Deep learning recognises antibiotic modes of action from brightfield images

Daniel Krentzel[1,7,8], Kelvin Kho[2,7], Julienne Petit[3,4,7], Nassim Mahtal[5], Spencer L. Shorte[5], Anne Marie Wehenkel[3,4], Ivo G. Boneca[2,8] and Christophe Zimmer[1,6,8]

[1]Institut Pasteur, Université Paris Cité, Imaging and Modeling Unit, F-75015 Paris, France

[2]Institut Pasteur, Université Paris Cité, CNRS UMR6047, INSERM U1306, Unité de Biologie et génétique de la paroi bactérienne, F-75015, Paris, France

[3]Institut Pasteur, CNRS UMR 3528, Université Paris Cité, Structural Microbiology Unit, F-75015 Paris, France

[4]Institut Pasteur, Université Paris Cité, Bacterial Cell Cycle Mechanisms Unit, F-75015 Paris, France

[5]Institut Pasteur, Université Paris Cité, Photonic Bio-Imaging, Centre de Ressources et Recherches Technologiques (UTechS-PBI, C2RT), F-75015 Paris, France

[6]Machine Biophotonics Lab, Rudolf Virchow Center, University of Würzburg, Würzburg, Germany

[7]These authors share first authorship: DK, KK and JP

[8]Corresponding authors: DK, IGB and CZ

Abstract

Antimicrobial resistance is a growing public health threat predicted to cause up to 10 million deaths a year by 2050. To circumvent existing bacterial resistance mechanisms, discovering antibiotics with novel modes of action (MoAs) is crucial. While growth inhibition assays can robustly identify antibiotic molecules, they miss promising compounds with subinhibitory phenotypes and do not inform on drug MoA. Microscopy-based cytological profiling of drug-treated bacteria with hand-crafted image descriptors or more recently deep learning (DL) provides complementary information on the MoA. However, current approaches are limited by the need for fluorescent labelling and drug exposure at inhibitory concentrations. It also remains unclear if cytological profiling enables the detection of drugs with novel MoAs. Here, we demonstrate an approach based on supervised DL to identify antibiotic drug MoA from microscopy images without fluorescent labelling. We train a convolutional neural network to predict treatment conditions from brightfield images of *Escherichia coli* exposed to reference drugs covering multiple MoAs. Our method can detect drug exposure at subinhibitory concentrations and distinguishes individual drug treatments with high accuracy (86.21%). The learned representations implicitly capture MoA-specific phenotypes, enabling perfect MoA recognition (100%) without retraining using a model trained on only 644 images. Our approach can identify the MoA of previously unseen drugs with good accuracy (77.8±3.3%), as long as the MoA is represented by at least one of the training compounds. Finally, we show that our approach can detect if a drug has a novel MoA with an area under the curve of 0.77±0.19, facilitating microscopy-based identification of novel classes of antibiotics. Our methods and results pave the way towards an automated pipeline for antibiotic drug discovery based on imaging and DL.

Introduction

Antimicrobial resistance (AMR) contributed to 5 million deaths in 2019 (1). However, recent estimates have drastically corrected this value upwards and some studies project the death toll to rise to as much as 10 million deaths a year by 2050 (2). During the last 40 years, very few antibiotics with novel modes of action (MoAs) have been discovered. New resistance mechanisms of various pathogenic bacteria have thus remained unchecked, thereby increasingly rendering the existing arsenal of antibiotics ineffective (3).

Phenotypic drug discovery (PDD) is an attractive approach to discover antibiotics with new MoAs, since no prior target hypothesis is required. A typical read-out in PDD screens is optical density (OD), which serves as a proxy for cell density. While this read-out allows to identify compounds that inhibit bacterial growth at concentrations exceeding the minimum inhibitory concentration (MIC), it precludes the identification of antibacterial compounds at sub-MIC concentrations that could be optimised by medicinal chemistry or be effective in combination therapies. Moreover, these single read-out screens cannot provide information on how an antibacterial compound inhibited bacterial growth, i.e. the MoA (4).

To overcome these limitations, recent antibacterial screening campaigns (5) have drawn inspiration from eukaryotic PDD approaches using microscopy images as read-outs (6). Nonejuie *et al.* (7) and Zoffmann *et al.* (8) acquired fluorescence microscopy images of bacteria and extracted hand-crafted quantitative descriptors (features) from segmented bacterial cells to perform bacterial cytological profiling (BCP). Importantly, compounds with the same MoA induce similar phenotypic changes, thus enabling the computational association of compounds by MoA directly from microscopy images. However, a limitation of early BCP is its reliance on cell segmentation, an error-prone step especially at high cell densities, and the potentially insufficient or unspecific information contained in hand-crafted features (4). Deep learning (DL) approaches, in contrast, hold the potential to extract highly specific phenotypic features from images in an end-to-end manner without segmentation (4). Prior work on microscopy images of eukaryotic cells labelled with multiple fluorescent dyes (Cell Painting) (9) has shown that features learned by a convolutional neural network (CNN) can outperform handcrafted features in discriminating and associating different and similar treatment conditions respectively (10). Recently, a CNN trained on over 5,000 images has

been shown to successfully classify the MoA of antibiotics at concentrations of 1-5×MIC from multi-channel images of fluorescently stained *Mycobacterium tuberculosis* bacteria (11).

Despite these encouraging efforts, several limitations and questions remain to be addressed before DL analysis of high-throughput microscopy images can be leveraged efficiently in future screens to discover antibiotic compounds with novel MoAs. First, it is unclear if antibiotic compounds can be identified from images at sub-MIC concentrations, which would greatly improve the sensitivity of large-scale screens that typically test thousands of compounds at a single concentration. Second, it remains to be determined to what extent unseen compounds can be assigned to their correct MoA and, critically, whether compounds with novel MoAs can be detected. Third, fluorescent labelling is expensive, significantly adds complexity to experimental protocols, can introduce batch-to-batch variability, and may interfere with the molecular mechanism of compounds. Whether fluorescent labelling can be dispensed with in favour of brightfield (BF) images in phenotypic screens of bacteria, as recently shown in eukaryotic cells (12), is unclear, since bacterial cells are roughly an order of magnitude smaller, potentially limiting the range of visible morphological phenotypes. Finally, the performance of DL typically comes at a steep cost in training data, as e.g. ~200 fields of view (FOV) per compound were used to train a CNN in (11), imposing limits to scaling up this approach to thousands of compounds and calling for more efficient approaches.

Here, using *Escherichia coli*, we develop and validate a DL approach to extract bacterial phenotypic features end-to-end directly from BF images, thus dispensing with the need for fluorescent staining. We show that from these BF images alone our model can reveal alterations induced by antibiotic compounds at concentrations well below the MIC and accurately recognise the MoA of compounds when trained on as few as seven images per well. Moreover, our approach can recognise the MoA of unseen compounds by comparing to reference antibiotics of the same MoA, and determine if a compound has a novel MoA, again from BF images only. Our results lay the foundations for a time- and cost-effective pipeline to identify chemical compounds with potentially novel antibiotic MoAs.

Results

Collecting a high-throughput image dataset of bacterial phenotypes

We collected a high-throughput imaging dataset of *E. coli* bacteria exposed to 22 widely used reference drugs (**Table 1**). These include eleven inhibitors of cell wall synthesis - encompassing four beta-lactamase inhibitors - shown to target the penicillin binding proteins PBP1A/B, PBP2 or PBP3, four antibiotics targeting ribosomes, three targeting DNA gyrase, two targeting membrane integrity, one targeting RNA polymerase, and one targeting DNA synthesis (**Table 1**). For each of these drugs, we determined the half maximal inhibitory concentration (IC50) (**Methods**), which ranged from 5.7 ng/ml to 64 µg/ml (**Table 1**). We distributed these antibiotics onto four 96-well plates with randomised layouts (**Supp. Fig. 1**) using an acoustic liquid handler (**Methods**) to expose bacteria to one of four subinhibitory concentrations (0.125×IC50, 0.25×IC50, 0.5×IC50, IC50). Then, we grew bacteria to exponential phase and seeded them onto the four 96-well plates containing the previously distributed antibiotics (**Methods**). After a 2h15min incubation followed by bacterial membrane staining with FM4-64FX for 15 min, we fixed bacteria with a combination of paraformaldehyde and glutaraldehyde whereupon we stained DNA with Hoechst. Next, we acquired high-throughput images (**Methods**) in three channels (BF, FM4-64 and Hoechst) using a high-content screening system in non-confocal mode with a 63× water-immersion objective (**Fig. 1a, 2a**).

Deep learning model and training strategy

We then designed a DL model (**Supp. File 1**) to map images of antibiotic-treated bacteria to chemical perturbations (**Methods**). We based our model on the CNN architecture EfficientNet-B0 (13), which was shown to achieve state-of-the-art performance on established image classification benchmarks with much fewer parameters than more widely used architectures like ResNet-50 (14) (e.g. 5.3 million parameters vs. 26 million parameters and 76.3% vs. 76.0% accuracy on ImageNet classification (13,15)), thereby keeping memory requirements and computation time moderate.

As input data we used images of entire FoVs, of 2,160×2,160 pixels (corresponding to an area of 139.32×139.32 µm). Images were passed through a data augmentation pipeline and divided into nine tiles which were processed in parallel using the same convolutional filters. Tile-wise

embedding vectors were aggregated before the final classification layer (**Fig. 1b**; **Methods**), thus diminishing the influence of tiles without bacteria on model training.

We trained our model to perform image classification, where each distinct antibiotic drug treatment, i.e. each combination of antibiotic and concentration (e.g. “Aztreonam” and “0.25×IC50”) was considered as a separate class, resulting in a total of 89 classes (4 concentrations × 22 compounds + 1 control) (**Methods**). In order to reduce our model’s sensitivity to possible technical confounders (e.g. plate-specific image features and well position-dependent effects independent of the drug-induced phenotypes) and increase its robustness to biological variability, we used three biological replicates where bacteria were cultured independently and grown on different days with randomised well positions (**Supp. Fig. 1**) to train a single model. Critically, to avoid biases and spurious correlations between training and testing data that could lead to an overestimation of classification performance, we evaluated each model’s classification performance on a biological replicate, using a fourth hold-out test plate with a different randomised layout.

Deep learning accurately recognises antibiotic mode of action from images

We first asked if and to what extent our model can predict the combination of both drug and concentration from images alone, considering all the available imaging modalities (BF, Hoechst, FM4-64FX). This task is a very difficult task, since our treatments include multiple compounds with the same targets (**Table 1**) and changes in drug concentration are not necessarily expected to induce different phenotypes (e.g. if concentrations are much below or above the IC50). Indeed, although some drug treatments led to easily recognisable phenotypes, distinguishing between all treatment conditions only by visual inspection from images remained extremely challenging (**Supp. Fig. 2**). If the model were unable to learn any features specific to the combination of drug and concentration, the expected random classification accuracy would be 1.12% (1/89). We assessed this accuracy on four rotations of the training and testing plates, with confusion matrices shown in **Supp. Fig. 3.** Remarkably, the classification accuracy ranged from 31.3% to 47.4% at the image level, i.e. 28 to 42 times better than random. This indicates that the model extracted information specific to drug-concentration combinations from images, although far from allowing perfect classification.

As the effect of antibiotics is likely to be more pronounced at higher concentrations, we next assessed our model (trained on all drug-concentration combinations) only on test data at the highest drug concentrations, i.e. the IC50, and assessed accuracy based only on the predicted antibiotic (thus ignoring the predicted concentration). On these data, our model achieved a very high per-image accuracy of 76.2% to 90.2%, compared to 4.3% (1/23) for random classification (**Supp. Fig. 4**). Per-well accuracies (obtained using the majority of predictions from all images in the same well) were slightly higher, namely 78.3% to 91.3%, indicating that at the IC50, the vast majority of the 22 reference drugs used here led to highly specific phenotypes, which our model was able to automatically learn from the training data (**Fig. 2b**).

Nevertheless, prediction of antibiotic identity was not perfect, as the model still tended to confuse certain compounds. Inspection of individual hold-out test replicates revealed that the model tended to systematically confuse a subset of drugs (**Fig. 2b**, **Supp. Fig. 4**). For example, Aztreonam was confused with Cefepime, Cefsulodin or Ceftriaxone for 3 out of 4, and Relebactam was confused with Meropenem in 2 out of 4 pairs of model and test data, and Norfloxacin was systematically mistaken for Ciprofloxacin in one hold-out test dataset. We noted that the confused compounds tended to have similar MoAs (**Table 1**). Moreover, when computing cosine distances between median embeddings for all pairs of antibiotics, we could observe clustering by MoA (**Fig. 2d**), as also illustrated by UMAPs (**Fig. 2e,f**), suggesting that our model extracted similar feature vectors from images for compounds targeting the same MoA despite only having been trained on drug labels.

This observation encouraged us to assess our model's ability to recognise the MoA, as opposed to the identity of the drugs themselves. For this purpose, we assigned each of the reference drugs to one out of six MoAs or target classes, namely: cell wall, ribosome, DNA gyrase, RNA polymerase, DNA synthesis and membrane integrity (**Table 1**). Because half of the antibiotics were cell wall inhibitors (11 out of 22), we further subdivided this class by their main target proteins, namely PBP 1A/B, 2 and 3 (**Table 1**), thereby leading to eight classes of targets with one to five antibiotics per target class. We then assessed the model's performance at recognising the MoA of each drug by pooling drugs and predictions according to their known MoA. Note that this did not involve any retraining on MoAs as we still used the model trained on combinations of antibiotics and

concentrations. Remarkably, our model achieved an accuracy of 84.1% to 99.3% in classifying drug MoA at the image-level and of 100% for three out of four replicates at the well-level (on the remaining plate the classification accuracy was 88.9%), indicating almost perfect recognition of MoA from images (**Fig. 2c**).

Thus, our data show that images of bacteria treated with reference antibiotics at IC50 concentrations contain highly specific information about drug MoA, allowing accurate MoA recognition from images alone. Moreover, our results indicate that established CNN architectures are well suited to extract this information in an end-to-end fashion when trained on drug labels alone, without the need for handcrafted features, cell segmentation and *a priori* knowledge of drug MoAs.

Brightfield images alone allow accurate recognition of drug MoA

In contrast to BF, fluorescent labelling of various bacterial cell compartments requires selection of dyes with compatible excitation and emission spectra, as well as the development of appropriate staining protocols. We therefore asked if all three imaging channels, namely BF, FM4-64FX and Hoechst (**Fig. 3a**) were necessary to accurately distinguish between antibiotic drugs or MoA. To quantify this, we trained several CNN models as described previously, but with all possible combinations of the three imaging channels as input and assessed classification accuracy on the hold-out test data using the same channels. In some cases, the accuracy varied significantly between the four hold-out plates, with accuracies ranging from 65.2% to 95.7% for the models trained only on the FM4-64FX channel. We speculate that these differences could be due to variations in staining efficiency between plates. Unsurprisingly, the median accuracy over the four replicates was highest (87.0%) for the model trained and tested on all three channels together (**Fig. 3b, Supp. Fig. 5**). To our surprise, however, ignoring one or even two of the channels in training and testing resulted in only small reductions of median accuracy when compared to the differences between plates (**Supp. Fig. 6**). This suggests that the three channels contained largely redundant information. Strikingly, when considering the BF channel alone (**Fig. 3d**), the median accuracy was 82.6%, only 4.4% below the three-channel accuracy, and was more consistent across plates (accuracies ranged from 73.9% to 82.6% vs. 78.3% to 91.3%). Pooling the predictions from the

BF-only model by MoA led to perfect MoA classification (**Fig. 3c,e**) and feature vectors obtained from the BF-only model also clustered by MoA (**Fig. 3f**).

These results show that accurate MoA recognition is possible from BF images alone, thereby removing the need for fluorescent labelling. This could drastically simplify experimental protocols and significantly reduce costs in large phenotypic screens.

CNN learns MoA-specific features from only seven images per condition

A major drawback of using DL algorithms compared to classical analysis methods based on hand-crafted features is the need for large amounts of training data. Assessing the minimum number of images required to train a DL model without losing classification performance can help to optimise and potentially reduce the burden of data collection. We therefore estimated the minimum number of images required to train our CNN without adversely affecting MoA classification and also assessed the impact of pretraining our model on a large public dataset of natural images (ImageNet (15)).

To ascertain the minimum number of required images to robustly distinguish between MoA classes, we trained our CNN using an EfficientNet-B0 convolutional backbone pretrained on ImageNet (**Fig. 1b**) with varying numbers of images per condition ranging from three to 120 and computed mean classification accuracies across four replicates. We observed that classification accuracy stopped increasing with the number of training images and fitted a two-parameter exponential function to determine the accuracy plateau and the number of images required to reach it (within 5%), at four different concentrations (**Fig. 3gh, Methods**). We found that between seven (IC50) and 15 images (0.25×IC50) per condition (644 and 1,380 images in total) were sufficient to achieve the maximum accuracy. Repeating the analysis with randomly initialised weights (**Fig. 3i**), we found that pretraining on ImageNet data reduced the number of required images by between 25.0% (0.125×IC50) and 52.6% (IC50).

These results show that near-perfect MoA classification performance can be achieved with only seven images per condition for *E. coli* bacteria exposed to compounds at the IC50 concentration and highlight the benefit of using pretrained models to reduce the required training data.

Detecting drug phenotypes at subinhibitory concentrations

The above results show that DL can reliably extract MoA-specific features from images of bacteria exposed to drugs at IC50 concentrations, which for the reference drugs considered here ranged from 5.7 ng/ml to 64 µg/ml (**Table 1**). Visual inspection of drug-treated bacteria shows that the cellular phenotypes tend to be less pronounced at lower concentrations, as expected (**Fig. 4a**). When screening large libraries of compounds, systematic titration to determine the IC50 for each compound is often not practical. Instead, chemical screens typically rely on a single concentration, which for most active compounds is generally lower or higher than the inhibitory concentration. We therefore asked if our DL method can detect or even classify drug-induced phenotypes at subinhibitory concentrations, and in particular concentrations at or below the IC50. Since at sub-MIC concentrations, new chemical compounds may induce only subtle phenotypic effects before optimisation by medicinal chemistry, the ability to detect such weak phenotypes could increase the sensitivity and hit rate of phenotypic screening for antibiotic drug discovery.

We obtained confusion matrices for sub-IC50 concentrations and observed that, unsurprisingly, our model's ability to correctly classify bacterial images by compound increased with higher concentrations (**Supp. Fig. 7**). We then analysed the embeddings computed by our BF-only model (trained on reference antibiotics at multiple concentrations as detailed above) on test data at increasing concentrations, from 0.125×IC50 to 1×IC50 (**Fig. 4b,c**). For most antibiotics, embeddings were similar to untreated controls (DMSO) at the lowest tested concentration (0.125×IC50) and tended to move away from the control and from one another as concentration increased. The exceptions were the PBP 2 inhibitors Avibactam and Mecillinam, as well as the DNA synthesis inhibitor Trimethoprim, for which embeddings already localised away from the controls at the lowest concentration, suggesting that these compounds induce distinctive phenotypic changes at concentrations well below inhibitory values. To quantify this observation, we computed the cosine distances between the median embeddings of drug-treated cell images and the DMSO images, at each concentration (**Fig. 4d**). We also defined a threshold distance as the mean plus three standard deviations of the cosine distance between individual DMSO images. Based on this threshold, we found that even at 0.125×IC50, compounds from three groups of antibiotics (inhibitors of DNA synthesis, ribosome, PBP 1 and PBP 2) exhibited sufficiently different phenotypes to be detected

(**Fig. 4e**). At 0.5×IC50, all but two groups of antibiotics (RNA polymerase and PBP 3 inhibitors) were distinguished from the negative control.

To assess drug exposure detection based on cosine distance to DMSO at different subinhibitory concentrations, we obtained receiver operating characteristic (ROC) curves cross-validated on four hold-out plates and computed the corresponding area under the curve (AUC) (**Fig. 4f**). We observed that, in line with **Fig. 4e**, drug exposure detection improved with increasing concentrations and that even at the lowest concentration (0.125×IC50) drug exposure was robustly detected with an AUC of 0.81±0.08, while at IC50 detection was perfect (AUC=1.00).

Overall, our data show that image embeddings learned by our CNN can be used to robustly detect subtle phenotypic changes at subinhibitory concentrations from BF images alone, thereby enabling highly sensitive detection of potentially active compounds in phenotypic screens.

Recognising known modes of actions of unseen drugs

Our analyses above showed that our model can recognise the identity of a compound or the MoA of several antibiotics after training on independent images obtained with the same antibiotics. However, in the context of a screening campaign, the goal is to determine the MoA of a previously unseen novel drug. Therefore, we next asked if our model can correctly recognise the MoA of an unseen compound, i.e. a compound that was entirely absent from the training data. To address this question, we performed a leave-one-compound-out (LOCO) analysis for compounds sharing a MoA with at least two other compounds.

We retrained our CNN using BF images only by systematically excluding all images of one compound from the training data (**Fig. 5a**). Then, we used this new model to obtain embeddings from images of bacteria exposed to the left-out compound at IC50 on a hold-out test plate. Next, we obtained reference embeddings using the same hold-out test plate from images of bacteria exposed to all other compounds at IC50. For each FoV, we identified the closest reference compound using a k-NN classifier (k=150 and cosine distance) on embedding vectors (**Fig. 5b**, **Methods**). For each left-out compound, we then assigned the MoA of the reference compound that was the closest for most FoVs. We found that at a concentration of IC50, the classification accuracy across plates ranged from 72.2% to 83.3% (**Fig. 5c**). Moreover, we observed that the k-NN classification on

embeddings performed better on average than directly using the MoA predictions of our model (75.0±3.5% vs. 63.8±5.4%, **Fig. 5c**). We then evaluated MoA assignment accuracy for a range of concentrations and found that accuracy tended to improve with higher concentration (**Fig. 5d**). Interestingly, drugs targeting PBP 2 were perfectly classified across the entire range of concentrations, indicating that these compounds induce highly distinctive phenotypes even at low concentrations.

At the highest concentration (IC50), we observed good to excellent MoA prediction accuracy for compounds inhibiting PBP 2 (100%) PBP 3 (88.89%), DNA gyrase (100%) and ribosome (75%), while compounds targeting PBP 1A/B were almost never correctly identified (1 out of 12 cases; **Fig. 5e**; **Discussion**).

We then investigated MoA assignments at the level of individual FoVs (**Fig. 5f**). For instance, in the case of Avibactam, out of the 50 FoVs, 45 were classified as Mecillinam and 4 as Meropenem. Since both compounds target PBP 2, Avibactam was correctly classified as a PBP 2 inhibitor. This classification is also consistent with the images, since both Avibactam and Mecillinam induce coccoid morphologies (**Fig. 5f**). Similarly, 20 FoVs of bacteria exposed to Ciprofloxacin were classified as Norfloxacin and 19 FoVs as Levofloxacin which is consistent with the elongated phenotypes induced by these compounds. By contrast, MoA prediction failed for the three PBP 1A/B inhibitors Sulbactam, Cefsulodin and Penicillin G. Penicillin G was incorrectly associated with the DNA synthesis inhibitor Trimethoprim, while Cefsulodin was incorrectly associated with the PBP 3 inhibitor Ceftriaxone (**Fig. 5g**). While these incorrect classifications may represent limitations of our approach, they are consistent with visual inspection of the images, which show that the morphology of bacteria exposed to Penicillin G, a PBP 1A/B inhibitor, resembles that of bacteria treated with Trimethoprim more so than that of bacteria treated with Cefsulodin (the other PBP 1A/B inhibitor that Penicillin G should ideally be associated with). A similar observation can be made about Cefsulodin, a PBP 1A/B inhibitor that produces morphologies more similar to those of the PBP 3 inhibitor Ceftriaxone than those of Penicillin G. Overall, the predicted MoA was correct in 56 out of 72 cases (**Fig. 5h**).

Thus, our data indicate that our CNN can learn to extract embeddings from BF images alone that enable accurate MoA prediction of an unseen drug as long as the MoA is represented by reference

compounds and the phenotypic changes induced by drugs within this MoA are consistent.

Detecting drugs with unseen modes of action

While the above results indicate that known MoAs can be assigned to novel compounds, addressing AMR requires the ability to identify compounds with novel MoAs. To address this, we assessed if our approach is capable of detecting compounds with a MoA that - unlike in our previous analysis - was not seen by our model during training. We designed a leave-one-MoA-out (LOMO) experiment where we singled out one MoA and trained our CNN on BF images of drug-treated bacteria after systematically leaving out all compounds from this MoA (**Fig. 6a**). We then computed outlier scores on image embeddings with a local outlier factor (LOF) algorithm (16) (**Fig. 6b**). To serve as negative controls for novel MoA detection, we randomly removed one compound for each MoA with at least three compounds (**Table 1**) from the training data.

After model training, we obtained embeddings from images of bacteria on a hold-out test plate treated with all compounds except those of the left-out MoA and negative controls at all four concentrations (grey dots in **Fig. 6c**). Next, we defined an image as showing a novel MoA if the outlier score of its embedding exceeded a threshold (and as a known MoA if it did not). We then computed “novelty probabilities” for each left-out compound (at the IC50) by considering the fraction of FoVs that were detected as exhibiting a novel MoA based on their embeddings (**Fig. 6d**). The compounds of the left-out MoA and the negative controls allowed us to quantify true and false positives, respectively. We determined ROC curves by cross-validation across four hold-out test plates and computed mean AUC values for each MoA (**Fig. 6e**).

All compounds targeting PBP 2, DNA gyrase, ribosome and membrane integrity exhibited median novelty probabilities above 0.8 and high or very high AUC scores (PBP 2: 0.98, DNA gyrase: 0.75, ribosome: 0.92, membrane integrity: 0.82), indicating that these compounds with these MoAs induce phenotypes that can be well differentiated from the MoAs seen by the model during training. In the case of PBP 1A/B (AUC=0.76), only Penicillin G and Sulbactam exhibited high novelty probabilities, whereas Cefsulodin did not (**Fig. 6d**). For the DNA gyrase inhibitors, median novelty scores across compounds were consistently high (AUC=0.75), but individual scores by plate exhibited a higher degree of spreading compared to Ribosome or PBP 2 inhibitors. Conversely,

PBP 3 inhibitors consistently showed low novelty probabilities (AUC=0.40), suggesting that these compounds induce phenotypes that do not differ from other MoAs sufficiently strongly to be determined as novel MoAs *de novo*. Note that the PBP 1 inhibitor Cefsulodin was predominantly classified as a PBP 3 inhibitor in our LOCO analysis (**Fig. 5h**) indicating that Cefsulodin induces PBP 3-like morphologies. The consistently low novelty scores observed for PBP 3 inhibitors could thus explain the low novelty scores observed for Cefsulodin.

Taken together, our results show that compounds from all but one of the studied MoAs induce phenotypes that are robustly detected as novel when excluded from the training data. Our approach could thus play a key role in identifying promising compounds with potentially novel MoAs in phenotypic screens.

Discussion

In this study, we introduced a deep learning (DL) approach based on convolutional neural networks (CNNs) that can extract meaningful phenotypic features from microscopy images of bacteria treated with antibiotic compounds and showed that these features can be used for multiple downstream tasks with high relevance for phenotypic drug screening. These tasks include the recognition of MoA or even compound identity directly from images (**Fig. 2**), the detection of weak phenotypes at subinhibitory drug concentrations (**Fig. 4**), the assignment of drug MoAs for novel compounds (**Fig. 5**) and the identification of compounds with novel MoAs (**Fig. 6**). Our work builds on prior studies in eukaryotic cells and bacteria that established cytological profiling to analyse drug-induced phenotypes and group drugs by MoA, whether using handcrafted features (7,8), or more recently, DL approaches (11). However, our study represents a significant extension of these early approaches in several ways.

First, we established that DL can extract both MoA- and compound-specific features from brightfield (BF) images alone, enabling perfect MoA recognition without any fluorescent labelling (**Fig. 3c,e**). This came to us as a surprise, since most phenotypic drug screening assays employ a mixture of fluorescent dyes to characterise subcellular features, as illustrated by the popular Cell Painting assay (9,17,18) or bacterial cell profiling assays (5,7,8,11). Our finding that fluorescent labelling of bacterial membranes and DNA is redundant with BF images is consistent with studies showing that

DL models can be trained to generate realistic fluorescent labelling or coloration (virtual staining) *in silico* from label-free images in eukaryotes (19) and in bacteria respectively, as illustrated e.g. by a recent report that Gram-staining can be predicted from darkfield images of *E. coli* and *Listeria innocua* (20). Our results extend previous work in eukaryotic cells showing that DL applied to BF images performs on par with classical CellProfiler features to determine drug MoA (12) to the case of bacteria. More importantly, our findings enable a drastic simplification of the experimental workflow of phenotypic imaging screens by removing the need for staining and protocol optimisation, as well as avoiding the associated costs and potential confounding factors like variations in fluorescent staining. The removal of staining also opens the door to phenotypic screening in live cells, potentially circumventing fixation issues and enabling the incorporation of live cell dynamics in phenotyping.

Second, although DL-based image recognition for natural images typically requires thousands of training images per class to reach state-of-the-art classification accuracies, we showed that on the order of seven FoVs per condition (i.e. compound and concentration) are sufficient to achieve high accuracies in MoA recognition and that acquiring more images does not yield significant improvements (**Fig. 3g-i**). The main reason for this surprisingly low number is probably the large number of cells contained in each FoV, and the relative simplicity of *E. coli* BF images compared to natural images like those present in the ImageNet dataset. Nevertheless, our observation that pretraining our model on ImageNet significantly reduces the required number of images suggests that many ImageNet features are also relevant for BF *E. coli* images. Our finding that recognition accuracy saturates with only a handful of FoVs should encourage future phenotypic imaging screens to reduce image acquisition time per well and thereby increase the rate of assayed compounds.

Third, we emphasise that our model was trained to predict the combinations of drugs and concentrations and thus has no explicit prior knowledge of the MoA. Despite this, we show that our model implicitly encodes similarities by MoA in its latent space. This is a welcome feature, as it means that a straightforward supervised learning strategy with well-defined and unambiguous labels (drugs and concentrations) can learn biologically meaningful representations from the image data alone.

Fourth, we have shown that DL can robustly detect weak phenotypes induced by drugs at concentrations well below the inhibitory concentrations (detection AUC=0.81±0.08 at 0.125×IC50). This result implies that large chemical libraries at a fixed concentration can be screened with high sensitivity, i.e. with a high likelihood of picking up active compounds that elicit subtle phenotypes in the images without affecting bacterial growth. In a second step, these hit compounds could then be titrated to determine their IC50 values, imaged at IC50, and analysed for MoA assignment with high accuracy at IC50 concentrations for most MoAs (PBP 2: 100%, PBP 3: 91.7%, DNA gyrase: 100%, ribosome: 75%) except PBP 1 A/B (8.3%, **Fig. 2c, 5e**).

Finally, we have addressed the critical task of determining if a novel compound also has a novel MoA. We have shown robust novelty detection for four out of five tested MoAs, with excellent performance (specificity and sensitivity) for PBP 2 and ribosome inhibitors (AUC=0.98 and 0.92, respectively), good to very good performance for inhibitors of PBP 1A/B, DNA gyrase and membrane integrity (AUC=0.76, 0.75, and 0.82, respectively), but acknowledge poor performance for PBP 3 inhibitors (AUC=0.40). Despite imperfections, our novelty detection opens the door to screening large libraries of compounds using high-throughput microscopy to identify compounds that are likely to act through novel MoAs.

We acknowledge several limitations. While MoA assignment in the LOCO analysis was good for compounds belonging to four MoA groups, it is notable that none of the PBP 1A/B inhibitors were correctly assigned. However, inspecting images of *E. coli* treated with different PBP 1A/B inhibitors (**Supp. Fig. 2a**) reveals that these misclassifications are likely due to inconsistent phenotypes induced by these compounds (**Fig. 5g**). For instance, bacteria treated with Penicillin G exhibit long filamentous morphologies similar to those induced by Trimethoprim, while bacteria treated with Cefsulodin exhibit shorter filaments and are thus more similar to morphologies observed for PBP 3 inhibitors. We speculate that these differences might be due to these compounds exhibiting different PBP affinities. Cefsulodin has been reported to have high affinity for PBP 1A (0.47 µg/mL) and PBP 1B (3.7 µg/mL), but does neither target PBP 2 (>250 µg/mL) nor PBP 3 (>250 µg/mL) (21). By contrast, Penicillin G has reportedly high affinity for a variety of PBPs (PBP 1A: 0.5 µg/mL, PBP 1B: 3.0 µg/mL, PBP 2: 0.8 µg/mL, PBP 3: 0.9 µg/mL and PBP 4: 1.0 µg/mL) (22) plausibly causing inconsistencies in phenotype. Alternatively, the PBPs of the *E. coli* strain used here might have

different affinities to Cefsulodin and Penicillin G than those reported in the literature. The affinity of PBPs to beta-lactams can indeed vary between strains due to their individual history of pre-exposure to beta-lactams. Additionally, our study included four beta-lactamase inhibitors (Sulbactam, Avibactam, Clavulanate and Relebactam) in the context of this study. Sulbactam has been shown to preferentially bind PBP 1A/B, while Avibactam and Clavulanate have been shown to target PBP 2. Relebactam, however, has not been studied in *E. coli* and our LOCO analysis indicates that this drug induces PBP 2-like phenotypes similar to Avibactam and Clavulanate (**Fig. 5h**). This stands in sharp contrast to prior work showing that Relebactam preferentially binds PBP 1B and PBP 4 in *Pseudomonas aeruginosa* (23). Thus, some or all of the MoA assignments in our LOCO analyses may potentially arise from differences in PBP binding affinities of different beta-lactamase inhibitors between species or even strains of the same species. Intriguingly, while Trimethoprim is commonly classified as a DNA synthesis inhibitor and specifically targets folate synthesis, some studies have shown that Trimethoprim also directly targets the peptidoglycan pathway (24,25). Indeed, peptidoglycan synthesis relies on nucleotide-activated sugars such as UDP-N-acetylglucosamine and UDP-N-acetyl muramic acid. Inhibition of nucleotide synthesis by Trimethoprim would thus also affect peptidoglycan synthesis and more generally cell wall biogenesis that relies on many nucleotide-activated sugars including Lipopolysaccharides. The assignment of Penicillin G to Trimethoprim appears consistent with these prior findings. Taken together, these observations highlight the limitation of grouping compounds into discrete MoA classes and using reference compounds that inhibit multiple targets. One promising avenue to overcoming this limitation is to predict drug targets by comparing their induced phenotypes to those of large mutant libraries. As another caveat, our results are limited to *E. coli*, which are relatively large and exhibit strong morphological changes in response to drug exposure. Future work should thus explore if our BF-only methodology extends to other bacterial species with more subtle morphological phenotypes such as *Staphylococcus aureus* or *Acinetobacter baumannii*.

Overall, our work is a step towards deploying state-of-the-art computational analysis methods, notably DL-based analysis of microscopy images, to bacterial phenotypic screens. We propose that our CNN approach and the downstream analysis methods described here could form the basis for a comprehensive phenotypic screening pipeline to discover novel classes of antibiotics to fight AMR.

Methods

Bacterial growth protocol for *E. coli*

E. coli K-12 MG1655 was grown in 5 mL of cation adjusted Mueller-Hinton Broth II (MHBII, Fisher Scientific cat. 11800482) overnight at 37°C until stationary phase. The culture was then diluted to 0.05 OD in fresh MHBII and grown at 37°C for 1h 45min until mid-exponential phase between 0.5 to 0.7 OD before further dilution. IC50s were determined via broth serial dilutions in 96-well plates with a starting inoculum density of 0.05 OD in a total volume 100µL. Micro dilution plates were incubated at 37°C, 180 rpm for 18h. The final cell density was measured on a Tecan Spark Multimode Plate Reader and plotted on GraphPad Prism to calculate the IC50 values.

For microscopy, antibiotics were dispensed into 96-well plates in nanoliter volumes using an Echo liquid handler (Beckman Coulter) before adding 100µL of bacteria from mid-exponential growth diluted to 0.05 OD. The plates were incubated at 37°C, 180 rpm for 2h 15 min. Then, 20µg/mL of FM4-64FX (Invitrogen cat. F34653) was added and was further incubated at 37°C, 180 rpm for another 15 min. The supernatant was then removed after centrifugation at 4000 x g for 5 min. 100µL of fixative (2.8% paraformaldehyde, 0.04% glutaraldehyde in PBS) was added and the plates were incubated at room temperature at 800 rpm for 30 min. Cells were washed twice with 200µL of PBS by centrifugation and resuspension. Then, cells were suspended in 100µL of PBS containing 10µg/mL of Hoechst 33342 (Invitrogen cat. H1399) and transferred to PhenoPlates (Revvity cat. 6055300) to determine the OD at 600nm. Finally, the entire plate was diluted based on the median OD of the untreated wells to a calculated OD of 0.036 in 100µL. The fixed and diluted cells were stored at 4°C until imaging on the Opera Phenix Plus (Revvity).

High-throughput imaging

Prior to imaging, plates stored at 4°C were left for at least 20 min at room temperature, and then centrifuged at 500 rpm for 10 min to bring floating bacteria to the well bottom. Images were then acquired using an automated Opera Phenix Plus High-Content Screening System (Revvity) with 63×/NA 1.15 water immersion objective in widefield mode without binning. Two imaging sequences were set up: one for the simultaneous acquisition of fluorescent markers through two different

cameras, and one for the BF acquisition via a dedicated camera. In total, 50 FoVs were acquired in non-overlapping randomly chosen positions of the whole area of each well. These random positions in each well were the same on individual plates, but varied between plates. Images were acquired with exposure times of 80 ms for BF, 400 ms for FM4-64FX and 250 ms for Hoechst.

Image preprocessing

Images for training were obtained by converting high-throughput images from 16-bit to 8-bit and clipping pixel intensities by percentile between 0.1 and 99.9 to remove spurious saturated pixels. Images from three out of four plates were then randomly split into training (80%) and validation (20%) set. During training, full-resolution images with 2,160×2,160 pixels were passed through a data augmentation pipeline: images were randomly flipped, rotated and random changes in brightness, contrast, saturation and hue (for images with more than one channel) were applied. To account for potential edge effects due to random rotation, full images were further centre cropped (1,528×1,528 pixels). Next, from this centre-crop, a random crop of size 1,500×1,500 pixels was obtained and images subdivided into nine tiles of size 500×500 pixels. Tiles were then resized to 256×256 pixels and pixel intensities normalised between -1 and 1.

Deep learning model training

The DL model was trained for 250 epochs using an Adam (26) optimiser with a learning rate of 0.001. A batch-size of 16 images was used. Note that the effective batch-size was 144 (16×9), since each image was further subdivided into nine tiles. During training, tiles were passed through an EfficientNet-B0 backbone, with weights initialised by pretraining on ImageNet[1] (except where indicated otherwise), and latent feature vectors corresponding to individual tiles aggregated by averaging to obtain a global feature vector for the full-resolution input image. The model was trained to predict the combination of antibiotic treatment and concentration as a multi-class classification task with a cross-entropy loss using L2 regularisation with $\lambda=0.001$ to avoid overfitting. Models were trained with PyTorch 2.3.1 (27) in Python 3.9 on the HPC cluster at Institut Pasteur Paris. See **Supp. File 1** for the full model architecture.

[1] https://download.pytorch.org/models/efficientnet_b0_rwightman-7f5810bc.pth

Estimating minimum number of required images

To estimate the minimum number of required training images, we fit an exponential $f(x) = a(1 - e^{-bx})$ (where x is the number of images per condition and $f(x)$ the hold-out test accuracy) to mean MoA accuracies obtained from hold-out test data with models trained on different numbers of images per well ranging from 3 to 120. From these exponential fits, the plateau a was obtained and the number of images per well estimated to achieve 95% of this MoA classification accuracy plateau. We used the curve fitting function from the scipy package in Python 3.9 (28).

MoA assignment with k-NN classifier

Embeddings from images of previously seen conditions were obtained from a hold-out test plate. A k-NN classifier was then applied to these embeddings corresponding to the highest concentration (IC50) using MoA labels as targets with the scikit-learn package (29) in Python 3.9. The number of k neighbours was set to 150 using cosine distance (1 minus the cosine similarity) as the metric and weighing points by the inverse of their distance thereby ensuring that closer neighbours had a greater influence on the MoA assignment than neighbours which were further away. To predict the MoA, embeddings of previously unseen compounds were obtained from the same hold-out test plate. This analysis was repeated for each model with a left-out compound across four hold-out test plates.

Novelty detection with local outlier factor algorithm

Embeddings from images of previously seen conditions were obtained from a hold-out test plate. A LOF classifier was applied to these embeddings using all available concentrations with the scikit-learn package in Python 3.9. The number of k neighbours was set to k=9 and cosine distance (1 minus the cosine similarity) was used as the distance metric. To determine if a compound was an outlier, i.e. had a novel MoA, the local outlier factor on embeddings from unseen conditions at the highest concentration (IC50) was computed and a threshold applied as described in (16). For each compound, novel MoA detection probabilities were computed as the fraction of detected outliers considering all FoVs. ROC curves were obtained by computing the local outlier factor for previously unseen compounds and embeddings corresponding to left-out negative control conditions (one per

MoA with at least three compounds) and positive controls (i.e. compounds with the left-out MoA). AUC values were computed with the scikit-learn package. This analysis was repeated for each model with a left-out MoA across four hold-out test plates.

Acknowledgements

We thank C. Thépenier, C. Spahn, S. Jang, D. Shum and N. Aulner for insightful comments on the manuscript. We thank J. Chiaravalli and A. Zettor for preparing plates with an acoustic liquid handler. D.K. was funded by the Pasteur-Paris University International doctoral program (PPU), the INCEPTION program (Investissement d'Avenir grant ANR-16-CONV-0005) and a Fondation pour la Recherche Médicale (FRM) Fin de Thèse (FDT) grant. K.K. was supported by the INCEPTION program, the Pasteur-Roux-Cantarini Postdoctoral Fellowship (Institut Pasteur) and the Fondation pour la Recherche Médicale (FRM) grant EQU202403018034. J.P. was funded by Fondation pour la Recherche Médicale (grant number EQU202303016284 to Pedro Alzari). Work in the Boneca lab was supported by the Fondation pour la Recherche Médicale (FRM) grant EQU202403018034. The UTechS Photonic BioImaging, C2RT, Institut Pasteur, is supported by the French National Research Agency (France BioImaging, ANR-24-INBS-0005 FBI (BIOGEN); Investments for the Future). We acknowledge Institut Pasteur and the Région Île-de-France (DIM1Health program) for the use of the Opera Phenix system. We also acknowledge the INCEPTION program for funding a GPU farm used in this work.

Bibliography

1. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399,** 629–655 (2022).
2. O'Neill, J. Tackling drug-resistant infections globally: final report and recommendations. (2016).
3. Brown, D. Antibiotic resistance breakers: can repurposed drugs fill the antibiotic discovery void? *Nat. Rev. Drug Discov.* **14,** 821–832 (2015).
4. Krentzel, D., Shorte, S. L. & Zimmer, C. Deep learning in image-based phenotypic drug discovery. *Trends Cell Biol.* **33,** 538–554 (2023).
5. Zampaloni, C. *et al.* A novel antibiotic class targeting the lipopolysaccharide transporter. *Nature* **625,** 566–571 (2024).
6. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).
7. Nonejuie, P., Burkart, M., Pogliano, K. & Pogliano, J. Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. *Proc. Natl. Acad. Sci.* **110,** 16169–16174 (2013).
8. Zoffmann, S. *et al.* Machine learning-powered antibiotics phenotypic drug discovery. *Sci. Rep.* **9,** 5013 (2019).
9. Bray, M.-A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. 18 (2016).
10. Moshkov, N. *et al.* Learning representations for image-based profiling of perturbations. *Nat. Commun.* **15,** 1594 (2024).
11. Quach, D. *et al.* Deep learning–driven bacterial cytological profiling to determine antimicrobial mechanisms in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci.* **122,** e2419813122 (2025).
12. Harrison, P. J. *et al.* Evaluating the utility of brightfield image data for mechanism of action prediction. *PLOS Comput. Biol.* **19,** e1011323 (2023).
13. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. at http://arxiv.org/abs/1905.11946 (2020)
14. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* 770–778 (2016). doi:10.1109/CVPR.2016.90

15. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.* 248–255 (2009). doi:10.1109/CVPR.2009.5206848
16. Breunig, M. M., Kriegel, H.-P., Ng, R. T. & Sander, J. LOF: identifying density-based local outliers. in *Proc. 2000 ACM SIGMOD Int. Conf. Manag. Data* 93–104 (Association for Computing Machinery, 2000). doi:10.1145/342009.335388
17. Cimini, B. A. *et al.* Optimizing the Cell Painting assay for image-based profiling. *Nat. Protoc.* **18,** 1981–2013 (2023).
18. Willis, C., Nyffeler, J. & Harrill, J. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. *SLAS Discov.* **25,** 755–769 (2020).
19. Cross-Zamirski, J. O. *et al.* Label-free prediction of cell painting from brightfield images. *Sci. Rep.* **12,** 10001 (2022).
20. Işıl, Ç. *et al.* Virtual Gram staining of label-free bacteria using dark-field microscopy and deep learning. *Sci. Adv.* **11,** eads2757 (2025).
21. Fontana, R., Cornaglia, G., Ligozzi, M. & Mazzariol, A. The final goal: penicillin-binding proteins and the target of cephalosporins. *Clin. Microbiol. Infect.* **6,** 34–40 (2000).
22. Curtis, N. A., Orr, D., Ross, G. W. & Boulton, M. G. Affinities of penicillins and cephalosporins for the penicillin-binding proteins of Escherichia coli K-12 and their antibacterial activity. *Antimicrob. Agents Chemother.* **16,** 533–539 (1979).
23. Montaner, M., Lopez-Argüello, S., Oliver, A. & Moya, B. PBP Target Profiling by β-Lactam and β-Lactamase Inhibitors in Intact Pseudomonas aeruginosa: Effects of the Intrinsic and Acquired Resistance Determinants on the Periplasmic Drug Availability. *Microbiol. Spectr.* **11,** e03038-22 (2022).
24. Daschner, F. Inhibition of Cell Wall Synthesis by Sulfonamides and Trimethoprim. *Chemotherapy* **22,** 12–18 (2009).
25. Yaeger, Luke N, Sychantha, David & Luong, Princeton. Metabolic connections between folate and peptidoglycan pathways in Pseudomonas aeruginosa inform rational design of a dual-action inhibitor. *bioRxiv* 2023--11 (2023).
26. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).

at <http://arxiv.org/abs/1412.6980>

27. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. 12
28. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17,** 261–272 (2020).
29. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON*
30. Kocaoglu, O. & Carlson, E. E. Profiling of β-Lactam Selectivity for Penicillin-Binding Proteins in Escherichia coli Strain DC2. *Antimicrob. Agents Chemother.* **59,** (2015).
31. Shirley, J. D., Nauta, K. M. & Carlson, E. E. Live-Cell Profiling of Penicillin-Binding Protein Inhibitors in Escherichia coli MG1655. *ACS Infect. Dis.* **8,** 1241–1252 (2022).
32. Moosdeen, F., Williams, J. D. & Yamabe, S. Antibacterial characteristics of YTR 830, a sulfone beta-lactamase inhibitor, compared with those of clavulanic acid and sulbactam. *Antimicrob. Agents Chemother.* **32,** 925–927 (1988).
33. Asli, A., Brouillette, E., Krause, K. M., Nichols, W. W. & Malouin, F. Distinctive Binding of Avibactam to Penicillin-Binding Proteins of Gram-Negative and Gram-Positive Bacteria. *Antimicrob. Agents Chemother.* **60,** (2016).
34. Davies, T. A. *et al.* Binding of Ceftobiprole and Comparators to the Penicillin-Binding Proteins of Escherichia coli, Pseudomonas aeruginosa, Staphylococcus aureus, and Streptococcus pneumoniae. *Antimicrob. Agents Chemother.* **51,** 2621–2624 (2007).
35. Pucci, M. J., Boice-Sowek, J., Kessler, R. E. & Dougherty, T. J. Comparison of cefepime, cefpirome, and cefaclidine binding affinities for penicillin-binding proteins in Escherichia coli K-12 and Pseudomonas aeruginosa SC8329. *Antimicrob. Agents Chemother.* **35,** 2312–2317 (1991).
36. Jardetzky, O. & Julian, G. R. Chloramphenicol Inhibition of Polyuridylic Acid Binding to E. coli Ribosomes. *Nature* **201,** 397–398 (1964).
37. 懸川友人, 小川達洋 & 廣瀬聖雄. Te-031の蛋白質合成阻害機作. *Chemotherapy* **36,** 123–128 (1988).
38. Chukwudi, C. U. & Good, L. Doxycycline inhibits pre-rRNA processing and mature rRNA formation in E. coli. *J. Antibiot. (Tokyo)* **72,** 225–236 (2019).
39. Misumi, M., Nishimura, T., Komai, T. & Tanaka, N. Interaction of kanamycin and related

antibiotics with the large subunit of ribosomes and the inhibition of translocation. *Biochem. Biophys. Res. Commun.* **84,** 358–365 (1978).

40. Vance-Bryan, K., Guay, D. R. P. & Rotschafer, J. C. Clinical Pharmacokinetics of Ciprofloxacin. *Clin. Pharmacokinet.* **19,** 434–461 (1990).
41. Tunitskaya, V. L., Khomutov, A. R., Kochetkov, S. N., Kotovskaya, S. K. & Charushin, V. N. Inhibition of DNA Gyrase by Levofloxacin and Related Fluorine-Containing Heterocyclic Compounds. *Acta Naturae* **3,** 94–99 (2011).
42. Shen, L. L., Kohlbrenner, W. E., Weigl, D. & Baranowski, J. Mechanism of quinolone inhibition of DNA gyrase: Appearance of unique norfloxacin binding sites in enzyme-DNA complexes. *J. Biol. Chem.* **264,** 2973–2978 (1989).
43. McClure, W. R. & Cech, C. L. On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.* **253,** 8949–8956 (1978).
44. Baker, D. j. *et al.* The binding of trimethoprim to bacterial dihydrofolate reductase. *FEBS Lett.* **126,** 49–52 (1981).
45. Sabnis, A. *et al.* Colistin kills bacteria by targeting lipopolysaccharide in the cytoplasmic membrane. *eLife* **10,** e65836 (2021).
46. Dixon, R. A. & Chopra, I. Polymyxin B and polymyxin B nonapeptide alter cytoplasmic membrane permeability in Escherichia coli. *J. Antimicrob. Chemother.* **18,** 557–563 (1986).

Tables

Table 1: Antibiotics used on E. coli

MoA		Antibiotic	IC50 (µg/ml)	MoA reference
Cell wall	***PBP 1A/B***	Cefsulodin	23.77	(22,30)
		Penicillin G	25.65	(22,31)
		Sulbactam	15.62	(32)
	PBP 2	Avibactam	19.20	(31)
		Mecillinam	0.3009	(22,30,33)
		Meropenem	0.1085	(31,34)
		Clavulanate	17.24	(32)
		Relebactam	64	N/A for *E. coli*
	PBP 3	Aztreonam	0.01568	(31)
		Cefepime	0.007517	(31,35)
		Ceftriaxone	0.004	(31)
Ribosome		Chloramphenicol	3.079	(36)
		Clarithromycin	22.24	(37)
		Doxycycline	0.3703	(38)
		Kanamycin	2.529	(39)
DNA gyrase		Ciprofloxacin	0.005723	(40)
		Levofloxacin	0.01089	(41)
		Norfloxacin	0.02538	(42)
RNA polymerase		Rifampicin	3.358	(43)
DNA synthesis		Trimethoprim	0.1335	(44)
Membrane integrity		Colistin	0.6465	(45)
		Polymyxin B	1.166	(46)

Figures

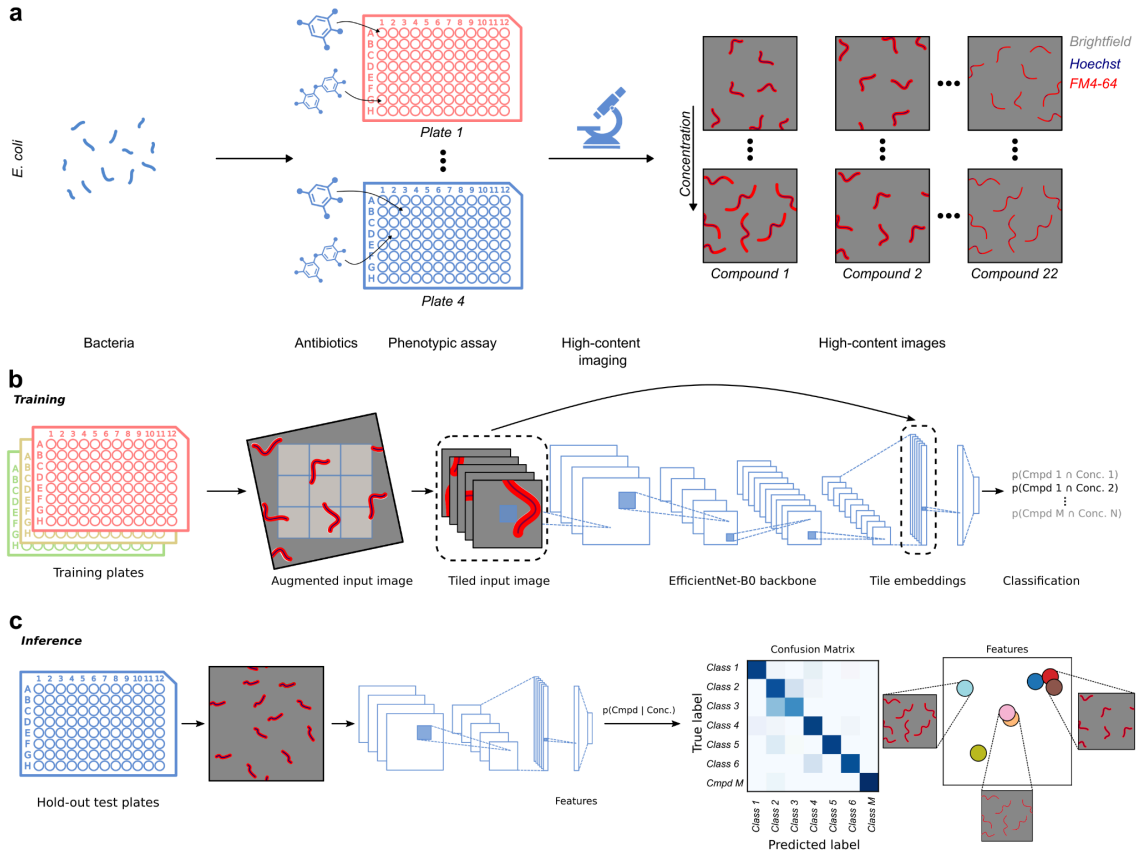

Fig. 1. Data acquisition, model training and inference. (**a**) *E.coli* bacteria were grown to exponential phase in culture, antibiotics were distributed onto four 96-well plates at four different concentrations with an acoustic liquid handler and bacteria dispatched onto the well plates. Bacteria were then grown with antibiotics for 2h15min and fixed. Plates were imaged on a high-content microscope with a 63× water-immersion objective in non-confocal mode with three channels (BF, FM4-64FX and Hoechst). (**b**) The model was trained on high-throughput images from three out of four plates to predict the combination of antibiotic and concentration with a cross-entropy loss. Images were first passed through a data augmentation pipeline and then tiled. Each tile was then passed through the same EfficientNet-B0 backbone whereupon embedding vectors from each image tile were aggregated to obtain a global embedding vector. (**c**) Testing was performed on a hold-out test plate with predictions computed at a given concentration and embedding vectors visualised in 2D with UMAP.

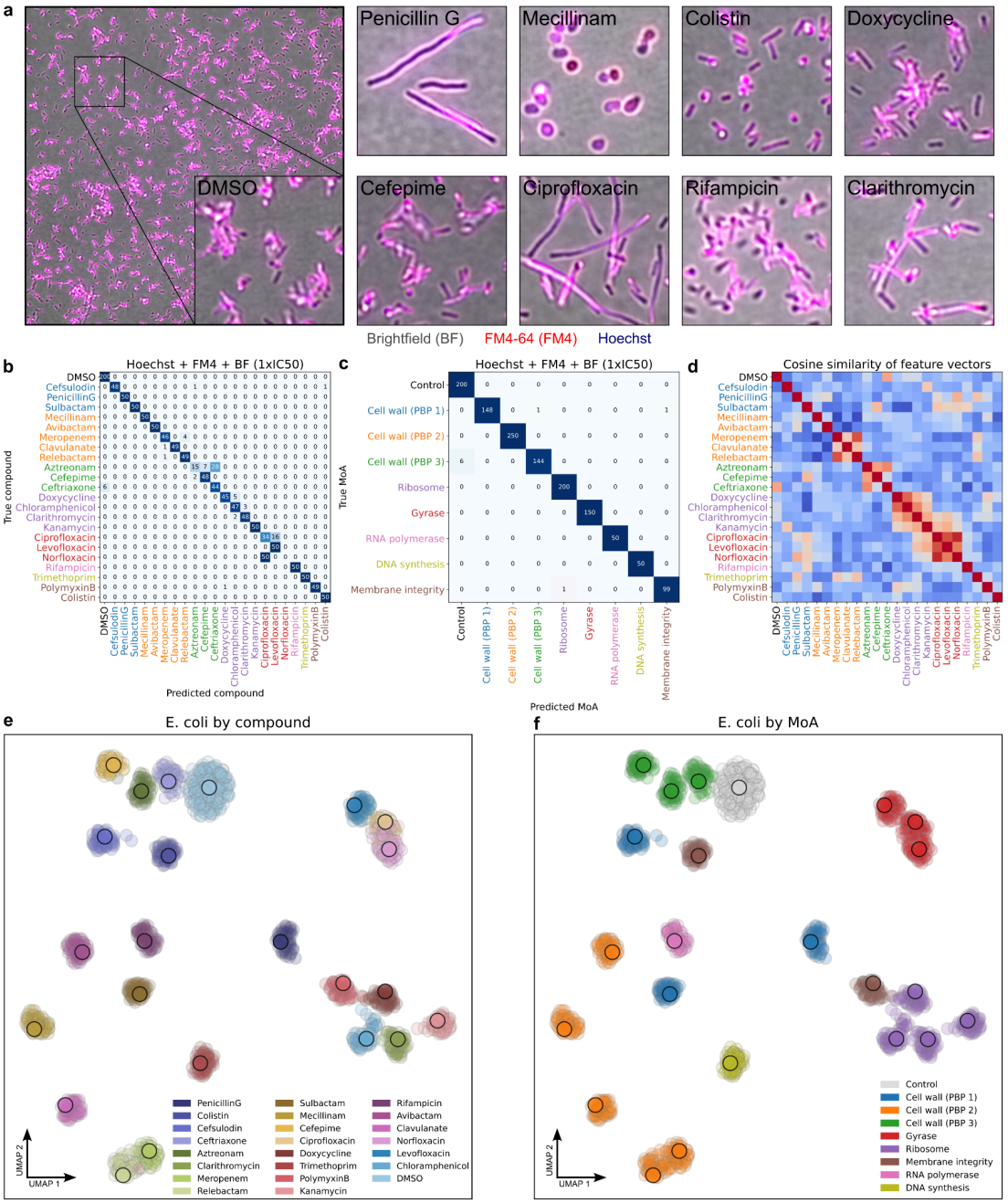

Fig. 2. Deep learning recognises compounds and their MoA from *E. coli* images. (**a**) Representative images from the test plate used to generate results with BF shown in grey, FM4-64FX in red and Hoechst in blue. See **Supp. Fig. 2a-f** for more examples. (**b**) Confusion matrix showing predictions from high-throughput images of bacteria treated with the same relative concentration (IC50) from a model trained on all imaging channels and concentrations. (**c**) Confusion matrix by MoA was obtained by pooling predictions of compounds belonging to the same MoA (**Table 1**). (**d**) Cosine similarities between embedding vectors obtained from bacterial high-throughput images were computed on median feature vectors. High cosine similarity (red) indicates phenotypic similarity between conditions and tends to occur for antibiotics belonging to the same MoA (e.g. Ciprofloxacin, Norfloxacin and

Moxifloxacin target the DNA Gyrase). (**e**) UMAP was used to project high-dimensional feature vectors to 2D. Small dots correspond to individual high-throughput images and large dots to median feature vectors. Dots are coloured by antibiotic compounds. (**f**) UMAP with dots coloured by MoA shows that antibiotics belonging to the same MoA tend to co-localise in latent space (e.g. PBP 3 inhibitors in green and DNA gyrase inhibitors in red).

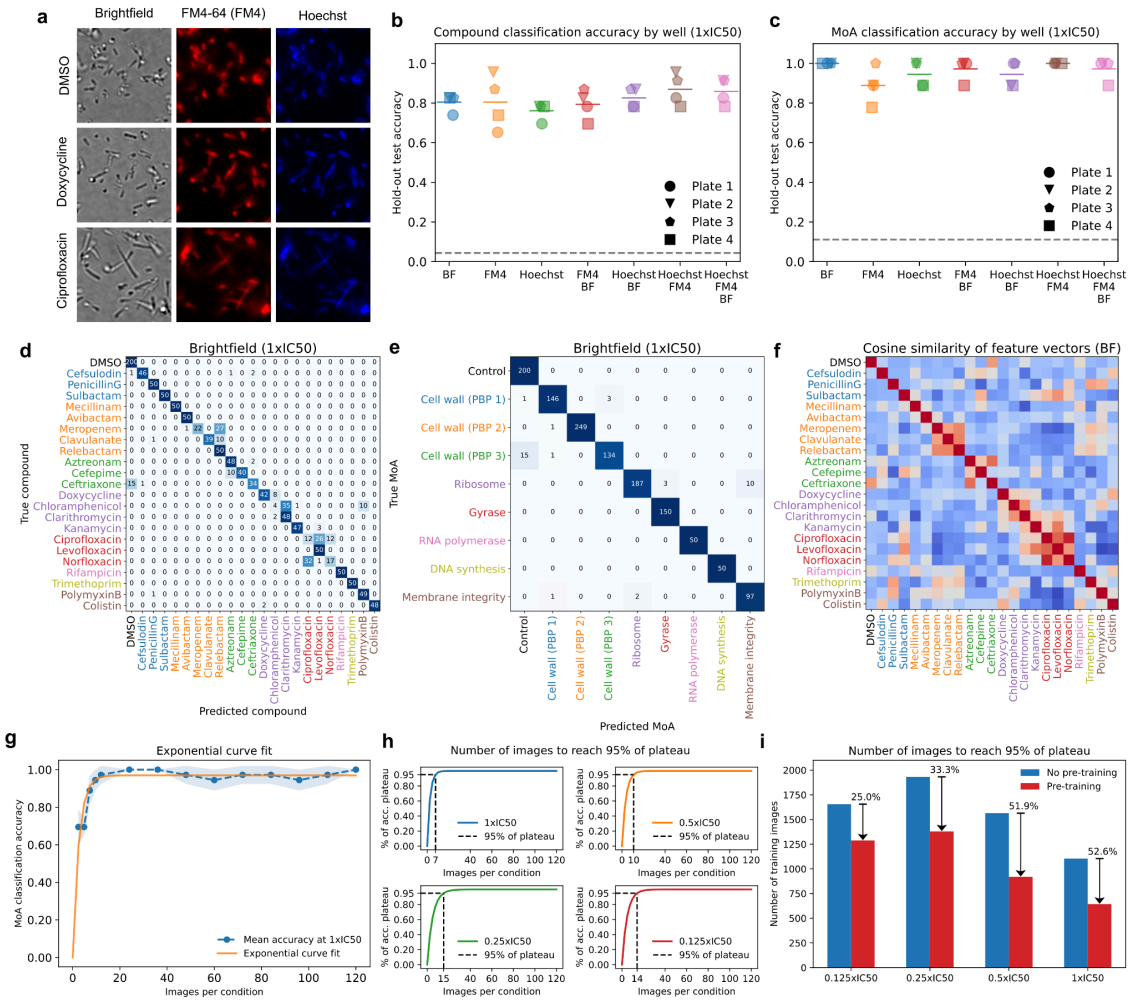

Fig. 3. Classification performance for different numbers of imaging channels and training images. (**a**) Example images split by imaging channel. (**b**) Models were trained in a cross-validation experiment with three plates using different combinations of imaging channels as input and tested on a hold-out test plate. Per-well compound classification accuracies were obtained by aggregating predictions from all images acquired in the same well of bacteria treated at the same relative concentration (IC50). The dashed line indicates random classification. (**c**) Per-well MoA classification accuracies were obtained by pooling per-well compound predictions. Models trained on BF images alone perform on par with models trained on Hoechst and BF images. The dashed line indicates random classification. (**d**) Confusion matrix showing predictions from high-throughput images of bacteria treated with the same relative concentration (IC50) from a model trained only on BF images. (**e**) Confusion matrix by MoA was obtained by pooling predictions of compounds belonging to the same MoA (**Table 1**). (**f**) Cosine similarities between embedding vectors obtained from only BF images were computed on median feature vectors. High cosine similarity (red) indicates phenotypic similarity between conditions and is observed for antibiotics belonging to the same MoA (e.g. Meropenem, Clavulanate and Relebactam target PBP 2). (**g, h**) MoA classification accuracy as function of the number of training images. The number of images required to reach 95% MoA

classification accuracy was estimated by fitting a two-parameter exponential function to averaged MoA classification curves with an example shown in (**g**) for predictions obtained from images of bacteria treated at IC50. (**h**) The number of images per condition required to reach 95% of the accuracy plateau was computed for each concentration showing that between seven and 15 images are required. (**i**) Bars show the number of training images required to reach 95% of the accuracy plateau with (red) or without (blue) pretraining on ImageNet evaluated for different concentrations by cross-validation on four hold-out test plates. The number of required training images ranged from 644 (IC50) to 1,380 (0.25×IC50) with pretraining. All models were trained on BF images only.

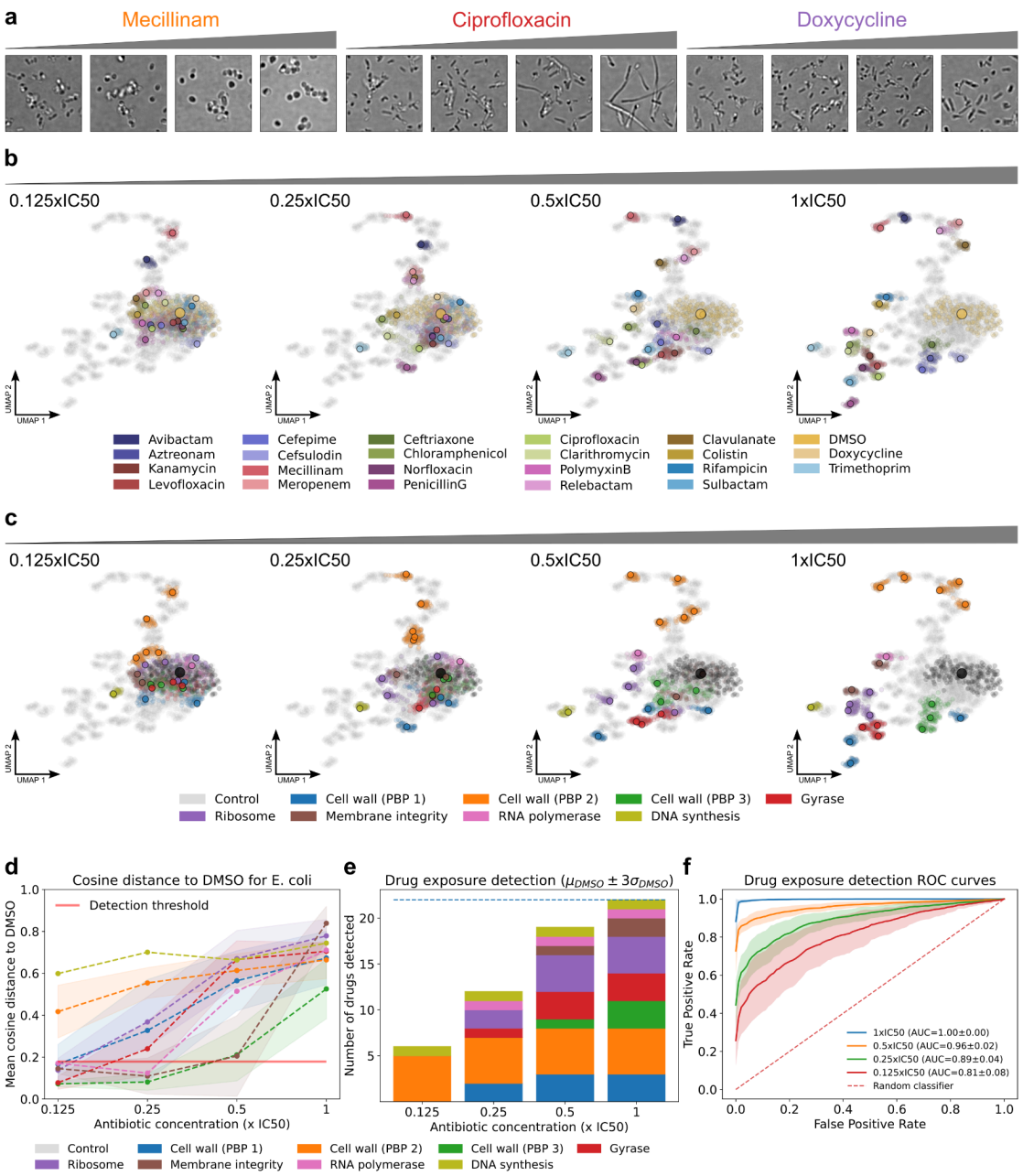

Fig. 4. Detecting drug exposure at subinhibitory concentrations. (**a**) Example images of bacteria treated at increasing concentrations show dose-dependent phenotypic changes. (**b**) Feature vectors of bacteria treated at four different concentrations were jointly projected with UMAP and dots corresponding to a given concentration highlighted in each plot. Small dots correspond to individual high-throughput images and larger dots correspond to median feature vectors. Dots are coloured by antibiotics. (**c**) UMAP plots coloured by MoA show that, except for antibiotics targeting PBP 2 and DNA synthesis, feature vectors tend to be further away from DMSO as concentration increases. (**d**) The cosine distance between median feature vectors and DMSO was computed for each antibiotic class at different concentrations. Average distances by MoA are shown with dashed lines and shaded

areas indicate the standard deviation. A detection threshold (red line) was computed from the intra-class cosine distances of DMSO feature vectors (mean plus three standard deviations). (**e**) Drug exposure detection was determined by evaluating cosine distances between median feature vectors and DMSO with respect to the detection threshold. The number of detected drugs at each relative concentration is shown and bars are coloured by the MoA of the respective antibiotics. The dashed line indicates the overall number of antibiotics. (**f**) ROC curves for drug exposure detection at different concentrations cross-validated on four hold-out test replicates with corresponding AUC values are shown.

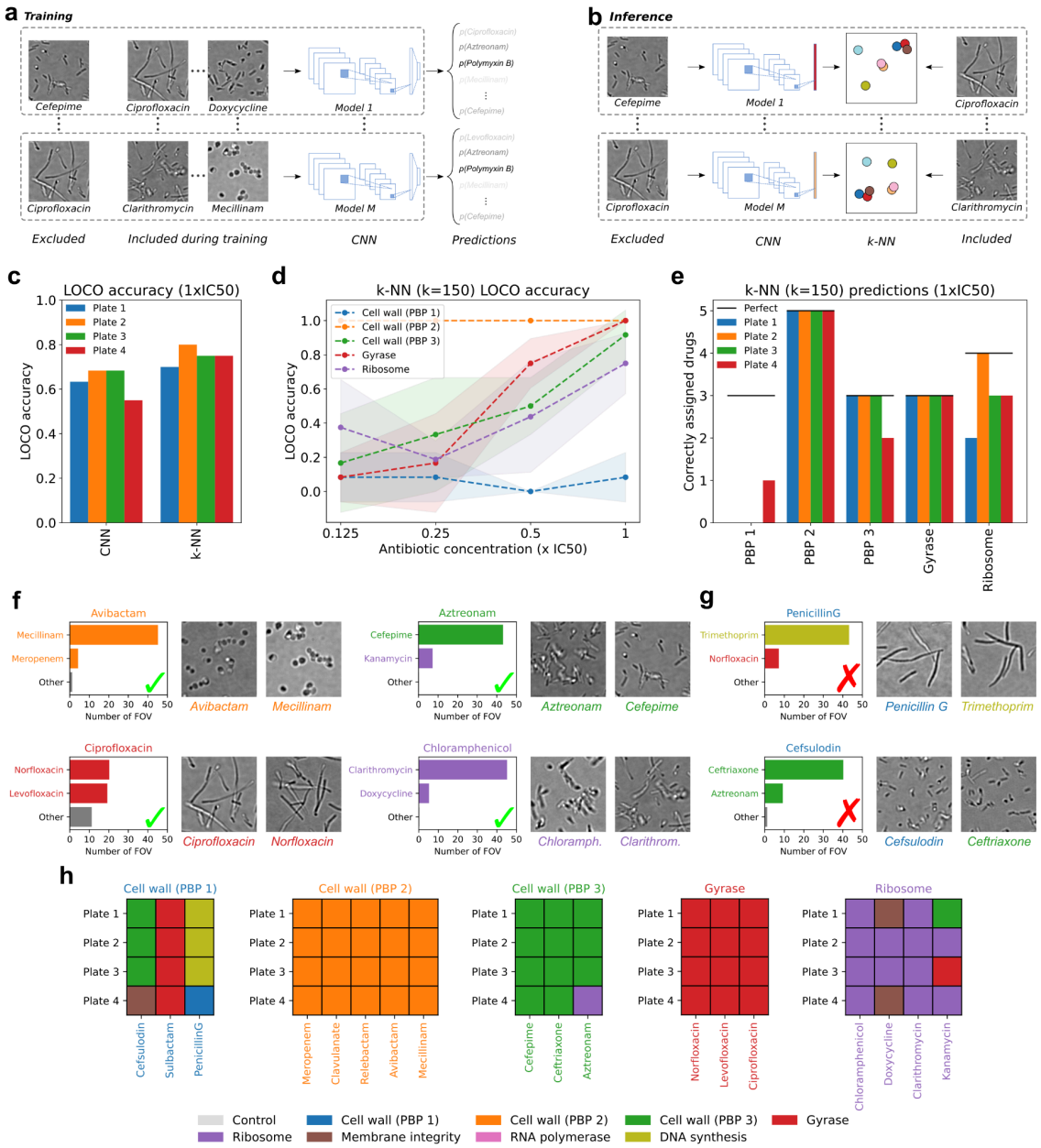

Fig. 5. Assigning MoA of unseen compounds. (**a**) All images belonging to one antibiotic were systematically left out from the training data and models were trained in a cross-validated setting on three out of four plates. (**b**) The MoA of images belonging to the left-out antibiotic class was then determined via k-NN classification using the image embeddings of previously seen antibiotics on a hold-out test replicate. (**c**) Bars show the leave-one-compound-out (LOCO) MoA prediction accuracy at IC50 for each hold-out test plate, using either the CNN predictions (left: 63.8±5.4%) or the k-NN classifier (right: 75.0±3.5%). (**d**) The LOCO MoA accuracy was computed for images of bacteria exposed to different antibiotic concentrations with the k-NN classifier. The dashed lines show the average LOCO accuracy for each MoA and shaded areas indicate the standard deviation. (**e**) The number of correctly assigned antibiotics using k-NN classification is shown for all four hold-out

test plates by MoA. The black horizontal lines indicate perfect assignment. (**f**) Examples of correctly assigned antibiotics showing the assignment of individual images to previously seen antibiotics with a k-NN classifier (k=150). Image crops of bacteria exposed to the left-out antibiotic and bacteria treated with antibiotics most commonly associated are shown. (**g**) Examples of incorrectly assigned antibiotics according to **Table 1** with image crops showing bacteria exposed to the left-out antibiotic and bacteria treated with the antibiotics most commonly associated. (**h**) LOCO predictions are shown for each hold-out test plate and antibiotics with colours indicating the assigned MoA.

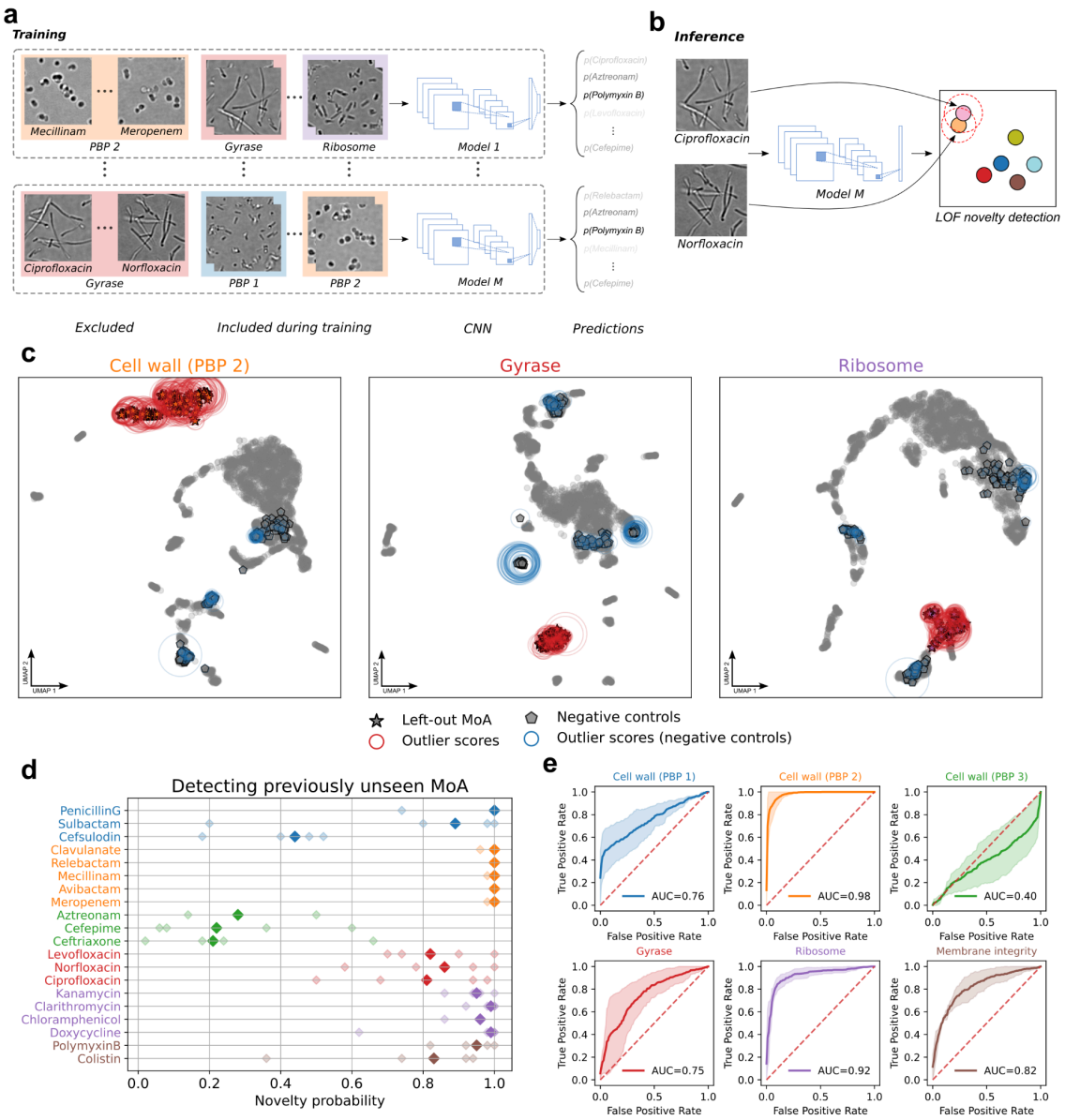

Fig. 6. Detecting antibiotics with novel MoAs. (**a**) All images belonging to the same MoA were systematically left out from the training data and models were trained in a cross-validated setting on three out of four plates. (**b**) Embeddings from images of the left-out MoA were then obtained and outlier scores assigned with a LOF algorithm. (**c**) UMAPs of embeddings for models trained on compounds from all MoAs except PBP 2 (left), DNA gyrase (centre) and ribosome (right). Unsupervised outlier detection was conducted using LOF on embeddings of bacteria exposed to all concentrations (grey dots). Embeddings of compounds from the held-out MoA are shown as stars, negative controls as pentagons and their outlier scores as red and blue circles respectively. (**d**) For each compound, probability scores of MoA novelty were computed by thresholding outlier scores and computing the fraction of detected embeddings (one for each MoA). Pale symbols show the novelty detection probability for each plate and the median is shown in bright colours. Colours indicate the MoA. (**f**) Cross-validated receiver operating characteristic (ROC) curves were obtained for

each held-out MoA on four hold-out test plates with mean AUC values shown.