 6,$ and 9 plates for each of the two strains at 0, 10, 25, and 50 mM copper, respectively. Data are presented as mean values \pm standard deviation.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

A custom Applescript used to control an Epson scanner to scan at specified intervals for a specified period of time. The plate reader data was collected with the Tecan i-control software, v 2.0.10.0.

Data analysis

Custom MATLAB code for analysis (polar and Fourier transformation of images, measurements on images, calculation of dynamics) was written and Python scripts were used for all deep learning.
 Versions used:
 MATLAB R2019A-R2022B
 Plotting(MathWorks FileExchange): Colorbrewer: v 3.2.1; Interpolate_cbrewer & Plot_brewer_cmap & cbrewer.m: V 1.0; linspecer: 1.4.0.0;
 Export_fig: v 3.1.6; latexTable: 1.181.0.0
 Statistics (MathWorks FileExchange): multiClassAUC: 1.0; extpowerStudent: v 1.11.0.0
 For dual-input image classification and regression prediction: Tensorflow: 2.8.0; OpenCV: CV2; Sklearn: 1.2
 For plate reader data: Excel: 16.66.1
 The VGG-11 U-Net segmentation models were implemented and evaluated in Python 3.7 with the library segmentation-models-pytorch (0.2.0) and its requirements, i.e. torchvision>=0.5.0, pretrainedmodels==0.7.4, efficientnet-pytorch==0.6.3, and timm==0.4.12, as well as the requirements of each of these, such as torch (1.7.0 used here), numpy (1.19.5 used here), and tqdm (4.59.0 used here). Additional libraries included pandas (1.2.3), cv2 (4.5.1), csv (1.0), albumentations (0.5.2), matplotlib (3.4.1), scikit-image (0.18.1), and openpyxl (3.1.0). The SwinTransformerTiny224 classification models were implemented and evaluated in Python 3.7 with the library tfswin (>= 2.0.0) and its requirements, i.e. tensorflow>=2.7.0, keras>=2.7.0, and numpy>=1.19.2. Additional libraries included google-colab (1.0.0), pandas (1.3.5), matplotlib (3.2.2), scikit-learn (1.0.2), and seaborn (0.11.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

A low-resolution version of the image datasets and other data generated for this paper have been uploaded at GitHub at <https://github.com/daninolab/proteus-mirabilis-engineered>. DOI: 10.5281/zenodo.7637609. Due to large file sizes, timelapses have been uploaded as videos. All raw image data is available to share upon reasonable request.

The *Proteus mirabilis* 7002 sequencing project accession code is JOVJ00000000.1.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size pre-determination methods were used because Petri dishes were typically generated in groups of three replicates or more following convention in the field. The number of separate plate images analyzed in each figure is described in its legend.
Data exclusions	Obvious outliers due to petri dishes which may have had uneven antibiotic mixing or uneven agar solidification were excluded from further analysis.
Replication	Sample size of n = 3 plates per condition within an experiment was typically used and endpoint experiments were typically repeated 2-3 times to generate enough plate data for deep learning where needed. For timelapses, the scanner size constrained an experiment to 4 plates (or 6 could be used but two plates would be cropped slightly), so where possible we repeated time-lapses at least once. For the most part, replication of experiments resulted in similar results; however, some variability between experiments in features such as exact colony radius was observed as expected due to the sensitivity of the system to slight differences in temperature, agar solidity, etc. We have described this expected variability, its implications for the applicability of our system, and our proposed mitigation methods in our Discussion section.
Randomization	For deep learning, shuffling of images was done with standard methods. Within the flgM dataset, training and test sets were split to ensure equal representation of the different classes within each set (i.e., that the test set would have images from both the started-on-bench and started-in-incubator groups), and then shuffled. For all other classification work, datasets were split randomly into train/validation/test sets using standard methods. For experiments, no random sample allocation methods were performed because they were not relevant to the type of experiments.
Blinding	Studies were not blinded as it was not relevant to the study, because the data collection involved scanning the plates exactly as they were and analysis was all numerical. There was no occasion on which bias would change the output analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Involvement |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Involvement |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |