bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# LEARNING THE RULES OF CELL COMPETITION WITHOUT PRIOR SCIENTIFIC KNOWLEDGE

**Christopher J. Soelistyo**[1,2] **Giulia Vallardi**[2] **Guillaume Charras**[1,3,4]

**Alan R. Lowe**[1,2,4,5*]

[1]Institute for the Physics of Living Systems
[2]Department of Structural and Molecular Biology
[3]Department of Cell and Developmental Biology
[4]London Centre for Nanotechnology
University College London, Gower St, London, WC1E 6BT, UK

[5]The Alan Turing Institute
Euston Rd, London NW1 2DB, UK

## ABSTRACT

Deep learning is now a powerful tool in microscopy data analysis, and is routinely used for image processing applications such as segmentation and denoising. However, it has rarely been used to directly learn mechanistic models of a biological system, owing to the complexity of the internal representations. Here, we develop an end-to-end machine learning model capable of learning the rules of a complex biological phenomenon, cell competition, directly from a large corpus of time-lapse microscopy data. Cell competition is a quality control mechanism that eliminates unfit cells from a tissue and during which cell fate is thought to be determined by the local cellular neighborhood over time. To investigate this, we developed a new approach ($\tau$-VAE) by coupling a variational autoencoder to a temporal convolution network to predict the fate of each cell in an epithelium. Using the $\tau$-VAE's latent representation of the local tissue organization and the flow of information in the network, we decode the physical parameters responsible for correct prediction of fate in cell competition. Remarkably, the model autonomously learns that cell density is the single most important factor in predicting cell fate – a conclusion that has taken over a decade of traditional experimental research to reach. Finally, to test the learned internal representation, we challenge the network with experiments performed in the presence of drugs that block signalling pathways involved in competition. We present a novel discriminator network that, using the predictions of the $\tau$-VAE, can identify conditions which deviate from the normal behaviour, paving the way for automated, mechanism-aware drug screening.

## Introduction

Cell competition is a phenomenon that results in the elimination of less fit cells from a tissue – a critical process in development, homeostasis and disease [1]. The viability of loser cells depends strongly on context: when they are cultured alone, they thrive, but when in a mixed population, they are eliminated by cells with greater fitness (**Fig 1a**). In development, competition acts as a quality control mechanism and also participates in pattern formation [2]. In cancer, competition has been hypothesised to underlie the heterogeneity in cell types present in tumours and promote the emergence of the most aggressive cells [3]. However, the rules that determine individual cell fate are poorly understood. A number of mechanisms of cell competition have been identified to date involving either biochemical competition (for

*To whom correspondence should be addressed: a.lowe@ucl.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



example through competition for pro-survival growth factors) or mechanical competition (for example a fast growing clone compresses cells in a slow growing clone, which results in cell extrusion for the now denser slow growing clone) [1, 4, 5]. While competition was initially thought to take place only at the interface between cell lineages, the discovery of mechanical competition revealed that this is not necessarily the case and that extrusion may take place several cell diameters away from this interface [4]. Over a decade of experimental research has suggested that local cell density is a key determinant of cell fate in mechanical competition [6, 4, 7, 8, 9]. However, the vast majority of studies have examined mechanisms of competition at the population level, owing to the difficulty of quantitatively describing the time evolution of an entire system of cells. As such, our understanding of the cell-scale topological and physical parameters that determine fate in competition remains incomplete.

In this study, we sought to examine a new scientific paradigm – using Artificial Intelligence (AI) to uncover the determinants of cell fate directly from a large corpus of time-lapse microscopy data. Specifically, we sought to explore the possibility of learning an interpretable and predictive model of competition using a minimally-supervised and unbiased approach.

Recent studies have shown that machine learning (ML) is adept at uncovering complex patterns in microscopy data. In conventional feature engineering approaches, prior knowledge is incorporated into a model by choosing features that represent the system, for example, by measuring image properties or adding relevant fluorescent cell signalling reporters. This has recently been used, with ML-enabled dimensionality reduction, to study transitions in human pluripotent stem cell populations [10]. However, choosing appropriate measurements becomes increasingly difficult with more complex features such as describing the local topology of tissues comprising multiple cell types and varying degrees of epithelialisation. One promising method is the use of unsupervised deep learning methods, such as variational autoencoders (VAE, [11]). A VAE learns a probabilistic approximation of the underlying distribution of data, meaning that the latent representation can be used as descriptive features of the system. Several recent studies have utilised autoencoders to encode complex cell shapes and other visual features in an interpretable manner [12, 13, 14, 15]. However, these studies have typically been performed on sparse, isolated cells and usually as single observations in time. Other models have attempted to explicitly incorporate time. For example, a recurrent neural network was used to predict lineage choices in hematopoietic stem cells [16]. However, this architecture does not lend itself to introspection and therefore does not directly provide any interpretable insight into the biology.

Here, we sought to learn a model of cell behaviour directly from time-lapse image data. We expand upon the use VAEs to encode cell shape and incorporate local tissue topology as well temporal features, to learn an explainable model of a complex, physiologically important biological phenomenon, cell competition. Finally, we introduce a novel discriminator network, that uses this learned model to identify drugs that affect the underlying mechanism of cell competition.

## Results and Discussion

### Data acquisition and training data

We used a well-described model of cell competition consisting of co-cultures of mammalian MDCK wild-type ($MDCK^{WT}$) and a variant expressing an shRNA targeting the polarity protein *scribble* that can be induced to become mechanical loser cells by addition of tetracycline ($scrib^{kd}$) [17]. To differentiate $scrib^{kd}$ from their $MDCK^{WT}$ counterparts, we expressed nuclear markers fused to different fluorescent proteins (e.g. H2B-GFP for $MDCK^{WT}$, and H2B-RFP for $scrib^{kd}$). Cells were seeded in different ratios, and then, using automated time-lapse microscopy we followed the evolution of the competition over periods of 80 h, taking images at 4 min intervals. We collected 111 independent movies, totaling 7,768 hours of competition experiments (**Fig 1a**).

From this dataset, we extracted single-cell trajectories, making sure that we could observe the entire lifespan of each cell including its fate (either mitosis or apoptosis). To do this, we segmented the time-lapse image data using a fully convolutional residual U-Net [18], then used a dedicated convolutional neural network (CNN) to classify each nucleus into one of five states (interphase, prophase, metaphase, anaphase or apoptotic) based on image features [8]. Then, we tracked all cells over time [19]. Next, we classified the fate of each track as either mitotic, apoptotic or unknown using a dedicated cell fate classification network (**Fig 1b**, **Supplementary Information**). We discarded trajectories with an unknown fate. We manually verified all apoptotic trajectories and a subset of mitotic trajectories to confirm that the fate labels could be used as ground truth for the purposes of training a model. In total we acquired 36,062 mitotic and 2,225 apoptotic trajectories, distributed across the two cell types (**Supplementary Information**). Since we do not know *a priori* what information is required to predict cell fate, we extracted glimpses [20] (**Fig 1c**) which capture three different spatial scales of the neighbourhood surrounding the cell of interest (Small, Mid and Large, corresponding to $21\times21\,\mu m$, $42\times42\,\mu m$ and $84\times84\,\mu m$ FOV respectively) but contain the same number of pixels ($64 \times 64$ px). Further, for the mid-scale view, we also applied masking (using segmentation masks) to artificially remove either the central

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



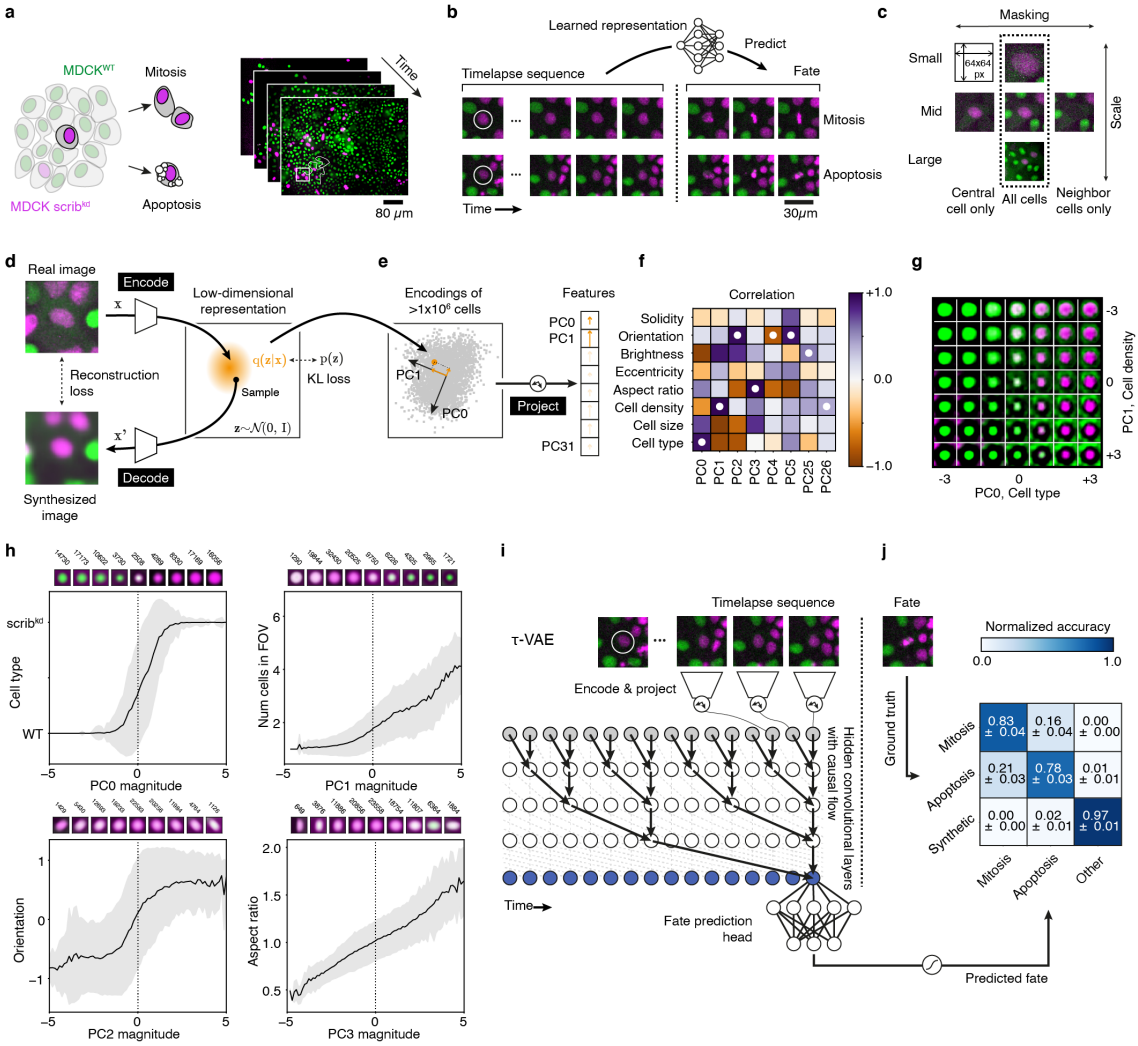



Figure 1: **Learning a meaningful representation of cell competition to predict the fate of cells.** (a) A single cell in a mixed population has two easily observable fates, mitosis or apoptosis. The central hypothesis is that the local tissue organization over time determines the fate. (b) Single-cell tracking is used to build a detailed training dataset of trajectories. The movies are truncated to remove images that encode the fate of the cell. The goal of the machine learning model is to learn a representation that can predict the fate of a cell (circled in white) given the local configuration during interphase. Images are taken at 4 minute intervals, MDCK$^{\text{WT}}$ cells appear in green and scrib$^{\text{kd}}$ in magenta. (c) Glimpse extraction and cell masking to determine the best image input for prediction. Three different scale windows are extracted. For the mid-scale, we also perform masking, by removing either the neighbor cells or the central cell to determine the important features for prediction. (d) A $\beta$-VAE is used to learn a low-dimensional representation of the image data. The orange region represents the multivariate normal distribution of the image encoding. A sample is taken from this distribution to generate the synthesized image. (e) Using a large corpus of images ($\sim$1.2 million) we perform PCA on encodings in latent space to explain the variability of the dataset. The encoding can then be projected using these vectors to yield the features. (f) Pearson correlation coefficient between principal components and measurable physical parameters. White dots denote the parameter with the highest correlation to each PC. Full matrix in the supplementary information. (g) Linear combinations of PC0 (cell type) and PC1 (cell density) over the range [-3, 3]. (h) Interpretability of the projected $\beta$-VAE latent space, showing correlation between PC0-3 and known physical parameters. Continuously varying the PC changes the state of the cell in an explainable manner. Above each plot are average images with binned encodings of the respective PC value. Numbers represent the distribution of these states in the dataset, calculated from a sample of images ($n = 100,000$). Errors represent SD. (i) Schematic of the $\tau$-VAE architecture. A TCN network uses the $\beta$-VAE encoded projection of the trajectories to predict the fate. The TCN network has a receptive field of 128 time steps ($\sim$8.5 h prior to the event) and seven hidden layers consisting of 64 convolutional kernels. Only five layers are shown for simplicity. The fate prediction head has three output layers corresponding to "apoptosis", "mitosis" and "other". (j) Confusion matrix for best performing model on scrib$^{\text{kd}}$ trajectories using randomly selected trajectories ($n = 300$) for testing across a 10-fold cross validation. Errors represent the SEM.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



cell or the neighbors from the image data (**Fig 1c**) to test which representation was most salient. These five datasets could then be used to determine the best performing models and determine where the information necessary for fate prediction is contained.

## An explainable model for cell fate prediction from image sequences

Having acquired a large training set, we next designed a machine learning framework to learn the features of a single cell and its neighbourhood over time, which act as strong predictors of cell fate. One of the design goals of this system was to have minimal human prior insight integrated into the model. First, we sought to learn an interpretable representation of the image data in an unsupervised manner. We trained a variational autoencoder ($\beta$-VAE, **Methods**, [11, 21]) to learn a compact latent representation of cell image data using 1.2 million different images of individual cells and their first neighbours (**Fig 1d**). The $\beta$-VAE learns a low dimensional representation of the image data using a probabilistic encoder. This low dimensional representation (bottleneck) should encode the image in a minimal number of parameters, in the same way that a human operator might describe the image in terms of the cell type, orientation of cells, local neighborhood and so forth. Next, a decoder attempts to reconstruct the real image using this low dimensional representation. The objective function guides the $\beta$-VAE to learn a representation of the image data which is expressive enough to reconstruct the original features, but where each of the latent dimensions are independent and interpretable. We modified the training objective to linearly increase the *capacity* of the bottleneck of the $\beta$-VAE during training. At low capacities the network is forced to prioritise matching the approximate posterior to the exact posterior distribution. At higher capacities, the network prioritises the quality of the reconstruction. During training the $\beta$-VAE first learns to represent gross level information such as cell type, before features such as cell shape and local topology (**Supplementary Information**).

By encoding a large set of images we were able to perform Principal Component Analysis (PCA) in the latent space, which yields linear features that explain the variance in the entire dataset (discussed later, **Fig 1e-h**). We use these principal components (PCs) as the inputs for making predictions.

Next, we built a prediction network that utilises the temporal sequence of images of each single-cell trajectory to output cell fate. Based on current biological knowledge, we reasoned that the biochemical commitment to apoptosis or division occurs hours before we observe the fate, so we trimmed each trajectory to remove observations that show recognizable morphological features of mitosis (such as DNA condensation in prometaphase and alignment of chromosomes during metaphase) or apoptosis (such as DNA fragmentation) (**Supplementary Information**). As such, the prediction network is forced to use only features from the time evolution of the interphase cell and its neighbours to make a prediction about the fate. Next, we encoded and projected each of these movies using the $\beta$-VAE (**Fig 1e**), and passed them to a temporal convolution network (TCN, [22]).

The TCN uses *causal convolutions* to ensure the prediction at time $t$ is only dependent on information from $x_{t-r} \ \dots \ x_t$, where $r$ is the receptive field (equivalent to a window of time before an event, $\sim$8.5 h in this case) of the TCN (**Supplementary Information**). The output of the TCN is connected to a dedicated prediction head, a densely connected network with three outputs corresponding to "apoptosis", "mitosis" and "other". A final softmax activation yields the prediction of the full network. Overall, The TCN acts as a sequence classifier, taking a sequence of observations of a single-cell in interphase and returning a prediction for the cell fate, *w*ithout ever observing the fate (**Fig 1i**). We refer to the variational encoder and projector coupled to the temporal convolution network as a $\tau$-VAE. In training the $\tau$-VAE, we supplemented the real data corresponding to mitotic and apoptotic classes, with a dynamically generated "synthetic" class to simulate trajectories which were neither apoptotic nor mitotic, corresponding to the "other" output class (**Methods**, **Supplementary Information**). In doing so, the prediction problem becomes a multiclass problem as opposed to a binary problem, thus ensuring that the model learns the features of both apoptotic and mitotic events. This is important as we do not want the model to learn only the features of mitosis, and predict apoptosis by exclusion, or *vice versa*.

We tested the five different representations of the image data (**Fig 1c**) as input to the prediction network. In all representations the image data was limited to the same number of pixels, but comprised either different spatial scales or lacked information about either the central cell or the neighbors. We trained the $\tau$-VAE network and measured the accuracy by comparing the predicted fate with the ground truth fate on a set of unseen test data consisting of 300 trajectories. For each dataset, we split the data by the cell type (MDCK$^{\text{WT}}$ or scrib$^{\text{kd}}$) and calculated separate confusion matrices to ensure that there was no systematic bias in the predictions. To account for any potential bias in the testing set, we performed $k$-fold cross-validation ($k = 10$) on each model (**Methods**), randomly choosing a subset of training and testing data in each validation. Overall, the best performing networks used only the central cell region to predict the fate, with an average fate prediction $F_1$-score of 0.87 ± 0.02, across both cell types as shown in the confusion plots (**Fig 1j**, **Supplementary Information**). For reference, a TCN trained using a set of human chosen image features (those shown in **Fig 1f**) achieves a lower $F_1$-score of 0.71, and is particularly poor at identifying apoptosis, with an accuracy

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



of 0.49. This demonstrates that the $\beta$-VAE is able to capture more salient image features, enabling a more accurate fate prediction, by learning them directly from the distribution of the data. We used the best performing network ("Small View, All Cells" model, *i.e.* cropped to the central cell at the highest magnification) for all further analyses. We concluded that the $\tau$-VAE network is able to accurately predict the cell fate based only on the interphase local tissue organization alone, having learnt features directly from the image data to enable this task. Next, we sought to introspect the model and to assign meaningful semantic labels to the learnt features.

## Interpreting the model

The goal of the $\tau$-VAE network is to learn an end-to-end model that requires minimal input from experts to predict the fate of cells in competition. The implicit hypothesis is that there is sufficient information in the observations of local tissue organization to enable this prediction. Having determined that our approach is able to accurately predict the fate of cells in a competitive system, we sought to interrogate the learnt features of the $\beta$-VAE. In contrast to other approaches, we do not perform feature engineering to select parameters that define the problem (such as cell density, number of neighbours, etc), but rather, we extract these automatically and directly from the data, based on the latent space of the $\beta$-VAE. We used several different approaches to interpret the model.

**Assigning physical parameters to the latent features** Since the training objective of the $\beta$-VAE encourages a continuous, but disentangled internal representation of the image data, we sought to assign meaningful semantic labels to those latent variables. Analysis of the latent features revealed that some parameters show covariance – for example, in the dataset, there is a correlation between cell type and nuclear area, since scrib$^{kd}$cells tend to have larger nuclei than MDCK$^{WT}$cells. Therefore, we performed PCA on the latent space ($\mathbf{z} \in \mathbb{R}^{32}$) yielding 32 principal components ordered by the magnitude of the variance explained by the component. We analysed the correlation of the components with parameters that could be measured from the images (**Fig 1g**). Inspecting these components shows that the first two (PC0 and PC1) account for 26.9% of the variability of the data, and seem to represent cell type and cell density (number of cells visible in the glimpse) respectively (**Fig 1f-h**). Component PC2 encodes an orientation parameter of the central nucleus, while component PC3 encodes nuclear aspect ratio (**Fig 1g-h**). Higher principal components, such as PC25 encode parameters such as fluorescence intensity of H2B-GFP/RFP but the correlation coefficient is weaker. We confirm these assignments by sampling images from the dataset with various values for these components (**Supplementary Information**). Later components broadly enable the network to encode the arrangement and identity of cells in the local neighbourhood (**Supplementary Information**). Strikingly, projecting the $\beta$-VAE latent space enabled us to learn an explainable model of the local tissue organization of cells in a completely unsupervised manner. Next, we sought to investigate the role of these principal components over time in the prediction of cell fate.

**Feature ablation studies to determine the minimal information required for prediction** To determine the minimal information required for cell fate prediction, we removed individual components in a systematic manner (replacing them with Gaussian noise at all time steps) and calculated the performance of the network after each component removal. Ablated networks were ranked according to their effect on the prediction accuracy. Through multiple iterations, we found that a single component (PC1 - nuclear area/cell density) could be used to predict cell fate with 43 ± 2% accuracy – significantly higher than random chance assuming an equal probability of choosing any fate (33.33%, **Fig 2a**). In the ablation approach, PC1 was the last component to be removed, suggesting the single highest contribution to the prediction accuracy. The ablation study reveals that the top five components (PC1, PC5, PC26, PC3, PC26) account for 64% of the prediction accuracy, with the remaining 27 components contributing a further 36%. Importantly, when all inputs are replaced by noise, all fates are predicted with equal probability, suggesting no inherent bias toward any fate in the network. Remarkably, this suggests that, in line with our current understanding of mechanical cell competition stemming from nearly a decade of experimental studies [17, 7, 23, 8, 9], our model has autonomously learnt that cell density is a strong predictor of cell fate, directly from the data with no expert input.

**Timescales of predictions and feature saliency to visualize network attention** Since the TCN is able to output a prediction after every additional time step, we can use the prediction head (**Fig 1i**) to visualize the time evolution of the prediction (**Fig 2b-c**). Across all of the data, we find that for the scrib$^{kd}$cells, the $\tau$-VAE network predicts apoptosis early (up to 8h before an event) and mitosis late ($\sim$2h before an event, **Supplementary Information**). Further, we can inspect the magnitude and contribution of each principal component to the prediction. In general, we found that PC1 (cell density) was often much larger in those trajectories undergoing apoptosis. To assess the model's use of components to make predictions, we utilise the error gradients during backpropagation in the network to calculate the saliency (**Methods**, [24]). Feature saliency is a method to determine which features of the input contribute most significantly to the classification accuracy of the $\tau$-VAE. We computed this in two ways (i) PC feature saliency, *i.e.* the PC input to the TCN network (**Fig 2b-c**) and (ii) raw pixel saliency *i.e.* the raw pixel input to the encoder network (**Supplementary Information**). The latter approach identifies the raw image pixels that contribute most to the eventual

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



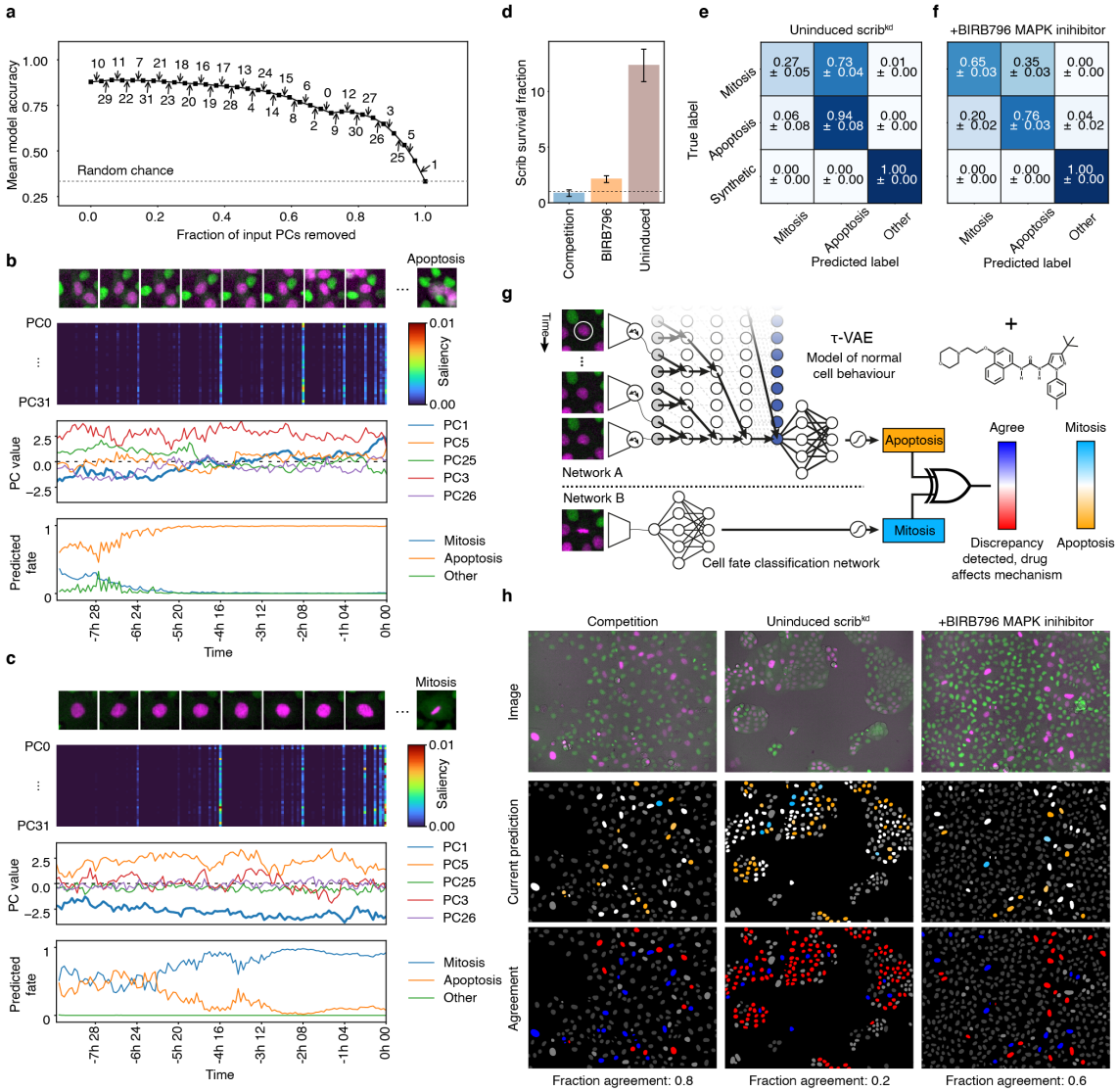



Figure 2: **An explainable internal representation of cell competition enables drug evaluation.** (a) Feature ablation demonstrates the role of each PC in the final prediction of the model. Each arrow indicates the cumulative replacement of a given PC with Gaussian noise. Error bars represent the standard deviation of model accuracy over all ten TCN models. (b) Example apoptotic trajectory with network predictions, and internal representation. Top row shows a sampled sequence of images from the trajectory. Second row shows the PC feature saliency over time calculated by the TCN. Third row is the values of the top-5 principal components over time. Final row is the prediction of cell fate over time. (c) Example mitotic trajectory with network predictions, and measured parameters, as in panel b. (d) Survival fraction for scrib$^{kd}$ cells in competition (MDCK$^{WT}$:scrib$^{kd}$), BIRB796 treated (MDCK$^{WT}$:scrib$^{kd}$ + 2 μM BIRB796) and uninduced (MDCK$^{WT}$:scrib$^{kd,\ tet-}$). Values below 1 (dotted line) indicate a cell population decrease, values greater than 1 indicate a population increase over the course of the experiment. (e) Confusion matrix of prediction accuracy for scrib$^{kd,\ tet-}$ cells ($n = 161$ real trajectories) showing many mitoses are incorrectly predicted as apoptoses. (f) Confusion matrix for BIRB796 treated cells ($n = 198$ real trajectories) showing a similar pattern to the scrib$^{kd,\ tet-}$ condition. (g) A discriminator network that uses two models to detect changes in cell behavior. Network A is the $\tau$-VAE model of learned cell behaviour and Network B is the cell fate classification network. A discriminator (shown as an XOR gate) determines discrepancies between the two outputs. (h) Example outputs of per-cell predictions and discriminator output for individual timepoints of competition, uninduced and BIRB796 treated timelapse movies, as in panel d. The top row shows the raw input image data. The middle row shows the current prediction (in this frame of the movie) of the $\tau$-VAE network for scrib$^{kd}$ cells in the FOV. Blue represents mitosis, orange represents apoptosis and white represents unknown fate or insufficient data to make a prediction at this time-point. The bottom row shows the discriminator output for each cell in the FOV. Red indicates that the $\tau$-VAE network did not agree with the fate classification, and blue indicates agreement. Grey cells have no predictions associated with them. The full movies can be found in the supplementary information.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



prediction. The PC feature saliency reveals the timescales of activations within the network that are used to make predictions. The high gradients reveal that many of the late time steps are used in making the predictions for apoptosis and mitosis. Empirically, the pixel saliency reveals that nearby cells and the geometry (aspect ratio, convexity) of the nucleus (**Supplementary Information**) have significant contributions to the prediction - however, it is difficult to assign quantitative meaning to these observations, owing to the fact that they are in image space rather than feature space.

### Challenging the learned representation with biochemical perturbations

Finally, to confirm that the network has learnt a model of mechanical cell competition, we sought to challenge the model with cells treated with different biochemical perturbations. For example, we performed experiments using cells that were uninduced (MDCK$^{WT}$:scrib$^{kd,\ tet-}$) such that there was no competition. In this case, the knockdown of the polarity protein Scribble is not induced with tetracycline (scrib$^{kd,\ tet-}$), so the cells do not engage in mechanical competition with MDCK$^{WT}$ cells, but can still be distinguished by their H2B-RFP marker. We acquired timelapse data of the cells and confirmed that the scrib$^{kd,\ tet-}$ cell count was consistent with a non-competitive scenario (**Fig 2d**, **Supplementary Information**). From this dataset, we randomly selected a set of full length trajectories with a known fate (either apoptosis, mitosis or neither), and, after removing the final images and discarding the trajectories where no event occurs, we passed them to the $\tau$-VAE network. We found that the fate prediction accuracy of the network dropped significantly for the scrib$^{kd,\ tet-}$ cells. Strikingly, many scrib$^{kd,\ tet-}$ mitotic trajectories were incorrectly predicted to be apoptotic as can been seen from the confusion matrix (**Fig 2e**). This is an important result – the $\tau$-VAE, using the available information predicts, correctly, that the scrib$^{kd,\ tet-}$ cells should die under these conditions if knock-down of Scribble had been induced to start competition. A human would arrive at the same conclusion, given the same information. The fact that, under an unseen biochemical perturbation that disturbs the competition, we subsequently observe that they do not die, lays the foundation for a method to identify systematic deviations from normal (*i.e.* predictable) behavior.

This variation in the performance of the $\tau$-VAE when applied to cells under different biochemical perturbations suggested that our model is sensitive to changes in gene expression and the biochemical mechanisms of competition. Therefore, we sought to determine whether the methodology could be used for identifying drugs that perturb competition without further modification to the prediction network. Recent studies have suggested that p38 kinase inhibitors may interfere with mechanical cell competition by inhibiting the stress response pathways that lead to apoptosis [7]. To test this hypothesis, and to determine whether the network was able to discriminate these events, we acquired timelapse data of the cell competition (MDCK$^{WT}$:scrib$^{kd}$) in the presence of $2\,\mu\mathrm{M}$ BIRB796, a p38 MAPK inhibitor [25]. We measured the loser cell count over the course of the experiments and noted that there was a higher survival fraction of the scrib$^{kd}$ cells, although they still grew significantly slower than MDCK$^{WT}$ (**Fig 2d**, **Supplementary Information**). From these data, we extracted single cell trajectories and used the $\tau$-VAE network to predict the fate of the cells as before. As with the uninduced dataset, the network predicted a significantly higher number of apoptoses where the true label was mitosis (**Fig 2f**).

Given that both the p38 MAPK inhibitor and the scrib$^{kd,\ tet-}$ condition interfere with the competition by limiting apoptosis, we would expect the scrib$^{kd}$ cells to reach higher densities in these conditions. Indeed, when analysing the network's representation, we noticed that the signatures of incorrectly predicted trajectories are more similar to the trajectories categorised as apoptotic under control conditions, especially with respect to the increased magnitude of PC1, that represents local cell density (**Supplementary Information**). This is consistent with the observations that the scrib$^{kd,\ tet-}$and BIRB796 treated scrib$^{kd}$cells reach higher densities, with significantly lower apoptotic rates. Overall, our results suggest that the $\tau$-VAE network, trained on the MDCK$^{WT}$:scrib$^{kd}$ data, has learnt a complex and predictive model of cell competition, that is sensitive to local changes in the tissue organization and the signalling pathways participating in competition.

**A method for automated drug screening** Having established that the $\tau$-VAE network is able to represent a complex model of cell behavior in an explainable manner, we sought to define a general approach to utilise such a predictive model for image-based drug screening [26]. To do so, we introduce a novel discriminator network (**Fig 2g**) that compares the outputs of two models (Networks A & B) to determine discrepancies indicative of drug activity. The two networks utilise different amounts of information from each single-cell trajectory. Network A (the $\tau$-VAE), a model of "normal" cellular behavior, uses only the early part of the trajectory to predict the fate of the cell. In contrast, Network B (the cell fate classification network, **Supplementary Information**) uses the entire trajectory to classify the actual fate of the cell. When these two models agree for a particular cell, it suggests that the fate is predictable and thus competition conforms to learned behavior. However, when the two models disagree, this suggests that the cell is deviating from normal learned behavior, presumably due to the influence of an added drug or other perturbation. We demonstrate the utility of the discriminator to evaluate individual cells in movies in three different conditions, MDCK$^{WT}$:scrib$^{kd}$, MDCK$^{WT}$:scrib$^{kd,\ tet-}$ and MDCK$^{WT}$:scrib$^{kd}$ + $2\,\mu\mathrm{M}$ BIRB796. First, we determined the performance of the cell fate

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



classification network, finding it to achieve an accuracy of 0.96 in determining cell fate ($n = 392$) across all of the data. Then we used the $\tau$-VAE to make predictions for these cells, and the discriminator to determine the agreement with the cell fate classification network. In control conditions, the two networks show a high level of agreement with a fraction of agreement $a$ of 0.8. The results show a similar pattern to the confusion matrices in **Fig 2e-f**, in that the uninduced Scribble ($a$=0.22) and and BIRB796 ($a$=0.62) treated cells show the lowest fraction of agreement between the $\tau$-VAE and the cell fate classification, indicating deviation from the normal model of behavior. Therefore, the discriminator network automatically identifies conditions that deviate from learned behavior. Further, this approach can be utilized to monitor the time evolution and spatial pattern of predictions for individual cells in their original context (**Fig 2h**).

## Conclusions

Deep learning is now a powerful tool in microscopy image analysis. However, the complex internal representations of many deep learning models, and the difficulty of analysing time dependent features, means that they have rarely been used to gain mechanistic insight into biological phenomena. Here, we developed an end-to-end machine learning model capable of discovering the physical parameters and rules of a complex biological phenomenon, cell competition, directly from image data. Starting *tabula rasa*, we demonstrate that our approach is able to learn a meaningful representation of cell behaviour in an automated and minimally-supervised manner. The model requires minimal human input to train and is able to correctly predict the fate of cells in mechanical cell competition. Strikingly, the model learns that local cell density is the single most important determinant of cell fate in mechanical competition, an observation that has taken scientists a decade of experimental research and data analysis to determine. Most exciting is that we are able to introspect the model to identify the physical features enabling prediction as well as the time-scale over which correct predictions are made. In the case of cell competition, we expect that these features will prove invaluable in formulating hypotheses about the nature of the mechanical changes detected during competition and the signalling pathways leading to loser cell death. This model can be used directly, and with no further modification, to investigate which candidate pathways participate in competition. Although we have demonstrated the utility of this approach using a model phenomenon and cell type, the approach is generalizable to many other systems, for example the study of the microenvironmental factors leading to differentiation of stem cells or embryonic development. Finally, we have shown that a novel discriminator network, based on the autonomously learned $\tau$-VAE model, can detect conditions which deviate from the normal cellular behaviour due to inhibition or silencing, signifying that it can be used as a screening tool. Once trained, this fully automated system is able to discriminate between normal cell behavior and perturbations without any further human intervention, paving the way for mechanism-aware AI-based drug discovery.

## Methods

**Cell culture, imaging assays, single-cell tracking and cell fate classification network** Detailed methods can be found in the supplementary information.

**Drug treatments** BIRB796 (Tocris 5989, [25]) was dissolved in DMSO and added to the cell culture 5 h before imaging at a final concentration of 2 μM. As a negative control, a similar volume of DMSO was added to some wells.

**Variational Autoencoder ($\beta$-VAE)** We built a convolutional variational autoencoder to learn a low-dimensional representation of the cell image data that could be used by a TCN for fate prediction. The encoder network consists of four convolutional layers with $3 \times 3$ kernels, Swish activations, with 32, 64, 128 and 256 kernels respectively. Each layer was pooled by a $2 \times 2$ max-pooling operation. The convolutional output was flattened and split into two arms with two fully connected layers for the $\mu$ and $\sigma^2$ estimators. We found that adding additional fully connected layers of 128 units between the estimators and the flattened encoder output improved the model performance. We used a random normal sampler to generate samples from the distribution. The decoder is inverse of the encoder, using nearest neighbor upsampling between the convolutional layers.

We use the revised objective function [21, 27] to train the network:

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}, \gamma, C) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \gamma \left| \mathrm{D}_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \,||\, p(\mathbf{z})) - C \right| \tag{1}$$

Our loss function is composed of two terms. The first term is the reconstruction loss term which penalises differences between the reconstruction (the decoded, encoded input) and the real input. In practice, we use the mean squared error between the input image ($\mathbf{x}$) and the output ($\mathbf{x}'$) of the $\beta$-VAE. The second term is the Kullback-Leibler divergence ($\mathrm{D}_{\mathrm{KL}}(\cdot \,||\, \cdot)$), which penalizes the latent space model variance from dropping to zero, by forcing the encoding to match

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



the Gaussian prior with a diagonal covariance matrix, $N(0, I)$. This has the effect of regularizing the latent space to promote a continuous representation of the underlying image data.

In our implementation, we dynamically adjust the bottleneck capacity ($C$) of the network during training. The value of $C$ is scaled linearly as a function of training iteration, reaching a maximum value $C_{\text{max}}$. This ensures that at early training iterations the network prioritizes the encoding, while at later iterations this is refined to optimize the decoding. The scaling constant $\gamma$ balances the two terms of the loss function.

We prepared a training set of 1.2 million images of cells by sampling individual cells from the time-lapse movies. A fraction of cells (random 10% of cells in frame) was selected from a random sample of frames (10% of frames in movie) and an ROI of varying size around each cell was extracted. These were then downsampled using nearest neighbor sampling, to the network input size of ($64 \times 64 \times 2$). We then trained the $\beta$-VAE network using $\gamma = 1000$, $C_{\text{max}} = 50$, latent vector $\mathbf{z} \in \mathbb{R}^n$. We found $n = 32$ to be the optimal value of the latent space dimensions. Optimization was performed using an Adam optimizer for 100 epochs and a minibatch size of 256. Training images were augmented *on-the-fly* by random flipping, rotations and simulated edge cropping.

**Trajectory Synthesis** In order to simulate a third class of trajectory, referred to as "synthetic", we utilised the generative property of the decoder network. The first frame of each synthetic trajectory is a real image, i.e. it is decoded by the decoder network from the latent-space encoding of an image that actually exists. For the next frame, the encoding is adjusted by adding to each latent variable a scalar that is sampled from a Gaussian distribution (with $\mu = 0.0$ and $\sigma = 0.2$). The image for this second frame is then the decoder output with this new encoding as input. This process is then repeated until all the frames of the synthetic movie are generated, with the trajectories taking a random walk in latent space.

**Principal Component Analysis (PCA)** We used PCA to analyse the learnt representation of the total dataset. The principal components of the latent features derived by the $\beta$-VAE have a higher degree of interpretability than the latent features themselves. PCA was applied to the latent features before analysis of the latent space was undertaken. We used the *PCA* function from *Scikit-learn* to perform the decomposition into 32 components, once for each of the $\beta$-VAE latent dimensions.

**Temporal Convolution Network (TCN)** The timelapse sequences encoded by the $\beta$-VAE become the input features ($n \times t$) of a temporal convolution network (TCN) [22]. The procedure for preparing a timelapse movie for input to the TCN is as follows. First, each frame in the timelapse movie is normalized such that the pixel values of that individual frame have zero mean and unit variance. This is performed on a per-channel basis, such that the RFP and GFP channels are normalized separately. Next, the normalized timelapse movie is fed through the encoder network of the trained $\beta$-VAE. The encoder yields three outputs for each latent variable - the mean, standard deviation, and Gaussian-sampled value. This is calculated for every timelapse movie in the dataset.

Once encoded, the timelapse sequences are transformed from latent space ($\mathbf{Z}$) into principal component space ($\mathbf{T}$) using the transform $\mathbf{T} = \mathbf{ZW}$, where $\mathbf{W}$ are the principal components. and then fed into the TCN for training. The TCN is formed of seven convolutional layers with respective dilations of 1, 2, 4, 8, 16, 32 and 64. Each of these layers has 64 convolutional filters. The output from the convolutional layers is fed into a fate prediction head. This network projects the output into a classification, and is composed of a fully connected layer with 128 units, and a final fully connected layer with three units that represent the possible classifications of the sequence ("apoptosis", "mitosis" or "synthetic"). We used a sparse categorical cross entropy loss function to train the network.

The TCN is trained with a batch size of 128 for 100 epochs. Optimization was performed using the RMSprop optimizer with a learning rate of 0.001. Training sequences were augmented *on-the-fly* by random frame removal, cropping and the addition of noise to the encoded sequences. Regularization was ensured by applying batch normalization and a dropout rate of 0.3 to the TCN layer of the network.

**Feature saliency** We can determine the salient features of the input images by calculating a saliency map ($M_c$) for an input ($x$). This is achieved by calculating the gradient of the output function $S_c$ with respect to the input during backpropagation: $M_c(x) = \partial S_c(x)/\partial x$. In practice, we average $n$ gradient calculations, each with a small added Gaussian noise ($\mu$=0.0, $\sigma = 0.1 \times |(\max(S_c) - \min(S_c)|$) component to the input.

**Software availability** A reference implementation of the $\tau$-VAE is available at: `https://github.com/lowe-lab-ucl/cellx-predict`

**Data availability** Data is available from the UCL data repository: `https://dx.doi.org/10.5522/04/16578959`.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



## Author contributions

ARL and GC conceived and designed the research. GV performed experiments. CJS developed and performed computational analysis. ARL wrote the image processing and cell tracking code. CJS, GV, GC and ARL evaluated the results and wrote the paper.

## Acknowledgments

This work was supported by a BBSRC LIDo AI PhD studentship to CJS. GV was supported by BBSRC grant BB/S009329/1. We thank Nathan Day, Jasmine Michalowska and Dan Smaje for help annotating data. We also thank members of the Lowe and Charras labs for discussions and technical support during the project. ARL wishes to acknowledge the Turing Fellowship from the Alan Turing Institute. ARL and GC wish to acknowledge the support of BBSRC grant BB/S009329/1.

## References

[1] R. Levayer and E. Moreno, "Mechanisms of cell competition: Themes and variations," *The Journal of Cell Biology*, vol. 200, no. 6, pp. 689–698, 2013.

[2] G. Morata and P. Ripoll, "Minutes: mutants of drosophila autonomously affecting cell division rate," *Dev. Biol.*, vol. 42, pp. 211–221, Feb 1975.

[3] T. Parker, E. Madan, K. Gupta, E. Moreno, and R. Gogna, "Cell competition spurs selection of aggressive cancer cells," *Trends in Cancer*, vol. 6, pp. 732–736, sep 2020.

[4] R. Levayer, B. Hauert, and E. Moreno, "Cell mixing induced by myc is required for competitive tissue invasion and destruction," *Nature*, vol. 524, pp. 476 EP –, 08 2015.

[5] J.-P. Vincent, A. G. Fletcher, and L. A. Baena-Lopez, "Mechanisms and mechanics of cell competition in epithelia," *Nature Reviews Molecular Cell Biology*, vol. 14, pp. 581 EP –, 08 2013.

[6] C. Hogan, S. Dupré-Crochet, M. Norman, M. Kajita, C. Zimmermann, A. E. Pelling, E. Piddini, L. A. Baena-López, J.-P. Vincent, Y. Itoh, H. Hosoya, F. Pichaud, and Y. Fujita, "Characterization of the interface between normal and transformed epithelial cells," *Nature Cell Biology*, vol. 11, pp. 460 EP –, 03 2009.

[7] L. Wagstaff, M. Goschorska, K. Kozyrska, G. Duclos, I. Kucinski, A. Chessel, L. Hampton-O'Neil, C. R. Bradshaw, G. E. Allen, E. L. Rawlins, P. Silberzan, R. E. Carazo Salas, and E. Piddini, "Mechanical cell competition kills cells via induction of lethal p53 levels," *Nature Communications*, vol. 7, pp. 11373 EP –, 04 2016.

[8] A. Bove, D. Gradeci, Y. Fujita, S. Banerjee, G. Charras, and A. R. Lowe, "Local cellular neighborhood controls proliferation in cell competition," *Molecular Biology of the Cell*, vol. 28, no. 23, pp. 3215–3228, 2017. PMID: 28931601.

[9] D. Gradeci, A. Bove, G. Vallardi, A. R. Lowe, S. Banerjee, and G. Charras, "Cell-scale biophysical determinants of cell competition in epithelia," *eLife*, vol. 10, May 2021.

[10] E. Ren, S. Kim, S. Mohamad, S. F. Huguet, Y. Shi, A. R. Cohen, E. Piddini, and R. C. Salas, "Deep learning-enhanced morphological profiling predicts cell fate dynamics in real-time in hPSCs," *bioRxiv*, aug 2021.

[11] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013. cite arxiv:1312.6114.

[12] C. K. Chan, A. Hadjitheodorou, T. Y.-C. Tsai, and J. A. Theriot, "Quantitative comparison of principal component analysis and unsupervised deep learning using variational autoencoders for shape analysis of motile cells," *bioRxiv*, jun 2020.

[13] A. Zaritsky, A. R. Jamieson, E. S. Welf, A. Nevarez, J. Cillay, U. Eskiocak, B. L. Cantarel, and G. Danuser, "Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma," *Cell Systems*, vol. 12, pp. 733–747.e6, jul 2021.

[14] Z. Wu, B. B. Chhun, G. Popova, S.-M. Guo, C. N. Kim, L.-H. Yeh, T. Nowakowski, J. Zou, and S. B. Mehta, "DynaMorph: self-supervised learning of morphodynamic states of live cells," *bioRxiv*, jul 2020.

[15] K. D. Yang, K. Damodaran, S. Venkatachalapathy, A. C. Soylemezoglu, G. V. Shivashankar, and C. Uhler, "Predicting cell lineages using autoencoders and optimal transport," *PLOS Computational Biology*, vol. 16, p. e1007828, apr 2020.

[16] F. Buggenthin, F. Buettner, P. S. Hoppe, M. Endele, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck, T. Schroeder, F. J. Theis, and C. Marr, "Prospective identification of hematopoietic lineage choice by deep learning," *Nature Methods*, vol. 14, pp. 403–406, feb 2017.

[17] M. Norman, K. A. Wisniewska, K. Lawrenson, P. Garcia-Miranda, M. Tada, M. Kajita, H. Mano, S. Ishikawa, M. Ikegawa, T. Shimada, and Y. Fujita, "Loss of scribble causes cell competition in mammalian cells," *Journal of Cell Science*, vol. 125, no. 1, pp. 59–66, 2012.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469554; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



[19] K. Ulicna, G. Vallardi, G. Charras, and A. R. Lowe, "Automated deep lineage tree analysis using a Bayesian single cell tracking approach," *Frontiers in Computer Science*, sep 2021.

[20] V. Mnih, N. Heess, A. Graves, and k. kavukcuoglu, "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds.), vol. 27, Curran Associates, Inc., 2014.

[21] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[22] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv*, 2016.

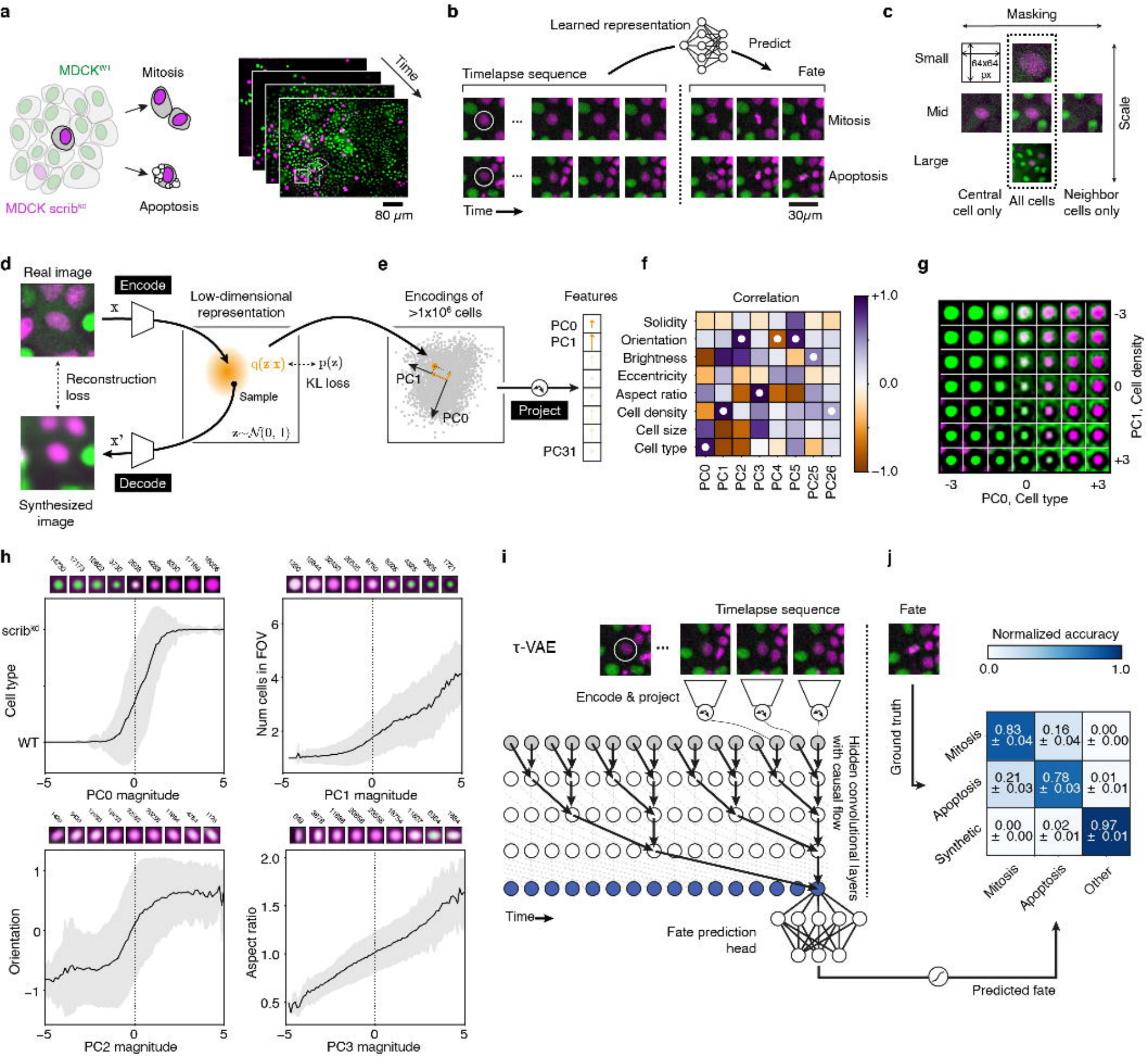
[23] R. Levayer, C. Dupont, and E. Moreno, "Tissue crowding induces caspase-dependent competition for space," *Current Biology*, vol. 26, pp. 670–677, mar 2016.

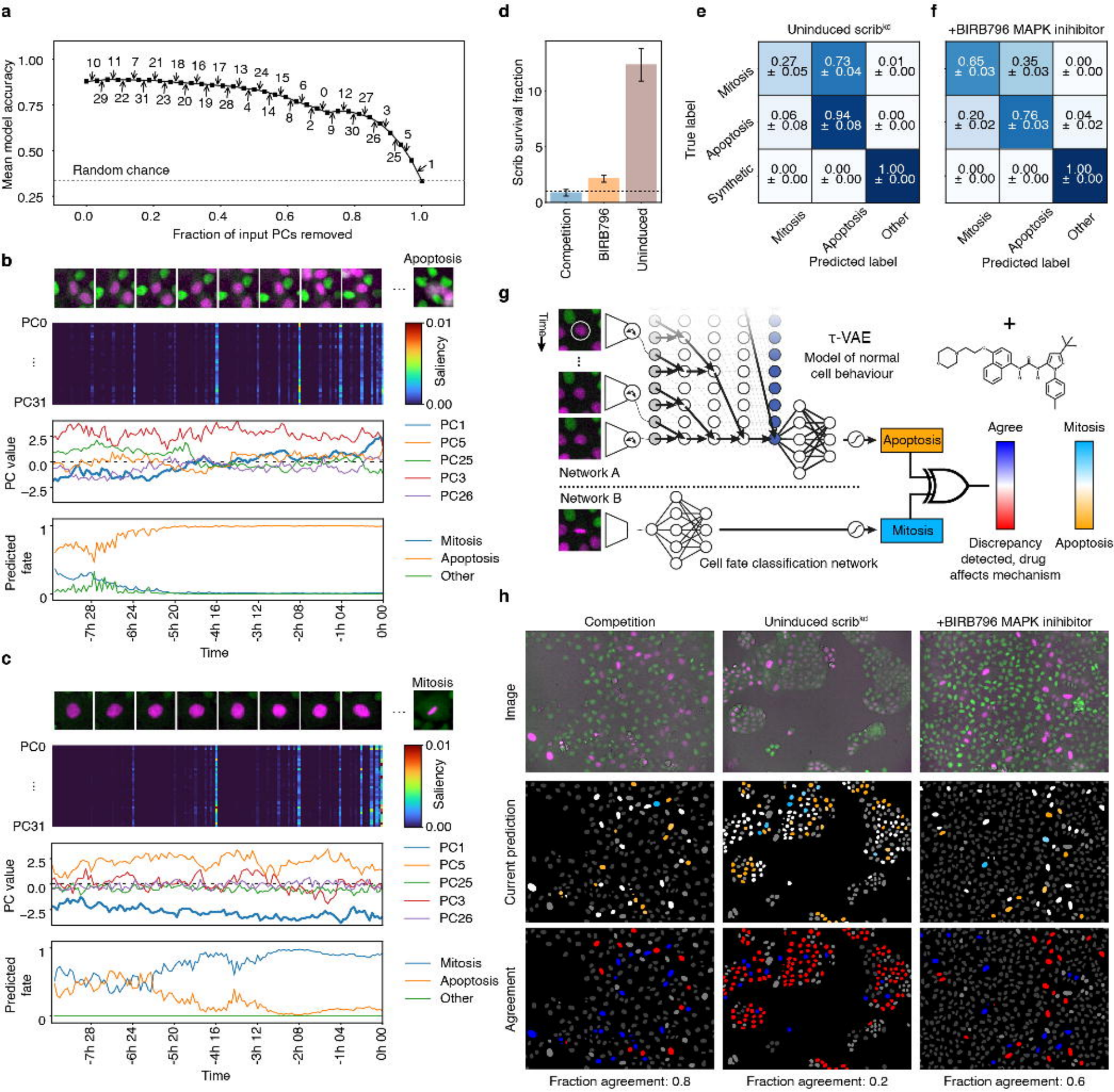
[24] D. Smilkov, N. Thorat, B. Kim, F. B. Viégas, and M. Wattenberg, "Smoothgrad: removing noise by adding noise," *CoRR*, vol. abs/1706.03825, 2017.

[25] Y. Kuma, G. Sabio, J. Bain, N. Shpiro, R. Márquez, and A. Cuenda, "BIRB796 inhibits all p38 MAPK isoforms in vitro and in vivo," *Journal of Biological Chemistry*, vol. 280, pp. 19472–19479, may 2005.

[26] S. N. Chandrasekaran, H. Ceulemans, J. D. Boyd, and A. E. Carpenter, "Image-based profiling for drug discovery: due for a machine-learning upgrade?," *Nature Reviews Drug Discovery*, vol. 20, pp. 145–159, dec 2020.

[27] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in $\beta$-vae," 2018.


a
MDCK^WT
Mitosis
MDCK scrib^kd
Apoptosis
Time
80 μm
b
Learned representation
Predict
Timelapse sequence
Fate
Mitosis
Apoptosis
Time
30μm
c
Masking
Small
64x64 px
Mid
Large
Scale
Central cell only
All cells
Neighbor cells only
d
Real image
Encode
x
Low-dimensional representation
q(z|x)
p(z)
KL loss
Sample
z~N(0, 1)
Reconstruction loss
x'
Decode
Synthesized image
e
Encodings of >1x10^6 cells
PC1
PC0
Project
Features
PC0
PC1
PC31
f
Correlation
Solidity
Orientation
Brightness
Eccentricity
Aspect ratio
Cell density
Cell size
Cell type
PC0
PC1
PC2
PC3
PC4
PC5
PC25
PC26
+1.0
0.0
−1.0
g
-3
0
+3
PC1, Cell density
PC0, Cell type
h
scrib^kd
WT
Cell type
PC0 magnitude
Num cells in FOV
PC1 magnitude
Orientation
PC2 magnitude
Aspect ratio
PC3 magnitude
i
τ-VAE
Timelapse sequence
Encode & project
Hidden convolutional layers with causal flow
Time
Fate prediction head
j
Fate
Normalized accuracy
0.0
1.0
Ground truth
Mitosis
Apoptosis
Synthetic
0.83 ± 0.04
0.16 ± 0.04
0.00 ± 0.00
0.21 ± 0.03
0.78 ± 0.03
0.01 ± 0.01
0.00 ± 0.00
0.02 ± 0.01
0.97 ± 0.01
Mitosis
Apoptosis
Other
Predicted fate


a
Mean model accuracy
Random chance
Fraction of input PCs removed
b
Apoptosis
PC0
PC31
Saliency
PC value
PC1
PC5
PC25
PC3
PC26
Predicted fate
Mitosis
Apoptosis
Other
Time
c
Mitosis
d
Scrib survival fraction
Competition
BIRB796
Uninduced
e
Uninduced scrib
True label
Predicted label
f
+BIRB796 MAPK inihibitor
g
τ-VAE
Model of normal cell behaviour
Network A
Network B
Cell fate classification network
Agree
Discrepancy detected, drug affects mechanism
h
Image
Current prediction
Agreement
Fraction agreement: 0.8
Fraction agreement: 0.2
Fraction agreement: 0.6