MELBA Journal Machine Learning for Biomedical Imaging

Sequence models for continuous cell cycle stage prediction from brightfield images

Andrea **Salati*** [1] Louis-Alexandre **Leger*** [1] Maxine **Leonardi*** [1] Martin **Weigert**† [1,2,3] Felix **Naef**† [1]

* These authors contributed equally
† Shared supervision
1 Institute of Bioengineering, School of Life Sciences, EPFL, Lausanne, Switzerland
2 ScaDS.AI, Dresden/Leipzig, Germany
3 TUD Dresden University of Technology

Abstract

The cell division cycle is a ubiquitous essential process across the tree of life. Understanding cell cycle dynamics is crucial for studying biological processes such as growth, development and disease progression. While fluorescent protein reporters like the FUCCI system allow live monitoring of cell cycle phases, they require genetic engineering and occupy additional fluorescence channels, limiting broader applicability in complex experiments. In this study, we conduct a comprehensive evaluation of deep learning methods for predicting continuous FUCCI signals using non-fluorescence brightfield imaging, a widely available label-free imaging modality. To that end, we generated a large dataset of 1.3 M images of dividing human RPE1 cells with full cell cycle trajectories to quantitatively compare the predictive performance of distinct model categories including single time-frame models, causal state space models and bidirectional transformer models. We show that both causal and transformer-based models significantly outperform single- and fixed frame approaches, enabling the prediction of visually imperceptible transitions like G1/S within 1 hour resolution. Our findings underscore the importance of sequence models for accurate predictions of cell cycle dynamics and highlight their potential for label-free imaging.

Keywords

Cell cycle prediction, label-free microscopy, sequence-models

Article informations

https://doi.org/10.59275/j.melba.2026-84ea

Volume 2026, Received: 2025-12, Published 2026-07

Corresponding author: felix.naef@epfl.ch

Special issue: Medical Imaging with Deep Learning (MIDL) 2025
Guest editors: Lisa Koch, Ronald M. Summers, Chen Chen, Yan Zhuang

1. Introduction

The cell cycle is the driving force behind the growth and development of all living organisms. This well-studied sequence of cellular events is tightly regulated and aberrations in such mechanisms can lead to genomic instability. It is divided into distinct phases—G1, S, G2, and M—that coordinate cell growth, DNA replication, and division, respectively, to ensure accurate transmission of genetic material to daughter cells. Each phase is controlled by specific checkpoints that monitor cellular integrity and proper progression through the cycle. This well-studied sequence of cellular events is tightly regulated and aberrations in such mechanisms can lead to genomic instability, a key driver of various diseases including cancer Kastan and Bartek (2004); Malumbres and Barbacid (2009). Live cell fluorescence microscopy has become a powerful tool for studying cell cycle progression in real time, particularly through the genetic engineering of fluorescent reporters like the FUCCI system Sakaue-Sawano et al. (2008); Stallaert et al. (2022). This system enables the distinction of cell cycle phases from single images by fluorescently tagging the two proteins Cdt1 and Geminin, whose expression changes distinctively with the cell cycle (Figure 1). Recently, such reporters

have even been instrumental in connecting live-cell imaging with single-cell transcriptomics to link cell cycle states with gene expression programs Bues et al. (2025). However, despite its utility, the classic FUCCI system and recent variants Sakaue-Sawano et al. (2017); Grant et al. (2018) are limiting in practice as they occupy two of the few available microscopy channels, reducing the ability to study other cellular processes simultaneously and requiring genetic modification that might interfere with the endogenous cell cycle regulation.

In contrast, brightfield microscopy is an easily accessible and label-free imaging modality that does not require genetic engineering, albeit providing limited specificity and imaging contrast. Although some cell cycle transitions, such as nuclear envelope breakdown, are marked by distinct morphological changes that are easily detectable, other cell cycle transitions are visually indiscernible in individual brightfield images of cells. While machine learning approaches have demonstrated some success in classifying cell states from single label-free snapshots Blasi et al. (2016); Eulenberg et al. (2017); Ounkomol et al. (2018), they often lack the temporal context required to resolve more subtle or continuous transitions between cell cycle phases. In this paper, we ask whether leveraging the *temporal information* in time-lapse brightfield microscopy images of cells would allow to predict *continuous* cell cycle states without the need for fluorescent reporters such as FUCCI. To address this, we study several sequence-based deep learning models, including transformers Vaswani et al. (2017) and recently proposed state-space models Gu et al. (2022); Gu and Dao (2024). In particular, we will investigate both *causal* sequence models that only use information from previous time points, as well as *non-causal* models that may ingest the entire sequence. Providing a new dataset of over 1 M images of segmented and tracked human RPE1 cells with accompanying ground truth FUCCI signals, we show that both causal and transformer-based non-causal models significantly outperform single-frame approaches, enabling the prediction of morphologically subtle cell state transitions like the G1/S transition within 1h resolution from live-cell brightfield imaging alone.

2. Related Works

The prediction of individual cell states from single microscopy images has found considerable interest in the literature. Deep learning models, particularly convolutional neural networks (CNNs), have emerged as a promising avenue to predict cell cycle phases directly from image data. For example, Rappez et al. (2020) classified cell cycle states using CNNs trained on Hoechst-stained images, while Narotamo et al. (2021); Li et al. (2024a) employed nuclear features from DAPI-stained images to classify *discrete* cell cycle interphase states such as the G1, S, or G2 phases from static multichannel images of cells. Label-free imaging modalities, such as brightfield or phase contrast imaging, present an attractive alternative to fluorescence-based methods due to their non-invasive nature and wider accessibility. However, accurately annotating discrete cell cycle stages is highly dependent on the imaging modality; while nuclear stains like DAPI or Hoechst provide clear features, phase or brightfield imaging requires substantially more manual annotation expertise. Despite this challenge, Eulenberg et al. (2017); Blasi et al. (2016) demonstrated the feasibility of using brightfield and darkfield imaging combined with machine learning to analyze cell cycle stages without fluorescence labeling. Jin et al. (2021) applied a ResNet approach to classify brightfield images, addressing challenges of imbalanced datasets using generative adversarial networks, while He et al. (2022) showcased how phase imaging with computational specificity (PICS) can accurately classify cell cycle staging. However, treating cells as static snapshots neglects the inherent temporal dynamics of the cell cycle, and misses that cells exhibit temporally correlated heterogeneity, for instance shape and size, which can confound prediction from single frames. To address this, recent methods have incorporated time, starting with approaches that implicitly encode temporal structure.

One effective strategy that encodes time without an explicit sequence encoder is time-contrastive learning. In these primarily unsupervised applications, the temporal index is leveraged to construct positive triplets, thereby temporally regularizing the learned embeddings Hirata-Miyasaki et al. (2025); Sermanet et al. (2018); Hyvarinen and Morioka (2016). Notably, Bonte et al. (2025) employed a time-regularized Variational Autoencoder to predict *discrete* cell cycle labels from the SiR-DNA fluorescent channel. To capture explicit temporal dependencies, *sequence models* have been widely adopted. While Recurrent Neural Networks (RNNs) Hochreiter and Schmidhuber (1997) established a strong baseline for modeling temporal data Hewamalage et al. (2021), a notable example being Jose et al. (2024), which uses RNNs for cell cycle prediction in morphologically stark intermitotic phases, they often struggle with computational bottlenecks when processing long sequences. Consequently, *State-Space Models* (SSMs), such as Mamba Gu and Dao (2024), have recently gained considerable interest. Unlike traditional approaches, SSMs offer linear scaling with sequence length, providing significant improvements in both training and inference efficiency for high-dimensional data Gu et al. (2022); Xiao et al. (2025); Li et al. (2024b). By design, these models are inherently causal, processing sequences unidirectionally unless explicitly adapted for bidirectional modeling. In contrast, *Transformers* Vaswani et al. (2017) have revolutionized sequence modeling by leveraging self-attention mechanisms to access global context

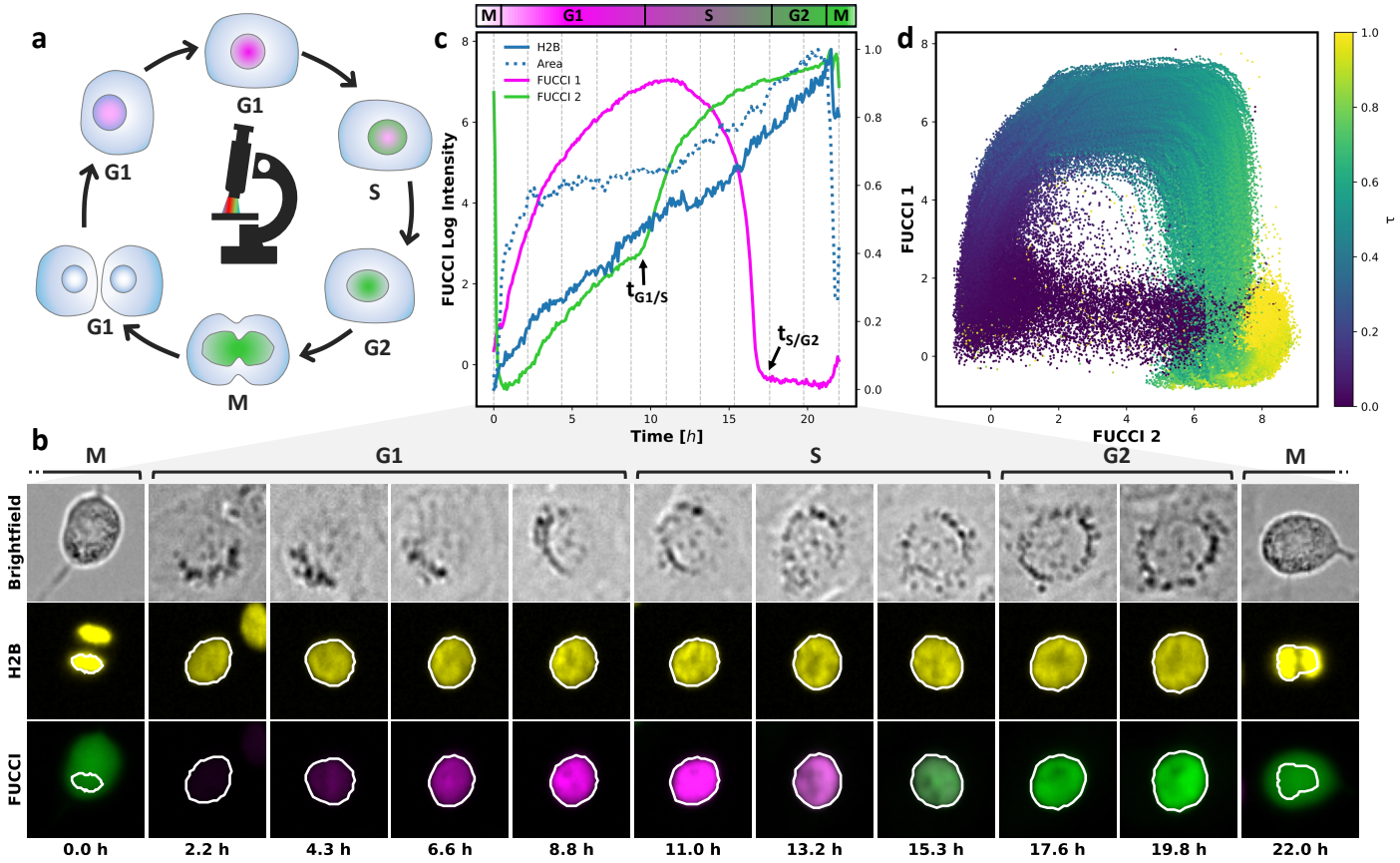

Figure 1: **Multi-modal imaging of Fucci-reporter cells reveals a continuous representation of cell cycle states. a)** Time-lapse imaging of FUCCI-reporting cells allows for precise quantification of cell cycle staging through the characteristic oscillations of fluorescent reporter intensities. **b)** Representative time-lapse images of brightfield, H2B, and FUCCI channels across one full (M-M) cell cycle. **c)** Quantification of integrated logarithmic fluorescence intensities from FUCCI reporters, normalized nuclear area, and normalized total H2B signal in a representative full M-M (mitosis to mitosis) track. Vertical lines mark the time points corresponding to the images shown in d. **d)** Log-transformed FUCCI manifold for continuous inference of cell cycle states.

across the entire sequence simultaneously. In the domain of time-series analysis Wen et al. (2023), this architecture is inherently non-causal (bidirectional), allowing the model to utilize future context for state estimation. However, when strict causality is required—such as in real-time forecasting—attention can be masked to prevent information leakage from future time steps Lim et al. (2020). This capability to model complex temporal dependencies has recently been adapted for single-cell dynamics, where Gallusser and Weigert (2024) and O'Connor and Dunlop (2025) successfully employed Transformer-based architectures to solve challenging cell tracking and state prediction tasks. Ultimately, as the cell cycle is a continuous, causal process, sequence models appear to be the most natural fit for our task. By aligning the computational framework with these intrinsic biological dynamics, we can effectively capture the progressive evolution of cell cycle states.

3. Methods

3.1 Dataset

We generated a large dataset of dividing human FUCCI RPE1 cells using combined brightfield and fluorescence time lapse microscopy. Movies spanning 72 hours were acquired at a 5 minute time resolution, capturing multiple cell cycles. In addition to the brightfield modality, we acquired a nuclear marker channel (Histone H2B) Kanda et al. (1998) and the two FUCCI channels ($\text{FUCCI}_{1/2}$) (see Appendix A.1-2 for the precise experimental protocol). Based on the H2B channel, we segmented the cell nuclei with a custom StarDist model Schmidt et al. (2018) and tracked them across frames using TrackMate Tinevez et al. (2017). Note that since the amount of H2B histones needed by cells to pack DNA doubles during S (DNA replication), the H2B channel does contain information on cell cycle progression (Figure 1c). Below, we leverage this as a control

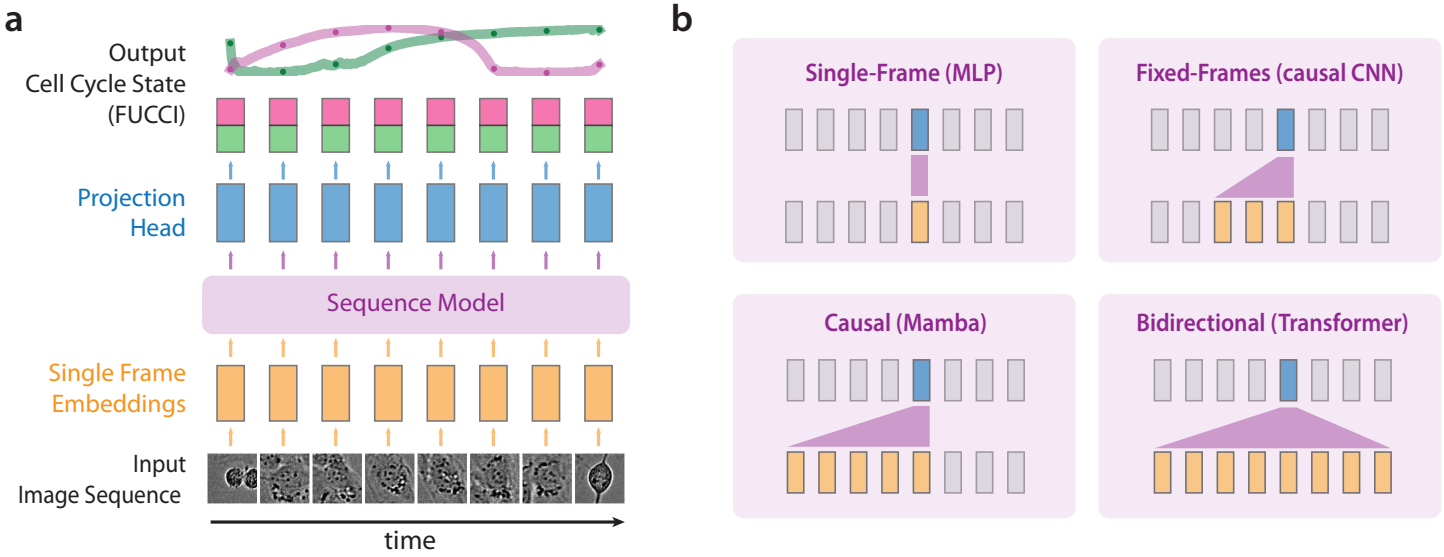

Figure 2: **Overview of the approach. a)** We use a ResNet-18 He et al. (2015) to extract single frame embeddings from an input sequences which are then fed into a sequence model that predicts both FUCCI channels. **b)** Sequence models explored in this paper: Single Frame MLP, Fixed-frame CNN, causal state-space models *e.g.* Mamba Gu and Dao (2024), bidirectional models *e.g.* transformers Vaswani et al. (2017).

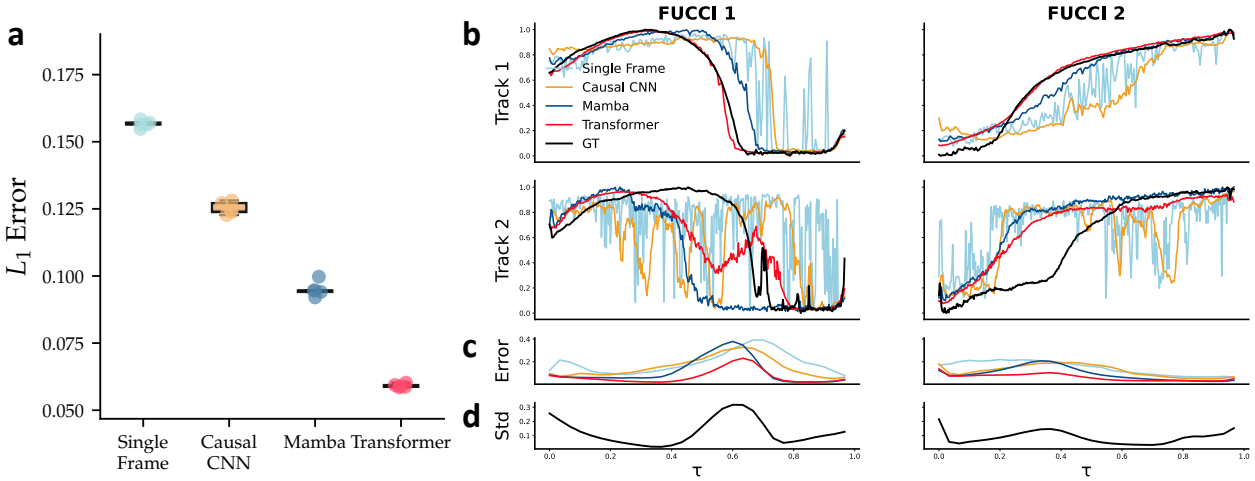

Figure 3: **Predictions on unperturbed RPE1 cells Regular. a)** Distribution of $L1$ errors across the different models. Each box summarizes, across 5 independent training runs (random seeds), the mean L_1 error over all tracks in the held-out test set; the spread reflects run-to-run variability from random initialization. **b)** Predictions of FUCCI signals on two example tracks: one with accurate and one with poor predictions. The ground truth signal is shown in black. **c)** Average prediction error, and **d)** variability of ground truth FUCCI signals as a function of normalized cell cycle time τ.

to assess predictive performance of brightfield *vs.* H2B. Full cell cycle tracks (from one mitosis to the next, M-M) were identified using K-Means clustering and ground-truth FUCCI signals were computed by normalizing the average FUCCI intensities measured across the segmented nuclear mask (Figure 1c,d). The training dataset comprises 5,188 full (M-M) cell cycle tracks with an average track length of 230 frames. Each track contains paired brightfield and H2B images of size 64×64 centered on the nucleus and the corresponding integrated FUCCI signals. To evaluate model performance, we created two additional test datasets: REGULAR, which contains 358 additional full tracks from RPE1 cells acquired at similar conditions as the training set, and DRUG, which comprises 73 complete tracks of cells treated with the cell cycle inhibitor *Palbociclib* that heavily distorts the cell cycle and which is used in the clinic to treat breast cancer. The split is performed at the level of whole movies (one 96-well plate well = one movie): all

Table 1: **Cell cycle prediction accuracy for brightfield on Regular**. Shown are mean and standard deviation of the L_1 error per FUCCI channel and Δ_{DTW} for full tracks across REGULAR. Δ_{DTW} when using H2B images as input for comparison. The $\pm$ values report the standard deviation of per-track errors across all test tracks.

	Brightfield			Histone H2B
Models	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	Δ_{DTW}	Δ_{DTW}
Single Frame	0.193 $\pm$ 0.066	0.146 $\pm$ 0.045	3.735 $\pm$ 0.863	2.595 $\pm$ 1.201
Causal CNN	0.157 $\pm$ 0.078	0.122 $\pm$ 0.049	2.468 $\pm$ 0.917	2.165 $\pm$ 1.210
Mamba	0.112 $\pm$ 0.072	0.091 $\pm$ 0.049	1.444 $\pm$ 0.898	1.426 $\pm$ 0.949
Transformer	**0.066 $\pm$ 0.038**	**0.062 $\pm$ 0.037**	**1.285 $\pm$ 0.553**	**1.155 $\pm$ 0.612**

tracked nuclei from a given movie are assigned exclusively to one set, so no movie contributes tracks to more than one split, which rules out leakage from cells sharing the same field-of-view, illumination, or local imaging artefacts (see Appendix A.4 for the full split composition and the per-movie normalization protocol). In total, this training and testing dataset consists of approximately 1.3 M images and FUCCI signals, all of which we make publicly available alongside this paper to facilitate future devolvements by the community[1].

3.2 Model

The cell cycle prediction task can be formalized as follows: Given a temporal sequence of N brightfield images $X \in \mathbb{R}^{N\times 64\times 64}$, each associated with a corresponding two-dimensional FUCCI signal $Y \in \mathbb{R}^{N\times 2}$, the goal is to train a model f that predicts the normalized FUCCI intensities across the entire sequence in a supervised manner, *i.e.* $f : \mathbb{R}^{N\times 64\times 64} \to \mathbb{R}^{N\times 2}$. Note that the length N of the input is not fixed and the sequences is not required to span the entire cell cycle, allowing us to analyze the impact of temporal context on prediction performance. For this cell cycle prediction task, we evaluated three conceptually distinct model classes: single-frame models, causal models and non-causal models (Figure 2). Each model uses first a ResNet-18 He et al. (2015) as feature extractor which for every input image in the sequence independently creates a 512-dimensional embedding. The backbone is initialized with ImageNet-pretrained weights (torchvision, `pretrained=True`) and fine-tuned end-to-end jointly with the sequence head; backbone and head share a single AdamW optimizer with the same learning rate (10^{-4}) and weight decay (0.01), with no differential learning rate or additional regularization on the backbone. The sequence of embeddings is then fed into the proper sequence model head that differs between the three classes: *Single-frame models* serve as a baseline, predicting the FUCCI signal from each image embedding independently without leveraging temporal information. We use a simple 4 layer MLP with hidden dimension 512. *Fixed-frames models* that use a fixed history of past frames for prediction and for which we use a causal convolutional neural network Oord et al. (2016) with a fixed causal temporal receptive field. *Causal models* that incorporate past and present temporal context in more flexible way and which are able to potentially capture arbitrary long temporal dependencies. In particular, we compare LSTMs Hochreiter and Schmidhuber (1997), Mamba Gu and Dao (2024), and causal transformers with masked attention. *Bidirectional models* process the entire sequence bidirectionally, using both past and future frames for inference (*i.e.* non-causal information). We use a standard transformer (4 layers) as representative architecture. To ensure a fair comparison, we choose all sequence heads to have the same number of parameters ($\approx 1M$). All transformer variants additionally use rotary positional embeddings Su et al. (2021) to encode the relative temporal position of each frame. We train each model for 150 epochs while randomly sampling subtracks of variable lengths, using L_1 loss. We optimize with AdamW with a cosine annealing learning rate schedule, an initial learning rate of 10^{-4}, weight decay of 0.01, and a batch size of 8; no gradient clipping or early stopping is applied. These hyperparameters were selected following standard sequence-modelling practice rather than systematic search, as the goal was a controlled architectural comparison rather than maximum absolute performance. We train 5 independent runs with different random seeds and report metrics from one randomly selected seed, chosen prior to evaluation. We use standard data augmentation such as random rotations and flips, which is applied to the whole track.

More concretely, during training a random contiguous sub-sequence is drawn from each track per batch. The window length is sampled per batch from {4, 16, 32, 64, 96, 128, 192 } frames and is held constant across all tracks within that batch. If a track is shorter than the sampled window, its frames are repeated via uniform resampling (`numpy.linspace` over the available frame indices) so that every sequence reaches exactly the target length; no zero-padding is used. At inference, full tracks are passed as-is without sub-sampling or truncation, and the evaluation batch size is 1, so no alignment or padding is required. There is no fixed architectural limit on sequence length: all temporal heads (transformer with RoPE, LSTM, causal

1. `https://zenodo.org/records/14987478`

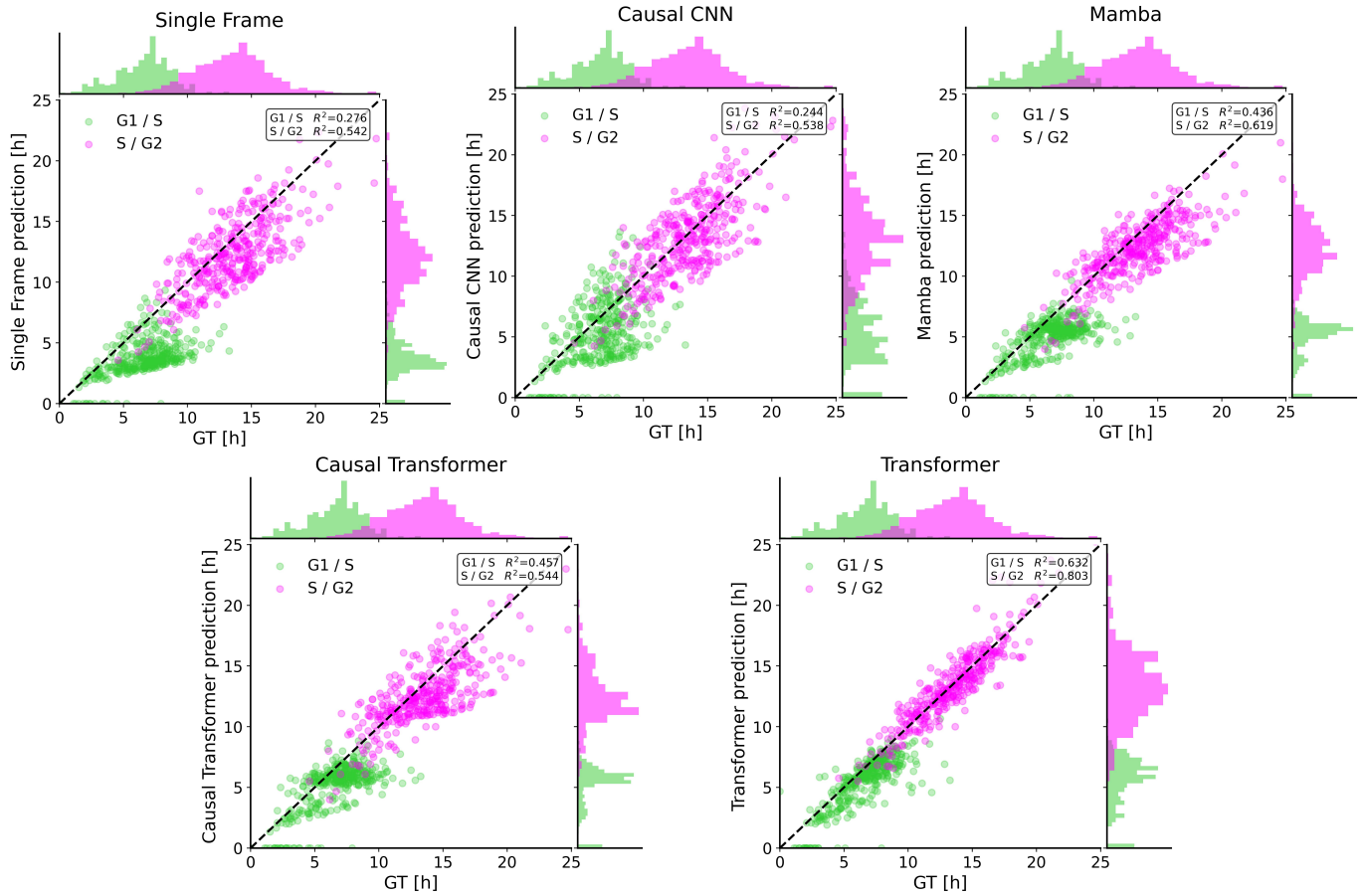

Figure 4: **Distribution of Phase Transition Timings in Ground Truth and Model Predictions Across Different Temporal encoders for Brightfield Imaging** The joint distribution of GT and predicted cell-cycle phase transition timings is here represented for the REGULAR data. The non-causal transformer is able to significantly outperform the other temporal encoders, in both G1/S and S/G2 transitions.

CNN) support arbitrary-length inputs. The longest tracks in our dataset are ∼400 frames (∼33 h at 5 min per frame); the practical upper bound at inference is GPU memory, particularly for the bidirectional transformer.

3.2.1 Compute and efficiency

All models were trained and benchmarked on a single NVIDIA H100 GPU (94 GB HBM3), on a node with dual AMD EPYC 9334 CPUs and 768 GB of system RAM (PyTorch 2.10 / CUDA 13.0); no multi-GPU parallelism was used. Because all architectures share the same ResNet-18 backbone, which dominates the per-step cost, the total training time over 150 epochs is comparable across models, ranging from ∼112 to ∼127 minutes (Supp. Table B 2). At inference, every model processes a full track in under 16 ms (Supp. Table B 3). Among causal architectures, Mamba scales best to longer tracks (5.8 ms at $N{=}512$), whereas the LSTM becomes substantially slower (15.5 ms); causal and bidirectional transformers incur essentially identical cost.

3.2.2 Performance metrics

We evaluated model prediction by computing the mean L_1 error for each FUCCI channel across a given track. Further, we use the *dynamic time warping distance* Δ_{DTW} between the predicted and ground truth FUCCI signal that takes into account both the signal prediction error as well as the temporal misalignment between the two signals. We use the default Δ_{DTW} distance implementation from the `dtaidistance` package Meert et al. (2020) with a penalty of 0.1. Additionally, we introduced two biologically meaningful cell cycle checkpoints (Figure 1) and measured the time difference between our predicted and observed checkpoints in minutes. The first checkpoint $t_{G1/S}$ is the onset of the Geminin (FUCCI 2) signal, marking the G1/S phase transition Sakaue-Sawano et al. (2017) (additional information about how the checkpoints were calculated, including a threshold sensitivity analysis, can be found in the Appendix; see Figure 13). While the classic FUCCI reporter does not provide an exact molecular landmark of the S/G2 transition,

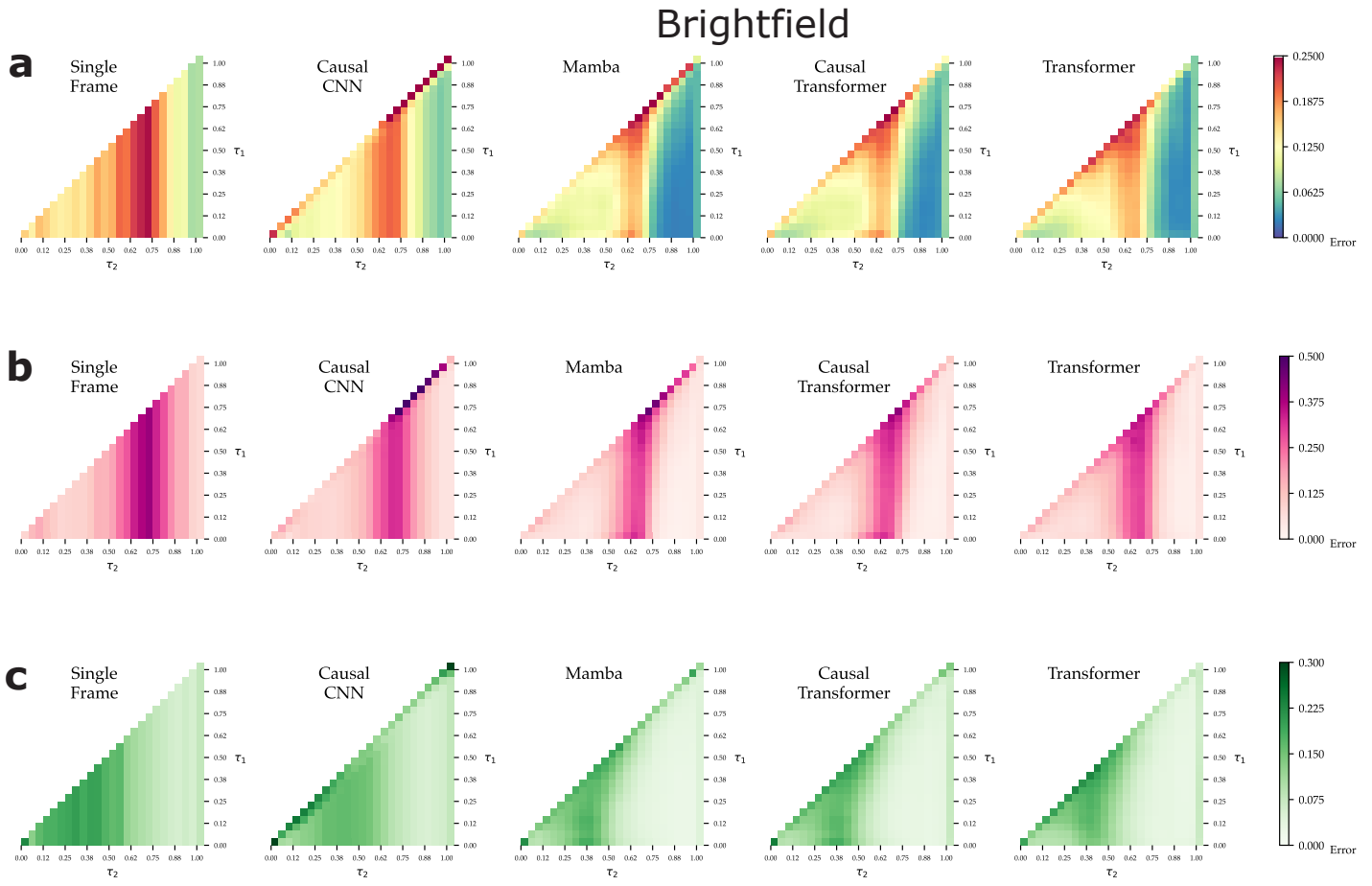

Figure 5: **Comparative performance on partial cell cycle tracks (brightfield).** Shown is the L_1 error of both FUCCI signals when using partial tracks as input, parametrized by their relative start and end time $\tau_1 \leq \tau_2 \in [0, 1]$. **a)** The average of the two FUCCI **b)** FUCCI 1 only: The highest error is obtained around the steep S/G2 transition where the signal declines steeply **c)** FUCCI 2 only: The highest error is obtained around the steep G1/S transition

we use the disappearance of the Cdt1 (FUCCI 1) signal as an approximate landmark for evaluating S/G2 transition predictions $t_{S/G2}$. We use these two landmarks to categorize cells into discrete G1, S, and G2 phase classes for which we compute the $F1$-score of predictions. S phase is more constrained than G1, as genome duplication depends on DNA polymerase rates and replication origins. G1 is the key regulator of cell cycle duration and is expected to be more variable. In REGULAR, the biological variability in phase duration ranges from a standard deviation of 1.6 hour for the S phase, to 2.2 hour for the G1 phase. Throughout this paper, box plots in figures summarize run-to-run variability across 5 independent training seeds, while the $\pm$ values in tables report the standard deviation of per-track errors across all test tracks from a single representative training run.

4. Results

4.1 Comparison of prediction accuracy across sequence models

Models	$\Delta t_{G1/S}$[min]	$\Delta t_{S/G2}$[min]	$G1$	S	$G2$
Single Frame	191.3	117.9	0.71	0.64	0.87
Causal CNN	146.8	113.3	0.78	0.64	0.86
Mamba	111.4	102.4	0.83	0.72	0.89
Transformer	**60.1**	**57.2**	**0.90**	**0.85**	**0.93**

Table 2: **Prediction accuracy of biological checkpoints and selected cell cycle states from brightfield images on Regular.**

We first compare the performance of the different sequence models on predicting the FUCCI signals from brightfield images for full (M-M) cell cycle tracks. As seen in Figure 3a, the single frame model predicts extremely noisy signals that deviate substantially from the ground truth FUCCI signal, whereas both causal and bidirectional models achieve qualitatively much better predictions on both

FUCCI channels and generally aligns with the expected trends (*cf.* Supp. Figure 9a,b, Supp. Figure 10). To quantitatively assess the performance of the different models, we show the mean L_1 error for each FUCCI channel across all tracks as well as the average Δ_{DTW} distance between the predicted and ground truth FUCCI signals for the different models in Table 1. As expected, the single frame model which operates without integrating temporal information performs the worst across all metrics ($\Delta_{DTW} = 3.735$), while integrating the full bidirectional (non-causal) sequence information via a transformer achieves the best prediction ($\Delta_{DTW} = 1.285$). Surprisingly, there is a notable difference between the performance of the fixed-frames model (causal CNN) and the state-space model (Mamba), with the former performing substantially worse than the latter ($\Delta_{DTW} = 2.468$ *vs.* 1.444). This suggests that models that allow information propagation across the entire sequence can be more effective than models that only use a fixed-size temporal context. We also evaluated LSTM and causal transformer as additional flexible causal architectures; their performance on REGULAR is comparable to Mamba, as detailed in Supp. Table B 4. We additionally computed Δ_{DTW} when training with the H2B channel as input modality, which a priori should be a substantially easier task as the H2B signal is biologically correlated with the cell cycle. Indeed, this is corroborated by the performance of the single frame model that vastly improves in this case ($\Delta_{DTW} = 2.595$). Interestingly, both state-space models as well as the transformer only marginally improve, suggesting that these models are able to extract temporal cues from the brightfield images comparable to the easier H2B modality. The predictions at biological checkpoints remain consistent with the performance observed across other evaluation metrics (Table 2, Supp. Figure 11). As observed in Figure 4, the timing of these checkpoints is highly variable. Sequence models effectively capture this variability, with the bi-directional transformer predictions being more accurate and therefore better aligned with the ground-truth distribution. In contrast, other methods tend to underestimate such variability. The bidirectional transformer requires the full sequence and is suited for offline/retrospective analysis; for real-time applications, only causal models apply. As expected, the majority of prediction errors measured by the L_1 metric occur near the $t_{G1/S}$ and $t_{S/G2}$ landmarks, reflecting abrupt transitions where the FUCCI signal exhibits higher variability (Figure 3 b, c).

4.2 Prediction on partial tracks

So far we focused on full (M-M) tracks of non-perturbed cells, all of which exhibit fairly stereotypical cell cycle trajectories and for which sequence models are able to base their predictions on a well defined starting points (*i.e.* the cell division event). We now evaluate the performance of the different models on partial tracks, where the starting point is not known a priori, which is a more challenging task.

In Figure 5, we show the average L_1 error with cropped partial tracks as input, indicated by their relative start and end time $\tau_1 \leq \tau_2 \in [0, 1]$. These partial tracks ranged from single-frame portions (along the diagonal) to entire tracks (lower-right element). Causal models still achieved better accuracy in predicting FUCCI values compared to non-temporal MLPs or fixed history CNNs. Surprisingly, the performance advantage of transformers over causal methods observed in full tracks diminishes on partial tracks. For all models, the error is maximal for segments that end near $t_{S/G2}$. Errors in FUCCI_1 show the same pattern while the errors in the FUCCI_2 arise mostly when taking segments from the beginning of the cell cycle).

4.3 Prediction on out-of-distribution perturbations

Finally, we evaluate the model performance on DRUG, *i.e.* biologically strongly perturbed cells that can be considered out-of-distribution. Specifically, these cells were treated with the drug *Palbociclib*, a CDK4-6 inhibitor, that increases the cell cycle duration almost two-fold from ∼20h to 40 h (Figure 6 a) and specifically the G1 phase duration, leading to a strongly distorted cell-cycle (Figure 6 b). As expected, almost all models demonstrated a significant drop in accuracy when predicting FUCCI signal in these unseen drug-treated cells, as indicated by all evaluation metrics (Table 3). The notable exception is the bidirectional transformer, that provides reasonable predictions and correctly captures distortions in the G1 phase (Figure 6 c) that all other models significantly underestimated. Interestingly, the MLP outperformed the other causal models on this distorted data, potentially as the latter overfitted on the training data. When performing the same analysis on H2B, we found slightly better predictive performance (Supp. Table B 4 Table B 5, Figure 7, Figure 12), which is expected due to the stronger correlation with the cell cycle.

Models	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	Δ_{DTW}
Single Frame	0.239 ± 0.082	0.182 ± 0.056	5.329 ± 1.147
Causal CNN	0.252 ± 0.113	0.161 ± 0.059	4.323 ± 1.302
Mamba	0.485 ± 0.090	0.259 ± 0.045	3.563 ± 1.918
Transformer	**0.147 ± 0.056**	**0.139 ± 0.048**	**3.022 ± 0.985**

Table 3: **L_1 error and Δ_{DTW} for Fucci channels on Drug Dataset.** The non causal transformer generalizes better to out of distribution data with a longer than usual G1 phase. The ± values report the standard deviation of per-track errors across all test tracks.

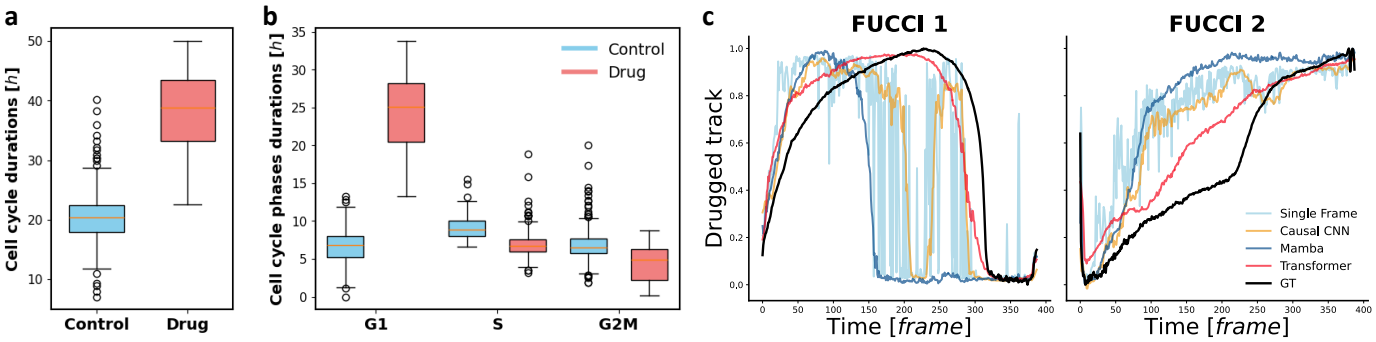

Figure 6: **Results on perturbed RPE1 cells Drug. a,b)** Effect of CDK4/6 inhibition on cell cycle durations and onset of G1, S, and G2/M phases, extending G1 duration while leaving S and G2/M unchanged. **c)** Example FUCCI predictions on DRUG. Causal models struggle to adapt to the longer tracks, underestimating the length of G1.

5. Comparison to Fluorescence

To establish a performance benchmark for our label-free approach, we trained our sequence models using images of Histone 2B (H2B)-labeled nuclei. H2B is a core chromatin component, widely used as a nuclear marker to visualize chromatin condensation and nuclear morphology Kanda et al. (1998). Biologically, H2B features are intrinsically linked to cell cycle progression; nuclear size scales with cell growth, and integrated intensity often correlates with the accumulation of Geminin (Fucci Green) during the S and G2 phases. Consequently, H2B-based prediction represents a highly informative "upper bound" for determining cell cycle state from morphology.

As expected, models trained on H2B sequences achieved the lowest overall reconstruction error (Table 4, Figure 7), confirming that direct observation of nuclear dynamics provides rich information for inferring cell cycle status. Consistent with our brightfield findings, the non-causal transformer architecture yielded the best performance, effectively leveraging future context to resolve ambiguous phases. However, the performance gap between H2B and brightfield inputs was remarkably narrow for untreated cells. This similarity suggests that, under normal growth conditions, the label-free brightfield signal captures a density of morphological cues that is nearly as informative as the explicit nuclear channel.

A divergence in performance was observed only under drug-perturbed conditions (Supp. Table B 5). While brightfield prediction accuracy degraded when cells were exposed to cell cycle inhibitors, H2B-based models remained more robust. This indicates that while general cellular morphology (captured in brightfield) can remain the same despite the strong effect of a perturbation, nuclear chromatin dynamics provide a more stable and resilient feature set for tracking cell cycle progression under stress.

6. Discussion and Conclusions

In this study, we generated and released a large time-lapse imaging dataset of cycling RPE1 cells under both normal and drug-treated conditions and used it to investigate the utility of sequence models to infer the continuous cell cycle state from label-free brightfield images. Our analysis demonstrates that temporal sequence models can significantly improve the cell cycle prediction accuracy and enable the assessment of cell cycle state from brightfield images to a level comparable when using the more informative H2B channel. However, the "black box" nature of these deep learning models raises important questions regarding interpretability. Future work should leverage emerging explainable AI (XAI) techniques, such as spatiotemporal saliency maps, to visualize the specific image regions driving the model's decisions Nabi et al. (2024). Beyond feature visualization, true interpretability also requires communicating the model's reliability. To this end, it would be valuable to move beyond point predictions and explore probabilistic architectures like variational autoencoders Zhao et al. (2019) Bonte et al. (2025). This suggests that brightfield morphological cues alone carry sufficient information for cell cycle inference. Importantly, we found that causal state-space models substantially outperform commonly used fixed-history convolutional networks, demonstrating their potential for real-time computer vision and smart microscopy applications, where such causal inference is essential Simon et al. (2025), Mahecic et al. (2022). Such real-time capabilities are particularly relevant for Event-Driven Acquisition (EDA) frameworks, which aim to automate microscope control based on the observed biological dynamics. The bidirectional transformer, by contrast, requires access to the complete sequence and is therefore suited exclusively to offline retrospective analysis (e.g., dataset annotation, post-hoc cell cycle reconstruction). This is an inherent trade-off, not a limitation of any single model. Recent work has highlighted the potential of "Hybrid-EDA" approaches, where label-free imaging serves

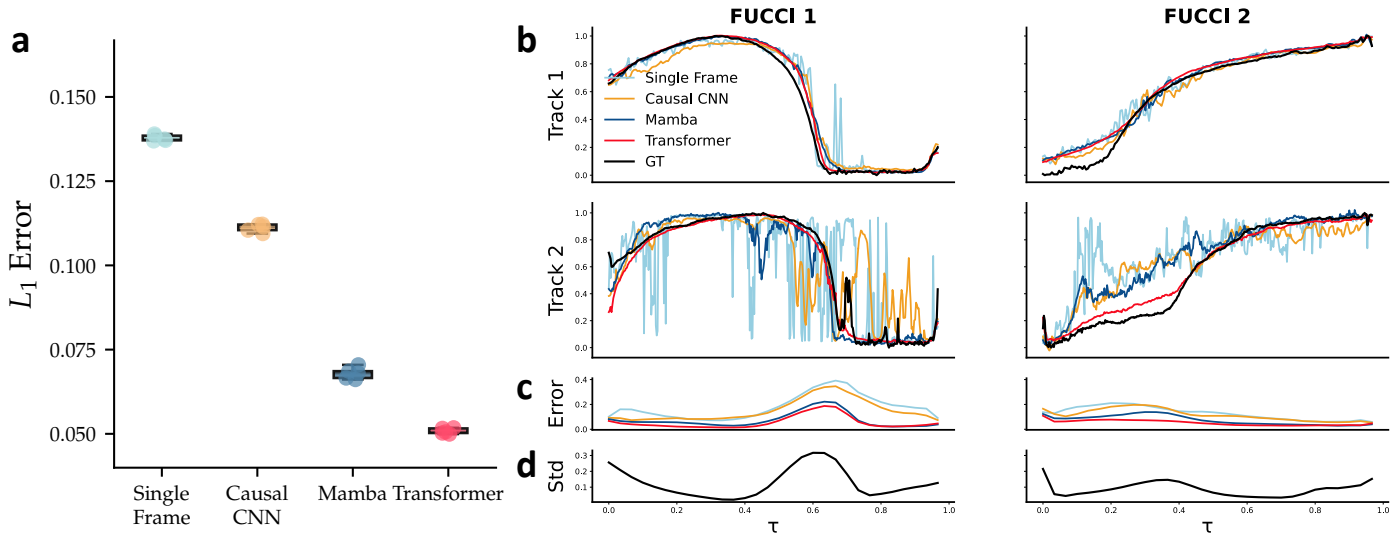

Figure 7: **Comparative Performance of Temporal Encoders in Predicting Continuous Cell Cycle States from H2B Imaging in Unperturbed RPE1 Cells. a)** Distribution of L1 errors across the different models. Each box summarizes, across 5 independent training runs (random seeds), the mean L_1 error over all tracks in the held-out test set; the spread reflects run-to-run variability from random initialization. **b)** Example predictions of FUCCI signals from different models on two tracks: one with accurate predictions and one with poor predictions. The ground truth signal is shown in black. **c)** Average prediction error and **d.** ground truth standard deviation are plotted in function of cell cycle phases.

as a continuous, low-toxicity surveillance channel to trigger fluorescence acquisition only when specific events occur Stepp et al. (2025). We note that the observed reduced accuracy on drug-treated cells underscores that creating general predictive cell state models for strong biological perturbations remains challenging. These results should be interpreted cautiously: DRUG covers a single CDK4/6 inhibitor, and generalisation claims would require validation across other inhibitor classes, drug concentrations, and cell lines. To address similar generalization hurdles, foundation models trained on massive, multi-modal repositories seem a promising avenue. Recent works, such as Gupta et al. (2025) and Kraus et al. (2023), demonstrate that self-supervised learning can extract robust, universal representations of cellular morphology. These large-scale models are particularly impactful for tasks like "virtual staining," where the network learns to infer unseen subcellular structures—such as nuclei or cytoskeletal components—directly from label-free modalities like brightfield. For broader applicability, the model needs exposure to more diverse cell types, microscopy setups, and biologically perturbed conditions. Capturing variations in morphology, imaging parameters, and drug-induced alterations will enhance robustness, allowing it to reliably detect atypical or rare cell cycle dynamics. Our findings demonstrate that sequence models can be effective predictors of cellular dynamics in more controlled settings, providing a powerful tool for studying proliferation, cancer dynamics, and addressing cell cycle-related confounding factors. With larger datasets, models could improve drug effect predictions, paving the way for screening and broader biomedical applications. For downstream use, we recommend the bidirectional transformer for retrospective analysis with full tracks, and Mamba for real-time or online applications, where it achieves the best causal performance with linear scaling in sequence length.

Acknowledgments

We thank Benjamin Gallusser for his precious help and expertise in the areas of cell segmentation and cell tracking.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

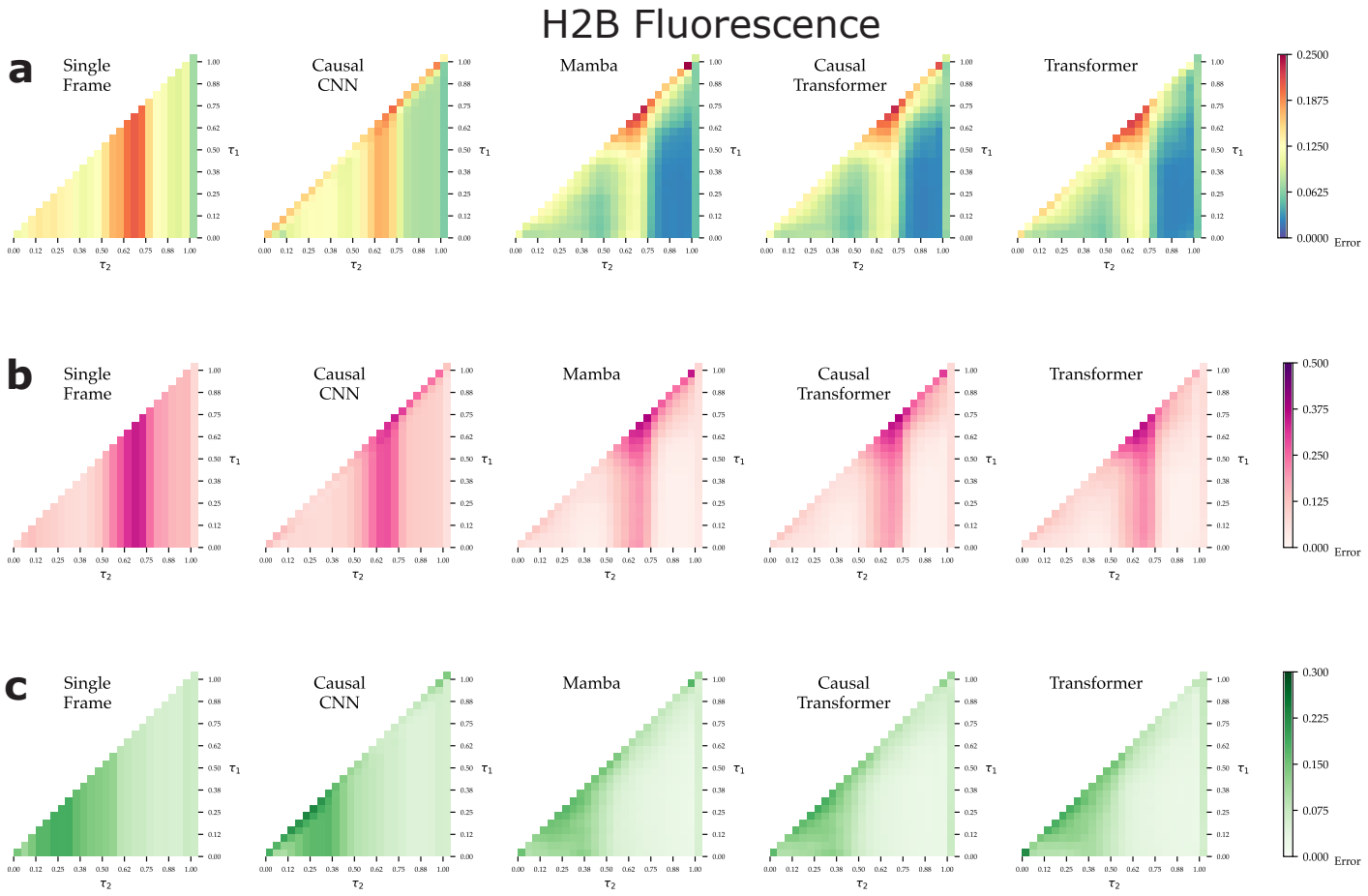

Figure 8: **Comparative performance on partial cell cycle tracks (H2B).** Shown is the average L_1 error of both FUCCI signals when using partial tracks as input, parametrized by their relative start and end time $\tau_1 \leq \tau_2 \in [0, 1]$.

The authors declare no conflicts of interest.

Data availability

The datasets generated during this study, comprising approximately 1.3 M images and FUCCI signals, are publicly available on Zenodo at `https://zenodo.org/records/14987478`. This includes the *Regular* test set (358 tracks from RPE1 cells) and the *Drug* test set (73 tracks from cells treated with Palbociclib).

References

Thomas Blasi, Holger Hennig, Huw D. Summers, Fabian J. Theis, Joana Cerveira, James O. Patterson, Derek Davies, Andrew Filby, Anne E. Carpenter, and Paul Rees. Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nature Communications*, 7(1):10256, January 2016. ISSN 2041-1723. . URL `https://www.nature.com/articles/ncomms10256`. Publisher: Nature Publishing Group.

Thomas Bonte, Oriane Pourcelot, Adham Safieddine, Floric Slimani, Florian Mueller, Dominique Weil, Edouard Bertrand, and Thomas Walter. A deep learning approach for time-consistent cell cycle phase prediction from microscopy data. *PLOS Computational Biology*, 21(12):e1013800, December 2025. ISSN 1553-7358. . URL `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1013800`. Publisher: Public Library of Science.

Johannes Bues, Joern Pezoldt, Camille Lucie Lambert, Benjamin David Hale, Elisa Bugani, Ramon Vinas Torne, Timothee Ferrari, Nadia Grenningloh, Vincent Gardeux, Shuo Wen, Caroline Wandinger, Maximilian Kohnen, Romina Augustin, Katharina Eckstein, Assia Ouanaya, Jillian Love, Sarthak Saha, Amirhossein Saba, Aviv Huttner, Maria Vittoria Impagliazzo, Jose Antonio Vasquez Porto Viso, Angel de Jesus Corria Osorio, Demetri Psaltis, Wouter Karthaus, Berend Snijder, Maria Brbic, and Bart Deplancke. Single-cell phenomics through integrated imaging and molecular profiling, December 2025. URL `https://www.biorxiv.org/content/10.`

Table 4: **Performance of Histone H2B modality at predicting Fucci channels.** Sequence encoders outperform the single frame method, with the Transformer model achieving the highest performance across all metrics. The ± values report the standard deviation of per-track errors across all test tracks.

Models	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	R^2	DTW
Single Frame	0.183 ± 0.104	0.130 ± 0.064	0.491 ± 0.431	2.595 ± 1.201
Causal CNN	0.154 ± 0.105	0.118 ± 0.061	0.586 ± 0.415	2.165 ± 1.210
LSTM	0.079 ± 0.065	0.075 ± 0.044	0.833 ± 0.265	1.467 ± 1.161
Causal Transformer	0.079 ± 0.057	0.079 ± 0.042	0.839 ± 0.214	1.552 ± 0.955
Mamba	0.074 ± 0.056	0.075 ± 0.040	0.853 ± 0.215	1.426 ± 0.949
Transformer	**0.056 ± 0.039**	**0.054 ± 0.033**	**0.912 ± 0.116**	**1.155 ± 0.612**

1101/2025.11.28.690954v1. ISSN: 2692-8205 Pages: 2025.11.28.690954 Section: New Results.

Philipp Eulenberg, Niklas Köhler, Thomas Blasi, Andrew Filby, Anne E. Carpenter, Paul Rees, Fabian J. Theis, and F. Alexander Wolf. Reconstructing cell cycle and disease progression using deep learning. *Nature Communications*, 8(1):463, September 2017. ISSN 2041-1723. . URL https://www.nature.com/articles/s41467-017-00623-3. Publisher: Nature Publishing Group.

Benjamin Gallusser and Martin Weigert. Trackastra: Transformer-based cell tracking for live-cell microscopy, July 2024. URL http://arxiv.org/abs/2405.15700. arXiv:2405.15700 [cs].

Gavin D. Grant, Katarzyna M. Kedziora, Juanita C. Limas, Jeanette Gowen Cook, and Jeremy E. Purvis. Accurate delineation of cell cycle phase transitions in living cells with PIP-FUCCI. *Cell Cycle*, 17(21-22):2496–2516, November 2018. ISSN 1538-4101, 1551-4005. . URL https://www.tandfonline.com/doi/full/10.1080/15384101.2018.1547001.

Albert Gu and Tri Dao. Mamba: Linear-Time Sequence Modeling with Selective State Spaces, May 2024. URL http://arxiv.org/abs/2312.00752. arXiv:2312.00752 [cs].

Albert Gu, Karan Goel, and Christopher Ré. Efficiently Modeling Long Sequences with Structured State Spaces, August 2022. URL http://arxiv.org/abs/2111.00396. arXiv:2111.00396 [cs].

Ankit Gupta, Zoe Wefers, Konstantin Kahnert, Jan N. Hansen, Mohini K. Misra, Will Leineweber, Anthony Cesnik, Dan Lu, Ulrika Axelsson, Frederic Balllllosera, Russ B. Altman, Theofanis Karaletsos, and Emma Lundberg. SubCell: Proteome-aware vision foundation models for microscopy capture single-cell biology, October 2025. URL https://www.biorxiv.org/content/10.1101/2024.12.06.627299v2. ISSN: 2692-8205 Pages: 2024.12.06.627299 Section: New Results.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. URL http://arxiv.org/abs/1512.03385. arXiv:1512.03385 [cs].

Yuchen R. He, Shenghua He, Mikhail E. Kandel, Young Jae Lee, Chenfei Hu, Nahil Sobh, Mark A. Anastasio, and Gabriel Popescu. Cell Cycle Stage Classification Using Phase Imaging with Computational Specificity. *ACS Photonics*, 9(4):1264–1273, April 2022. . URL https://doi.org/10.1021/acsphotonics.1c01779. Publisher: American Chemical Society.

Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427, January 2021. ISSN 0169-2070. . URL https://www.sciencedirect.com/science/article/pii/S0169207020300996.

Eduardo Hirata-Miyasaki, Soorya Pradeep, Ziwen Liu, Alishba Imran, Taylla Milena Theodoro, Ivan E. Ivanov, Sudip Khadka, See-Chi Lee, Michelle Grunberg, Hunter Woosley, Madhura Bhave, Carolina Arias, and Shalin B. Mehta. DynaCLR: Contrastive Learning of Cellular Dynamics with Temporal Regularization, July 2025. URL http://arxiv.org/abs/2410.11281. arXiv:2410.11281 [cs].

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997. ISSN 0899-7667. . URL https://ieeexplore.ieee.org/abstract/document/6795963. Conference Name: Neural Computation.

Aapo Hyvarinen and Hiroshi Morioka. Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA, May 2016. URL http://arxiv.org/abs/1605.06336. arXiv:1605.06336 [stat].

Xin Jin, Yuanwen Zou, and Zhongbing Huang. An Imbalanced Image Classification Method for the Cell Cycle Phase. *Information*, 12(6):249, June 2021. ISSN 2078-2489. . URL https://www.mdpi.com/2078-2489/12/

6/249. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Abin Jose, Rijo Roy, Daniel Moreno-Andrés, and Johannes Stegmaier. Automatic detection of cell-cycle stages using recurrent neural networks. *PloS One*, 19(3):e0297356, 2024. ISSN 1932-6203. .

Teru Kanda, Kevin F. Sullivan, and Geoffrey M. Wahl. Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Current Biology*, 8(7):377–385, March 1998. ISSN 0960-9822. . URL https://www.sciencedirect.com/science/article/pii/S0960982298701563.

Michael B. Kastan and Jiri Bartek. Cell-cycle checkpoints and cancer. *Nature*, 432(7015):316–323, November 2004. ISSN 1476-4687. . URL https://www.nature.com/articles/nature03097. Publisher: Nature Publishing Group.

Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, Maciej Sypetkowski, Chi Vicky Cheng, Kristen Morse, Maureen Makes, Ben Mabey, and Berton Earnshaw. Masked Autoencoders are Scalable Learners of Cellular Morphology, November 2023. URL http://arxiv.org/abs/2309.16064. arXiv:2309.16064 [cs].

Gang Li, Eva K. Nichols, Valentino E. Browning, Nicolas J. Longhi, Conor Camplisson, Brian J. Beliveau, and William Stafford Noble. Predicting cell cycle stage from 3D single-cell nuclear-stained images, September 2024a. URL https://www.biorxiv.org/content/10.1101/2024.08.30.610553v2. Pages: 2024.08.30.610553 Section: New Results.

Kunchang Li, Xinhao Li, Yi Wang, Yinan He, Yali Wang, Limin Wang, and Yu Qiao. VideoMamba: State Space Model for Efficient Video Understanding, March 2024b. URL http://arxiv.org/abs/2403.06977. arXiv:2403.06977 [cs].

Bryan Lim, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, September 2020. URL http://arxiv.org/abs/1912.09363. arXiv:1912.09363 [stat].

Dora Mahecic, Willi L. Stepp, Chen Zhang, Juliette Griffié, Martin Weigert, and Suliana Manley. Event-driven acquisition for content-enriched microscopy. *Nature Methods*, 19(10):1262–1267, October 2022. ISSN 1548-7105. . URL https://www.nature.com/articles/s41592-022-01589-x. Publisher: Nature Publishing Group.

Marcos Malumbres and Mariano Barbacid. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer*, 9(3):153–166, March 2009. ISSN 1474-1768. . URL https://www.nature.com/articles/nrc2602. Publisher: Nature Publishing Group.

Wannes Meert, Kilian Hendrickx, Toon Van Craenendonck, Pieter Robberechts, Hendrik Blockeel, and Jesse Davis. DTAIDistance, August 2020. URL https://zenodo.org/records/7158824.

Ivan R. Nabi, Ben Cardoen, Ismail M. Khater, Guang Gao, Timothy H. Wong, and Ghassan Hamarneh. AI analysis of super-resolution microscopy: Biological discovery in the absence of ground truth. *The Journal of Cell Biology*, 223(8):e202311073, August 2024. ISSN 1540-8140. .

Hemaxi Narotamo, Maria Sofia Fernandes, Ana Margarida Moreira, Soraia Melo, Raquel Seruca, Margarida Silveira, and João Miguel Sanches. A machine learning approach for single cell interphase cell cycle staging. *Scientific Reports*, 11(1):19278, September 2021. ISSN 2045-2322. . URL https://www.nature.com/articles/s41598-021-98489-5. Publisher: Nature Publishing Group.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio, September 2016. URL http://arxiv.org/abs/1609.03499. arXiv:1609.03499 [cs].

Chawin Ounkomol, Sharmishtaa Seshamani, Mary M. Maleckar, Forrest Collman, and Gregory R. Johnson. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods*, 15(11):917–920, November 2018. ISSN 1548-7105. . URL https://www.nature.com/articles/s41592-018-0111-2. Publisher: Nature Publishing Group.

Owen M. O'Connor and Mary J. Dunlop. Cell-TRACTR: A transformer-based model for end-to-end segmentation and tracking of cells. *PLOS Computational Biology*, 21(5):e1013071, May 2025. ISSN 1553-7358. . URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1013071. Publisher: Public Library of Science.

Stephan Preibisch, Stephan Saalfeld, and Pavel Tomancak. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics*, 25(11):1463–1465, June

2009. ISSN 1367-4803. . URL https://doi.org/10.1093/bioinformatics/btp184.

Luca Rappez, Alexander Rakhlin, Angelos Rigopoulos, Sergey Nikolenko, and Theodore Alexandrov. DeepCycle reconstructs a cyclic cell cycle trajectory from unsegmented cell images using convolutional neural networks. *Molecular Systems Biology*, 16 (10), October 2020. ISSN 1744-4292, 1744-4292. . URL https://onlinelibrary.wiley.com/doi/10.15252/msb.20209474. ZSCC: 0000000.

Asako Sakaue-Sawano, Hiroshi Kurokawa, Toshifumi Morimura, Aki Hanyu, Hiroshi Hama, Hatsuki Osawa, Saori Kashiwagi, Kiyoko Fukami, Takaki Miyata, Hiroyuki Miyoshi, Takeshi Imamura, Masaharu Ogawa, Hisao Masai, and Atsushi Miyawaki. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*, 132(3):487–498, February 2008. ISSN 0092-8674, 1097-4172. . URL https://www.cell.com/cell/abstract/S0092-8674(08)00054-8. Publisher: Elsevier.

Asako Sakaue-Sawano, Masahiro Yo, Naoki Komatsu, Toru Hiratsuka, Takako Kogure, Tetsushi Hoshida, Naoki Goshima, Michiyuki Matsuda, Hiroyuki Miyoshi, and Atsushi Miyawaki. Genetically Encoded Tools for Optical Dissection of the Mammalian Cell Cycle. *Molecular Cell*, 68(3):626–640.e5, November 2017. ISSN 10972765. . URL https://linkinghub.elsevier.com/retrieve/pii/S1097276517307505.

Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell Detection with Star-convex Polygons. volume 11071, pages 265–273. 2018. . URL http://arxiv.org/abs/1806.03535. arXiv:1806.03535 [cs].

Pierre Sermanet, Corey Lynch, Yevgen Chebotar, Jasmine Hsu, Eric Jang, Stefan Schaal, Sergey Levine, and Google Brain. Time-Contrastive Networks: Self-Supervised Learning from Video. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1134–1141, May 2018. . URL https://ieeexplore.ieee.org/document/8462891. ISSN: 2577-087X.

Franck Simon, Maria Colomba Comes, Tiziana Tocci, Louise Dupuis, Vincent Cabeli, Nikita Lagrange, Arianna Mencattini, Maria Carla Parrini, Eugenio Martinelli, and Herve Isambert. CausalXtract, a flexible pipeline to extract causal effects from live-cell time-lapse imaging data. *eLife*, 13:RP95485, January 2025. ISSN 2050-084X. . URL https://doi.org/10.7554/eLife.95485. Publisher: eLife Sciences Publications, Ltd.

Wayne Stallaert, Katarzyna M. Kedziora, Colin D. Taylor, Tarek M. Zikry, Jolene S. Ranek, Holly K. Sobon, Sovanny R. Taylor, Catherine L. Young, Jeanette G. Cook, and Jeremy E. Purvis. The structure of the human cell cycle. *Cell Systems*, 13(3):230–240.e3, March 2022. ISSN 2405-4712. . URL https://www.sciencedirect.com/science/article/pii/S240547122100418X.

Willi L. Stepp, Emine Berna Durmus, Santiago N. Rodriguez Alvarez, Juan C. Landoni, Giorgio Tortarolo, Kyle M. Douglass, Martin Weigert, and Suliana Manley. Smart hybrid microscopy for cell-friendly detection of rare events, April 2025. URL https://www.biorxiv.org/content/10.1101/2025.04.04.647219v1. Pages: 2025.04.04.647219 Section: New Results.

Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. RoFormer: Enhanced Transformer with Rotary Position Embedding, April 2021. URL http://arxiv.org/abs/2104.09864. arXiv:2104.09864 [cs] version: 1.

Jean-Yves Tinevez, Nick Perry, Johannes Schindelin, Genevieve M. Hoopes, Gregory D. Reynolds, Emmanuel Laplantine, Sebastian Y. Bednarek, Spencer L. Shorte, and Kevin W. Eliceiri. TrackMate: An open and extensible platform for single-particle tracking. *Methods*, 115:80–90, February 2017. ISSN 1046-2023. . URL https://www.sciencedirect.com/science/article/pii/S1046202316303346.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, June 2017. URL http://arxiv.org/abs/1706.03762. arXiv:1706.03762 [cs] version: 1.

Martin Weigert and Uwe Schmidt. Nuclei Instance Segmentation and Classification in Histopathology Images with Stardist. In *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pages 1–4, March 2022. . URL https://ieeexplore.ieee.org/abstract/document/9854534.

Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. Transformers in Time Series: A Survey, May 2023. URL http://arxiv.org/abs/2202.07125. arXiv:2202.07125 [cs].

Chaodong Xiao, Minghan Li, Zhengqiang Zhang, Deyu Meng, and Lei Zhang. Spatial-Mamba: Effective Visual State Space Models via Structure-aware State Fusion, February 2025. URL http://arxiv.org/abs/2410.15091. arXiv:2410.15091 [cs].

Qingyu Zhao, Ehsan Adeli, Nicolas Honnorat, Tuo Leng, and Kilian M. Pohl. Variational AutoEncoder

For Regression: Application to Brain Aging Analysis, July 2019. URL http://arxiv.org/abs/1904.05948. arXiv:1904.05948 [cs].

Appendix A. Supplementary Methods

A.1 Cell culture

FUCCI-RPE1 cells, kindly provided by Battich et al. [2020], were cultured at 37°C with 5% CO2 in DMEM/F12 medium (Gibco 11320033), supplemented with 1 % non-essential amino acids (NEAA) (Gibco 11140-035), 1% penicillin-streptomycin (Sigma-Aldrich G6784), and 10% fetal bovine serum (FBS) (Gibco 10437-028). In addition, the H2B-iRFP marker, driven by a PGK promoter, was introduced into the cells using the second-generation lentiviral system with a commercially available plasmid (Addgene: 90237).

A.2 Imaging

For imaging, H2B-FUCCI-RPE1 cells were seeded into 96-well plates and cultured under the conditions described above, with Fluorobrite medium (Gibco A1896701) replacing DMEM/F12. For the perturbation experiments, cells were treated with 10 nM Palbociclib (CDK4-6 inhibitor). Images from four channels—Brightfield, H2B (far red), Cdt1 (red), and Geminin (green)—were acquired every 5 minutes using a PerkinElmer Operetta Microscope with a 20x/0.80 objective (wide-field microscopy). Four or nine tiles per well were captured for each channel, with a 15% overlap for subsequent stitching. In the images, 1 pixel equates to $0.5979761\mu m$. The laser intensities and time of exposure for each channel are shown in the table below.

Channel	Laser intensity	Exposure time
FUCCI Green	25%	30ms
FUCCI Red	15%	10ms
H2B Far red	30%	30ms
Brightfield	50%	5ms

A.3 Image preprocessing

Image preprocessing involved stitching the tiles Preibisch et al. (2009) and applying background subtraction to fluorescent channels using a rolling ball algorithm. Cell nuclei were segmented on the H2B channel with a custom StarDist model Weigert and Schmidt (2022) , and tracked across frames using TrackMate Tinevez et al. (2017). Full cell cycle tracks (M-M, tracks encompassing one complete cell division cycle from one mitosis (M) to the next) were isolated using K-Means clustering of interpolated FUCCI signals. Our groundtruth labels were obtained by averaging the fluorescent channels over the nuclei area and taking the logarithm of this signal, with a smooth noise removal. The raw fluorescent FUCCI signal is not normalized for background noise (starting at 2^5) and expresses a greater dynamic range in log scale as previously shown in DeepCycle Rappez et al. (2020). However taking the logarithm increases the dynamic range of the background noise, leading to interesting questions about the proper scale of these tracks. We average the pixels present in the nucleus for each FUCCI marker and then for the background noise normalization, we express the signal shifted and in units of an ϵ, where ϵ can be $k-$th percentile of the signals distribution. We use the 1st percentile $\epsilon = P_1 = (P_{1,f_1}, P_{1,f_2})$. Then to deal with the increased dynamic range from the log, we take the Softplus with $\beta = 1$ of our new units of ϵ before applying the logarithm:

$$F = (f_1, f_2), \quad \overline{F} = \frac{F}{A} = (\overline{f_1}, \overline{f_2})$$

$$F' = \frac{\overline{F} - \epsilon}{\epsilon}, \quad \text{FUCCI} = log_2(Softplus(F'))$$

Where F is the raw FUCCI values from imaging of each nucleus pixel, A is the area of the nucleus and $\text{Softplus}(x) = \frac{1}{\beta} * log(1 + exp(\beta * x))$.

A.4 Data splits and leakage control

All data used in this study come from a single imaging experiment. The unit of split is the whole time-lapse movie (one 96-well plate well = one movie): all tracked nuclei from a given movie are assigned exclusively to one of the three sets, and no movie contributes tracks to more than one split. This prevents leakage from cells sharing the same field-of-view, illumination, or local imaging artefacts. The resulting composition is:

- **Training:** 8 movies, control condition, 5,188 full M–M tracks.
- **Regular:** 1 movie, control condition, 358 full tracks.
- **Drug:** 2 movies, treated with 10 nM Palbociclib (CDK4/6 inhibitor) at the same dose, 73 full tracks.

The split was performed *before* any normalization or preprocessing step. All signal and image normalizations are computed locally, per movie or per track, and never share statistics across splits: per-movie FUCCI label normalization, per-image z-score on brightfield and H2B channels, per-track min–max scaling of FUCCI labels, and an additional per-track z-score on image intensities applied at load time. No statistic derived from the training set is ever applied to the test sets.

A.5 Identification of biological checkpoints

To identify the time frame at which phase transitions occur in a sequence of intensities (both in real data and in the model's predictions), we devised a simple threshold-based method that accurately detects the onset of the FUCCI green signal (G1/S) and the disappearance of the red signal (S/G2).

- The "linear" signal (not log-transformed) of both channels is normalized between 0 and 1 to ensure comparability across tracks.
- The signal is smoothed using a convolution with a window size of 20.
- For the green signal, we identify the first time point where the intensity crosses above the 0.05 threshold (5% of its maximum intensity).
- For the red signal, we determine the transition point as the first time it drops below the 0.05 threshold (5% of its maximum intensity), marking its disappearance.

Appendix B. Supplementary Tables and Figures

Model	Parameters
MLP	1.32×10^6
Causal CNN	0.94×10^6
LSTM	1.28×10^6
Mamba	1.11×10^6
Transformer	1.12×10^6

Table B 1: Number of parameters for each sequence model head.

Model	Total time	Per epoch
Single frame (MLP)	112 min	44.9 s
Causal CNN	114 min	45.5 s
LSTM	115 min	46.2 s
Mamba	121 min	48.2 s
Causal transformer	127 min	50.5 s
Transformer (bidir.)	125 min	50.1 s

Table B 2: **Training cost across architectures.** Total and per-epoch wall-clock time on a single NVIDIA H100 GPU (brightfield, 150 epochs). All architectures share the same ResNet-18 backbone, which dominates the per-step cost; total training time is therefore comparable across models.

Model	ms/track (N=128)	ms/track (N=512)	μs/frame (N=512)
Single frame (MLP)	1.9	4.3	8.4
Causal CNN	2.5	4.4	8.7
LSTM	4.5	15.5	30.3
Mamba	5.2	5.8	11.2
Causal transformer	8.4	8.5	16.5
Transformer (bidir.)	8.0	8.1	15.9

Table B 3: **Inference latency per full track and per frame.** Single NVIDIA H100 GPU; N is the track length in frames. Timings include GPU warm-up and CUDA synchronisation. Among the causal models Mamba scales best to longer tracks, while the LSTM scales linearly and becomes the slowest at N=512.

	Brightfield				Histone H2B			
Models	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	R^2	DTW	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	R^2	Δ_{DTW}
Single Frame	0.193 ± 0.066	0.146 ± 0.045	0.459 ± 0.271	3.735 ± 0.863	0.183 ± 0.104	0.130 ± 0.064	0.491 ± 0.431	2.595 ± 1.201
Causal CNN	0.157 ± 0.078	0.122 ± 0.049	0.608 ± 0.294	2.468 ± 0.917	0.154 ± 0.105	0.118 ± 0.061	0.586 ± 0.415	2.165 ± 1.210
LSTM	0.108 ± 0.069	0.087 ± 0.047	0.749 ± 0.266	1.527 ± 0.814	0.079 ± 0.065	0.075 ± 0.044	0.833 ± 0.265	1.467 ± 1.161
Causal Transformer	0.121 ± 0.073	0.094 ± 0.048	0.720 ± 0.279	1.728 ± 0.811	0.079 ± 0.057	0.079 ± 0.042	0.839 ± 0.214	1.552 ± 0.955
Mamba	0.112 ± 0.072	0.091 ± 0.049	0.739 ± 0.282	1.444 ± 0.898	0.074 ± 0.056	0.075 ± 0.040	0.853 ± 0.215	1.426 ± 0.949
Transformer	**0.066 ± 0.038**	**0.062 ± 0.037**	**0.892 ± 0.111**	**1.285 ± 0.553**	**0.056 ± 0.039**	**0.054 ± 0.033**	**0.912 ± 0.116**	**1.155 ± 0.612**

Table B 4: **Side by side performance comparison of BF and H2B modalities at predicting Fucci channels on Regular.** Both data modalities present similar results: sequence encoders outperform the single frame method. Moreover H2B only shows modestly better performance than BF. The ± values report the standard deviation of per-track errors across all test tracks from a single representative training run, reflecting per-track heterogeneity rather than variability across seeds.

Palbociclib	Brightfield				Histone H2B			
Models	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	R^2	Δ_{DTW}	$L_{1,FUCCI_1}$	$L_{1,FUCCI_2}$	R^2	Δ_{DTW}
Single Frame	0.239 ± 0.082	0.182 ± 0.056	-0.297 ± 1.064	5.329 ± 1.147	0.183 ± 0.050	0.107 ± 0.048	0.260 ± 0.466	3.285 ± 0.820
Causal CNN	0.252 ± 0.113	0.161 ± 0.059	-0.353 ± 1.459	4.323 ± 1.302	0.149 ± 0.042	0.125 ± 0.040	0.401 ± 0.376	3.077 ± 0.840
LSTM	0.424 ± 0.101	0.229 ± 0.045	-1.663 ± 1.727	3.685 ± 1.678	0.140 ± 0.051	0.115 ± 0.040	0.503 ± 0.404	**2.750 ± 0.838**
Causal Transformer	0.326 ± 0.104	0.214 ± 0.049	-0.728 ± 1.558	5.159 ± 1.562	0.132 ± 0.034	0.111 ± 0.037	0.628 ± 0.207	3.154 ± 0.872
Mamba	0.485 ± 0.090	0.259 ± 0.045	-2.244 ± 1.949	3.563 ± 1.918	0.185 ± 0.078	0.134 ± 0.043	0.255 ± 0.651	2.896 ± 0.903
Transformer	**0.147 ± 0.056**	**0.139 ± 0.048**	**0.408 ± 0.478**	**3.022 ± 0.985**	**0.074 ± 0.029**	**0.095 ± 0.031**	**0.789 ± 0.131**	2.896 ± 1.201

Table B 5: **Performance metrics for both brightfield and histone H2B modalities on Drug.** The ± values report the standard deviation of per-track errors across all test tracks from a single representative training run, reflecting per-track heterogeneity rather than variability across seeds.

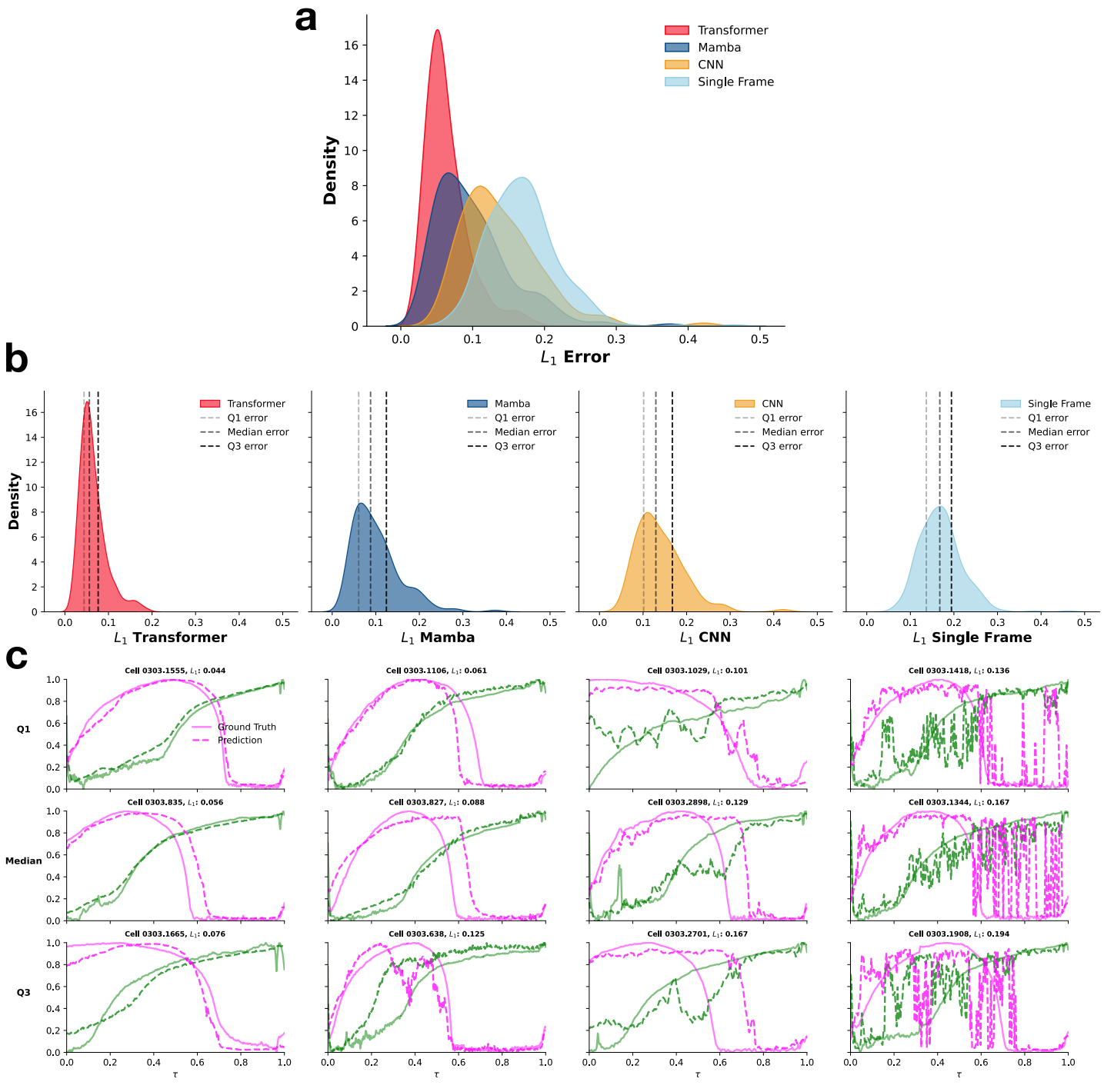

Figure 9: **Error Distribution of Predictions of Fucci on Test Set with Brightfield. a.** Distribution of L1 errors across the different models **b.** Error Distributions with Q1, Median and Q3 Percentiles overlayed **c.** Q1, Median and Q3 Error Predictions visualized per model.

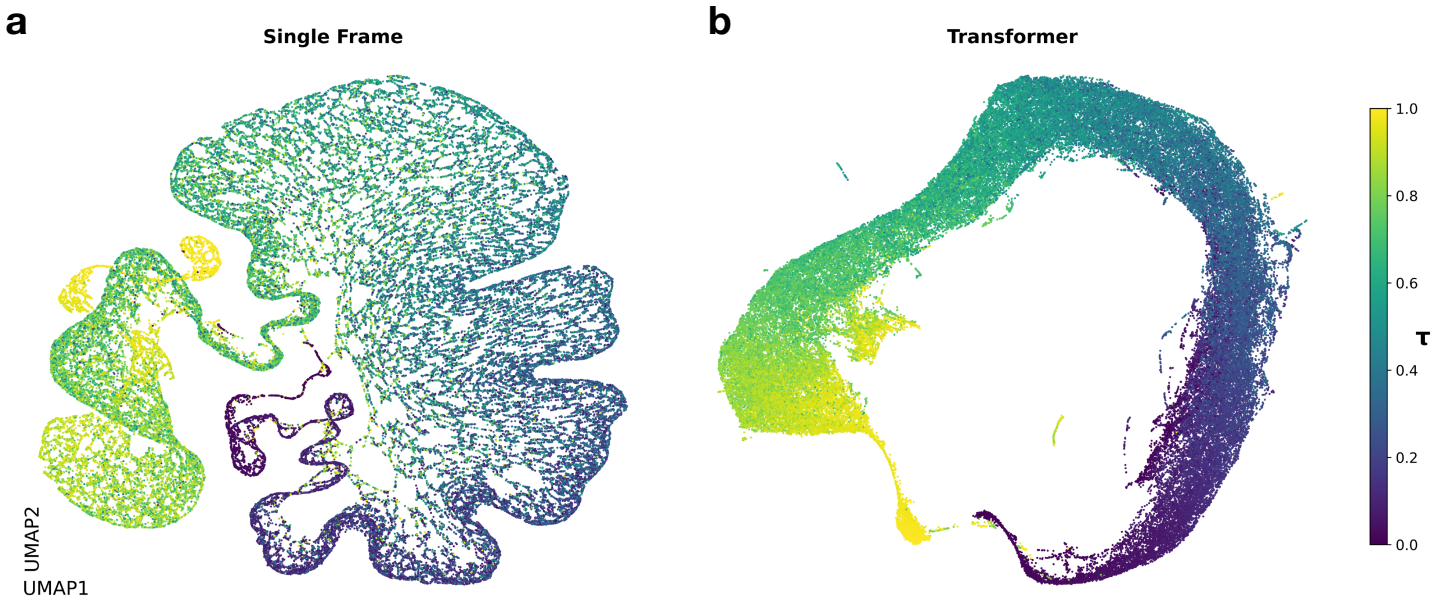

Figure 10: **Learned Latent Space Representations (UMAP).** Each frame of a track is represented as a dot in umap space, the coloring is the normalized time **a.** Single Frame (no history). **b.** Transformer (full sequence).

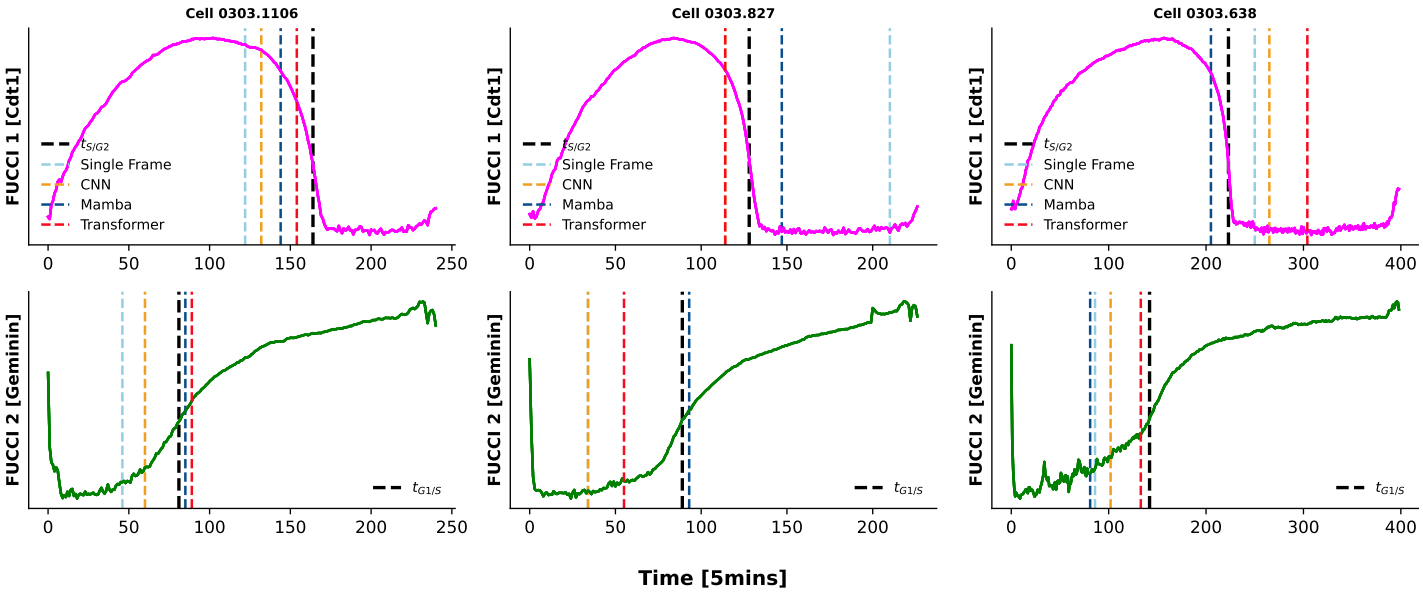

Figure 11: **Predicted $\Delta t_{G1/S}$ and $\Delta t_{S/G2}$ from BF images for the different models.**

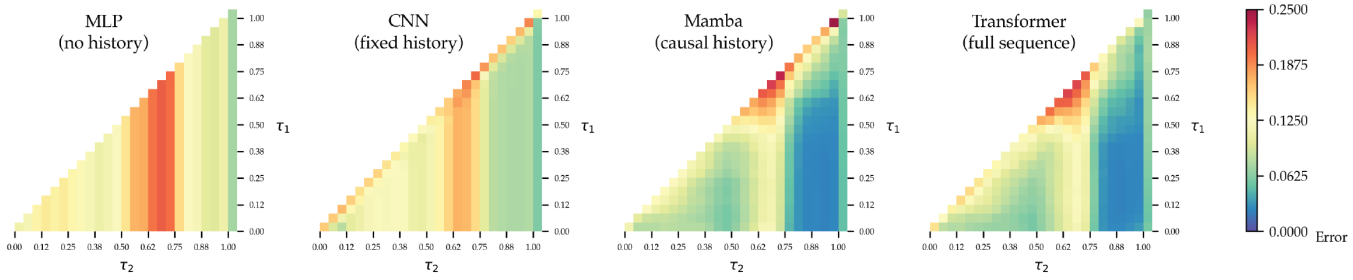

Figure 12: **Comparative Performance of Temporal Encoders in Predicting Continuous Cell Cycle States from H2B in partial cell cycle tracks.** Error maps showing the prediction error of the different models, assessed on the last frame of segments from the M-M track, spanning indices τ_1 to τ_2.

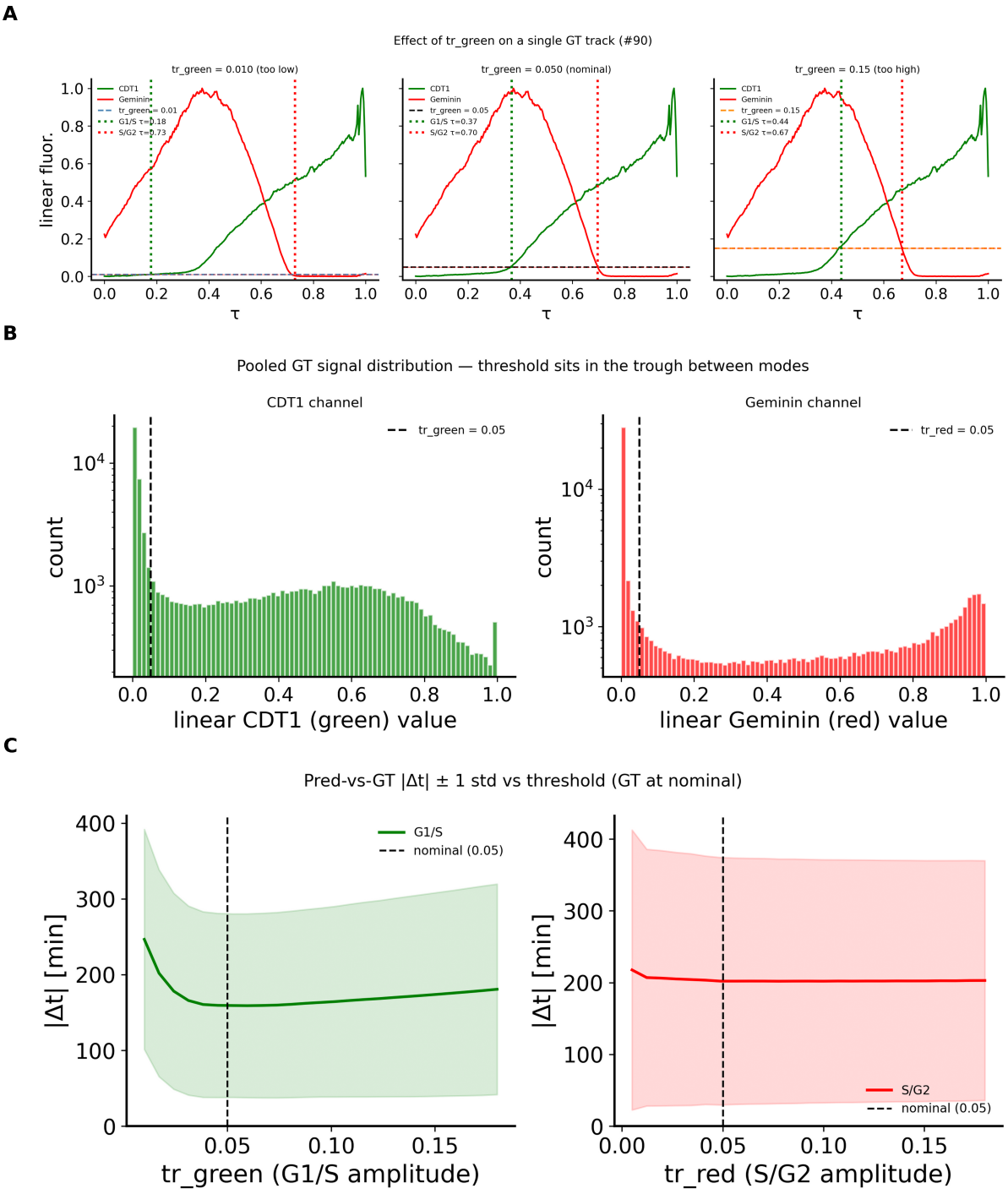

Figure 13: **Justification and sensitivity analysis for checkpoint detection thresholds. (A)** Effect of the CDT1 (green) amplitude threshold θ_{green} on a representative ground-truth track shown in linear fluorescence units. From left to right: a threshold set too low (0.010) misidentifies noise fluctuations as a G1/S transition; the nominal value (0.050) correctly captures the onset of CDT1 rise; a threshold set too high (0.150) delays or misses the transition. The Geminin (red) threshold is held at the same value in each panel. **(B)** Pooled distribution of linear-space fluorescence values across all ground-truth tracks and time points, for the CDT1 (left) and Geminin (right) channels (log-scale y-axis). Both channels exhibit a bimodal structure with a sharp noise-floor mode near zero and a broader signal-on mode at higher values. The nominal threshold of 0.05 (dashed line) falls within the trough between the two modes, confirming that it discriminates between signal-off and signal-on states without encroaching on either distribution. **(C)** Prediction-versus-ground-truth timing error $|\Delta t|$ (mean $\pm$ 1 s.d., in minutes) as a function of the amplitude threshold, sweeping θ_{green} for the G1/S checkpoint (left) and θ_{red} for the S/G2 checkpoint (right). Ground-truth crossings are held at the nominal threshold; only the predicted crossings are recomputed at each swept value. The Geminin channel yields a nearly flat curve over the full range tested, reflecting the sharpness of the S/G2 transition. The CDT1 channel shows moderate sensitivity, but $|\Delta t|$ remains stable in a wide neighbourhood around the nominal value (dashed line), supporting the robustness of the chosen parameters.