# LiveCellMiner: A New Tool to Analyze Mitotic Progression

Daniel Moreno-Andrés[1,*,†], Anuk Bhattacharyya[2], Anja Scheufen[1], Johannes Stegmaier[2,*,†]

**1** Institute of Biochemistry and Molecular Cell Biology, Medical School, RWTH Aachen University, Aachen, Germany
**2** Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany

* Correspondence: dmoreno@ukaachen.de, johannes.stegmaier@lfb.rwth-aachen.de †: These authors contributed equally.

## Abstract

Live-cell imaging has become state of the art to accurately identify the nature of mitotic and cell cycle defects. Low- and high-throughput microscopy setups have yield huge data amounts of cells recorded in different experimental and pathological conditions. Tailored semi-automated and automated image analysis approaches allow the analysis of high-content screening data sets, saving time and avoiding bias. However, they were mostly designed for very specific experimental setups, which restricts their flexibility and usability. The general need for dedicated experiment-specific user-annotated training sets and experiment-specific user-defined segmentation parameters remains a major bottleneck for fully automating the analysis process. In this work we present LiveCellMiner, a highly flexible open-source software tool to automatically extract, analyze and visualize both aggregated and time-resolved image features with potential biological relevance. The software tool allows analysis across high-content data sets obtained in different platforms, in a quantitative and unbiased manner. As proof of principle application, we analyze here the dynamic chromatin and tubulin cytoskeleton features in human cells passing through mitosis highlighting the versatile and flexible potential of this tool set.

## Introduction

State of the art cell cycle and mitosis research stongly relies on advanced live-cell microscopy for recording cells in model organisms and tissue cultures. With the help of different labelling techniques (target protein fusion with fluorescent proteins, direct fluorescent labelling of cellular targets, organelle specific fluoresecent probes) subcellular structures and cell cycle markers can be followed by high-content screening (HCS) approaches. This allows investigating the different steps of life and fate of single cells and cell populations. Such experiments generate massive amounts of data, which help to pinpoint the nature of mitotic and cell cycle defects and to accurately identify and characterize key molecular factors in different experimental conditions and clinically relevant situations (for review, see [1–4]).

Analyzing high-content data sets is a formidable task where supervised machine learning methods have been so far crucial [5–12]. More recently, convolutional neural networks (CNN) further improved the possibilities for automatic detection, segmentation and classification tasks [13–16]. Supervised machine learning approaches,

however, are still time-consuming because they rely on user-curated phenotype definitions and analysis. To overcome the limitations, unsupervised machine learning without the need of manually-annotated morphology and phenotype-based classifications recently demonstrated first promising results [17–20].

Supervised and unsupervised phenotypic profiling of live-cell microscopy data is emerging as a powerful tool for clinicians, pharmaceutical industry and biology research in general. Multiparametric data analysis at single-cell level allows the integration of up to hundreds of quantitative features to describe and distinguish phenotypes going far beyond traditional approaches based on the analysis of one or a few features in single-cell images. However, in most of the cases where numerous features are measured and integrated to generate high-dimensional phenotypic profiles, it is unclear whether all of them are required for describing the phenotypic classes. The generated phenotypic classes work as closed entities, where no gradual time-dependent changes can be accurately quantified. Furthermore, the quantification of simple morphological and shape parameters, *e.g.*, of the cell nucleus, has potential diagnostic value [21, 22]. In this regard, most available analysis tools are able to measure health-related morphology profiles of tuneable complexity in static images of fixed cell samples ( [23]; for review, see [3]). Therefore, tools for unbiased and comprehensive analysis of image features directly reflecting time-dependent live-cell shape and morphology are urgently needed both in basic and transnational research [24].

Another common limitation in the available HCS tools is the lack of user-independent segmentation. The selection of an important number of object detection parameters relies on user choices, which may ultimately introduce bias in the phenotype classes determined in supervised approaches. As first step, automated cell identification has recently been achieved using deep learning, where CNNs are particularly applicable for the instance segmentation task [16, 25, 26]. These tools perform well when being trained with sufficiently different modalities [16] but tend to behave unpredictably when being applied to data that significantly deviates from images seen during the training phase.

Here, we introduce LiveCellMiner, a new open-source fully-automated HCS software tool for the quantitative analysis and comparison of 2D+t microscopy images of fluorescently-labeled cells. The software allows automatic segmentation and tracking and extracts quantitative features for all tracked objects. It enables automatic temporal synchronization of extracted tracks and offers comprehensive data visualization and selection possibilities. LiveCellMiner was primarily developed to analyze mitotic phenotypes in cells but can be easily extended to other scenarios. Since mitosis is characterized by a succession of distinctive chromatin morphologies, prophase, prometaphase, metaphase, early anaphase, late anaphase and telophase, it offers an excellent multi-level benchmark for the study of cytologic and temporal phenotypes. In early mitosis, the nuclear envelope breaks down and the chromatin condenses generating individualized and rod-shaped chromosomes, which are captured by the mitotic spindle and segregated to sister chromatin masses during anaphase (for review, see [27, 28]). During telophase and early G1, the chromatin masses decondense allowing the reassembly of functional nuclei able of gene transcription and genome replication (for review, see [29–31]). Here, we reanalyze published data sets of cells passing mitosis generated with different live-cell imaging platforms showing high reproducibility between the systems. We confirm previously described mitotic alterations after downregulation of the Lysine Specific Demethylase (LSD1), RecQ-like Helicase 4 (RecQL4) and the Protein Phosphate PP2A-complex but identify also phenotypes that so far escaped our attention. We use the reanalyzed data set also to demonstrate the visualization tools of LiveCellMiner for presenting a high-content screen data sets.

# Design and Implementation

We developed LiveCellMiner, a new open-source software tool for studying the cytology of cell division under different experimental conditions. LiveCellMiner is an extension package for the general-purpose data mining MATLAB toolbox SciXMiner [32]. In the following sections, we present the individual modules of LiveCellMiner and show how they are used for data import, feature extraction, cell trajectory synchronization, data selection and visualization. The presented proof of principle applications are based on existing data sets from previous publications [33–35] where image data were acquired using 2D+t widefield and confocal microscopy in different platforms (see Note S1 for details on the experimental setup).

## Data Import

LiveCellMiner expects time series of 2D images of cells with chromatin-labeled nuclei and optionally additional markers in other channels. The first step comprises the detection, segmentation and tracking of all cell nuclei. We adapted the Laplacian-of-Gaussian-based object detection method implemented in XPIWIT [36, 37] to perform automatic detection of nuclei centroids using a set of predefined processing pipelines adjusted for various image resolutions. Detected centroids are tracked using the methods described in [38]. Briefly, tracking is performed in a time-reversed manner starting with the last frame and by sequentially linking objects to their predecessors using hierarchical clustering with Ward's linkage criterion [39]. The cluster cut-off can either be explicitly specified based on prior knowledge or it can be determined heuristically as half the average distance of each object to its eight spatially nearest neighbors (default setting). If two objects with a different tracking id end up in the same cluster, a cell division event is annotated. After tracking is performed, mitotic trajectories are extracted that fulfil user-defined constraints like a minimum number of successfully tracked frames before and after the cell division occurred (by default we set this parameters to 30 and 60 frames before and after the cell division, respectively, which corresponds to 90 minutes and 180 minutes with the 3 minutes sampling intervals used in all our experiments). All detected centroids are then used to initialize the automatic segmentation of the cells. We provide both a classical and a deep learning-based solution to the segmentation. The classical segmentation method crops a square region surrounding the current detection. First, the image is median filtered for noise reduction ($5 \times 5$ window size) and then binarized using the arithmetic mean of a threshold identified by Otsu's method [40] and the minimum intensity observed in the center part of the patch. The modified version of Otsu's threshold is used to avoid degenerate segmentations where a dim cell residing in the center of the patch could potentially be removed if it is surrounded by more bright objects. As the cell of interest is located in the center of the image patch, we initialize a seeded watershed with two seeds, one for the center cell and another one for the background and neighboring cells. The seeded watershed is applied on an inverted Euclidean distance map with intensity minima located in the centers of the nuclei. As an alternative to the classical segmentation pipeline, we integrated an interface to the recently published Cellpose algorithm [16]. Cellpose is automatically started and parameterized to extract all segments in each of the images. We then use the centroids of successfully tracked objects and crop the results of the Cellpose segmentation using the same region size as for the classical approach. In each patch, we only keep the single central cell to constrain the feature extraction to this region. Although Cellpose generally provides more accurate segmentation results, we found that directly using Cellpose to replace the LoG-based object detection yielded less reliable tracking results, as Cellpose frequently failed to properly segment rarely occurring cell shapes observed in meta- and anaphase.

Thus, we additionally included a fallback option in cases where segmentations provided by Cellpose were missing and in these cases switch back to the classical segmentation method for individual cells. A quantitative assessment of the segmentation and tracking quality is provided in Table S1.

## Extraction of Quantitative Features and Project Fusion

The segmented image patches are subsequently used for feature extraction. In addition to classical 2D features like area, centroid, major and minor axes, orientation, circularity and intensity statistics, we extracted a set of Haralick texture features from the gray level co-occurrence matrix [41] (see Table S2 and Table S3 for an overview of all available features). We empirically set the number of gray levels to 64, removed the background to foreground transitions from the co-occurrence matrices and computed average values obtained for the neighbor relations $[0, 1]$, $[1, 0]$, $[1, 1]$, $[-1, 1]$. Among others, the Haralick features comprise measures of texture entropy, correlation, contrast and variance (we refer to [41] for a complete definition and explanation of the individual features). Finally, we apply a GoogLeNet pretrained on ImageNet database on each image patch to obtain CNN-features that are used for automatic synchronization of the cell trajectories [42]. In addition to the raw image snippets of all available channels, the corresponding segmentations, GoogLeNet features and feature time series of all valid trajectories are stored in a SciXMiner-compatible format. Additional meta information like microscope, experiment ID, plate number and experimental conditions are saved as well and can later be used by the flexible and powerful data selection possibilities of SciXMiner. Individual projects obtained for different positions can be fused to a single SciXMiner project, to analyze even large projects in a single and consistent project file. After projects have been imported to the SciXMiner format, additional features can be derived from the time series and single features. In addition to all available feature transformations that are available by default in SciXMiner [32, 43], we incorporated dedicated features for the analysis of cell behavior.

Time series can be smoothed with a variable window size using any of the methods implemented in MATLAB's smooth function to remove small deviations in the extracted feature values. Moreover, absolute feature values can be normalized to predefined events of the cell cycle, such as the average interphase feature value or the feature value of the first late anaphase frame. As different microscopes or acquisition settings produce notably different absolute feature values, these normalization procedures are beneficial to make time series comparable among different experiments and to compute relative recovery time series of cell properties like fluorescence intensity after mitosis. The rate of change for selected features at a particular time point (e.g., to estimate the initial recovery rate of time series features immediately after cell division) can be approximated by a linear regression of the feature values in a small temporal window. The slope of these regression curves is stored as a single feature for each cell. Finally, to obtain a proxy for interphase recovery, we added a dedicated recovery feature that measures the absolute percentage deviation of one or more features to their respective interphase mean value with 100% indicating full recovery (average percentage deviation if multiple features are selected). While arbitrary features can be combined to a custom recovery feature, we used the features area, minor axis length, mean intensity, intensity standard deviation for all figures in this paper.

## Cell Trajectory Synchronization

To compensate for different duration of pro-, prometa- and metaphase, as observed also in our data sets, synchronization can be adapted to every particular cell cycle event under study. The LiveCellMiner toolbox provides different ways of synchronizing the

individual trajectories. We chose two characteristic events as the synchronization anchors, to obtain properly aligned time series for quantitative comparisons. To this end, we identify interphase to prophase transition (IP) or metaphase/early anaphase to late anaphase transition (MA) as reference mark for alignment of interphase or postmitotic frames, respectively. There are currently three options for automatic alignment. The first approach uses the classical object features area, circularity, mean intensity and intensity std. dev. to identify the IP transition by searching for two clusters that minimize the within-class variance in the frames before the chromatin masses separate using the temporally constrained combinatorial clustering (TC3) method [17]. If the division time point that was identified during tracking corresponds to early anaphase, the software can automatically reposition the MA transition. Detecting early anaphase is accomplished by a heuristic that checks if the centroid distance of the chromatin masses of both daughter cells exceeds a user-defined threshold. As the classical method is originally applied to all trajectories, it may happen that the project still contains invalid trajectories. The second method is similar to the first method, but uses an additional auto-rejection of erroneous tracks. This is accomplished with a trainable LSTM network [44] that assesses the validity of each trajectory as a whole. The third method uses another LSTM network that was trained on sequences of CNN features that were obtained from the pretrained GoogLeNet to predict the state sequence for all-time points, as well as identifying which of the cell tracks are valid/invalid. The predicted synchronization time points are post-processed with a Hidden Markov Model (HMM) that only allows valid state transitions (*e.g.*, state sequences $1 \rightarrow 2 \rightarrow 3$ for a valid track or 0 for an invalid track) [45]. Transition probabilities are manually specified and based on the predicted states of the LSTM, and we use the Viterbi algorithm to identify the most likely hidden state sequence [46].

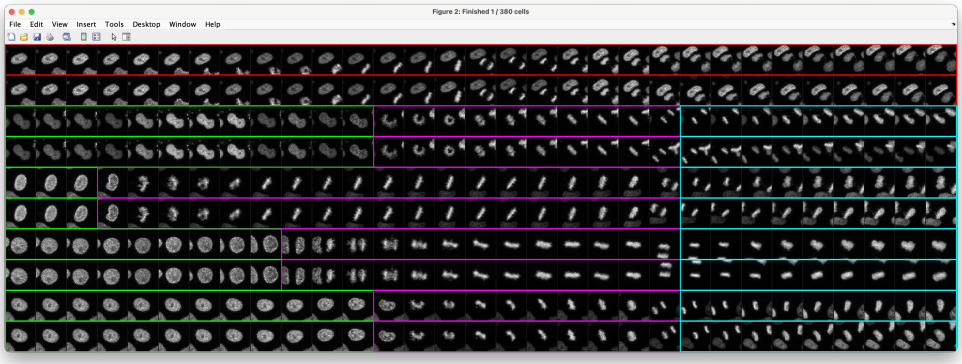


**Fig 1. Graphical User Interface for Trajectory Synchronization.** Using an intuitive annotation scheme, users can verify and potentially correct the cell synchronization. Moreover, erroneous tracks that do not contain a mitotic event can be discarded. An initial synchronization can be automatically obtained using both classical and machine learning-based synchronization methods as detailed in the main text. Colors indicate interphase frames (green), pro-, prometa-, meta- and early anaphase frames (magenta), late ana- and telophase (cyan) and erroneous trajectories (red).

To inspect and optionally correct the automatic synchronization results, we provide a simple graphical user interface to manually identify the state transitions (Fig. 1). This facilitates man-machine feedback, decreasing the size of the classifiers, and training time, if needed. It displays a set of cells, where two cells above one another are daughters and image snippets are preloaded to smoothly interact with the GUI. A manual annotation of two daughter cells can be accomplished with two clicks by

identifying the last frame considered as interphase to mark the IP transition and the early anaphase frame to mark the MA transition. All intermediate frames are classified accordingly, and the annotations of one of the daughter cells are directly copied to the other daughter to have a consistent alignment. The GUI also allows rejecting entire trajectories, *e.g.*, if no mitotic event is present or due to erroneous tracking. The manually synchronized cells can additionally be used for retraining the LSTMs of the automatic synchronization methods, and it is possible to specify separate models for different experimental conditions. We provide thorough validation of the different synchronization possibilities in Fig. S1, Fig. S2, Fig. S3 and in Table S4. Once all cells are properly aligned, both qualitative and quantitative comparisons between different experiments can be performed, as detailed in the next sections.

## Data Selection and Visualization Capabilities

An important aspect for the analysis of a particular subset of cells is data selection. This can be accomplished by using class-based selection procedures that allow to group the data according to imported metadata. For instance, it is possible to select experiments that were acquired with a particular microscope, a subset of treatments, a specific experiment or individual positions. It is also possible to use multiple properties in combination, to specify a feature range for selection, and to use the basic functionality of SciXMiner to add additional groupings derived from the individual or time series features [32]. Subsequent visualization, quantification and manual corrections are then automatically constrained to the selected cells.

In addition to the standard visualizations available in SciXMiner, we provide dedicated visualizations for the LiveCellMiner toolbox. Time series can be visualized as heatmaps, mean time series and combined line plots (Fig. 2A-C). In the heatmap visualization, each line represents a feature time series of a single cell, with feature values indicated by the color code. The identified synchronization time points are used to properly align the cell tracks below each other. Rather than displaying each cell separately, the mean ± std. dev. plots average the results of a particular position, experiment or microscope (Fig. 2B). Averaging is performed on the aligned tracks to ensure that only corresponding mitotic stages are compared. In addition to presenting a single plot per selected group, it is also possible to combine all line plots including error bars in a single plot for better comparison (Fig. 2C). Extracted single features can be visualized as box or violin plots (Fig. 2D) and as histograms (Fig. 2E).

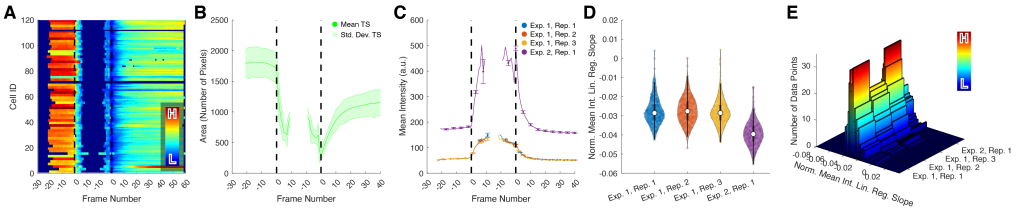


**Fig 2. Visualization Options of LiveCellMiner.** The LiveCellMiner extension provides multiple ways of data visualization. (A) Temporally aligned heatmaps of feature time series (color encodes the feature value, *e.g.*, the area in numbers of pixels as in this example), (B) mean ± std. dev. curves summarizing all trajectories of a particular position or experiment in a separate subplot, (C) mean curves of multiple experiments with error bars plotted in a single axis for better comparability, (D) violin or box plots of single feature values and (E) histograms of individual data points grouped according to the current selection.

All visualizations and subplots can be adjusted according to the selected grouping of

the data. Aside from plotting all individual results in separate subplots, this allows combining related experiments, e.g. visualizing the average response of a particular treatment across experiments, summarizing different repititions of the same experiment or averaging responses across experiments. An example of three possible grouping scenarios is depicted in Fig. 3.

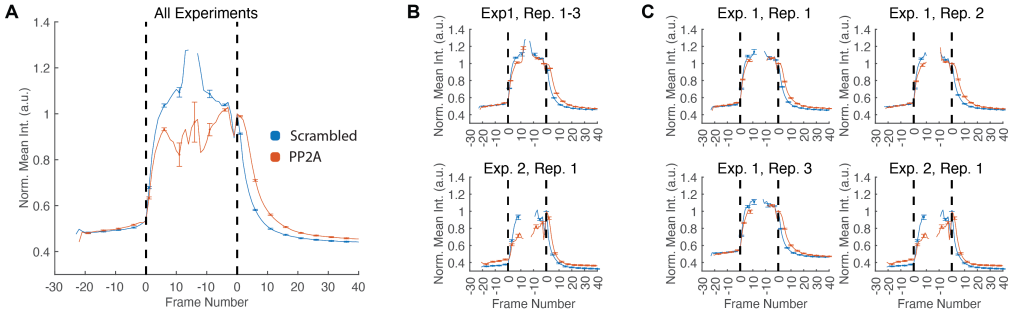


**Fig 3. Grouped Data Visualizations.** The three panels illustrate the different grouped visualization possibilities and were obtained using four experiments showing the normalized mean intensity for two oligos (Scrambled, PP2A). Exp. 1, Rep. 1-3 are three repetitions of the same experiment and Exp. 2, Rep. 1 is one separate experiment that was acquired using a different modality (confocal instead of a widefield microscope). The settings for combining the experiments (from left to right) are: (A) average time series of all experiments and repeats, (B) time series averaged over the repeats with separate plots per experiment and (C) individual plots for all experiments and repeats. Error bars indicate one standard deviation and the vertical bars represent the IP and the MA transitions.

In addition to the targeted visualizations, LiveCellMiner also provides an easy way to obtain a ranked overview of all extracted features including statistical readouts like minimum, maximum, mean, standard deviation and median in table form as well as graphically using heat maps and combined line plots for all features. To identify which features are potentially suitable for characterizing phenotypic differences, we compute the n-fold change of each feature from the interphase average to the average of the first two prophase frames and the first two anaphase frames (window for averaging can be changed by the user). The generated report is exported in HTML format and can be conveniently displayed with any conventional internet browser (see File S1 and File S2 for a demonstration of automatically generated reports). Last but not least, LiveCellMiner can be used for statistical analysis of selected cells based on the single features or based on time series features of each cell. The methods for comparing single features across different groups comprise both parametric (two-sample t-test, ANOVA) and non-parametric tests (Wilcoxon, Kruskal-Wallis) as well as a two-way ANOVA (treatment $\times$ time) to be applied on selected time series. Results are exported as easily accessible spreadsheet files including the test results and the p-values of the performed tests.

# Results

## Cross-Platform Reproducibility of LiveCellMiner Readouts

To test the capabilities of LiveCellMiner on the output of common light microscopy systems, we have reexamined the mitotic progression of cells after RNAi-mediated downregulation with negative and positive controls from previously published [33–35]

and unpublished data sets (LSM710 confocal, see Note S1). Our routine positive control is the RNAi-mediated downregulation of three subunits of the heterotrimeric PP2A complex: PPP2CA (catalytic subunit alpha), PPP2R1A (scaffold subunit alpha) and PPP2R2A (a regulatory subunit B55 alpha). This PP2A complex is involved in the control of mitotic spindle assembly [47], and the spindle assembly checkpoint [48] promotes mitotic exit, disassembly of the spindle-pole associated microtubules in anaphase, resumption of nucleo-cytoplasmic transport, reclustering Golgi apparatus and chromatin decondensation ( [49]; see [50] and [51] for review).

When this specific PP2A complex is downregulated, various defects of mitotic progression are observed: prolongation of early mitosis (prophase, prometaphase and metaphase), partial late anaphase arrest and delayed reestablishment of nucleo-cytoplasmic transport [49]. In addition, PP2A downregulation leads to faulty chromatin decondensation and results in a partial telophase arrest [34]. The analysis of phenotypes using LiveCellMiner demonstrates that the biological effects of a given treatment can be quantitatively extracted by measuring a basic set of image features (Fig. 4), without the need for extensive training of phenotypic classifiers. This is done in a reproducible manner across different live-cell microscopy platforms. In this case, area and intensity measures were used as proxy for chromatin decondensation (Fig. 4A-C), which is, as reported, delayed upon PP2A knockdown. In a similar way, deviations from control values in nuclear geometrical descriptors, *i.e.*, major and minor axes (Fig. 4D), might indicate deformations and irregularities due to altered cytoskeleton or chromatin regulation [52]. LiveCellMiner automatically detects interphase to prophase and metaphase to anaphase transitions as well as the degree of rotation of the chromatin mass (Fig. 4F, G). These readouts allow detecting important errors in early mitotic progression. Delayed anaphase onset might arise from persistent chromosome misalignment and/or deficient spindle function as well as spindle misorientation, which often lead to chromosomal instability (CIN), a common feature in many pathologies including cancer ( [53, 54], for review see [55, 56]) .

## Quantitative Characterization of Multiple Mitotic Phenotypes

To test LiveCellMiner for the quantitative study of a broad spectrum of mitotic phenotypes, we reanalyzed changes of chromatin and tubulin cytoskeleton appearance after LSD1 and RecQL4 RNAi-mediated downregulation in human cells using this tool.

We have previously described the Lysine Specific Demethylase, LSD1 (also known as KDM1A), as a crucial factor for reassembly of a functional nucleus at the end of mitosis [34]. Our recent work has shown that RecQ-like helicase 4 (RecQL4), whose mutations are causative of the Rothmund–Thomson syndrome, is important for stable chromosome alignment during mitosis [33]. These live-cell imaging experiments (Note S1) were carried out in HeLa cells stably expressing H2B-mCherry, as chromatin marker, and eGFP-Tubulin for the spindle apparatus, which is the molecular machinery in charge of organizing and exerting the necessary forces to segregate chromatin.

Without the need of training experiment-specific phenotypic classifiers for the chromatin morphology or spindle apparatus, LiveCellMiner corroborates reduced chromatin decondensation rates, as area, mean intensity and interphase recovery in LSD1 downregulated cells (Fig. 5A-C, E). In turn, these image features that describe chromatin compaction state are unaffected by the RecQL4 downregulation (Fig. S6A-C). The nuclear morphology can be also analyzed. In LSD1 and PP2A downregulated cells, after cell division, the nuclei become rounder (Fig. 5D), as indicated by minor vs. major axis ratios. By contrast, after RecQL4 downregulation, elongated nuclear shape is evident (Fig. 5J). These anomalies might indicate an unbalance in the plethora of dynamic processes, factors and structures that reform the nuclear compartment during late mitosis (see [57] and [31] for review). The molecular reasons and consequences of

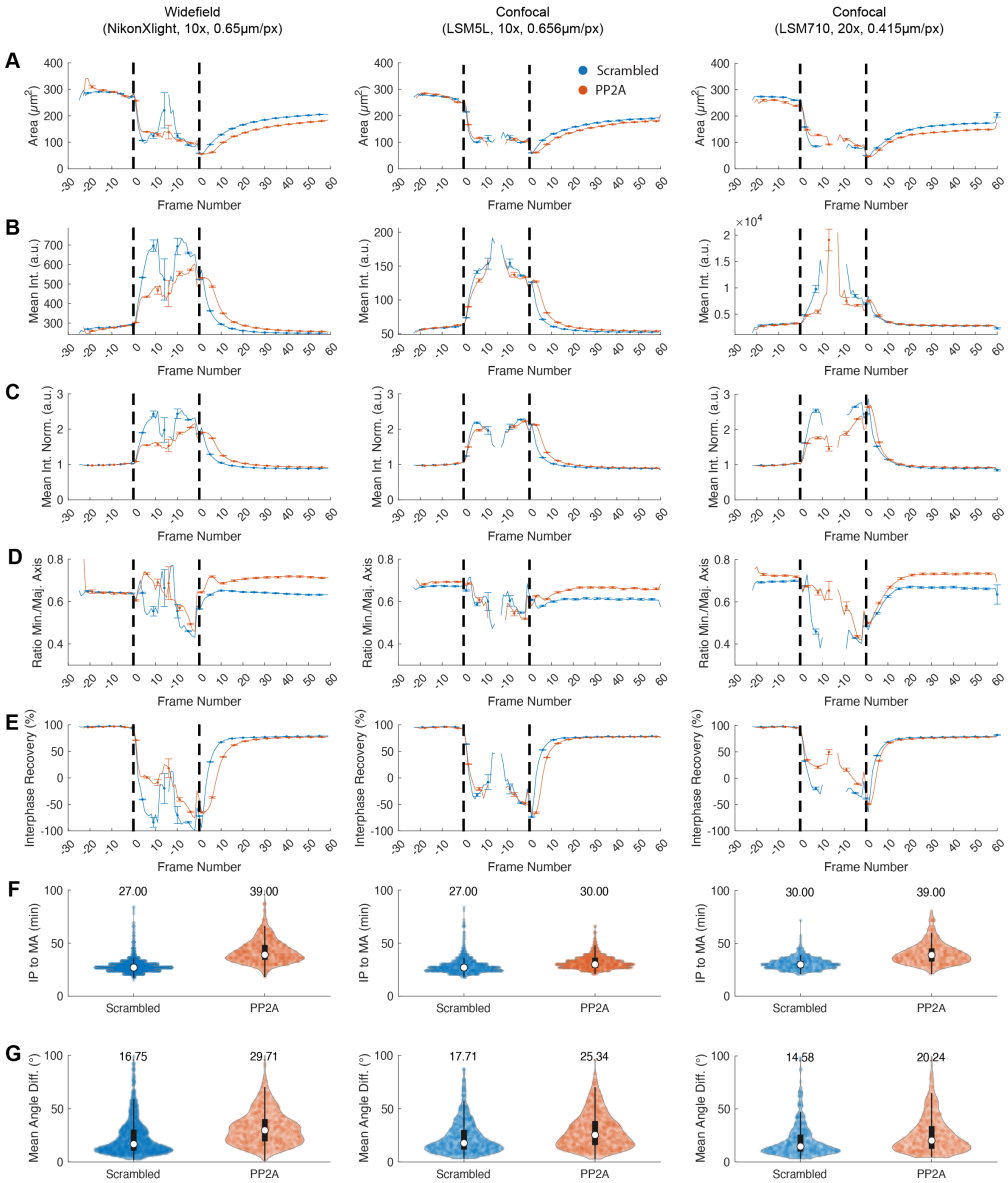


**Fig 4. Platform Comparison.** Reproducibility study with different microscope systems. The columns show exemplary quantifications of the same experiment conducted on different microscopy platforms. We compare Scrambled (control, blue) vs. PP2A knockdown cells (orange). The time series features involve the chromatin area ($\mu m^2$) (A), the chromatin mean intensity (a.u.) (B), the normalized mean intensity (absolute intensity values divided by the interphase mean intensity of each cell, a.u.) (C), the minor axis vs. major axis ratio (D) and interphase-recovery feature as detailed in Table S2 (E). The violin plots show the duration between interphase-prophase and meta-anaphase transition in minutes (F) and the sum of the absolute angular changes in degrees (G). Widefield 10×: $N_{\text{scrambled}} = 1262$, $N_{\text{PP2A}} = 1198$; Confocal 10×: $N_{\text{scrambled}} = 2830$, $N_{\text{PP2A}} = 1008$; Confocal 20×: $N_{\text{scrambled}} = 792$, $N_{\text{PP2A}} = 668$.

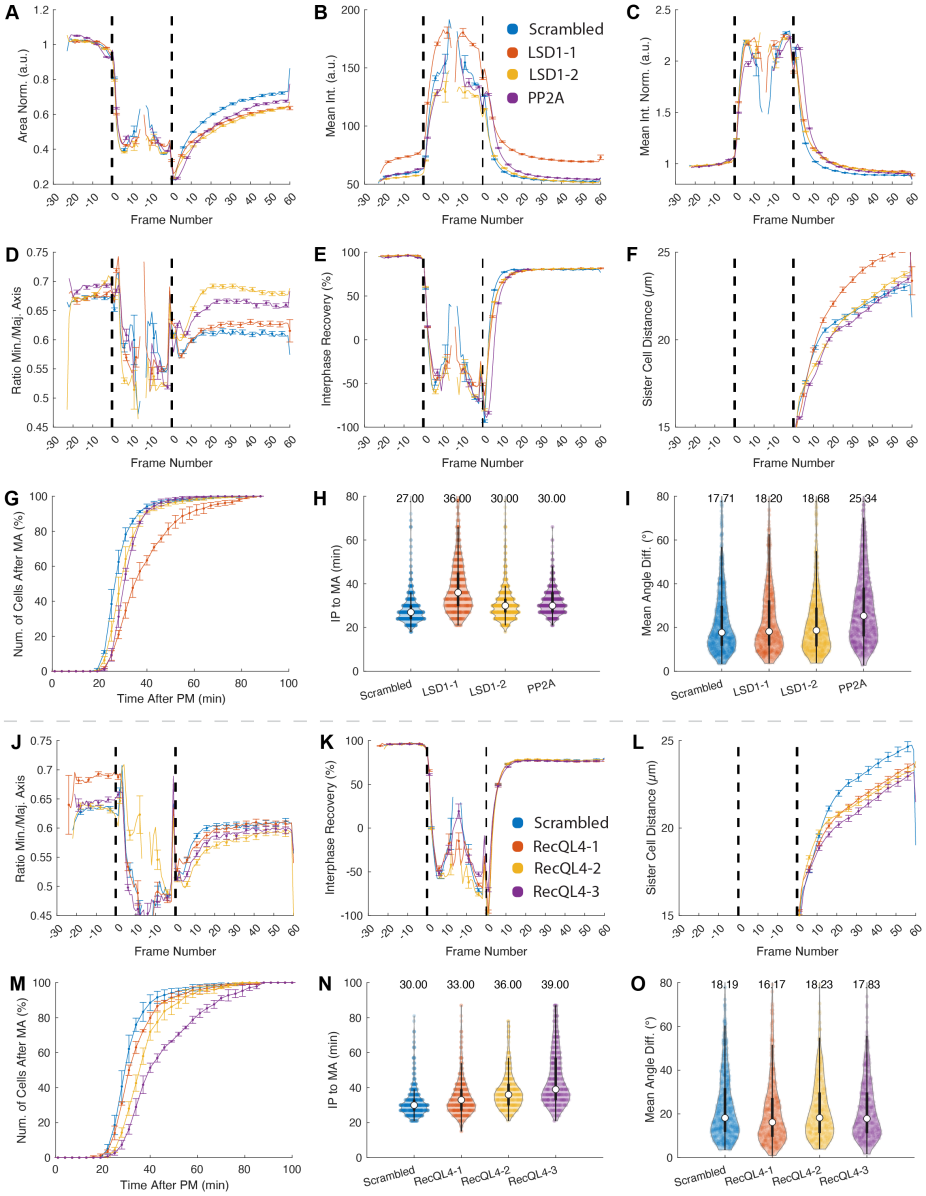


**Fig 5. Analysis of LSD1 and RecQL4 Knockdowns.** Panels (A)-(I) show control (Scrambled) vs. LSD1-1, LSD1-2 and PP2A, whereas panels (J)-(O) show control (Scrambled) vs. RecQL4-1, RecQL4-2 and RecQL4-3 knockdown cells. The features involve the normalized area (A), the mean intensity (B), the normalized mean intensity (absolute intensity values divided by the interphase mean intensity of each cell, C), the minor axis vs. major axis ratio (D, J), the interphase recovery ratio (E, K), distance between sister chromatin masses (F, L) and cumulative histograms for the time in early mitotic progression until anaphase onset (G, M). The violin plots show the duration between interphase-prophase and meta-anaphase transition in minutes (H, N) and the mean angular difference in degrees (I, O). Images of panels (A)-(I) were acquired with a confocal microscope (LSM5L, 10×, 0.656$\mu m$/pixel). The plots combine extracted trajectories from three independent repeats with a total number of $N_{\text{scrambled}} = 1262$, $N_{\text{LSD1-2}} = 970$, $N_{\text{LSD1-6}} = 1332$, $N_{\text{PP2A}} = 1198$ cells. Images of panels (J)-(O) below the dashed line were acquired with a confocal microscope (LSM5L, 20$X$, 0.328$\mu m$/pixel). The plots combine extracted trajectories from three independent repeats with a total number of $N_{\text{scrambled}} = 1094$, $N_{\text{RecQL4-1}} = 814$, $N_{\text{RecQL4-3}} = 842$, $N_{\text{RecQL4-4}} = 786$ cells. See Table S2 and Table S3 for details on the depicted features.

these, previously unnoticed, alterations in the nuclear morphology are not clear yet. However, abnormalities in the nuclear shape and architecture are widely observed in pathological conditions and ageing ( [58] for review), which hint new paths for biomedical research regarding these protein targets.

Previous work indicate that downregulation of PP2A, LSD1 or RecQL4 delays early mitotic progression. LiveCellMiner analysis shows the expected increase in the average time spent from prophase to anaphase onset after RNAi mediated downregulation of these targets (Fig. 5G-H, M-N). However, a high degree of metaphase plate rotation is only observed in PP2A-downregulated cells, as the mean angle difference follows the axis of cell division (Fig. 4G and Fig. 5I,O). This points to the different roles of PP2A and LSD1/RecQL4 in the control of the spindle apparatus during mitosis, consistent with current knowledge. The early mitotic delay after PP2A downregulation reflects broad defects in spindle function at the level of microtubule–kinetochore attachment [48] and bipolar spindle formation [47]. In turn, the abnormal fluctuations in the orientation of metaphase chromatin indicates faulty function of the cortical network and/or defective astral microtubules emanating from the spindle poles ( [53, 54]; see [59] for review). In the case of LSD1 downregulation, where no metaphase chromatin rotation is observed, the early mitotic delay might arise from defects in chromatin methylation impairing correct chromosome segregation ( [60]; see [61] for review), and might reflect unbalances in the expression of three major players of the mitotic control: PLK1 [62], BUBR1 and MAD2 [63], whose transcriptional regulation is influenced by LSD1. Likewise, the absence of rotation of the metaphase plate in RecQL4-downregulated cells suggests that the delay in early mitotic progression is not due to malfunction of the astral microtubules and/or cortical network.

LiveCellMiner is able to measure the time-dependent distancing of daughter chromatin masses after anaphase onset. This readout is a proxy for chromosome and spindle dynamics during later mitotic stages. Consistent with the known alterations in spindle dynamics during mitotic exit after PP2A downregulation, a delayed chromatin masses separation is observed early after anaphase onset (Fig. 5F, Fig. S5D). In turn, RecQL4 downregulated cells follow similar kinetics than control cells right after anaphase onset but decelerate the separation of the daughter nuclei later (Fig. 5F,L). This might indicate that RecQL4 is involved in the complex regulation of cytoskeleton dynamics during mitotic exit and cytokinesis (see [64–66]for review), and opens a new research avenue.

Expressing fluorescent reporters in cells could perturb the phenomena under study. For example, overexpression of highly charged core histones can replace the surfactant effect that Ki-67 exerts in prometaphase in order to avoid excessive chromosome clustering [67]. Additionally, cell-to-cell variability can introduce artefacts affecting the dynamic range of the measure and preventing the detection of anomalies. To overcome these problems, LiveCellMiner permits single-cell temporal normalization of the extracted features. This allows extracting rates of change in comparable conditions (Fig. 5B-C, Mean Int. vs. Mean Int. Norm, see also Fig. S6B-C). Furthermore, to avoid or minimize the impact of cell-to-cell differences in expression of the reporters, LiveCellMiner includes a module to select cell sub-populations fulfilling certain criteria, *e.g.*, a limited intensity range of the H2B-mCherry chromatin marker in a given part of the cell cycle. In our example within the LSD1 data set, the oligo LSD1-1, but not the LSD1-2, increases considerably the amount of cells with extremely high H2B-mCherry signals (Fig. 5A, Mean Int.). LiveCellMiner enables us to discard effects from the unequal reporter expression by confining the analysis, *i.e.*, to cells with low chromatin intensity values before entry into mitosis (Fig. S5).

The use of various fluorescent markers with different spectral properties is particularly useful for studying complex phenotypes. For mitosis-related events,

studying the mitotic spindle by fluorescent labelling of different tubulin isoforms, *e.g.*, stable expression of eGFP-alpha-Tubulin, helps unraveling whether formation and/or function of the spindle apparatus is affected. Here, LiveCellMiner can extract spindle dynamics data by morphological dilation of the primary segmentation from the chromatin channel with a disk-shaped structuring element with a 15 pixel radius (see Table S3 for available features). PP2A but not RecQL4 or LSD1 RNAi-treated cells show a reliable increase of the spindle intensity staining during mitotic exit (Fig. S6D,G) as well as a delayed disappearance of the astral spindle signal (detected as radial displacement of the intensity maximums in the GFP channel towards the polar regions) (Fig. S6E,H). These measurements are consistent with previous findings which suggest that the PP2A-B55 holoenzyme regulates spindle-pole associated microtubules disassembly in late anaphase [49]. Furthermore, LSD1 but not PP2A (Fig. S6D,F) or RecQL4 (Fig. S6G,I) RNAi-treated cells show stronger spindle intensity during early mitotic progression. This previously unnoticed phenotype could hint to additional roles of LSD1 for spindle stability, a finding, which needs to be confirmed in the future.

## Reanalysis of Mitotic Phenotypes Mediated by siRNA against VPS72, H2A.Z and the ATPase Subunits from the Chromatin Remodeling Complexes EP400, SRCAP and Ino80

We have recently shown that downregulation of VPS72, also referred to as YL-1, extends telophase in cells [35] by using live cell imaging combined with CecogAnalyzer 1.5.2. [10] analysis. VPS72 is part of the EP400 and the Snf2-related CBP-activator protein (SRCAP) chromatin remodeling complexes, where it works as a chaperon for H2A.Z. We also screened for telophase phenotypes after siRNA-mediated downregulation of EP400, SRCAP and H2A.Z, including as control the chromatin remodeling complex INO80, which does not contain VPS72. Here, we use the published data set to demonstrate the flexibility of LiveCellMiner for the analysis of high-content screening (HCS) data sets (Note S1). Box-and-whisker or violin plots allow for comprehensive visualization of biologically relevant image features grouped by treatment, replicate number, etc. to enable direct comparison between dozens of experiments and allowing identification of samples with phenotypic deviations from a huge data set.

Reanalysis of the data set with LiveCellMiner revealed, for example, that downregulation of INO80, SRCAP, EP400 and H2A.Z consistently lengthens early mitotic progression compared to the scramble control, similar to the PP2A positive control (Fig. S4A). In contrast, this is not observed upon VPS72 downregulation. These observations support previous findings that INO80 [68] and SRCAP [69] associate with the spindle apparatus and are required for proper mitotic progression. To our knowledge EP400 has not been linked to spindle function but the analysis raises the possibility that the EP400 chromatin remodeling complex is also involved in mitotic processes. Furthermore, the slight length extension of early mitosis in H2A.Z depleted cells might indicate defective chromosome capture by the spindle apparatus due to chromosome centromeres alteration, where H2A.Z acts as organizer [70].

In a similar way, the detailed analysis of single features like chromatin area at different times (here, 21 and 42 min after anaphase onset) reveals that in PP2A- and VPS72-downregulated cells, nuclei at the end of mitosis are consistently smaller than in control cells (Fig. S4B,C). In turn, PP2A siRNA-treated cells show additionally a delay in the rate of recovery of interphase average area (Fig. S4D,E), which is consistent with the function of PP2A in spindle-pole associated microtubules disassembly and reinitiating nucleo-cytoplasmic transport in anaphase [49].

With the comprehensive display, intra- and inter-experimental variability can be analyzed (Fig. S4). For example, inter-experimental variability (for EXP1 and EXP2;

see (Fig. S4D,E) legend) for chromatin area and area recovery rates at 21 and 41 min are observed for the treatments with the PP2A-siRNA but not for scramble and the other siRNAs.

In the future, we will validate the hypotheses formulated here and investigate the molecular mechanisms behind the newly described early mitotic phenotypes. In essence, the reevaluation of the live cell imaging data sets confirmed our previous findings regarding late mitotic phenotypes. These examples illustrate the power of LiveCellMiner to screen complex phenotypes in a quantitative manner using simple to complex live-cell imaging experiments. Our examples focus on mitotic chromatin, but with modest modifications, LiveCellMiner would also be applicable to other fluoresently labeled subcellular structures.

# Availability and Future Directions

The LiveCellMiner extension package can be obtained from the following repository `https://github.com/stegmaierj/LiveCellMiner`. We provide detailed installation instructions and an overview of all LiveCellMiner-specific functions on the landing page of the repository. As potential improvements will be made available via the repository, a snapshot of the latest version upon the submission time point can be obtained here `https://zenodo.org/badge/latestdoi/269630703`.

While LiveCellMiner provides already all tools to perform in-depth analyses of high-content screens, there are a few points we will address in future versions of the software. In the current implementation, LiveCellMiner searches for mitotic events and uses the anaphase onset as a reference for extracting the remaining frames for the analysis. In future versions, we will also add the possibility to analyze non-mitotic tracks and adapt the synchronization tools to other scenarios as well. The current detection and segmentation methods sequentially process individual images one at a time. To speedup processing for larger screens these steps could potentially be parallelized to fully exploit multicore CPUs and GPUs as available in the respective system. As a temporary workaround, one can run multiple instances of LiveCellMiner and thereby distribute the processing of independent projects, *e.g.*, on different cluster nodes. The deep learning-based segmentation relies currently on the external pretrained software tool Cellpose [16]. The advanced segmentation methods could be implemented directly in MATLAB to streamline the processing with as little additional dependencies as possible. Finally, the trajectory synchronization module of LiveCellMiner currently involves a few semi-automatic steps that can become time-consuming for very large and highly diverse screens. A future avenue of research will thus be improving the reliability of unsupervised approaches like [17] to ultimately analyze high-content screens in a fully-automatic fashion.

# Acknowledgments

We thank Wolfram Antonin for valuable comments and proofreading of the manuscript.

# Supporting Information

## Note S1 Experimental Details.

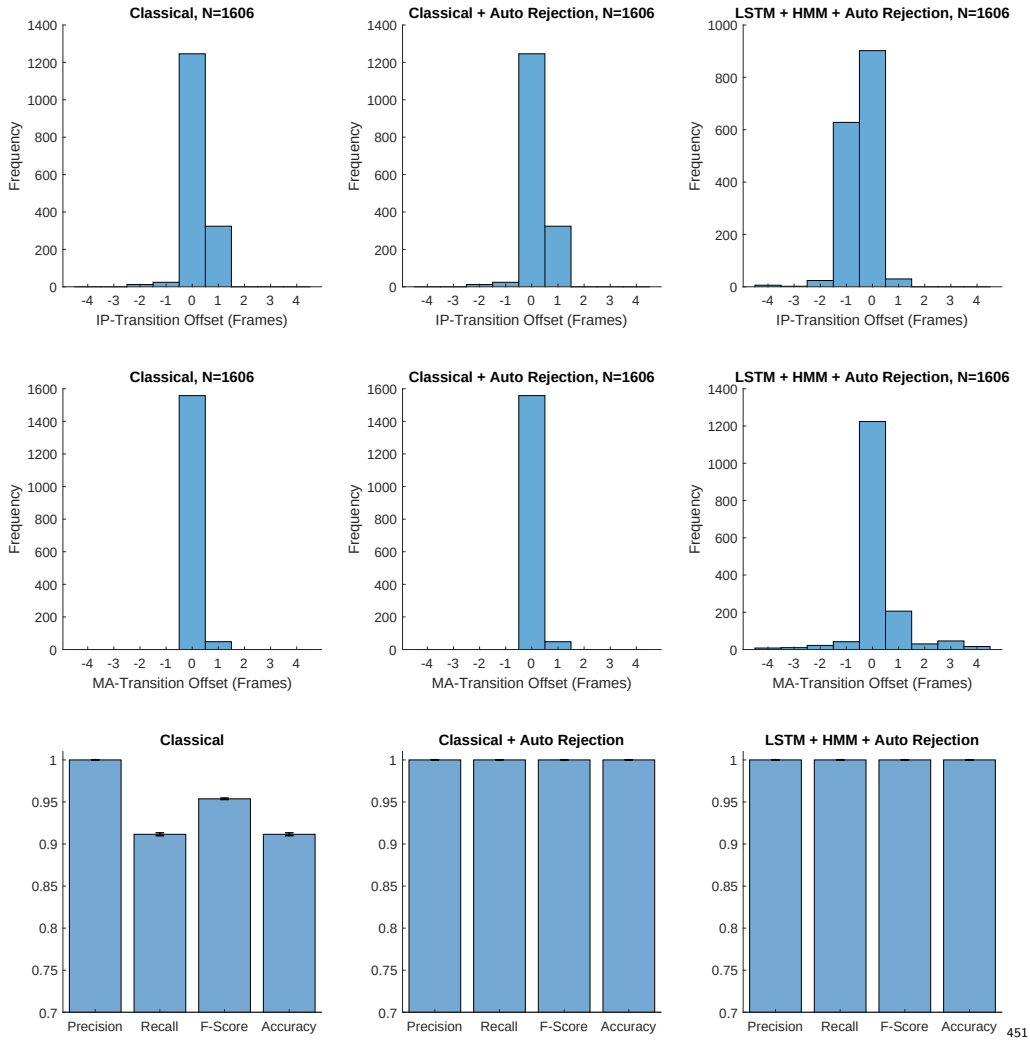


**Fig. S1 Validation of the Synchronization Performance for a Confocal Data Set.** The manually annotated data comprise three experiments with two positions each (20× objective, physical spacing $0.415\mu m$, 3 minute time intervals and $N = 832$ trajectories in total). Columns show the results of the three different synchronization methods (Classical, Classical+Auto Rejection and LSTM+HMM+Auto Rejection) as described in the main text. The first two rows show histograms of the frame offset of the automatically identified synchronization time points with respect to a manually annotated ground truth (IP: interphase to prophase transition and MA: metaphase to anaphase transition). For instance, a value of 1 indicates that the synchronization time point is set one frame too late and 0 indicates a perfect match. The last row quantifies precision, recall, f-score and accuracy of the auto rejection module that is intended to discard erroneous tracks. Validation was performed using a 6-fold cross validation by retraining the LSTM-based classifier on each split and testing the classifier on each of the remaining test data.

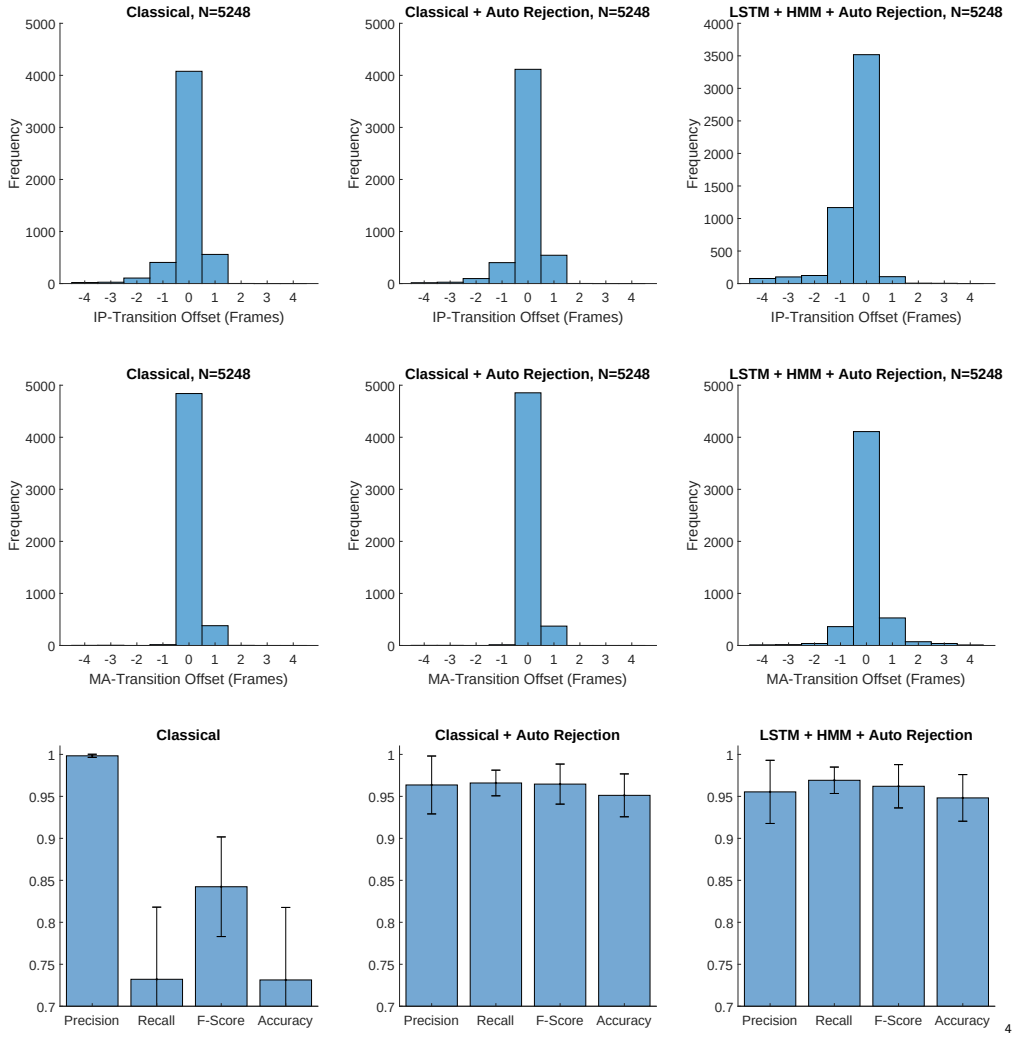


**Fig. S2 Validation of the Synchronization Performance for a Widefield Data Set.** The manually annotated data comprise three experiments with two positions each (10× objective, physical spacing $0.65\mu m$, 3 minute time intervals and $N = 832$ trajectories in total). Columns show the results of the three different synchronization methods (Classical, Classical+Auto Rejection and LSTM+HMM+Auto Rejection) as described in the main text. The first two rows show histograms of the frame offset of the automatically identified synchronization time points with respect to a manually annotated ground truth (IP: interphase to prophase transition and MA: metaphase to anaphase transition). For instance, a value of 1 indicates that the synchronization time point is set one frame too late and 0 indicates a perfect match. The last row quantifies precision, recall, f-score and accuracy of the auto rejection module that is intended to discard erroneous tracks. Validation was performed using a 6-fold cross validation by retraining the LSTM-based classifier on each split and testing the classifier on each of the remaining test data.

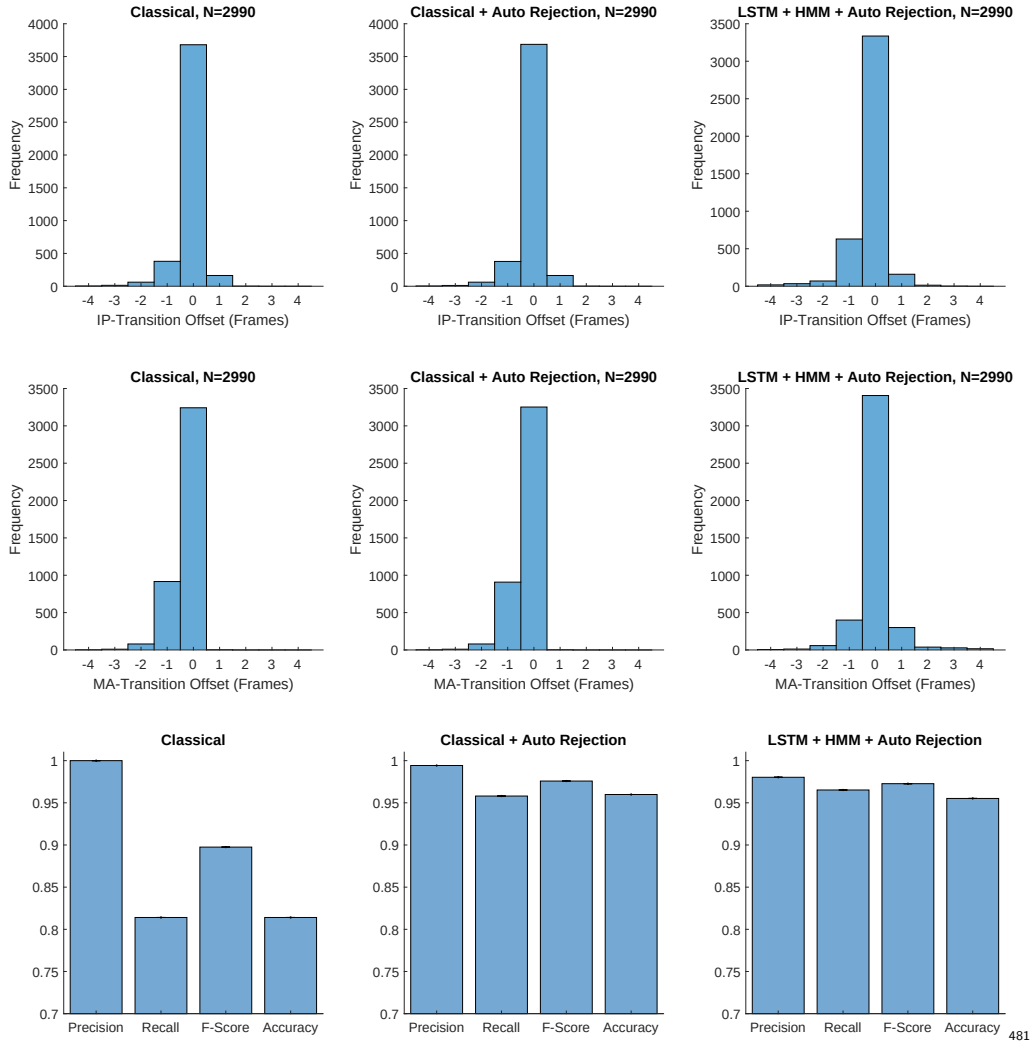


**Fig. S3 Validation of the Synchronization Performance for the LSD1 Data Set.** The manually annotated data comprise three experiments with 16 positions each ($10\times$ objective, physical spacing $0.656\mu m$, 3 minute time intervals and $N = 5878$ trajectories in total). Columns show the results of the three different synchronization methods (Classical, Classical+Auto Rejection and LSTM+HMM+Auto Rejection) as described in the main text. The first two rows show histograms of the frame offset of the automatically identified synchronization time points with respect to a manually annotated ground truth (IP: interphase to prophase transition and MA: metaphase to anaphase transition). For instance, a value of 1 indicates that the synchronization time point is set one frame too late and 0 indicates a perfect match. The last row quantifies precision, recall, f-score and accuracy of the auto rejection module that is intended to discard erroneous tracks. Validation was performed by training on 1328 trajectories that were evenly distributed among all positions and by applying it to a remaining set of 2990 trajectories.

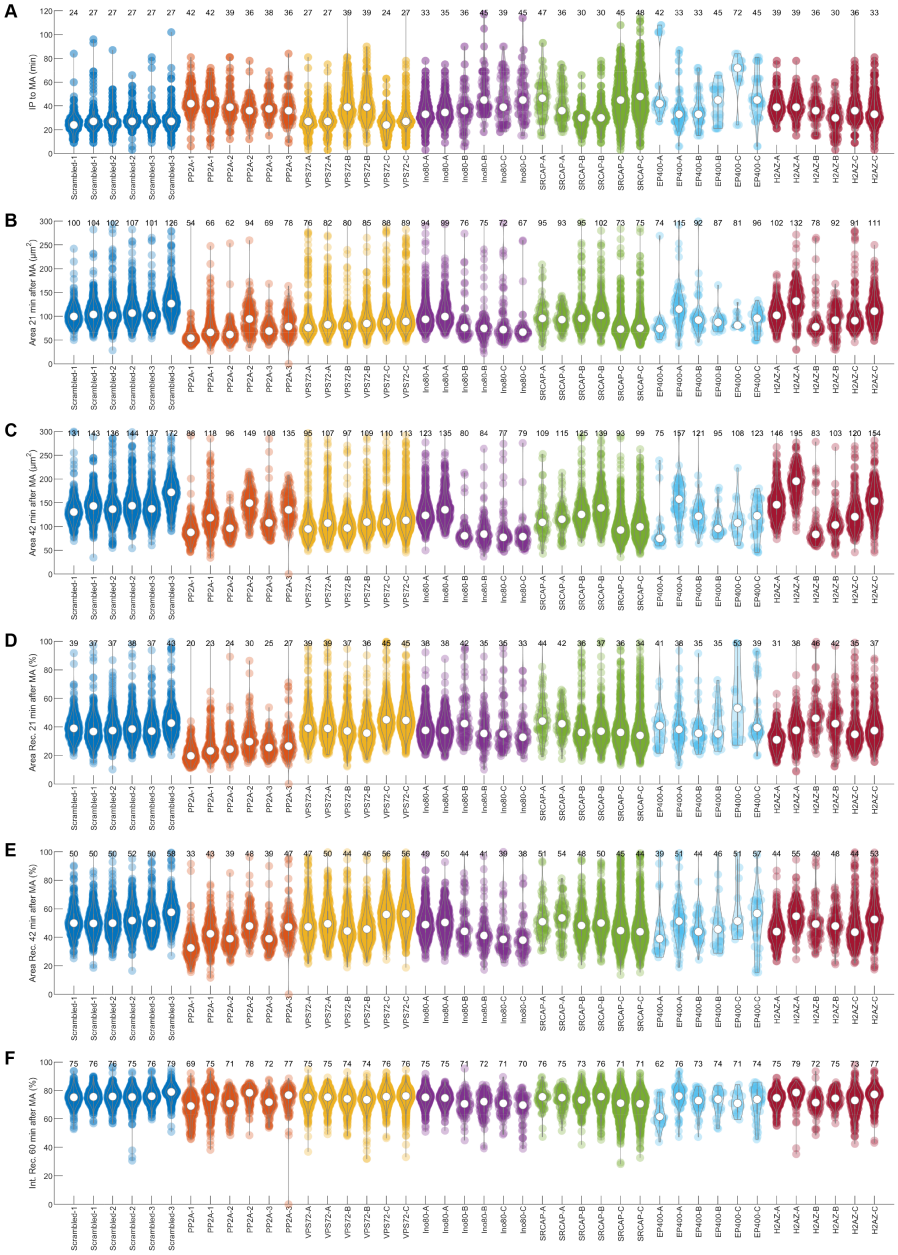


**Fig. S4 Exemplary Readouts from a High-Content Screen.** The different panels show violin plots of the IP to MA duration in minutes (A), the area at 21 and 42 minutes after anaphase onset (B, C), the area recovery compared to the level at interphase in % at 21 and 42 minutes after the anaphase onset (D, E) as well as a combined recovery measure comprised of area, minor axis length, mean intensity and intensity standard deviation at 60 minutes after anaphase onset (F). See Table S2 for a more detailed description of the individual features. Numbers above each violin are the respective median values.

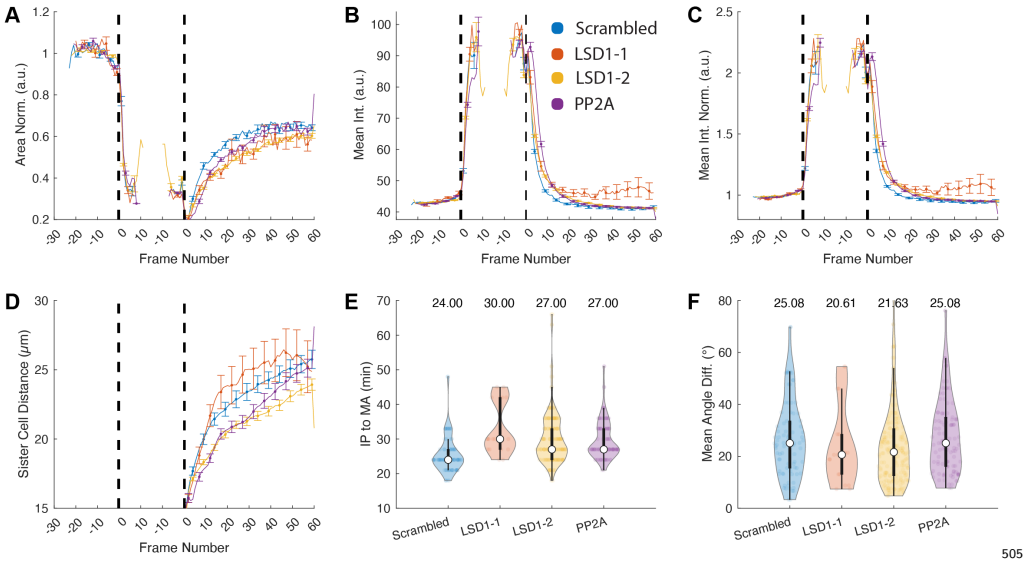


**Fig. S5 Analysis of LSD1 with constrained intensity range (40-46).** Images were acquired with a confocal microscope (LSM5L, 10×, 0.656$\mu m$/pixel). We compare control (Scrambled), LSD1-1, LSD1-2 and PP2A knockdown cells. The basic features involve the normalized area (A), the mean intensity (B), the normalized mean intensity (C) and the sister cell displacement (D). The violin plots show the duration between interphase-prophase and metaphase-anaphase transition in minutes (E) and the mean orientation angle difference in degrees (F). The selection was constrained to cells exhibiting an interphase mean intensity in the range of $40 - 46$, which yielded a set of $N_{\text{scrambled}} = 84$, $N_{\text{LSD1-1}} = 18$, $N_{\text{LSD1-2}} = 190$, $N_{\text{PP2A}} = 106$ cells.

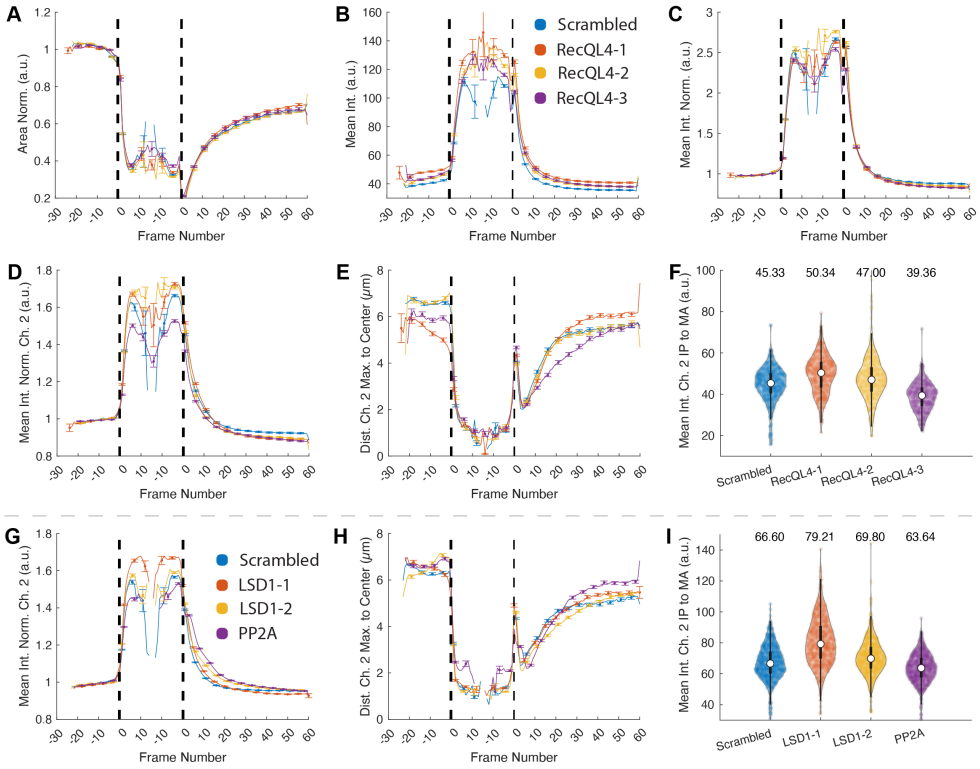


**Fig. S6 Analysis of LSD1 and RecQL4 Knockdowns (Additional Features).** Panels (A)-(F) show control (Scrambled) vs. RecQL4-1, RecQL4-2 and RecQL4-3, whereas panels (G) - (I) show control (Scrambled) vs. LSD1-1, LSD1-2 and PP2A knockdown cells. The features involve the normalized area (A), the mean intensity (B), the normalized mean intensity (absolute intensity values divided by the interphase mean intensity of each cell, C). Panels (D-I) exemplify features that were extracted from the second fluorescence channel and include the normalized mean intensity (D,G), the distance of the intensity maximum to the segmentation centroid (E,H) and the average mean intensity between the IP and MA transitions (F,I). Images of panels (A)-(F) above the dashed line were acquired with a confocal microscope (LSM5L, $20X$, $0.328\mu m$/pixel). The plots combine extracted trajectories from three independent repeats with a total number of $N_{\text{scrambled}} = 1094$, $N_{\text{RecQL4-1}} = 814$, $N_{\text{RecQL4-3}} = 842$, $N_{\text{RecQL4-4}} = 786$ cells. Images of panels (G)-(I) below the dashed line were acquired with a confocal microscope (LSM5L, $10\times$, $0.656\mu m$/pixel). The plots combine extracted trajectories from three independent repeats with a total number of $N_{\text{scrambled}} = 1262$, $N_{\text{LSD1-2}} = 970$, $N_{\text{LSD1-6}} = 1332$, $N_{\text{PP2A}} = 1198$ cells. See Table S2 and Table S3 for details on the depicted features.

**Table S1 Quantification of the Detection, Segmentation and Tracking Performance.** Measurements were performed on two manually annotated time series of the `Fluo-N2DL-HeLa` data set that is part of the Cell Tracking Challenge [71] and was most similar to our application scenario (acquisition was performed with an Olympus IX81 microscope using a $10 \times /0.4$ objective lens, a physical spacing of $0.645 \times 0.645 \mu m$ and 30 minute time intervals). The used metrics are DET (detection score), SEG (segmentation score), TRA (tracking score), $\mathrm{OP}_{\mathrm{CTB}}$ (average of SEG and TRA) and $\mathrm{OP}_{\mathrm{CSB}}$ (average of DET and SEG). All metrics have a value range between 0 and 1 with 1 being the optimum (see [71] for details on the measures). While Cellpose [16] yields better segmentation scores, it misses cells occasionally. On the other hand, our classical approach based on multi-scale Laplacian-of-Gaussian blob detection is able to find most cells with a slightly worse segmentation accuracy. Combining both approaches, *i.e.*, complementing the Cellpose segmentation with additional cells that were only found by the classical method, yielded consistently the best results (highlighted in bold face letters). The experiments described in the main paper were consistently acquired with 3 minute time intervals, which further improves the tracking performance. Please note that neither parameter tuning of the classical method nor any retraining of Cellpose was performed. Thus, the top-scoring methods listed on the Cell Tracking Challenge leader board are still slightly higher but we expect our method to generalize better to unseen data sets.

| **Method** | **Series** | **DET** | **SEG** | **TRA** | $\mathbf{OP_{CTB}}$ | $\mathbf{OP_{CSB}}$ |
|---|---|---|---|---|---|---|
| Classical | 01 | 0.969 | 0.687 | 0.941 | 0.814 | 0.828 |
| Cellpose | 01 | 0.903 | 0.681 | 0.879 | 0.780 | 0.792 |
| Cellpose+Classical | 01 | **0.978** | **0.766** | **0.960** | **0.863** | **0.872** |
| Classical | 02 | 0.939 | 0.700 | 0.909 | 0.805 | 0.820 |
| Cellpose | 02 | 0.929 | 0.806 | 0.906 | 0.856 | 0.868 |
| Cellpose+Classical | 02 | **0.948** | **0.821** | **0.927** | **0.874** | **0.885** |
| Classical | Average | 0.954 | 0.694 | 0.925 | 0.810 | 0.824 |
| Cellpose | Average | 0.916 | 0.743 | 0.893 | 0.818 | 0.830 |
| Cellpose+Classical | Average | **0.963** | **0.793** | **0.952** | **0.873** | **0.878** |

**Table S2 Description of the relevant time series (TS) and single features (SF) used for characterizing cell behavior based on the chromatin channel.**

| Feature Name | Type | Description |
|---|---|---|
| xpos | TS | Spatial x coordinate of the centroid of each nucleus per time point. |
| ypos | TS | Spatial y coordinate of the centroid of each nucleus per time point. |
| Area | TS | Area of the segmented nucleus measured in $\mu m^2$. |
| MajorAxisLength | TS | Length of the major axis the segmented nucleus measured in $\mu m$ [72]. |
| MinorAxisLength | TS | Length of the minor axis the segmented nucleus measured in $\mu m$ [72]. |
| Orientation | TS | Angle between the positive x-axis and the major axis measured in degrees [72]. |
| Circularity | TS | Measure for the roundness of a nucleus and defined as $(4 * \text{Area} * \pi)/(\text{Perimeter}^2)$. A perfect circle has a value of 1 [72]. |
| MeanIntensity | TS | Average intensity measured in the segmented area. |
| StdIntensity | TS | Standard deviation of the intensity measured in the segmented area. |
| StdIntensityGradMag | TS | Standard deviation of a gradient magnitude measured in the segmented area. |
| RecoveryFeature | TS | Interphase recovery in percent (100% indicates full recovery). Measures the absolute percentage deviation of one or more features to their respective interphase mean value (average percentage deviation if multiple features are selected). Default used for the figures in this paper: Area, Circularity, MeanIntensity, StdIntensity. |
| AngularSecondMoment | TS | Angular Second Moment (Energy) derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| Contrast | TS | Contrast derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| Correlation | TS | Correlation derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| Variance | TS | Variance derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| Homogeneity | TS | Inverse Difference Moment (Homogeneity) derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| SumAverage | TS | Sum Average derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| SumVariance | TS | Sum Variance derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| SumEntropy | TS | Sum Entropy derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| Entropy | TS | Entropy derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| DifferenceVariance | TS | Difference Variance derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| DifferenceEntropy | TS | Difference Entropy derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| InformationMeasureofCorrelationI | TS | Information Measure of Correlation I derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| InformationMeasureofCorrelationII | TS | Information Measure of Correlation II derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| MaximalCorrelationCoefficient | TS | Maximal Correlation Coefficient derived from the graylevel co-occurrence matrix as described in [41] and implemented in [73]. |
| manualSynchronization | TS | Synchronization stage assigned either automatically or manually. Allowed values are $-1$ (invalid trajectory), 0 (unlabeled trajectory), 1 (interphase), 2 (prophase, metaphase, early anaphase), 3 (late anaphase, telophase). |
| manuallyConfirmed | SF | Boolean value for each trajectory that indicates if the synchronization time points were manually checked. Only manually checked trajectories will be used for classifier training. |
| IPToMALength_Frames | SF | Duration between the IP and MA synchronization time points measured in number of frames. |
| IPToMALength_Minutes | SF | Duration between the IP and MA synchronization time points measured in minutes. |
| InterphaseMeanIntensity | SF | Average of the mean intensity during interphase. |
| AccumulatedOrientationDiffPMA | SF | Sum of the angular change of the major axis orientation for the entire time interval measured in degrees. |

**Table S3 Description of the relevant time series (TS) and single features (SF) that can be extracted from the second channel (optionally, if available).** Note that all features mentioned in Table S2 can also be extracted for the secondary channel and are not listed here. Moreover, each of the listed features can be computed on a grown/shrunk/toroidal region obtained via morphological dilation/erosion of the chromatin mass segmentation with a user-defined radius and structuring element.

| Feature Name | Type | Description |
|---|---|---|
| Ch1-Ch2-MI-Ratio | TS | Ratio of the mean intensities between Ch1 and Ch2. |
| Ch2-Outer-Inner-Mean-Ratio | TS | Mean value of the brightest 25% pixels in the secondary channel (measured in a toroidal region around the chromatin segmentation) and the mean value of the 25% of the brightest pixels of the secondary channel within the region of the chromatin segmentation. |
| Ch2-Max-Int-Displacement | TS | Euclidean distance between the intensity-weighted centroid of the brightest 25% pixels in the secondary channel measured in a toroidal region around the chromatin segmentation and the centroid of the chromatin segmentation. |
| %TSNAME%_%TIME%FramesAfterMA | SF | Extracts the value of a selected time series %TSNAME% at time point %TIME% as a single feature. |

**Table S4 Quantification of the Synchronization Performance.** Measurements were performed on a confocal data set (top, 20× objective, physical spacing $0.415\mu m$, 3 minute time intervals and $N = 1606$ trajectories in total) and a widefield data set (bottom, 10× objective, physical spacing $0.65\mu m$, 3 minute time intervals and $N = 5248$ trajectories in total). True positives (TN), true negatives (TN), false positives (FP) and false negative (FN) are summed over all folds of the 6-fold cross-validation. Precision, recall, accuracy and F-Score were computed individually on each fold and the displayed values are the mean ± SD. The last two columns contain the mean ± SD values of the synchronization time point offsets for the interphase prophase transition (IP) and the metaphase to anaphase transition (MA) measured in frames (a deviation of 1 frame corresponds to 3 minutes for both data sets). Bold-face letters indicate the best values for each measure and data set.

| Microscope / Method | TP | TN | FP | FN | Precision | Recall | Accuracy | F-Score | Avg. Dev. IP | Avg. Dev. MA |
|---|---|---|---|---|---|---|---|---|---|---|
| Conf. / Classical | 1464 | 0 | 0 | 142 | **1.00 ± 0.00** | 0.91 ± 0.00 | 0.91 ± 0.00 | 0.95 ± 0.00 | 0.17 ± 0.47 | 0.03 ± 0.17 |
| Conf. / Classical + Auto Rej. | 1464 | 142 | 0 | 0 | **1.00 ± 0.00** | **1.00 ± 0.00** | **1.00 ± 0.00** | **1.00 ± 0.00** | 0.17 ± 0.47 | 0.03 ± 0.17 |
| Conf. / LSTM + HMM + Auto Rej. | 1464 | 142 | 0 | 0 | **1.00 ± 0.00** | **1.00 ± 0.00** | **1.00 ± 0.00** | **1.00 ± 0.00** | -0.48 ± 0.86 | 0.19 ± 0.91 |
| Widef. / Classical | 3838 | 0 | 6 | 1404 | **1.00 ± 0.00** | 0.73 ± 0.09 | 0.73 ± 0.09 | 0.84 ± 0.06 | -0.11 ± 1.03 | 0.07 ± 0.35 |
| Widef. / Classical + Auto Rej. | 3714 | 1278 | 130 | 126 | 0.96 ± 0.03 | **0.97 ± 0.02** | **0.95 ± 0.03** | **0.96 ± 0.02** | -0.10 ± 0.97 | 0.07 ± 0.34 |
| Widef. / LSTM + HMM + Auto Rej. | 3684 | 1292 | 160 | 112 | 0.96 ± 0.04 | **0.97 ± 0.02** | **0.95 ± 0.03** | 0.96 ± 0.03 | -0.54 ± 1.55 | 0.04 ± 1.25 |
| LSD1 / Classical | 2434 | 0 | 0 | 556 | **1.00 ± 0.00** | 0.81 ± 0.00 | 0.81 ± 0.00 | 0.90 ± 0.00 | -0.09 ± 0.69 | -0.40 ± 1.44 |
| LSD1 / Classical + Auto Rej. | 2420 | 450 | 14 | 106 | 0.99 ± 0.00 | 0.96 ± 0.00 | **0.96 ± 0.00** | **0.98 ± 0.00** | **-0.09 ± 0.68** | -0.40 ± 1.43 |
| LSD1 / LSTM + HMM + Auto Rej. | 2386 | 470 | 48 | 86 | 0.98 ± 0.00 | **0.97 ± 0.00** | **0.96 ± 0.00** | 0.97 ± 0.00 | -0.27 ± 1.20 | **-0.06 ± 1.71** |

**Table S5 Statistical analysis.** Statistical test results for the violin plots of Fig. 4, Fig. 5, Fig. S4, Fig. S5 and Fig. S6 summarized in a supplementary spreadsheet file.

**Table S6 Statistical analysis.** Statistical test results for the time series plots of Fig. 4, Fig. 5, Fig. S4, Fig. S5 and Fig. S6 summarized in a supplementary spreadsheet file.

**File S1 Generated Report for the LSD1 Data Set.** All existing single features and time series features are contained and accessible from an HTML-based overview file. Extract the archive to a folder of your choice and open the HTML file in the root directory using any web browser.

**File S2 Generated Report for the RecQL4 Data Set.** All existing single features and time series features are contained and accessible from an HTML-based overview file. Extract the archive to a folder of your choice and open the HTML file in the root directory using any web browser.

# References


1. Usaj MM, Styles EB, Verster AJ, Friesen H, Boone C, Andrews BJ. High-Content Screening for Quantitative Cell Biology. Trends in Cell Biology. 2016;26(8):598–611.

2. Li S, Xia M. Review of High-Content Screening Applications in Toxicology. Archives of Toxicology. 2019;93(12):3387–3396.

3. Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE. Image-based Profiling for Drug Discovery: Due for a Machine-Learning Upgrade? Nature Reviews Drug Discovery. 2021;20(2):145–159.

4. Rajewsky N, Almouzni G, Gorski SA, Aerts S, Amit I, Bertero MG, et al. LifeTime and Improving European Healthcare Through Cell-Based Interceptive Medicine. Nature. 2020;587(7834):377–386.

5. Boland MV, Murphy RF. A Neural Network Classifier Capable of Recognizing the Patterns of All Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells. Bioinformatics. 2001;17(12):1213–23.

6. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. Genome Biology. 2006;7(10):R100.

7. Neumann B, Held M, Liebel U, Erfle H, Rogers P, Pepperkok R, et al. High-Throughput RNAi Screening by Time-Lapse Imaging of Live Human Cells. Nature Methods. 2006;3(5):385–90.

8. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn GW, Tao CY, et al. Integrating High-Content Screening and Ligand-Target Prediction to Identify Mechanism of Action. Nature Chemical Biology. 2008;4(1):59–68.

9. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, et al. Scoring Diverse Cellular Morphologies in Image-Based Screens with Iterative Feedback and Machine Learning. Proceedings of the National Academy of Sciences. 2009;106(6):1826–31. doi:10.1073/pnas.0808843106.

10. Held M, Schmitz MHa, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: Time-Resolved Phenotype Annotation in High-Throughput Live Cell Imaging. Nature Methods. 2010;7(9):747–54.

11. Piccinini F, Balassa T, Szkalisity A, Molnar C, Paavolainen L, Kujala K, et al. Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. Cell Systems. 2017;4(6):651–655.

12. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, et al. ilastik: Interactive Machine Learning for (Bio)Image Analysis. Nature Methods. 2019;16(12):1226–1232.

13. Caicedo JC, McQuin C, Goodman A, Singh S, Carpenter AE. Weakly Supervised Learning of Single-Cell Feature Embeddings. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 9309–9318.

14. Kusumoto D, Seki T, Sawada H, Kunitomi A, Katsuki T, Kimura M, et al. Anti-Senescent Drug Screening by Deep Learning-based Morphology Senescence Scoring. Nature Communications. 2021;12(1):257.

15. Wang R, Wang D, Kang D, Guo X, Guo C, Dongye M, et al. An Artificial Intelligent Platform for Live Cell Identification and the Detection of Cross-Contamination. Annals of Translational Medicine. 2020;8(11):697.

16. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: A Generalist Algorithm for Cellular Segmentation. Nature Methods. 2021;18(1):100–106.

17. Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW. Unsupervised Modeling of Cell Morphology Dynamics for Time-Lapse Microscopy. Nature Methods. 2012;9(7):711.

18. Failmezger H, Frohlich H, Tresch A. Unsupervised Automated High-Throughput Phenotyping of RNAi Time-Lapse Movies. BMC Bioinformatics. 2013;14:292.

19. Sommer C, Hoefler R, Samwer M, Gerlich DW. A Deep Learning and Novelty Detection Framework for Rapid Phenotyping in High-Content Screening. Molecular Biology of the Cell. 2017;28(23):3428–3436.

20. Sailem HZ, Rittscher J, Pelkmans L. KCML: A Machine-Learning Framework for Inference of Multi-Scale Gene Functions from Genetic Perturbation Screens. Molecular Systems Biology. 2020;16(3):e9083.

21. Fischer EG. Nuclear Morphology and the Biology of Cancer Cells. Acta Cytol. 2020;64(6):511–519.

22. Katayama A, Toss MS, Parkin M, Sano T, Oyama T, Quinn CM, et al. Nuclear Morphology in Breast Lesions: Refining its Assessment to Improve Diagnostic Concordance. Histopathology. 2021;.

23. Way GP, Kost-Alimova M, Shibue T, Harrington WF, Gill S, Piccioni F, et al. Predicting Cell Health Phenotypes using Image-based Morphology Profiling. Molecular Biology of the Cell. 2021;32(9):995–1005.

24. Wang W, Douglas D, Zhang J, Kumari S, Enuameh MS, Dai Y, et al. Live-Cell Imaging and Analysis Reveal Cell Phenotypic Transition Dynamics Inherently Missing in Snapshot Data. Science Advances. 2020;6(36):eaba9319.

25. Eschweiler D, Rethwisch M, Koppers S, Stegmaier J. Spherical Harmonics for Shape-Constrained 3D Cell Segmentation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE; 2021. p. 792–796.

26. Mandal S, Uhlmann V. SplineDist: Automated Cell Segmentation with Spline Curves. In: Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2021. p. 1082–1086.

27. Mierzwa B, Gerlich DW. Cytokinetic Abscission: Molecular Mechanisms and Temporal Control. Developmental Cell. 2014;31(5):525–538.

28. Batty P, Gerlich DW. Mitotic Chromosome Mechanics: How Cells Segregate Their Genome. Trends in Cell Biology. 2019;29(9):717–726.

29. Antonin W, Neumann H. Chromosome Condensation and Decondensation during Mitosis. Current Opinion in Cell Biology. 2016;40:15–22.

30. Schellhaus AK, De Magistris P, Antonin W. Nuclear Reformation at the End of Mitosis. Journal of Molecular Biology. 2016;428(10):1962–1985.

31. Liu S, Pellman D. The Coordination of Nuclear Envelope Assembly and Chromosome Segregation in Metazoans. Nucleus. 2020;11(1):35–52.

32. Mikut R, Bartschat A, Doneit W, Ordiano JÁG, Schott B, Stegmaier J, et al. The MATLAB Toolbox SciXMiner: User's Manual and Programmer's Guide. arXiv preprint. 2017; p. arXiv:1704.03298.

33. Yokoyama H, Moreno-Andres D, Astrinidis SA, Hao Y, Weberruss M, Schellhaus AK, et al. Chromosome Alignment Maintenance Requires the MAP RECQL4, Mutated in the Rothmund-Thomson Syndrome. Life Science Alliance. 2019;2(1).

34. Schooley A, Moreno-Andres D, De Magistris P, Vollmer B, Antonin W. The Lysine Demethylase LSD1 is Required for Nuclear Envelope Formation at the End of Mitosis. Journal of Cell Science. 2015;128(18):3466–77.

35. Moreno-Andres D, Yokoyama H, Scheufen A, Holzer G, Lue H, Schellhaus AK, et al. VPS72/YL1-Mediated H2A.Z Deposition Is Required for Nuclear Reassembly after Mitosis. Cells. 2020;9(7).

36. Stegmaier J, Otte JC, Kobitski A, Bartschat A, Garcia A, Nienhaus GU, et al. Fast Segmentation of Stained Nuclei in Terabyte-Scale, Time Resolved 3D Microscopy Image Stacks. PLOS ONE. 2014;9(2):1–11.

37. Bartschat A, Hübner E, Reischl M, Mikut R, Stegmaier J. XPIWIT - An XML Pipeline Wrapper for the Insight Toolkit. Bioinformatics. 2016;32(2):315–317.

38. Eschweiler D, Stegmaier J. Algorithms used for the Cell Segmentation Benchmark Competition at ISBI 2019 by RWTH-GE. arXiv preprint arXiv:190406890. 2019;.

39. Ward Jr JH. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. 1963;58(301):236–244.

40. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics. 1979;9:62–66.

41. Haralick RM, Shanmugam K, Dinstein IH. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics. 1973;6:610–621.

42. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper With Convolutions. Proceedings of the IEEE Computer Society Conference On Computer Vision and Pattern Recognition. 2015; p. 1–9.

43. Schott B, Traub M, Schlagenhauf C, Takamiya M, Antritter T, Bartschat A, et al. EmbryoMiner: A New Framework for Interactive Knowledge Discovery in Large-Scale Cell Tracking Data of Developing Embryos. PLOS Computational Biology. 2018;14(4):1–18.

44. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735–1780.

45. Durbin R, Eddy SR, Krogh A, Mitchison G. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press; 1998.

46. Forney GD. The Viterbi Algorithm. Proceedings of the IEEE. 1973;61(3):268–278.

47. Liu Y, Zhang Z, Liang H, Zhao X, Liang L, Wang G, et al. Protein Phosphatase 2A (PP2A) Regulates EG5 to Control Mitotic Progression. Scientific Reports. 2017;7(1):1630.

48. Espert A, Uluocak P, Bastos RN, Mangat D, Graab P, Gruneberg U. PP2A-B56 Opposes Mps1 Phosphorylation of Knl1 and Thereby Promotes Spindle Assembly Checkpoint Silencing. Journal of Cell Biology. 2014;206(7):833–42.

49. Schmitz MH, Held M, Janssens V, Hutchins JR, Hudecz O, Ivanova E, et al. Live-Cell Imaging RNAi Screen Identifies PP2A-B55alpha and Importin-Beta1 as Key Mitotic Exit Regulators in Human Cells. Nature Cell Biology. 2010;12(9):886–93.

50. Wlodarchak N, Xing Y. PP2A as a Master Regulator of the Cell Cycle. Critical Reviews in Biochemistry and Molecular Biology. 2016;51(3):162–84.

51. Holder J, Poser E, Barr FA. Getting Out of Mitosis: Spatial and Temporal Control of Mitotic Exit and Cytokinesis by PP1 and PP2A. FEBS Letters. 2019;593(20):2908–2924.

52. Tamashunas AC, Tocco VJ, Matthews J, Zhang Q, Atanasova KR, Paschall L, et al. High-Throughput Gene Screen Reveals Modulators of Nuclear Shape. Molecular Biology of the Cell. 2020;31(13):1392–1402. doi:10.1091/mbc.E19-09-0520.

53. Bendre S, Rondelet A, Hall C, Schmidt N, Lin YC, Brouhard GJ, et al. GTSE1 Tunes Microtubule Stability for Chromosome Alignment and Segregation by Inhibiting the Microtubule Depolymerase MCAK. Journal of Cell Biology. 2016;215(5):631–647.

54. Tame MA, Raaijmakers JA, Afanasyev P, Medema RH. Chromosome Misalignments Induce Spindle-Positioning Defects. EMBO Reports. 2016;17(3):317–25.

55. Brown A, Geiger H. Chromosome Integrity Checkpoints in Stem and Progenitor Cells: Transitions upon Differentiation, Pathogenesis, and Aging. Cellular and Molecular Life Sciences. 2018;75(20):3771–3779.

56. Jo M, Kusano Y, Hirota T. Unraveling Pathologies Underlying Chromosomal Instability in Cancers. Cancer Science. 2021;112(8):2975–2983.

57. Ungricht R, Kutay U. Mechanisms and Functions of Nuclear Envelope Remodelling. Nature Reviews Molecular Cell Biology. 2017;18(4):229–245.

58. Karoutas A, Akhtar A. Functional Mechanisms and Abnormalities of the Nuclear Lamina. Nature Cell Biology. 2021;23(2):116–126.

59. di Pietro F, Echard A, Morin X. Regulation of Mitotic Spindle Orientation: An Integrated View. EMBO Reports. 2016;17(8):1106–30.

60. Heit R, Rattner JB, Chan GK, Hendzel MJ. G2 Histone Methylation is Required for the Proper Segregation of Chromosomes. Journal of Cell Science. 2009;122(16):2957–68.

61. Schmitz ML, Higgins JMG, Seibert M. Priming Chromatin for Segregation: Functional Roles of Mitotic Histone Modifications. Cell Cycle. 2020;19(6):625–641.

62. Dalvi PS, Macheleidt IF, Lim SY, Meemboor S, Muller M, Eischeid-Scholz H, et al. LSD1 Inhibition Attenuates Tumor Growth by Disrupting PLK1 Mitotic Pathway. Molecular Cancer Research. 2019;17(6):1326–1337.

63. Lv S, Bu W, Jiao H, Liu B, Zhu L, Zhao H, et al. LSD1 is Required for Chromosome Segregation during Mitosis. European Journal of Cell Biology. 2010;89(7):557–63.

64. Afonso O, Figueiredo AC, Maiato H. Late Mitotic Functions of Aurora Kinases. Chromosoma. 2017;126(1):93–103.

65. Vukušić K, Buda R, Ponjavić I, Risteski P, Tolić IM. Mitotic Spindle Elongation is Driven by Joint Microtubule Sliding Action of Kinesins KIF4A and EG5. bioRxiv. 2019; p. 863381.

66. Vukusic K, Ponjavic I, Buda R, Risteski P, Tolic IM. Microtubule-Sliding Modules based on Kinesins EG5 and PRC1-Dependent KIF4A Drive Human Spindle Elongation. Developmental Cell. 2021;56(9):1253–1267 e10.

67. Cuylen-Haering S, Petrovic M, Hernandez-Armendariz A, Schneider MW, Samwer M, Blaukopf C, et al. Chromosome Clustering by Ki-67 Excludes Cytoplasm during Nuclear Assembly. Nature. 2020;587(7833):285–290.

68. Hur SK, Park EJ, Han JE, Kim YA, Kim JD, Kang D, et al. Roles of Human INO80 Chromatin Remodeling Enzyme in DNA Replication and Chromosome Segregation Suppress Genome Instability. Cellular and Molecular Life Sciences. 2010;67(13):2283–96.

69. Messina G, Prozzillo Y, Delle Monache F, Santopietro MV, Atterrato MT, Dimitri P. The ATPase SRCAP is Associated with the Mitotic Apparatus, Uncovering Novel Molecular Aspects of Floating-Harbor Syndrome. BMC Biology. 2021;19(1):184.

70. Greaves IK, Rangasamy D, Ridgway P, Tremethick DJ. H2A.Z Contributes to the Unique 3D Structure of the Centromere. Proceedings of the National Academy of Sciences. 2007;104(2):525–30.

71. Ulman V, Maška M, Magnusson KE, Ronneberger O, Haubold C, Harder N, et al. An Objective Comparison of Cell-Tracking Algorithms. Nature Methods. 2017;14(12):1141–1152.

72. The MathWorks I. MATLAB Documentation: regionprops; 2021. `https://de.mathworks.com/help/images/ref/regionprops.html`.

73. Monzel R. MATLAB Central: haralickTextureFeatures; 2018. `https://de.mathworks.com/matlabcentral/fileexchange/58769-haralicktexturefeatures`.