This work has been published in Nature Methods, August 8, 2010 [epub ahead of print]. DOI: 10.1038/nmeth.1486

CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging

Michael Held[1, 2], Michael H A Schmitz[1, 2], Bernd Fischer[3], Thomas Walter[4], Beate Neumann[5], Michael H Olma[1], Matthias Peter[1], Jan Ellenberg[4], and Daniel W Gerlich*[1, 2]

1. Institute of Biochemistry, Swiss Federal Institute of Technology Zurich (ETHZ), Zurich, Switzerland

2. Marine Biological Laboratory (MBL), Woods Hole, USA

3. Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

4. Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

5. Advanced Light Microscopy Facility, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

*Correspondence should be addressed to D. W. G. (daniel.gerlich@bc.biol.ethz.ch).

Keywords: Live cell imaging, RNAi screening, GFP, Machine Learning, Image analysis, hidden Markov model, Cdc20, mitotic exit

Abbreviations: RNAi: RNA interference; siRNA: small interfering RNA

Summary

Fluorescence time-lapse imaging has become a powerful tool to investigate complex dynamic processes such as cell division or intracellular trafficking. Automated microscopes generate time-resolved imaging data at high throughput, yet tools for quantification of large-scale movie data are largely missing. Here, we present *CellCognition*, a computational framework to annotate complex cellular dynamics. We developed a machine learning method that combines state-of-the-art classification with hidden Markov modeling for annotation of the progression through morphologically distinct biological states. The incorporation of time information into the annotation scheme was essential to suppress classification noise at state transitions, and confusion between different functional states with similar morphology. We demonstrate generic applicability in a set of different assays and perturbation conditions, including a candidate-based RNAi screen for mitotic exit regulators in human cells. *CellCognition* is published as open source software, enabling live imaging-based screening with assays that directly score cellular dynamics.

Introduction

The availability of RNAi technology for high-throughput gene inactivation experiments, fluorescent protein labeling, and automated microscopy has opened a new era of screening possibilities in higher eukaryotes[1]. Indeed, imaging-based RNAi and chemical compound screening has become one of the most important discovery tools for the identification of new gene function, for example in the regulation of DNA damage and repair[2], endocytosis[3], mitosis[4-6].

Imaging-based screens typically assay altered incidence of cells with specific features within a population of fixed, fluorescently labeled cells. The development of computational methods for the automated annotation of high-throughput imaging data was key to establish microscopy-based screening as a routine technology in a wide research community. Especially machine learning for supervised classification of cellular morphologies is one of the most powerful annotation strategies[7-12].

Many biological processes depend on stochastic events and occur in an unsynchronized and transient manner, which limits the applicability of single time point assays. Particularly, complex dynamic processes such as cell division or intracellular trafficking demand for time-resolved live cell imaging[13]. Automated microscopes now enable live imaging with high throughput and spatio-temporal resolution[1, 7, 14]. Computational analysis of such data is challenging and existing machine learning and classification approaches do not provide sufficient accuracy to correctly annotate cellular trajectories with multiple time points. Published live imaging-based RNAi screens scored phenotypes either exclusively at the cell population level[6, 7], or relied on visual evaluation of single cell dynamics[4]. However, cell population analysis cannot detect stochastic and transient phenotypes, and visual interpretation of morphological dynamics is very time consuming and often unreliable.

To improve the classification accuracy of machine learning methods, the temporal context can be taken into account. For example, if the biological process underlying an assay is well known, a biological model can be explicitly defined in an error correction scheme that suppresses illegitimate stage transitions. This has been applied to the pattern of mitotic chromatin morphology changes[11, 12]. However, temporal error correction based on biological *a priori* models limits the detection of unexpected

phenotypic variations, and the adaptation to different biological questions requires re-implementation of the underlying models by the user for each new assay.

Here, we present *CellCognition*, an integrated computational strategy that combines machine learning methods for supervised classification and hidden Markov modeling to measure morphological dynamics in live cell microscopic movies. Our error correction method does not require *a priori* definition of the temporal progression, which enables its application to a wide range of assays and phenotypic variations. We demonstrate efficiency and sensitivity of the methodology in various assays and perturbation conditions.

Results

High-throughput imaging of cellular dynamics

To visualize morphological dynamics of various cellular structures, we generated a collection of human HeLa reporter cell lines stably expressing different combinations of fluorescent markers. All cell lines expressed a red chromatin marker (core histone 2B (H2B) fused to mCherry). In this background, we co-expressed markers for microtubules (mEGFP-α-tubulin), the Golgi apparatus (Galactosyl transferase (GalT) fused to EGFP), or DNA replication factories (proliferating cell nuclear antigen (PCNA) fused to mEGFP). This diverse set of secondary markers (Fig. 1a) provides a well-suited test case for the implementation of a generic annotation method. With these cells, we performed multi-location time-lapse imaging on an automated wide-field epifluorescence microscope[14]. We typically recorded 96 movies in parallel, with a temporal resolution less than 5 min over a total duration of 24 h, generating datasets of about 100,000 images, or 200 Gigabyte, per day and microscope. The analysis of such a single experiment requires annotation of up to 25 million cellular morphologies derived from about 260,000 objects per movie with a 10x microscope objective.

Machine learning and classification of morphologies

Timing measurements in live cell imaging data are often based on the progression through distinct morphologies that relate to specific biological states. An excellent example for this is mitosis, for which the chromatin morphology can be used to annotate the canonical mitotic stages (Fig. 1b; Supplementary Movie 1). We decided to use this classic assay as a test case to measure timing events at the single cell level.

We first implemented a canonical strategy for automated annotation of morphological classes[7-9, 15], based on object detection, multivariate feature extraction, and supervised machine learning (Fig. 1c). We used local adaptive thresholding[7], followed by a watershed split-and-merge segmentation error correction[16] to detect individual cells at an accuracy of 95.7% (n = 1876 objects; 2.6% over-segmented (falsely cut objects); 1.7% under-segmented (falsely merged objects)). A set of 186 quantitative features[17,

[18] (Supplementary Table 1, Supplementary Fig. 1) describing texture and shape was then calculated for each object. Next, a support vector machine classifier[19] was trained for the discrimination of 8 different object morphologies (Fig. 1b; interphase, six different mitotic stages, and apoptosis). These classes were defined by manual annotation of 28 to 195 example objects. The match between human and computer annotation was 94.6% (mean of all classes; five-fold cross-validation), ranging between 75.0% for the early anaphase class, and 99.0% for interphase class (Fig. 1d). This performance is similar to a number of previously reported supervised machine learning applications [7, 9, 11, 20]. Next, individual cells were tracked over time by a nearest-neighbor algorithm that supports trajectory splitting (e.g., cell division) and merging (e.g., cell-to-cell fusion). The automated tracking matched 99.8% of the human annotated object-over-frame connections (n = 1942), again comparable to the performance of previous studies on cell tracking[11, 21].

The overall accuracy of the individual computational steps appears high. However, considering >500 frames per cell trajectory for our time-resolved datasets, almost no error-free trajectories were obtained by this approach (Fig. 1e, Supplementary Movie 2).

Detecting scarce events in long-term movies

Mitotic events are scarce in comparison to the much longer duration of interphase (Fig. 1e). To improve the sensitivity for mitotic stage annotation, we automatically selected mitotic events based on a morphology class sequence motif of prophase-prometaphase. This yielded a sub-graph highly enriched for mitotic events (Fig. 2a; Supplementary Movie 3; 81.5% of all mitotic events were automatically extracted; n = 294 mitotic events in three movies). This set of trajectories contained 2.1% misclassifications per object (*a posteriori* compared with human annotation).

Untrained biological users may annotate the classifier training set less reliably. To test the sensitivity of the support vector machine towards annotation errors, we randomized the labels on an increasing fraction of training objects, and measured the overall classification accuracy (Supplementary Fig. 2). Surprisingly, randomization of the labels on 50% of the training objects reduced the overall annotation accuracy only

slightly below 90%. This demonstrates that classification by support vector machine is relatively insensitive to annotation errors.

Hidden Markov model for time-lapse imaging

Single object-based machine learning and classification does not take the temporal context into account. However, objects with ambiguous morphologies occur within a typical context of preceding and following morphologies, which could help to derive correct annotation. This could be particularly relevant for gradual morphology changes at stage transitions, where single object-based classification is relatively inaccurate (e.g., interphase - prophase - interphase - interphase - prophase - prometaphase, see Fig. 2b or Supplementary Movie 3)

We reasoned that taking the history of a cell into account might provide a means to correct for such noise at stage transitions, as well as confusion between closely related morphology classes. We assumed that the true state of a cell at a given time point (the mitotic stage in this assay) is not known, but that it correlates with an observed state (the morphology class prediction probabilities). We further assumed that the progression to the next state entirely depends on a given present state. This fulfils the criteria for a hidden Markov model, which can be used for error correction in time-resolved data[22].

We built a model with five components: 1) hidden states, representing the true morphology classes (for example, mitotic stages), 2) observed states (the class prediction probability vectors of the support vector machine), 3) probabilities of hidden state transitions, 4) observation probabilities, and 5) initial probabilities of hidden states. All elements of this model were computationally derived from the data without further user interaction. The hidden states were defined by the initial class annotation, as described above (Fig. 1b). The observed states were derived from the support vector machine as a vector of class prediction probabilities for each time point. The hidden state probabilities were initialized at the first time point by the support vector machine predictions. Transition probabilities between hidden states were calculated based on the support vector machine prediction probabilities of all cellular trajectories per experimental condition (Fig. 2c, d), and the observation probabilities between hidden and observed states were estimated based on the

confusion matrix of the support vector machine. We derived the overall maximum likelihood path for the progression through mitosis by the Viterbi algorithm[23] (thick black line in Fig. 2e). This increased the overall per-object accuracy to 99.0%. Iterative learning of transition probabilities by the expectation-maximization algorithm[24, 25] did not improve prediction accuracy (98.1% after five iterations). We suspected that the confusion matrix overestimates observation probabilities, as classes that are difficult to discriminate (prophase and early anaphase) were over-represented in the annotation data. We therefore tested the performance of temporal error correction with lower error rates in the observation probabilities (0.1% for all transitions). Indeed, this eliminated noise at state transitions and corrected single frames of misclassified objects even more efficiently, yielding overall accuracy of 99.4% per object, and 91% completely error-free trajectories (n = 100 trajectories; 4,000 objects; Fig. 2f, Supplementary Fig. 3; Supplementary Movie 4).

We next tested if incorporation of *a priori* biological knowledge on state transitions further increases the annotation accuracy. Specifically, we constrained the state transition graph to the forward direction of three consecutive classes, and defined apoptosis as a terminal state (Supplementary Fig. 4a, b). The probability matrix for constrained state transitions improved the error correction performance of the hidden Markov model to 99.7% per object, yielding 94% completely error-free trajectory annotations (n = 100 trajectories; 4,000 objects; Supplementary Fig. 4c).

Temporal error correction by the hidden Markov model is expected to depend on good estimates of the predicted morphology classes. We therefore investigated the robustness of temporal error correction towards simulated classification noise. We randomized the class prediction probability vectors of an increasing fraction of objects, then learnt the hidden Markov model on the noisy trajectories, and applied it to correct classification errors (Supplementary Fig. 5). Comparison with manually annotated data demonstrated that the hidden Markov-based error correction improved the overall accuracy at all noise levels.

We also tested if the temporal error correction was sensitive to changes in the time-lapse interval by generating trajectories sampled to every 2nd up to every 6th time point (Supplementary Fig. 6). Comparison with the manually annotated labels showed that the hidden Markov model increased the overall annotation accuracy at all sampling intervals.

In conclusion, hidden Markov modeling provides a robust and efficient means to eliminate misclassifications and noise at morphology state transitions. The combination of mitotic event selection and hidden Markov error correction reduced the per-object error rate about 10-fold below single time point-based classification.

Generic strategy for annotation of cellular dynamics

We next used our tools for other assays and fluorescent markers. We were particularly interested in simultaneous analysis of multiple markers in the same cell, for example, to address temporal coordination of mitotic processes. We defined cytoplasmic areas based on their relative position to the chromatin marker, using non-overlapping region growing of the contours derived from the chromatin channel (Supplementary Fig. 7a, b). While this may be less precise than segmenting in the secondary channel, it proved to be robust over many different assays and was insensitive to temporal dynamics (see Fig. 3, 4, and below). Tracking results of the primary channel were applied to the secondary channel, and all subsequent analysis of temporal dynamics was performed independently for primary and secondary channels, as outlined above (see Supplementary Fig. 7c for workflow).

We first applied our methods to movies from cells expressing mEGFP-α-tubulin to annotate mitotic spindle assembly and disassembly (Fig. 3a and Supplementary Movie 5), and to movies from cells expressing GalT-EGFP to study mitotic breakdown and reassembly of the Golgi apparatus (Fig. 3b, and Supplementary Movie 6). We trained classifiers for six (α-tubulin), or five (GalT) distinct morphology classes. The mean accuracy of object class predictions was 96.5% for mEGFP-α-tubulin, and 97.3% for GalT-EGFP (5-fold cross-validation, computational versus visual scoring). This yielded 55% (α-tubulin), or 38% (GalT) completely error-free trajectories. By hidden Markov model error correction, the accuracy increased to 89% completely error-free trajectories for α-tubulin (Fig. 3d and Supplementary Movie 7), and 90% for GalT (Fig. 3e and Supplementary Movie 8; n = 100 for both assays; corresponding H2B-mCherry annotations are shown in Fig. 3g, h).

To apply our methods to non-mitotic cellular dynamics, we next annotated the timing of S-phase progression. We imaged a HeLa cell line stably expressing H2B-mCherry and EGFP-PCNA, a marker for DNA replication foci, which visualizes a

characteristic pattern of morphology changes during S-phase progression (Fig. 3c and Supplementary Movie 9). We trained classifiers for six distinct PCNA morphology classes, and established a hidden Markov model for error correction. This yielded 98.2% correctly annotated objects and 90% completely error-free trajectories (n = 100 trajectories containing 15,000 objects; Fig. 3f and Supplementary Movie 10, see Fig. 3i for H2B annotations of same cells). The high performance in this diverse set of assays demonstrates a generic applicability of our computational methods.

Quantitative phenotyping and kinetic measurements

Our methods were designed for the detection of timing phenotypes. We therefore established perturbation conditions that are known to delay or shorten particular stages of mitosis. First, we used the microtubule-depolymerizing drug Nocodazol, which arrests cells in prometaphase by permanent activation of the spindle checkpoint (Fig. 4a; Supplementary Movie 11). This was reliably detected by our computational tools (96.2% completely error-free annotated trajectories, n = 154; Fig. 4b).

Next, we depleted the essential spindle checkpoint component *Mad2* by RNAi, which is known to accelerate the timing from mitotic entry until anaphase onset in HeLa cells by about two-fold[26] (Fig. 4a; Supplementary Movie 12). We evaluated the accuracy of automated timing measurements, scoring the time from prometaphase until anaphase onset based on the chromatin marker (cells that did not segregate chromosomes were omitted). Automated measurements of 47.2 ± 20.0 min (mean ± s.d.; n = 195) in control cells did not significantly differ from manual annotation of the same dataset (48.5 ± 18.0 min; two-sided Mann-Whitney-Wilcoxon test: p = 0.12). Automated timing measurements in *Mad2* RNAi cells demonstrated mitotic acceleration (13.0 ± 3.6 min), again matching well measurements by manual annotation (12.4 ± 3.4 min; two-sided Mann-Whitney-Wilcoxon test: p = 0.23). As expected from the known biological function of *Mad2*, the mitotic acceleration in *Mad2* RNAi cells was mainly due to a shortened metaphase stage (1.6 ± 1.1 min in *Mad2* RNAi cells; 36.5 ± 16.6 min in control; Fig. 4b).

Simultaneous measurements of morphological dynamics and the state of regulatory factors provide a powerful approach for mechanistic dissection of perturbation phenotypes. Here, we combined the annotation of mitotic stages with kinetic

measurements of Securin degradation, which is required for anaphase initiation[27] (Fig. 4a; Supplementary Movies 11-13). In the normalized degradation kinetic profiles (Fig. 4c), we found that the Securin-mEGFP degradation in control cells initiated briefly before anaphase (compare Fig. 4b and c), consistent with spindle checkpoint inactivation at this stage. In nocodazol-arrested cells, almost Securin-mEGFP remained stable within the measurement period of 138 min, consistent with an efficient and permanent activation of the spindle checkpoint. Securin-mEGFP degradation in *Mad2* RNAi cells initiated directly after mitotic entry, at a stage where chromosomes were still in prometaphase configuration, indicating that the anaphase-promoting complex was activated before complete chromosome congression, as expected for a compromised spindle checkpoint function. In conclusion, these experiments demonstrate accurate timing phenotype annotation in RNAi- and drug-perturbed cells.

RNAi screen for mitotic exit regulators

To test the sensitivity and performance of our computational methods in a high-throughput application, we performed a screen for regulators of mitotic exit. Specifically, we aimed to identify regulators of post-anaphase stages of mitosis, for which RNAi phenotypes have not been reported so far. Mitotic exit control is well understood in budding yeast, yet it is unclear if homologues of the yeast factors also control mitotic exit in higher eukaryotes[28]. We therefore designed a library of 283 siRNA targeting 93 candidate regulators, including all known human genes with homology to budding yeast mitotic exit regulators and some additional genes known to be involved in mitotic regulation (see Supplementary Table 2). As an assay for mitotic exit timing, we scored the timing from anaphase onset, based on the chromatin marker H2B-mRFP, until postmitotic nuclear envelope reassembly, based on the nuclear import substrate IBB-EGFP (Fig. 5a; Supplementary Movie 14).

For solid-state transfection of siRNAs into HeLa cells, we used a high-density transfection array with 300 spots of different siRNA transfection solutions printed to the glass surface of a chambered coverslip[7]. We seeded the cells onto this array and 20 h later started parallel imaging of 108 movies per experiment, for a total duration of 46 h and with 3.7 min time resolution. We automatically annotated the mean

mitotic exit timing per experimental condition within the 1.6 TeraByte data containing 646'754 images and 16'314 mitotic events. Only one siRNA delayed mitotic exit above a z-score threshold of 3.0 (Fig. 5b, Supplementary Fig. 8a; 6.8 ± 2.0 min mean ± sd; n = 50 mitotic events). This oligo depleted the anaphase promoting complex co-activator *Cdc20*, as validated by Western Blotting (Supplementary Fig. 8b). The specificity of the phenotype was confirmed in two additional replicas with standard liquid phase transfection, and with an additional siRNA (Fig. 5c).

To test if *Cdc20* was required for other cellular reorganization processes during mitotic exit, we assayed chromosome decondensation and mitotic spindle disassembly. High resolution confocal time-lapse imaging of cells co-expressing H2B-mCherry and mEGFP-α-tubulin (Fig. 5d, e, and Supplementary Movies 15 and 16) showed that 100% (n = 30) of control cells started chromosome decondensation within 14 minutes after chromosome segregation, whereas only 54% (n = 36) did so after Cdc20 depletion. 31% (n = 36) of Cdc20-depleted cells started kinetochore fiber spindle disassembly 7 minutes post anaphase onset, in contrast to 87% (n = 30) in control cells. These data suggest a requirement of Cdc20 for various cellular processes leading to postmitotic reassembly of interphase cells. This is unexpected given that Cdc20 has so far been thought to act mainly at pre-anaphase stages of mitosis, and it has not been noticed in previous phenotypic analysis of Cdc20 RNAi cells[29].

Discussion

In this study, we present *CellCognition,* a computational framework for time-resolved single-cell assays in high-throughput imaging applications. Building on existing machine learning methodologies, the design of a generic workflow for annotation of morphological dynamics faced two main challenges. First, the classification noise at continuous morphology stage transitions impairs coherent trajectory annotation. Second, some biologically distinct classes appear morphologically similar, which leads to high classification confusion. By hidden Markov modeling, our methods efficiently correct both types of errors based on the temporal context. The hidden Markov models are learned individually for each experimental condition, without any

human supervision. This allows the software to automatically adapt the error correction scheme to phenotypic deviations.

Biological *a priori* knowledge to suppress state transitions that are assumed to be impossible can also be used to improve annotation accuracy[11, 12]. Such explicit error correction schemes cannot be applied to new markers or assay systems without adaptation, and they may not apply to phenotypes with potentially altered stage progression. We find that the gain in accuracy by biological *a priori* constraints on the temporal progression is only minor. Our hidden Markov implementation models time series analysis in a high dimensional feature space with an intrinsic class-discriminant dimensionality reduction. This preserves context-specific structures, in contrast to principle component analysis as used in [12, 30], which may explain the large gain in accuracy compared to the previous implementations (see Supplementary Tables 3 and 4). Compared to the models by [11, 12], our model is the only one able to handle arbitrary relationships between phenotypic cell classes, providing a powerful and generic solution for time-resolved cellular phenotyping.

Using a variety of different structural markers, we demonstrate that our analysis methods can be used for a broad range of biological assays. We are not aware of any constraints that would preclude the use of our methods in other biological context, e.g., apoptosis or cellular differentiation. However, the texture and shape features implemented into our software do not enable assays relying on absolute object counts, for example in centrosome duplication assays. Also, assays scoring rapid intracellular dynamics would require integration of motion feature extraction methods into our published software source code.

Supervised machine learning as in this study requires user-defined morphology classes. It is therefore not possible to detect aberrant phenotypic morphologies that do not occur in the control conditions used for annotating the classifier training set. This limitation may be overcome in future studies by implementing unsupervised machine learning methods for the analysis of image time series.

In conclusion, we present a powerful computational strategy for high-throughput phenotyping of single cell dynamics. Our methods are integrated into the platform-independent software package *CellCognition,* with graphical user interface and supporting high-throughput batch processing on computer clusters. *CellCognition* is

published as open source software (current version 1.0.7 in Supplementary Software), along with high quality reference image data on http://www.cellcognition.org/. With the increased availability of live cell screening microscopes, we anticipate that time-resolved imaging assays will soon dominate a significant fraction of high content screening and systems biology applications.

Methods

Cell culture, RNAi and cell transfection arrays, and Western Blotting

HeLa 'Kyoto' cells were cultured in DMEM (Gibco) supplemented with 10% fetal calf serum (PAA Laboratories) and 1% Penicillin/Streptomycin (Invitrogen), and grown on LabTek chambered coverslips (Nunc) for live microscopy. All experiments were performed with monoclonal cell lines stably expressing combinations of the fluorescent markers as indicated throughout the manuscript. Live imaging was in DMEM containing 10% fetal calf serum and 1% Penicillin/Streptomycin, but without phenolred and riboflavin to reduce autofluorescence of the medium. Cell transfection arrays for live cell RNAi screening were produced and used as described in [7, 31]. All other RNAi interference experiments were performed using single RNAi duplexes (Qiagen) that were liquid phase transfected with either Oligofectamine (Invitrogen) or HiPerfect (Qiagen) as transfection reagent according to the manufacturers protocols. Final siRNA concentrations were 50 nM for Oligofectamine or 10 nM for HiPerfect. Cdc20 siRNA validation oligos were obtained from Qiagen with the following target sequences: AACCTTGTGGATTGGAGTTCT (Cdc20_1), CACCACCATGATGTTCGGGTA (Cdc20_2). Total HeLa cell lysates for SDS/Page analysis were prepared according to standard procedures. Rabbit-anti-human Cdc20 antibody (diluted 1:5000) was from Bethyl laboratories.

Fluorescent reporter plasmid constructs

For efficient generation of cell lines stably expressing fluorescently tagged marker proteins, the genes were subcloned into pIRES-puro2 and pIRES-neo3 vectors (Clontech) that allow expression of resistance genes and tagged proteins from a single transcript. For details on the plasmids, see Supplementary Table 5.

Stably expressing cell lines

For generation of stably expressing cell lines, HeLa Kyoto cells were first transiently transfected using FuGENE6 (Roche) following the manufacturer's instructions. Cells were then seeded to clonal density and grown in culture medium supplemented with

500 µg/ml Geneticin (Invitrogen) and/or 0.5µg/ml Puromycin (Merck/Calbiochem) for three weeks. Individual colonies of resistant cells were picked, expanded, and validated for homogeneous expression levels and correct sub-cellular localization of fluorescent proteins. All cell lines used in this study had a normal morphology and cell cycle progression as compared to the maternal line. For details on the stable cell lines, see Supplementary Table 6.

Live microscopy

Automated microscopy with reflection-based laser auto focus was performed on a Molecular Devices ImageXpressMicro screening microscope equipped with 10x 0.5 N.A. and 20x 0.8. N.A. S Fluor dry objectives (Nikon), and recorded as 2D time-series. The microscope was controlled by in-house developed Metamorph macros (*PlateScan* software package, available at http://www.bc.biol.ethz.ch/people/groups/gerlichd). Cells were maintained in a microscope stage incubator at 37 ºC in humidified atmosphere of 5% CO_2 throughout the entire experiment. We adjusted illumination conditions such that cell death rate was below 5% in untreated control cells[14]. Confocal microscopy was performed on a customized Zeiss LSM 510 Axiovert microscope using a 63x, 1.4 N.A. Oil Plan-Apochromat objective (Zeiss). The microscope was equipped with piezo focus drives (piezosystemjena), custom-designed filters (Chroma), and EMBL incubation chamber (European Molecular Biology Laboratory), providing a humidified atmosphere at 37°C with 5% CO_2.

Image analysis

Cell nuclei were detected by local adaptive thresholding[7], which is robust towards variable expression levels of the fluorescent chromatin marker in individual cells, and inhomogeneous illumination typical for wide-field microscopy. To improve segmentation accuracy, we implemented a split-and-merge approach. First, we split objects containing directly adjacent nuclei, using watershed transformation based on object contours. In some cases, this incorrectly split single objects. Thus we implemented object merging based on *a priori* definition of size and circularity criteria[16]. Regions of interest for the secondary marker were derived by region

growing of the chromatin segmentation to a fixed size, but constrained by regions of neighboring cells. Depending on the marker, we defined nuclear, cytoplasmic, or total cellular areas. This segmentation strategy turned out to be more precise than direct segmentation in the secondary channel, as many secondary markers dramatically changed in intensity levels or pattern throughout the time course of the experiment. Texture and shape features[17, 18] (see Supplementary Table 1) were extracted from the two channels and all regions individually. For secondary region classification, only texture features were used since the shape information only depended on the chromatin segmentation.

Samples for morphology classes were manually annotated on the original images overlaid with the segmentation contours, to establish a training set for supervised classification. Support vector classification with radial-based kernel and probability estimates[32] was then computed with libSVM. Classification performance was calculated with five-fold cross-validation. Samples and feature plots for all classifiers used in this study can be accessed online through a web browser interface (see resource section).

Tracking cells over time was achieved by a constrained nearest-neighbor approach based on the Euclidian distance between objects[21]. Since tracks might be lost due to segmentation errors or migration of cells into the field of view the tracking must be able to create new tracks for all objects without incoming edges. To detect cell division events, or potential cell-to-cell fusion events, the tracking algorithm needed to support both splitting and merging. This yielded a hierarchical directed graph of isolated tracks for each cell over time. Tracking errors resulted mostly from segmentation errors and lead to wrong edges between the cell tracks. Secondary objects are tracked indirectly by the primary objects associated with them. Mitotic motifs were detected in this graph structure by the transition from prophase to prometaphase. Sub-graphs (mitotic trajectories) were extracted by considering a pre-defined number of frames preceding and following this mitotic motif, resulting in synchronized mitotic trajectories of equal length, as displayed in the figures.

Hidden Markov model and statistical analysis

A hidden Markov model λ is defined as $\lambda = (X, A, Y, B, \pi)$, where X is the set of hidden states, A is a matrix of transition probabilities from one state to another, Y is the set of observable variables per state, B is a matrix of observation probabilities storing the probability of observation k being produced from state j (also termed emission or observation probability), and π is a vector of probabilities of the initial state (first time point) in the trajectory.

The hidden states X are the true cellular stages expressed by the class labels (8 classes for fluorescent H2B, see Fig. 1b). The hidden Markov model is learned by maximum likelihood estimates from the aligned trajectories of estimated prediction probabilities of the support vector machine, which is a three-dimensional array over trajectories, time points, and classes. Transition probabilities A are learned from the prediction probabilities along the trajectories on the underlying graph structure. In a free model all transitions between morphology classes were allowed (Fig. 2c). In a constrained model some transitions were suppressed based on biological *a priori* knowledge (transition probabilities were set to 0 for edges missing in the graph; Supplementary Fig. 6a). For the initial probabilities π the prediction probabilities of all trajectories at the first time point are considered. The observables Y are the class labels. The observation probabilities were either set to an error rate of 0.1%, or derived from the confusion matrix of support vector machine training.

Using the Viterbi algorithm, each trajectory was corrected based on its sequence of support vector machine probability estimates and the trained hidden Markov model for a given experimental condition (decode problem). This correction scheme was calculated individually for each marker and experimental perturbation condition.

To detect the onset of nuclear envelope breakdown and nuclear envelope reformation the time series of IBB-EGFP intensity ratios of individual cells were analyzed. We computed the ratio by a shrunken area of the chromatin object and a ring around. The onset was defined as the time point where the ratio was 1.5 fold increased above the ratio at the time point of chromosome segregation.

For data normalization of Fig. 5b we computed the z-scores of mitotic exit timing for all siRNA conditions (mean over all values of one condition). The z-score was computed by the mean of negative controls and the standard deviation of the entire data set.

Implementation and performance

The basic image processing was implemented in C++ using VIGRA (http://hci.iwr.uni-heidelberg.de/vigra) and in house-developed extensions. The C++ code was then wrapped for Python, which is a programming language particularly well suited for handling complex data structures and integration of external modules. Statistical analysis and plots were performed with the R-project (http://www.r-project.org). The entire software package is platform-independent, and was compiled for Mac OS X and Windows environments.

Computation of each movie required 4-20 s per image and processor node, consuming 500-1500 MB RAM, depending of the number of frames and objects per frame. As an example, a single movie of Fig. 2 with 206 frames and ~37,000 objects required a total processing time of 34 min on a single processor node. For high-throughput analysis, we implemented distributed computing on a farm of desktop computers (four MacPro 2.2GHz, 28 cores total).

Software and data resources

CecogAnalyzer is a platform-independent graphical user interface, which covers the entire workflow presented in this paper. The software is publicly available in source and binary versions and was tested on MacOS X Leopard/SnowLeopard and Windows XP/7. We use a subversion repository for concurrent software development by remote contributors, and tracking of software changes. Our website is based on the project management tool TRAC (http://trac.edgewall.org/), which allows co-ordination of this open-source project by milestones, tickets, wiki pages and browsing of code changes.

The software, a subset of raw images presented here, the classifiers and parameters used for generating the figures are available online at http://www.cellcognition.org. The classifiers data sets consisting of annotated samples and extracted features are interactively visualized by Adobe Flex and can be browsed online at http://flex.cellcognition.org.

The MetaMorph journals developed for fast and robust acquisition of the time-lapse experiments presented here are available on our group website: http://www.bc.biol.ethz.ch/people/groups/gerlichd.

Acknowledgments

The authors thank Christian Conrad and Wolfram H. Gerlich for critical comments on the manuscript, F. Oliver Gathmann for helpful discussions about software engineering, Nicolas Graf for outstanding IT support, and Gabor Csucs, the ETHZ Light Microscopy Center (LMC), and the ETHZ RNAi Screening Center (RISC) for excellent technical support. We thank Jack Rohrer (University Zurich, Switzerland) for providing GalT-EGFP plasmid, and Jonathon Pines (Gurdon Institute, Cambridge, UK) for providing Securin-EYFP and cyclin B1-EGFP plasmids, and Katja Beck and Ulrike Kutay for providing images of cells expressing fluorescent α-tubulin and H2B, and Qing Zhong for generating the plot for Supplementary Fig. 1. Work in the Gerlich laboratory is supported by Swiss National Science Foundation (SNF) research grant 3100A0-114120, SNF ProDoc grant PDFMP3_124904, a European Young Investigator (EURYI) award of the European Science Foundation, an EMBO YIP fellowship, and a MBL Summer Research Fellowship to D.W.G., an ETH TH grant, a grant by the UBS foundation, a Roche Ph.D. fellowship to M.H.A.S, and a Mueller fellowship of the Molecular Life Sciences Ph.D. program Zurich to M.H. M.H. and M.H.A.S are fellows of the Zurich Ph.D. Program in Molecular Life Sciences. B.F. was supported by European Commission's seventh framework program project Cancer Pathways. Work in the Ellenberg laboratory is supported by a European Commission grant within the Mitocheck consortium (LSHG-CT-2004-503464). Work in the Peter laboratory is supported by the ETHZ, Oncosuisse, SystemsX.ch (LiverX) and the SNF.

Author contributions

M.H. designed the image analysis workflow, implemented the software, performed imaging experiments, and prepared the paper manuscript. M.H.A.S. established stable cell lines, performed most imaging and all RNAi experiments. B.F. designed and implemented the hidden Markov model. T.W. designed parts of the feature extraction and of the image analysis workflow. B.N. and J.E. generated the siRNA cell transfection array. M.H.O and M.P. established live imaging of EGFP-PCNA. D.W.G. designed assays and the general strategy for image processing, and wrote the paper.

Figure legends

Figure 1. Supervised machine learning and classification of morphologies. (**a**) Confocal images of live HeLa cells stably expressing a chromatin marker (H2B-mCherry), together with GalT-EGFP to visualize the Golgi apparatus, with mEGFP-α-tubulin, or with the replication factory marker EGFP-PCNA. The images show maximum intensity projections of five z-sections. (**b**) Live imaging of HeLa cells expressing H2B-mCherry at different cell cycle stages, or apoptosis (2D time series imaged with wide field epifluorescence 20x dry objective, see Supplementary Movie 1). The color scheme relates to H2B-mCherry morphology classifications of subsequent figures. (**c**) Object detection (contours) and classification (colors) of cellular morphologies corresponding to predefined mitotic stages as shown in (b). Cells were tracked over time (arrows). See Supplementary Movie 2. (**d**) Classification performance of support vector machines with radial basis functions. The confusion matrix displays the matching of human versus machine annotation, identical annotations are on the diagonal. (**e**) Automated annotation of cell trajectories over time by the workflow shown in (c). 80 randomly selected trajectories (rows) over 40 time frames (columns) are displayed (time-lapse: 4.6 min). Colors refer to morphology classes as labeled in (b). Tick marks indicate sampled time points. Mitotic events are rare, and the trajectories contain many single frames of mitotic annotations, likely due to classification errors. Scale bars: 10 μm.

Figure 2. Hidden Markov modeling of progression through morphology stages. (**a**) Automated extraction of mitotic events. Cells were synchronized *in silico* to the prophase - prometaphase transition. The plot displays a random selection of 100 mitotic events (from a total set of 172 mitotic events out of 8 movies; time-lapse: 4.6 min; see Supplementary Movie 1). Predicted morphology classes were color-labeled as in Fig. 1b. Asterisks: classification errors. Black frame indicates region of interest displayed by contour overlays on image data. For complete data, see Supplementary Movie 3. (**b**) Single cell and corresponding trajectory of class labels. Asterisks: classification errors. (**c**) Graph for all possible transitions between classes. Node 0 is start node, all other nodes are color-labeled as in Fig. 1b. (**d**) Learned class transition

probabilities based on the trajectories shown in (a). Normalization of probabilities was per node. (**e**) Trellis diagram showing all class prediction estimates for the cell shown in (c). Vertical columns correspond to single time points, aligned to the images in (c). Rows correspond to morphology classes, labeled as in Fig. 1b. Probability estimates derived from the support vector machine are coded by size. The Viterbi algorithm was used to decode the overall most likely sequence (thick black line). Thin black lines indicate the most likely preceding state of a label at each given time point. (**f**) Error correction as in (e) was performed for all trajectories shown in (a). See also Supplementary Fig. 3 and Supplementary Movie 4. Scale bars: 10 μm.

Figure 3. Automated annotation of mitotic spindle and Golgi dynamics, and replication factory patterns during S-phase progression. (**a**) Live imaging of mitotic spindle dynamics of a cell expressing H2B-mCherry and mEGFP-α-tubulin; 20x objective; 4.6 min time-lapse; see Supplementary Movie 5. Automated hidden Markov model-corrected classification of spindle morphology, was color labeled as indicated. (**b**) Live imaging of mitotic Golgi dynamics in a cell line expressing H2B-mCherry and GalT-EGFP; 10x objective; 2.8 min time-lapse; see Supplementary Movie 6. Colors indicate automated hidden Markov model-corrected annotation of Golgi morphologies. (**c**) Live imaging of DNA replication factory dynamics in a cell line expressing H2B-mCherry and PCNA-EGFP; 10x objective; 5.9 min time-lapse; see Supplementary Movie 9. Colors indicate automated hidden Markov model-corrected annotation of S-phase progression based on PCNA morphology. (**d**) Automated annotation of a high-throughput imaging dataset. 100 randomly selected mitotic events were derived and *in silico* synchronized to the prophase - prometaphase transition based on the H2B-mCherry annotation (see Fig. 2). The secondary channel annotation was calculated independently from the H2B-mCherry channel, as indicated in (a). See Supplementary Movie 7. (**e**) Automated annotation of Golgi dynamics, processed as in (d). See Supplementary Movie 8. (**f**) Automated annotation of S-phase progression. Cells were *in silico* synchronized to the G1 – early S transition based on the EGFP-PCNA classification. See Supplementary Movie 10. (**g-i**) Hidden Markov model-corrected annotations of H2B-mCherry morphologies for the cells shown in (d-f). Colors label classes as in Fig. 1b. Scale bars: 10 μm.

Figure 4. Timing phenotypes and kinetic measurements. (**a**) Mitotic progression assayed by H2B-mCherry morphology, and degradation of Securin-mEGFP. Examples are shown for untreated control cell (larger region of original data shown in Supplementary Movie 13), a cell with *Mad2* RNAi-inactivated spindle checkpoint (si*Mad2*; larger region of original data shown in Supplementary Movie 12), and a cell arrested in prometaphase by a Nocodazol (Noc; larger region of original data shown in Supplementary Movie 11). Time-lapse: 2.7 min. (**b**) Automated classification of mitotic stage progression as in Fig. 2f for the three experimental conditions shown in (a). (**c**) Securin-mEGFP degradation kinetics for the same cells shown in (b). Normalization was per trajectory to the first prometaphase frame. Scale bar: 10 μm.

Figure 5. RNAi screen for mitotic exit regulators. (**a**) Assay for mitotic exit timing. Live imaging of a cell line expressing H2B-mCherry and IBB-EGFP. The timing from anaphase onset (red bar) until onset of nuclear accumulation of IBB-EGFP (green bar) was used to define mitotic exit timing (arrow). Time is in min:s. Larger region of original data shown in Supplementary Movie 14. (**b**) Mitotic exit timing in an RNAi screen for 300 different RNAi conditions. 108 movies of different siRNA transfections were recorded in parallel over 46 h, to collect the entire dataset in four experiments. Time-lapse: 3.7 min; see Supplementary Movie 14. Each point in the graph indicates the z-score for one siRNA (for calculation of z-scores, see methods). Dashed lines indicate z-score threshold, solid line indicates mean of the entire dataset. Each gene was targeted by three different siRNA oligos (For full list of oligos, see Supplementary Table 1). (**c**) Cumulative percentage of cells exiting mitosis after onset of chromosome segregation (t = 0 min). The curves represent all mitotic events from two experimental replica. Cells were transfected in liquid phase with two different siRNA targeting Cdc20, or a non-targeting oligo for control, as indicated in the legend. (**d**) Confocal time-lapse imaging of a cell stably expressing H2B-mCherry and mEGFP-α-tubulin. Time is in min:s, maximum intensity projection of five z-slices. See Supplementary Movie 15. (**e**) Confocal imaging as in (a) for a Cdc20 RNAi cell. See Supplementary Movie 16. Scale bars: 10 μm.

References

1. Conrad, C. & Gerlich, D.W. Automated microscopy for high-content RNAi screening. *J Cell Biol* **188**, 453-461 (2010).
2. Doil, C. *et al.* RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136**, 435-446 (2009).
3. Collinet, C. *et al.* Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**, 243-249 (2010).
4. Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434**, 462-469 (2005).
5. Goshima, G. *et al.* Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science* **316**, 417-421 (2007).
6. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721-727 (2010).
7. Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat Methods* **3**, 385-390 (2006).
8. Loo, L.H., Wu, L.F. & Altschuler, S.J. Image-based multivariate profiling of drug responses from single cells. *Nat Methods* **4**, 445-453 (2007).
9. Conrad, C. *et al.* Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res* **14**, 1130-1136 (2004).
10. Glory, E. & Murphy, R.F. Automated subcellular location determination and high-throughput microscopy. *Dev Cell* **12**, 7-16 (2007).
11. Harder, N. *et al.* Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time. *Genome Res* (2009).
12. Zhou, X., Li, F., Yan, J. & Wong, S.T. A novel cell segmentation method and cell phase identification using Markov model. *IEEE Trans Inf Technol Biomed* **13**, 152-157 (2009).
13. Gerlich, D. & Ellenberg, J. 4D imaging to assay complex dynamics in live specimens. *Nat Cell Biol.* **5**, S14-S19 (2003).
14. Schmitz, M.H. & Gerlich, D.W. Automated live microscopy to study mitotic gene function in fluorescent reporter cell lines. *Methods Mol Biol* **545**, 113-134 (2009).
15. Boland, M.V. & Murphy, R.F. … the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* (2001).
16. Wahlby, C., Sintorn, I.M., Erlandsson, F. & Borgefors, G. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue …. *Journal of Microscopy* (2004).
17. Walker, R. & Jackway, P. Statistical geometric features-extensions for cytological textureanalysis. *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* **2** (1996).
18. Haralick, R., Dinstein & Shanmugam Textural features for image classification. *IEEE Transactions on Systems* (1973).
19. Boser, B.E., Guyon, I. & Vapnik, V. A training algorithm for optimal margin classifiers. *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory* (1992).
20. Wang, M. *et al.* in Bioinformatics, Vol. 24 94-101 (2008).

21. Chen, X., Zhou, X. & Wong, S.T. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *Biomedical Engineering, IEEE Transactions on* **53**, 762-766 (2006).
22. Durbin, R., R. Eddy, S., Krogh, A. & Mitchison, G. Biological sequence analysis: probabilistic models of proteins and nucleic acids‡. 356 (1998).
23. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on* **13**, 260 - 269 (1967).
24. Baum, L.E., Petrie, T., Soules, G. & Weiss, N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist* **41**, 164-171 (1970).
25. Dempster, A.P., Laird, N.M. & Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B* **39**, 1-3 (1977).
26. Meraldi, P., Draviam, V.M. & Sorger, P.K. Timing and checkpoints in the regulation of mitotic progression. *Dev Cell* **7**, 45-60 (2004).
27. Hagting, A. *et al.* Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *J Cell Biol* **157**, 1125-1137 (2002).
28. Bollen, M., Gerlich, D.W. & Lesage, B. Mitotic phosphatases: from entry guards to exit guides. *Trends Cell Biol* (2009).
29. Wolthuis, R. *et al.* Cdc20 and Cks direct the spindle checkpoint-independent destruction of cyclin A. *Mol Cell* **30**, 290-302 (2008).
30. Wang, M., Zhou, X., King, R.W. & Wong, S.T. Context based mixture model for cell phase identification in automated fluorescence microscopy. *BMC Bioinformatics* **8**, 32 (2007).
31. Erfle, H. *et al.* Reverse transfection on cell arrays for high content screening microscopy. *Nat Protoc* **2**, 392-399 (2007).
32. Wu, T.F., Lin, C.J. & Weng, R.C. Probability estimates for multi-class classification by pairwise coupling. *The Journal of Machine Learning Research* (2004).
33. Snapp, E.L. *et al.* Formation of stacked ER cisternae by low affinity protein interactions. *J Cell Biol* **163**, 257-269 (2003).
34. Dultz, E. *et al.* Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells. *J Cell Biol* **180**, 857-865 (2008).
35. Schaub, B.E., Berger, B., Berger, E.G. & Rohrer, J. Transition of galactosyltransferase 1 from trans-Golgi cisterna to the trans-Golgi network is signal mediated. *Mol Biol Cell* **17**, 5153-5162 (2006).
36. Leonhardt, H. *et al.* Dynamics of DNA replication factories in living cells. *J Cell Biol* **149**, 271-280. (2000).
37. Steigemann, P. *et al.* Aurora B-mediated abscission checkpoint protects against tetraploidization. *Cell* **136**, 473-484 (2009).

Figure 1

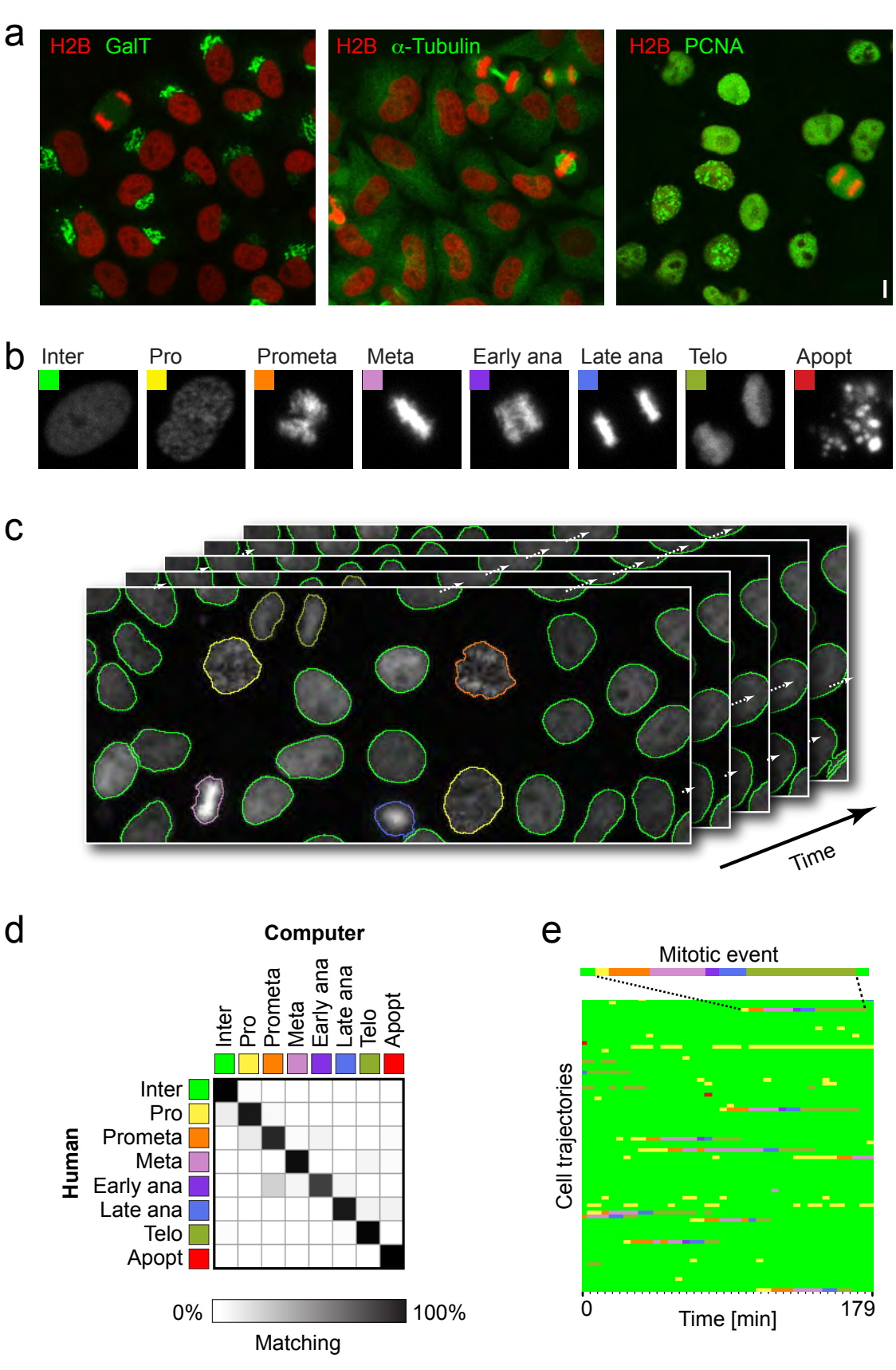

a
H2B GalT
H2B α-Tubulin
H2B PCNA
b
Inter
Pro
Prometa
Meta
Early ana
Late ana
Telo
Apopt
c
Time
d
Computer
Inter
Pro
Prometa
Meta
Early ana
Late ana
Telo
Apopt
Human
0%
100%
Matching
e
Mitotic event
Cell trajectories
0
Time [min]
179

Figure 2

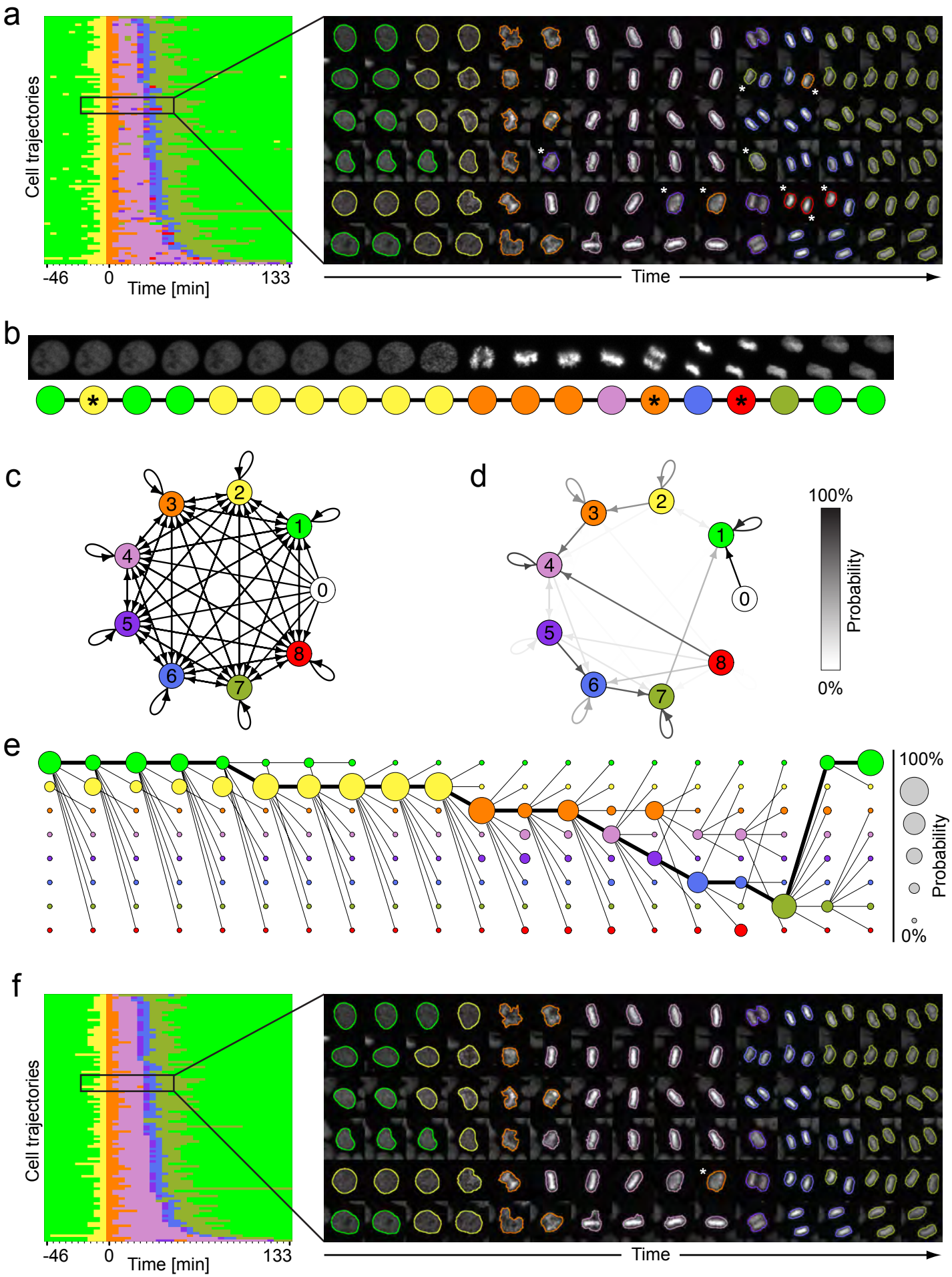

a
Cell trajectories
-46
0
Time [min]
133
Time
b
c
d
100%
Probability
0%
e
100%
Probability
0%
f
Cell trajectories
-46
0
Time [min]
133
Time

Figure 3

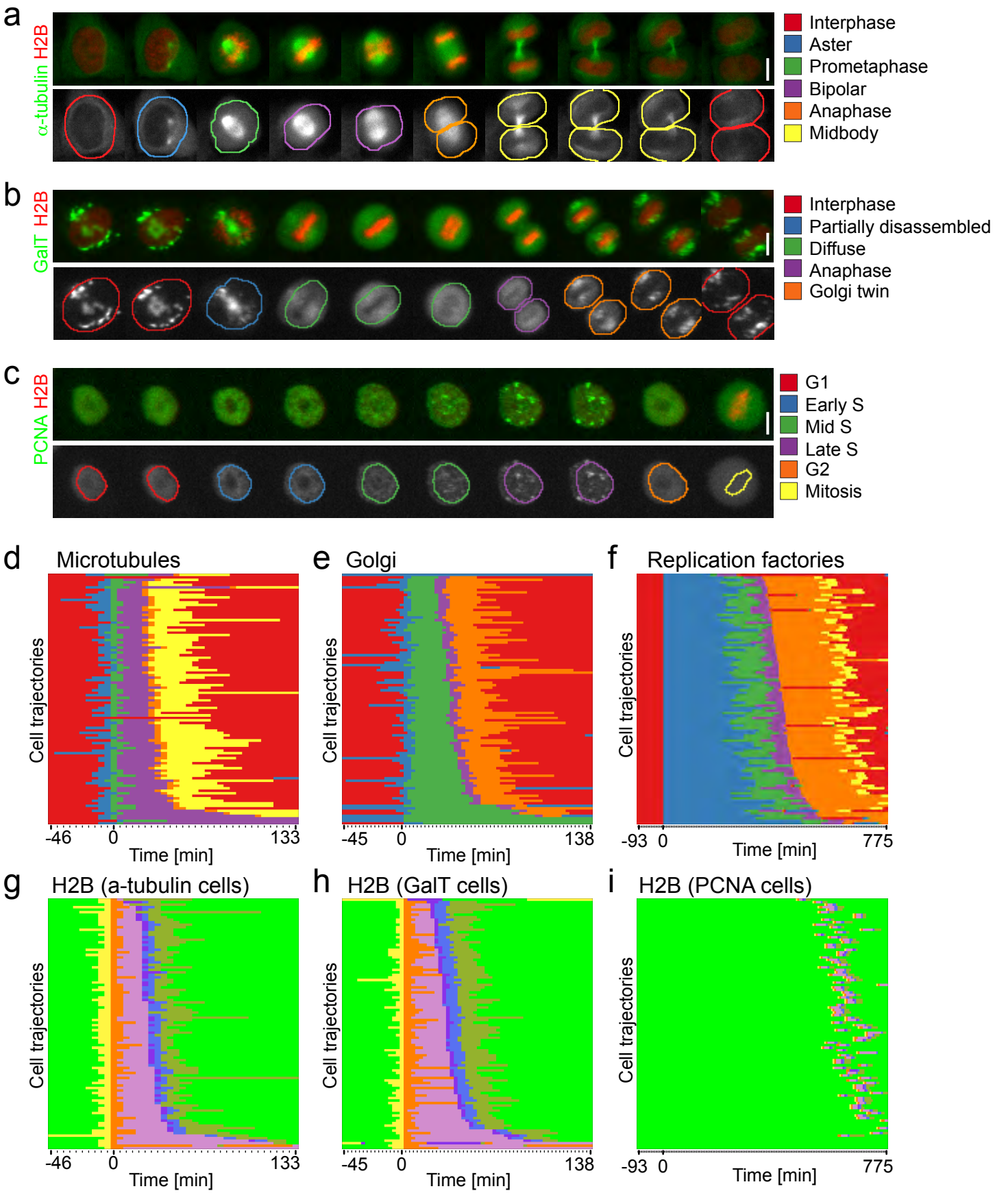

a
α-tubulin H2B
Interphase
Aster
Prometaphase
Bipolar
Anaphase
Midbody
b
GalT H2B
Interphase
Partially disassembled
Diffuse
Anaphase
Golgi twin
c
PCNA H2B
G1
Early S
Mid S
Late S
G2
Mitosis
d
Microtubules
Cell trajectories
-46
0
133
Time [min]
e
Golgi
Cell trajectories
-45
0
138
Time [min]
f
Replication factories
Cell trajectories
-93
0
775
Time [min]
g
H2B (a-tubulin cells)
Cell trajectories
-46
0
133
Time [min]
h
H2B (GalT cells)
Cell trajectories
-45
0
138
Time [min]
i
H2B (PCNA cells)
Cell trajectories
-93
0
775
Time [min]

Figure 4

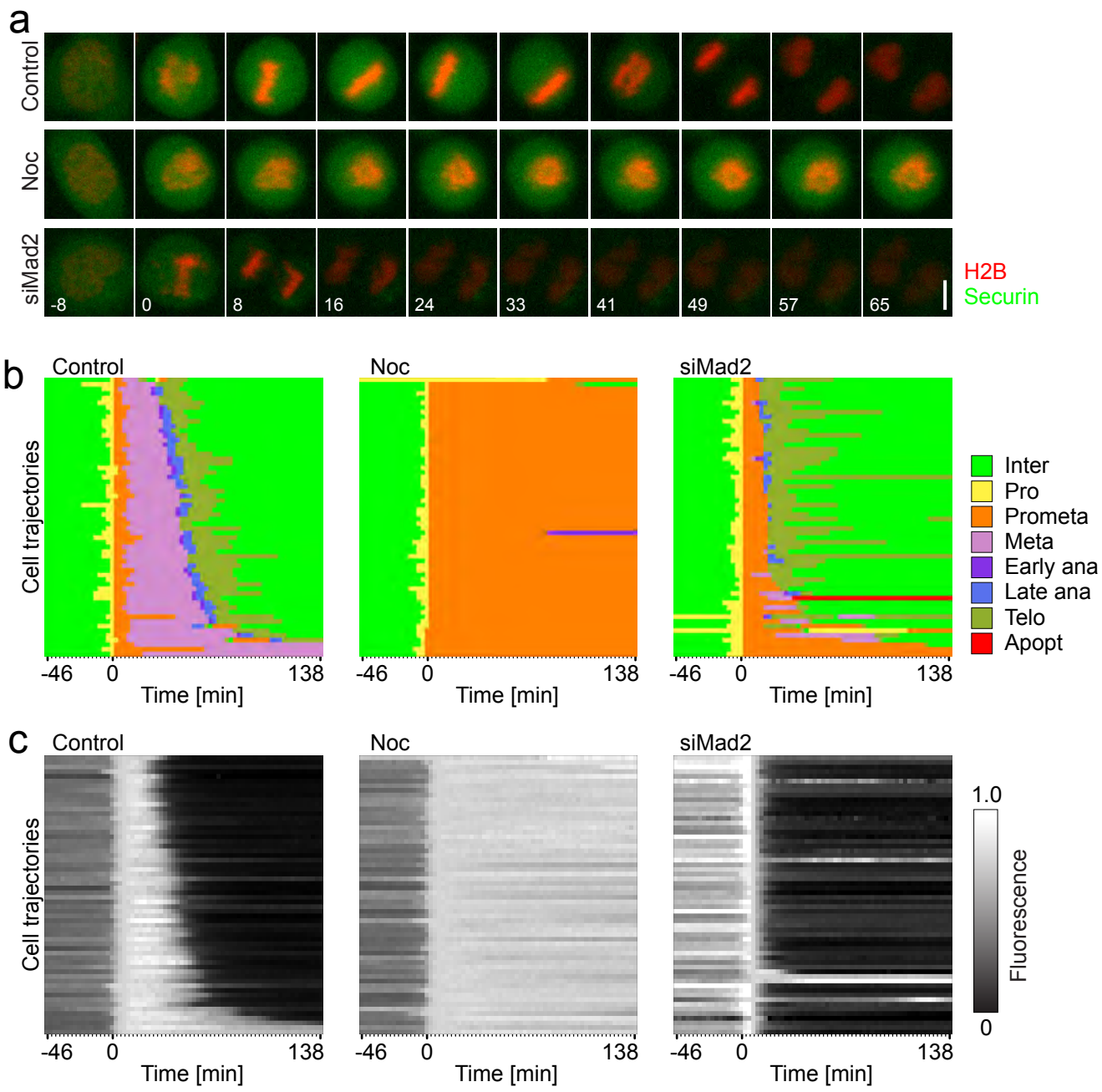
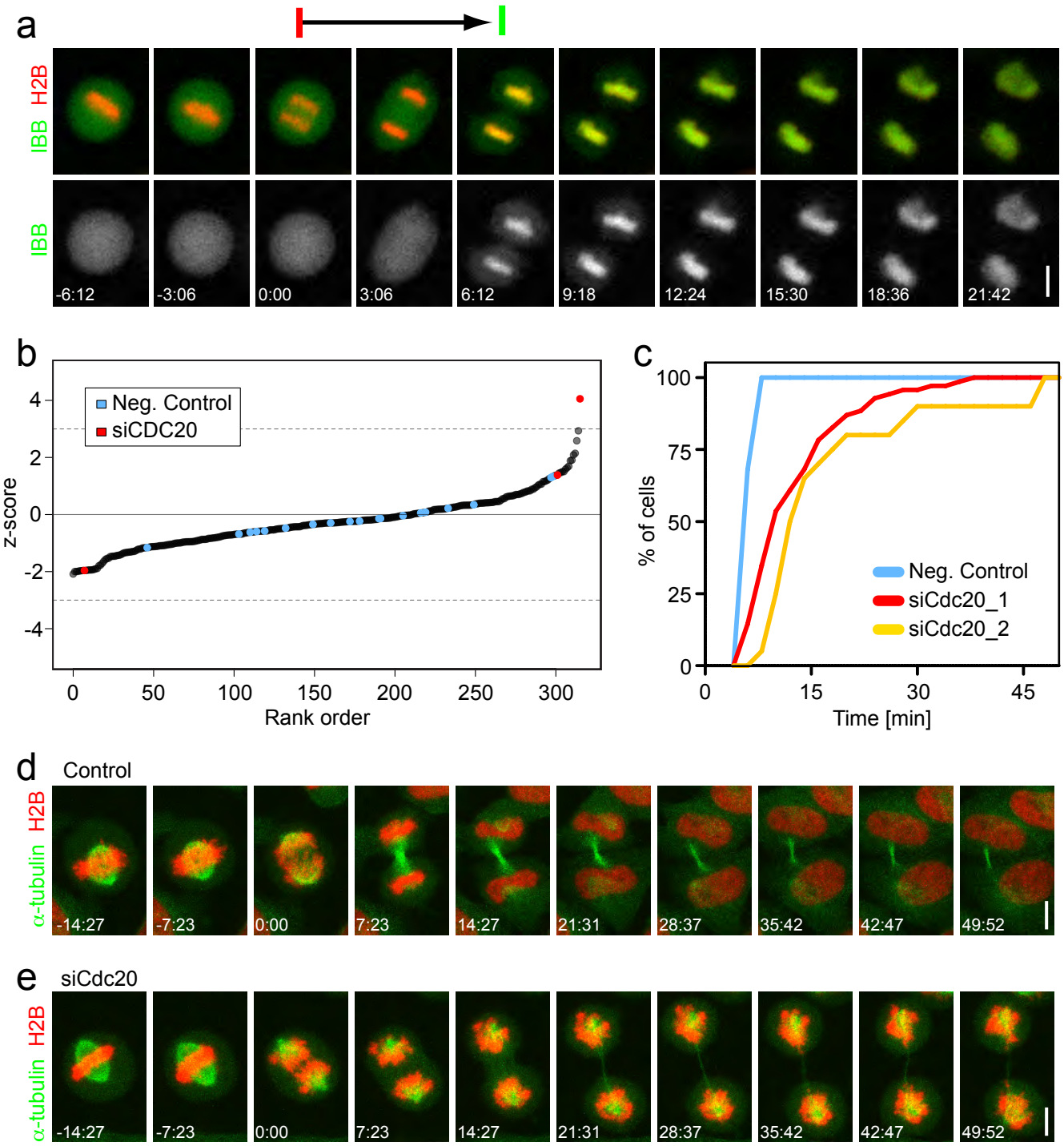
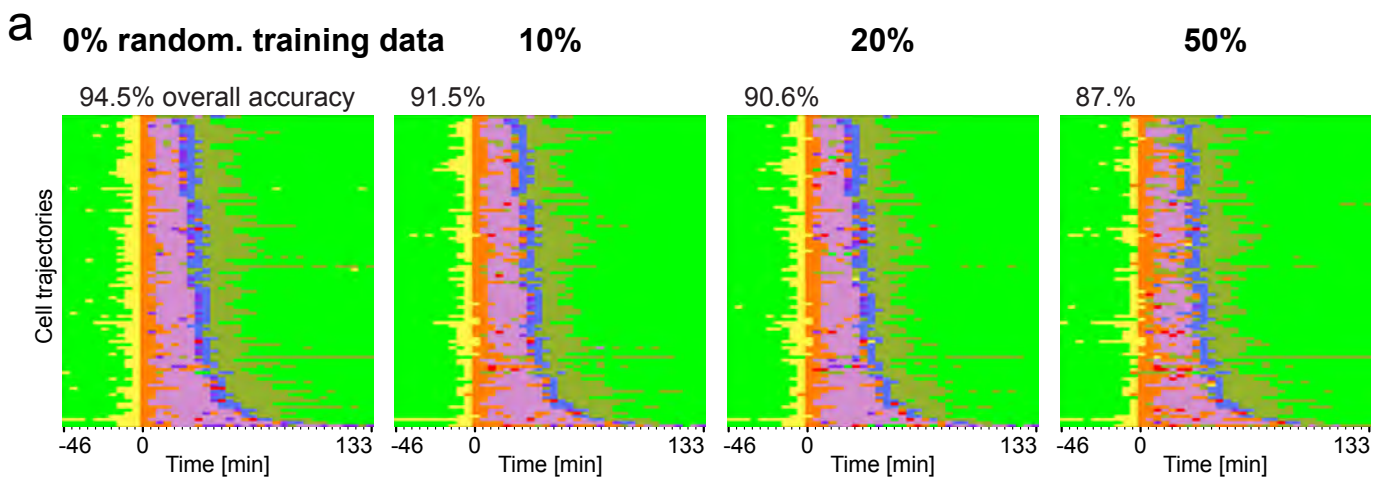
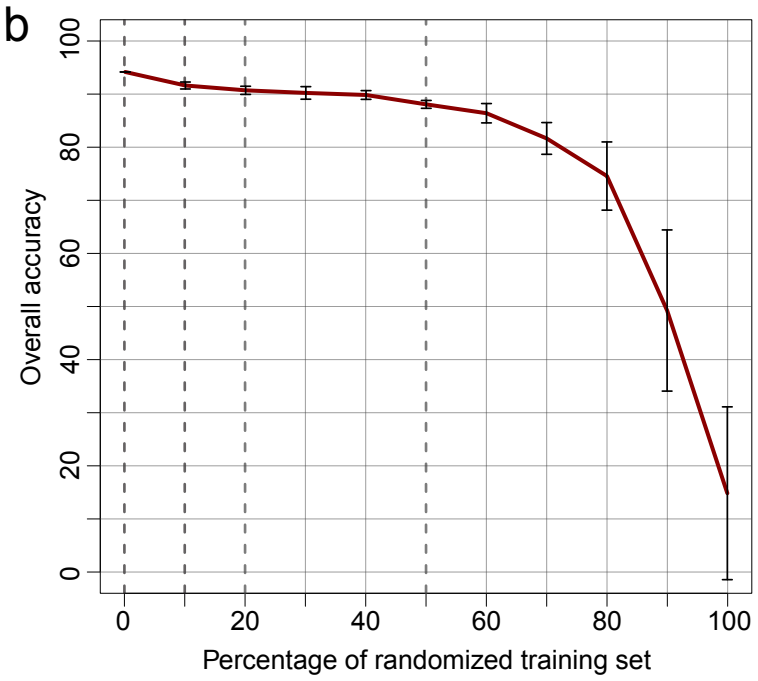

a
Control
Noc
siMad2
-8
0
8
16
24
33
41
49
57
65
H2B
Securin
b
Control
Noc
siMad2
Cell trajectories
-46
0
138
Time [min]
Inter
Pro
Prometa
Meta
Early ana
Late ana
Telo
Apopt
c
Control
Noc
siMad2
Cell trajectories
-46
0
138
Time [min]
1.0
0
Fluorescence

Figure 5


a
H2B
IBB
IBB
-6:12
-3:06
0:00
3:06
6:12
9:18
12:24
15:30
18:36
21:42
b
Neg. Control
siCDC20
z-score
4
2
0
-2
-4
0
50
100
150
200
250
300
Rank order
c
% of cells
100
75
50
25
0
0
15
30
45
Time [min]
Neg. Control
siCdc20_1
siCdc20_2
d
Control
α-tubulin
H2B
-14:27
-7:23
0:00
7:23
14:27
21:31
28:37
35:42
42:47
49:52
e
siCdc20
α-tubulin
H2B
-14:27
-7:23
0:00
7:23
14:27
21:31
28:37
35:42
42:47
49:52

Supplementary Figure 1



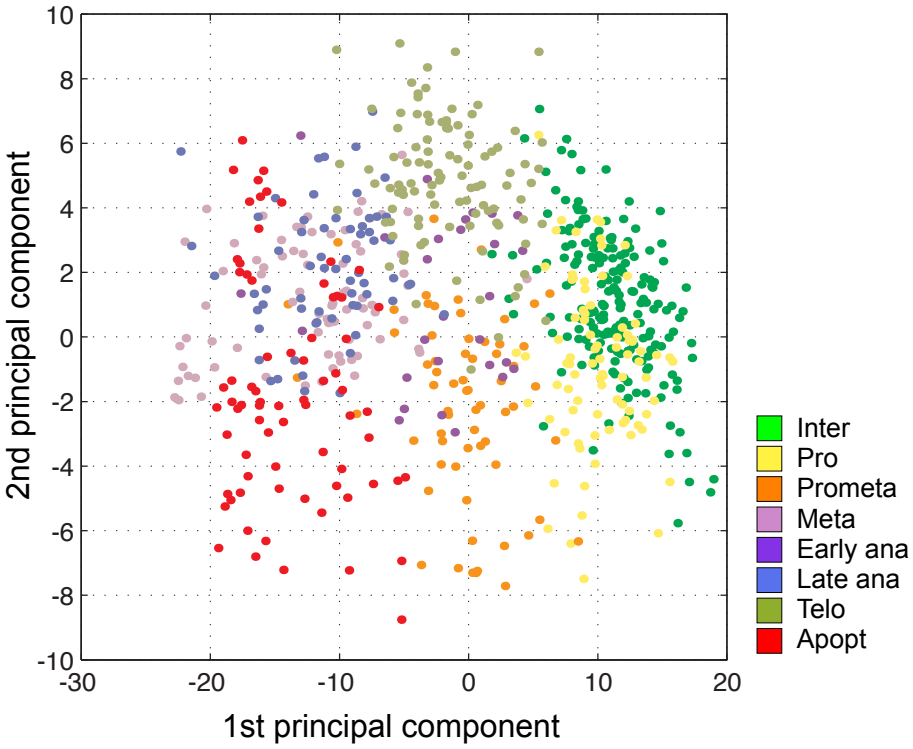


Effect of annotation errors on support vector machine accuracy

(a) Annotation of trajectories shown in Fig. 2a using support vector machine classifiers trained with partially randomized training data. A fraction of training objects was assigned with random class labels (percentage as indicated in the header line; uniform distribution; 8 classes). Support vector machines were trained by grid search and 5-fold cross-validation. The overall prediction accuracy (correct predictions / total predictions) was measured by comparison with original error-free manual annotation.

(b) Plot visualizing the overall prediction accuracy relative to the percentage of randomized training data as in (a). Data-points show the mean and error-bars the standard deviation of 8 repetitions.

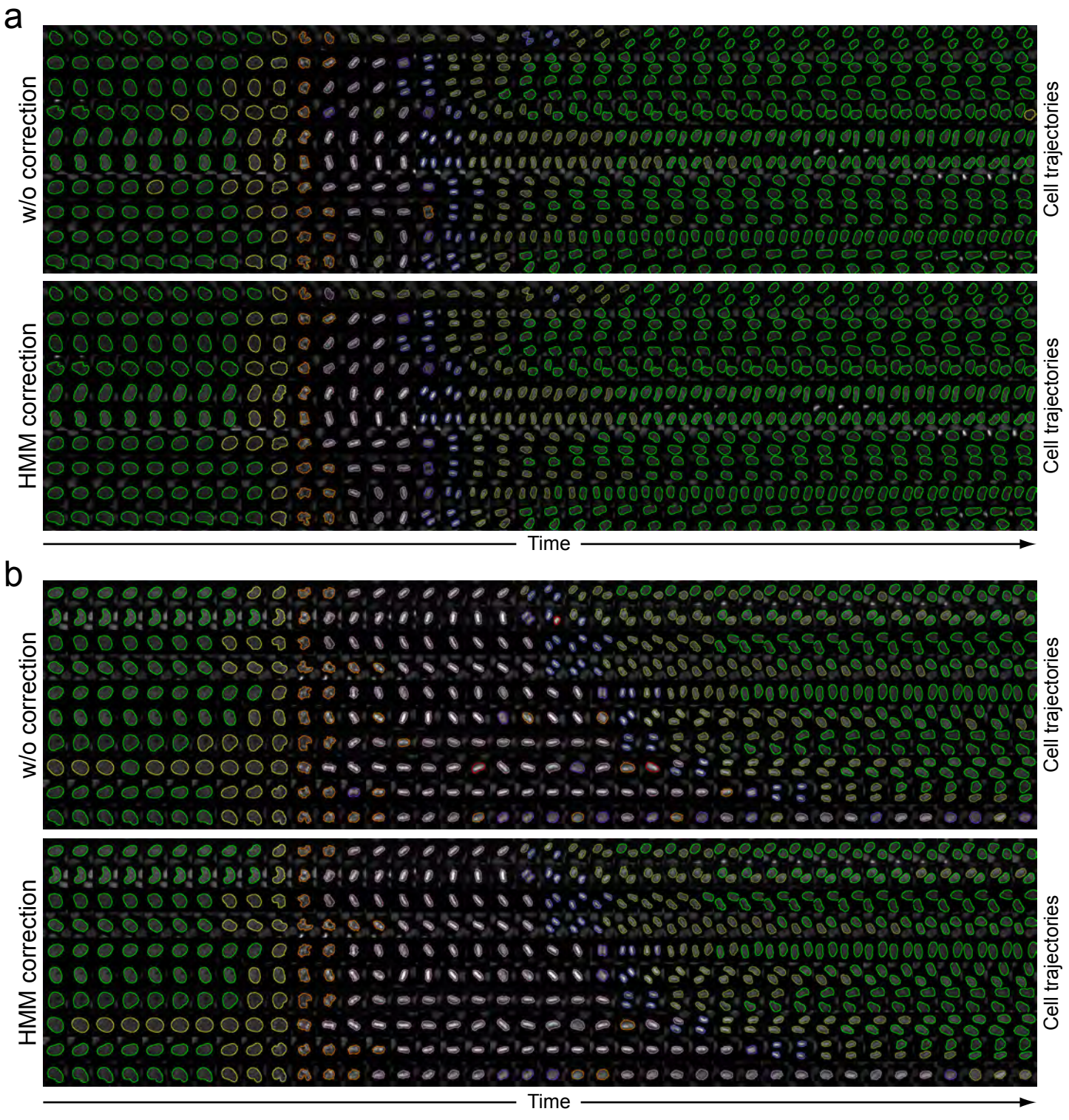
Supplementary Figure 2



Feature visualization in the subspace of the first two principal components

Data points visualize the first two principal components computed from 186 features and 689 manually annotated objects of the support vector machine classifier shown in Fig. 1d (color labels indicate morphology classes). Each object corresponds to a single H2B-labeled cell of the training set. Classes form overlapping clusters, which prevents perfect separation at this point.

Supplementary Figure 3

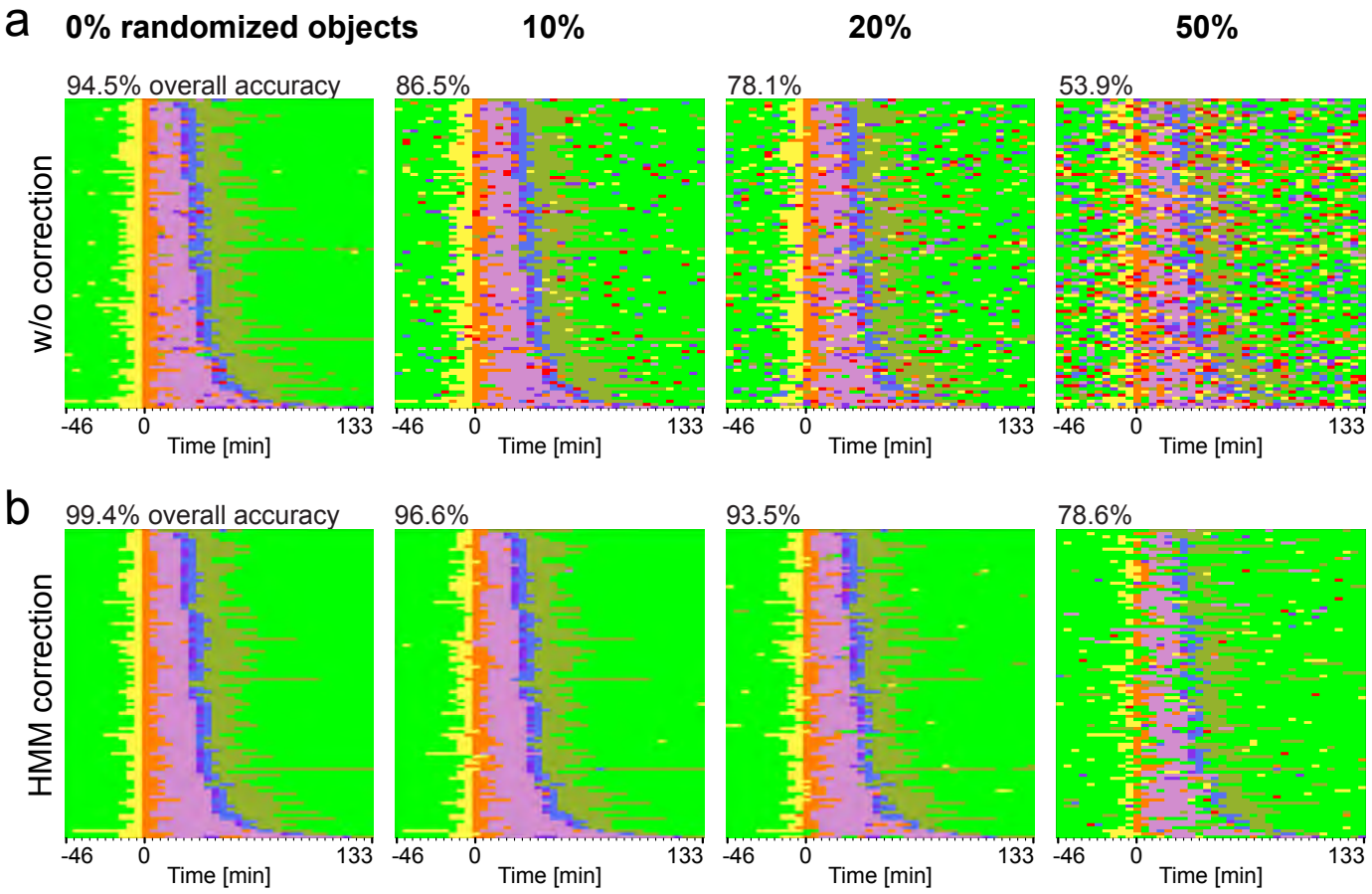


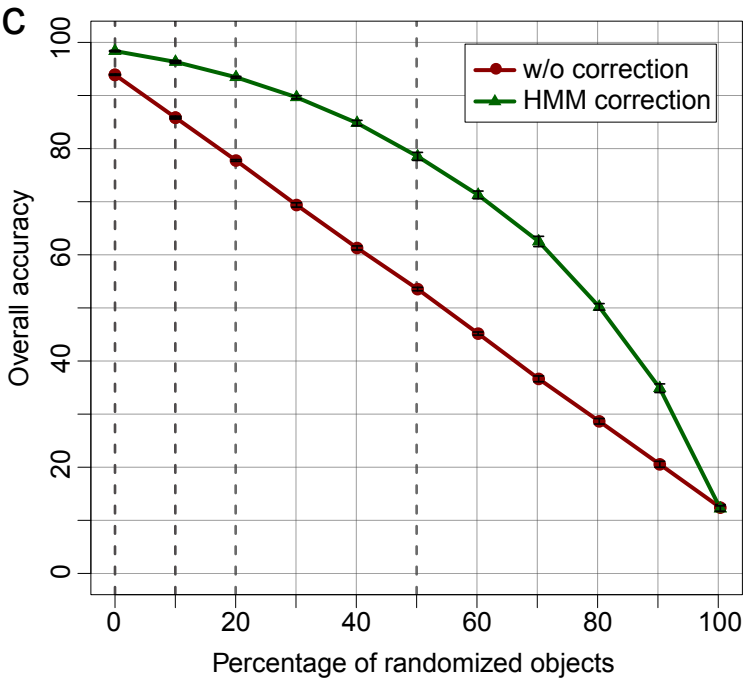
Annotation of the fastest and slowest trajectories

(a) Trajectories of the 10% cells progressing fastest through mitosis, from the cells shown in Fig. 2. Contour colors indicate the class label predictions by the support vector machine (upper row; w/o correction; see Fig. 2a), and the hidden Markov model-corrected class labels (lower row; HMM correction; see Fig. 2f).

(b) Trajectories of the 10% cells progressing slowest through mitosis as in (a).

Supplementary Figure 4





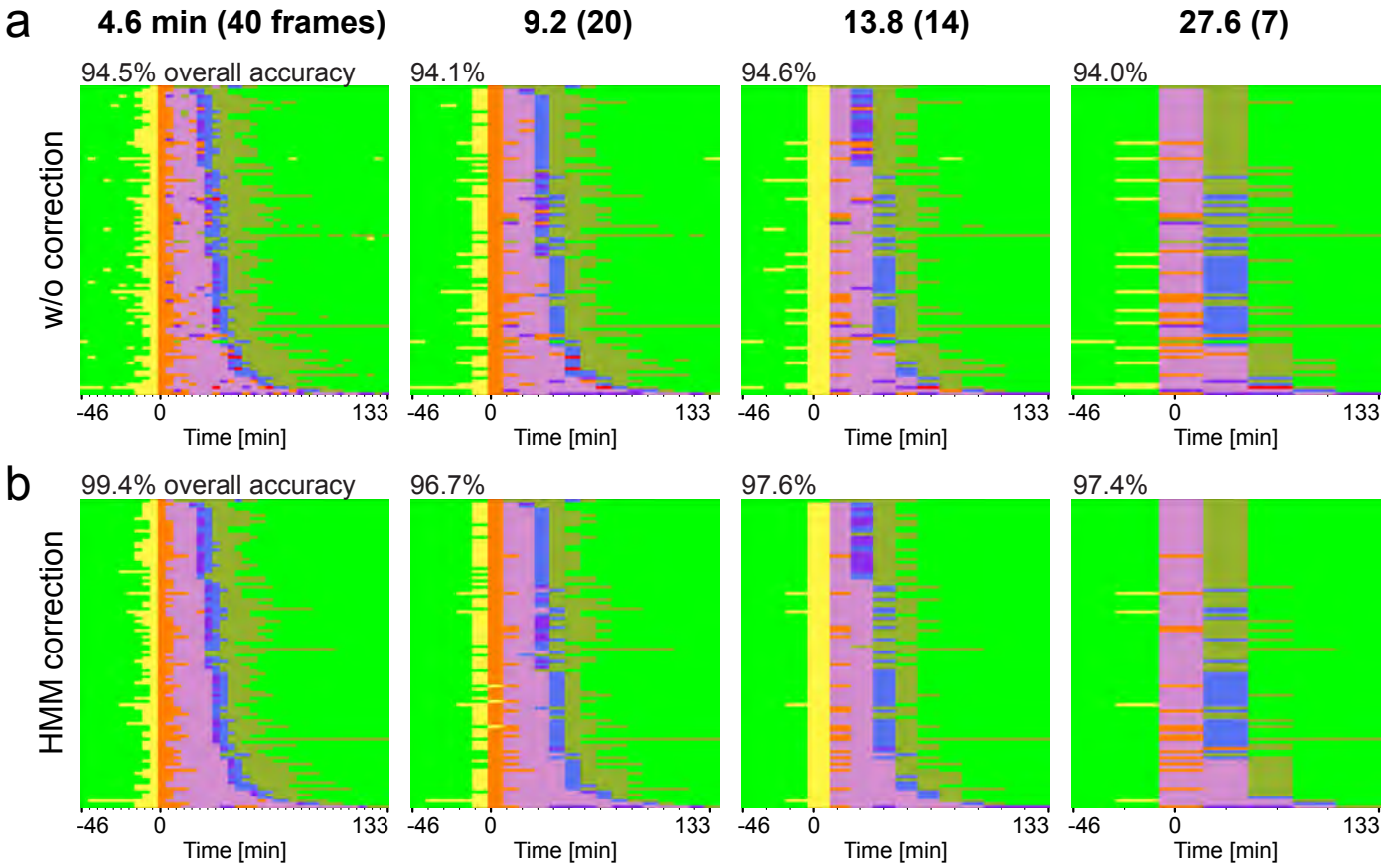
Effect of support vector machine class prediction noise on hidden Markov model error correction

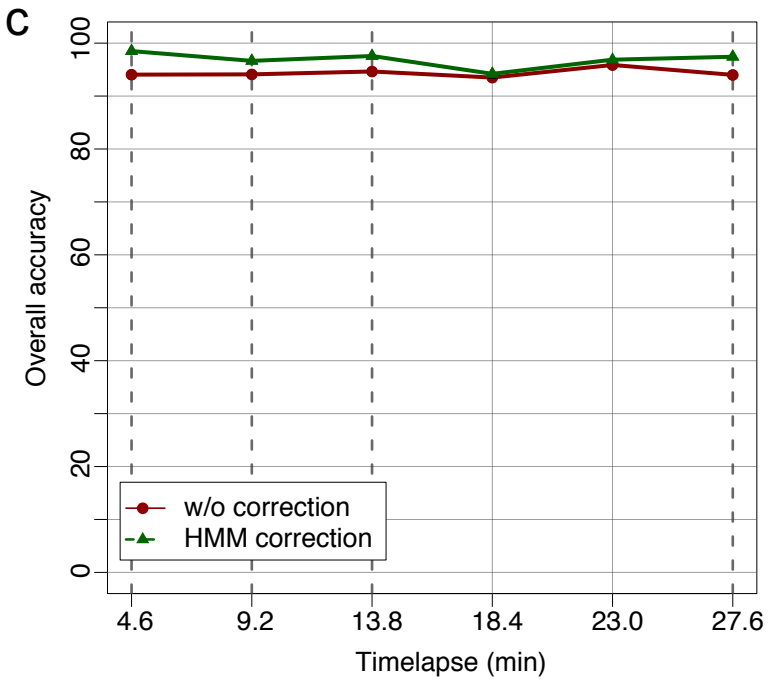
(a) Support vector machine prediction probabilities (8-vector; color corresponds to class label with highest probability) of Fig. 2a were randomly sampled and replaced by random numbers (uniform distribution; sum 1.0). Fractions of the 100x40 objects were randomized from 0% to 100% in 10% steps and 8 repetitions. The overall prediction accuracy (correct predictions / total predictions) was measured based on the manual annotation. Trajectory panels are shown for 0%, 10%, 20%, and 50% randomized data w/o correction (upper panel) and with HMM correction (lower panel).

(b) A hidden Markov model was trained on the partially randomized data shown in (a) and errors corrected on the randomized trajectories.

(c) Plot visualizing the overall prediction accuracy w/o correction (red line) and with hidden Markov model correction (green line) relative to the percentage of randomized data. Average and standard deviation of 8 repetitions shown.

Supplementary Figure 5





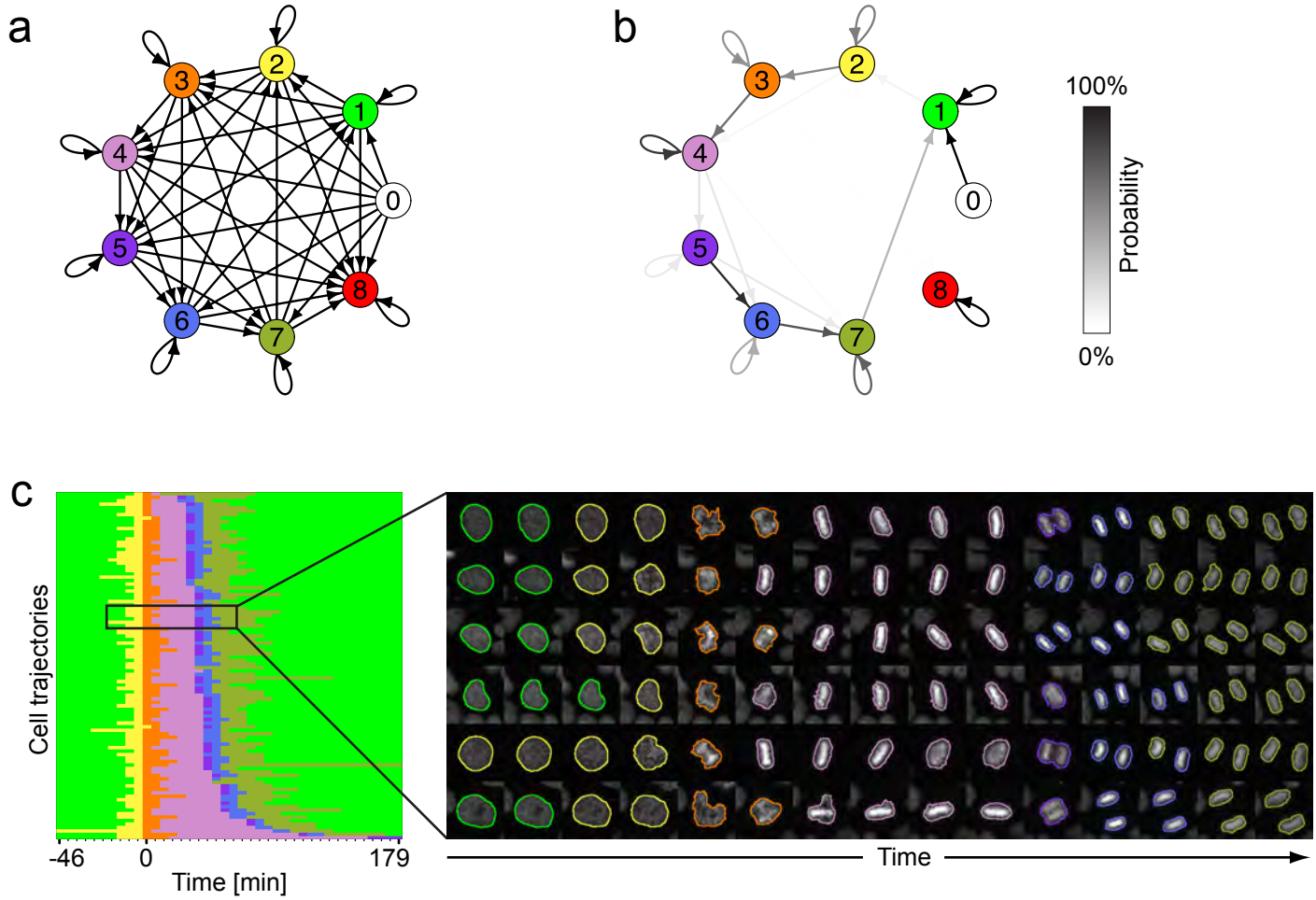
Effect of different time-lapse intervals on hidden Markov error correction

(a) Every n^{th} frame (from 1 to 6) of the trajectories from Fig. 2a was selected to simulate different time-lapse intervals (4.6 min to 27.6 min). The overall accuracy was measured by comparison with manual annotation.

(b) Same data and procedure as in (a) but with hidden Markov model trained on and applied to the subsampled trajectories.

(c) Plot visualizing the overall accuracy w/o correction (red line) and with hidden Markov model correction (green line) for differently sampled time-lapse. Note: A higher time-lapse might decrease tracking accuracy, which is not reflected by this plot.

Supplementary Figure 6



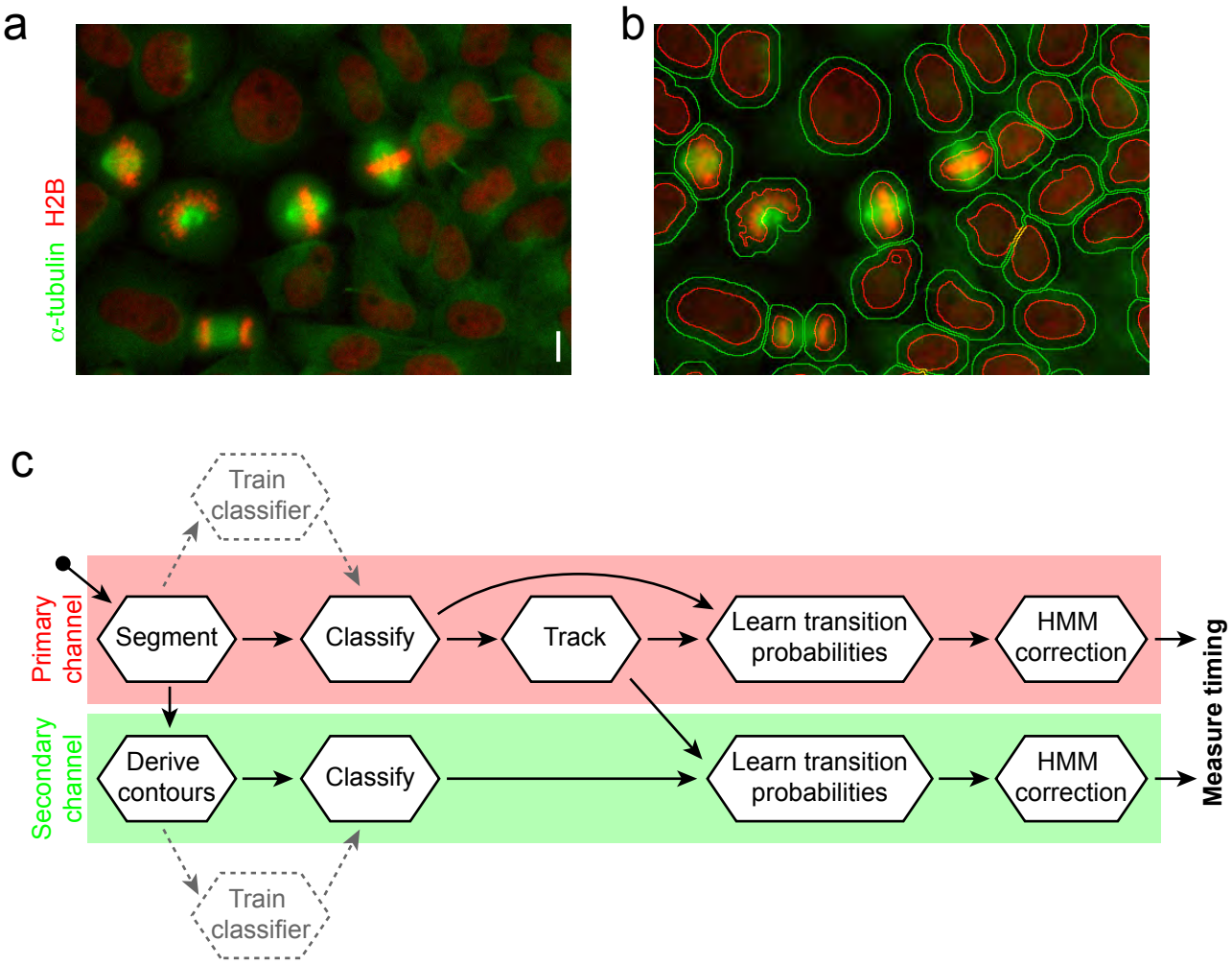
Hidden Markov model error correction based on biological a priori knowledge

(a) Class transitions were constrained to the forward direction of 3 consecutive frames, and apoptosis was defined as terminal state.

(b) Learned stage transition probabilities for the constrained model based on the same data shown in Fig. 2A.

(c) Error correction of the data shown in Fig. 2A using the constrained hidden Markov model.

Supplementary Figure 7



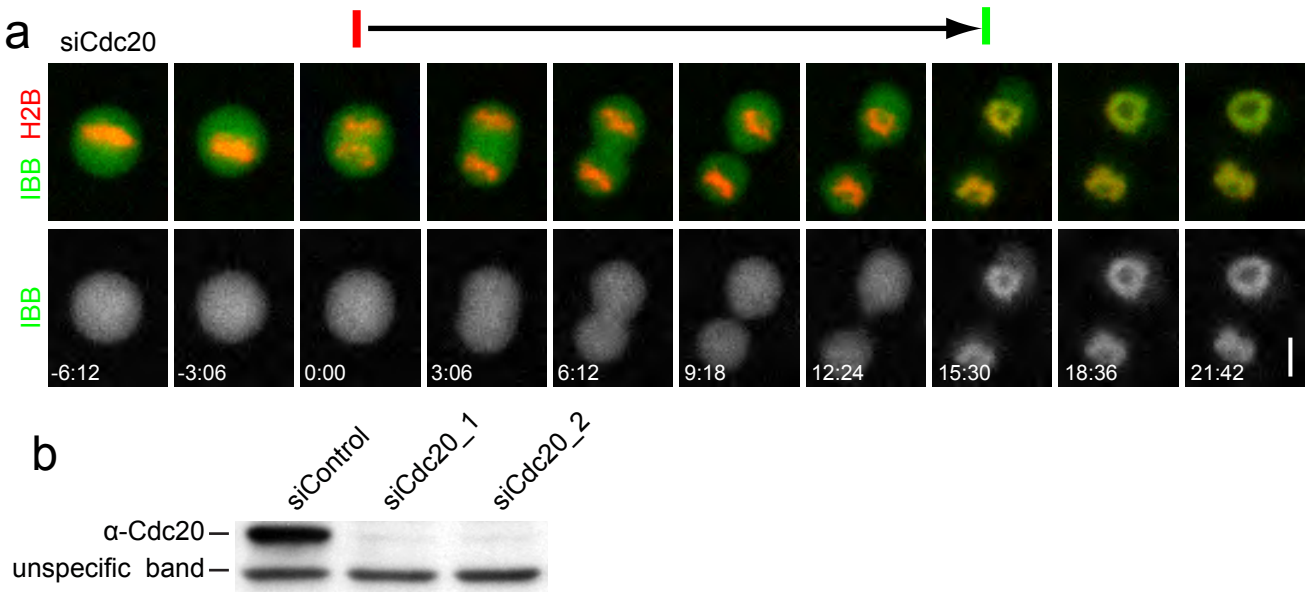
Strategy for annotation of multi-channel assays.

(a) Single frame of a movie from a cell expressing H2B-mCherry and mEGFP-a-tubulin.

(b) Segmentation of secondary channel. Cells were first segmented by the H2B-mCherry (red contours), which was dilated to derive cytoplasmic regions (green contours).

(c) Workflow schematic for processing of two-channel experiments. The training of the classifier (dashed lines) applies only once per experiment, all other worksteps are automated.

Supplementary Figure 8



Cdc20 RNAi phenotype

(a) Mitotic exit timing in Cdc20 RNAi cell expressing H2B-mCherry and IBB-EGFP. The arrow indicates timing from anaphase onset (red bar) until onset of nuclear accumulation of IBB-EGFP (green bar). Cells were imaged as 2D time series with widefield epifluorescence 10x dry objective; time is in min:s. Bar: 10 μm

(b) Validation of Cdc20 RNAi. Western Blotting of whole cell lysates 60 h after transfection of two different siRNA oligos targeting Cdc20, or a non-silencing siRNA oligo for negative control. The unspecific band detected by the anti-Cdc20 antibody served as a loading control.

Suppl. Table 1. Shape and texture features

Shape and size	circularity dist_max dist_min dist_ratio irregularity irregularity2 perimeter roisize
Haralick [1] (32 gray levels, not normalized, distances 1,2,4 and 8 pixels rotation invariant)	h1_ASM h1_CON h1_COR h1_COV h1_DAV h1_ENT h1_IDM h1_PRO h1_SAV h1_SET h1_SHA h1_SVA h1_VAR h1_average h1_variance h2_ASM h2_CON h2_COR h2_COV h2_DAV h2_ENT h2_IDM h2_PRO h2_SAV h2_SET h2_SHA h2_SVA h2_VAR h2_average h2_variance h4_ASM h4_CON h4_COR h4_COV h4_DAV h4_ENT h4_IDM h4_PRO h4_SAV h4_SET h4_SHA h4_SVA h4_VAR h4_average
	h4_variance h8_ASM h8_CON h8_COR h8_COV h8_DAV h8_ENT h8_IDM h8_PRO h8_SAV h8_SET h8_SHA h8_SVA h8_VAR h8_average h8_variance
Haralick [1] (32 gray levels, normalized to min/max gray values per object, distances 1,2,4 and 8 pixels rotation invariant)	h1_2ASM h1_2CON h1_2COR h1_2COV h1_2DAV h1_2ENT h1_2IDM h1_2PRO h1_2SAV h1_2SET h1_2SHA h1_2SVA h1_2VAR h1_2average h1_2variance h2_2ASM h2_2CON h2_2COR h2_2COV h2_2DAV h2_2ENT h2_2IDM h2_2PRO h2_2SAV h2_2SET h2_2SHA h2_2SVA h2_2VAR h2_2average h2_2variance h4_2ASM h4_2CON h4_2COR h4_2COV h4_2DAV h4_2ENT

	h4_2IDM h4_2PRO h4_2SAV h4_2SET h4_2SHA h4_2SVA h4_2VAR h4_2average h4_2variance h8_2ASM h8_2CON h8_2COR h8_2COV h8_2DAV h8_2ENT h8_2IDM h8_2PRO h8_2SAV h8_2SET h8_2SHA h8_2SVA h8_2VAR h8_2average h8_2variance
Statistical Geometric Features [2]	ls0_CAREA_avg_value ls0_CAREA_max_value ls0_CAREA_sample_mean ls0_CAREA_sample_sd ls0_DISP_avg_value ls0_DISP_max_value ls0_DISP_sample_mean ls0_DISP_sample_sd ls0_INTERIA_avg_value ls0_INTERIA_max_value ls0_INTERIA_sample_mean ls0_INTERIA_sample_sd ls0_IRGL_avg_value ls0_IRGL_max_value ls0_IRGL_sample_mean ls0_IRGL_sample_sd ls0_NCA_avg_value ls0_NCA_max_value ls0_NCA_sample_mean ls0_NCA_sample_sd ls0_TAREA_avg_value ls0_TAREA_max_value ls0_TAREA_sample_mean ls0_TAREA_sample_sd ls1_CAREA_avg_value ls1_CAREA_max_value ls1_CAREA_sample_mean ls1_CAREA_sample_sd ls1_DISP_avg_value ls1_DISP_max_value ls1_DISP_sample_mean
	ls1_DISP_sample_sd ls1_INTERIA_avg_value ls1_INTERIA_max_value ls1_INTERIA_sample_mean ls1_INTERIA_sample_sd ls1_IRGL_avg_value ls1_IRGL_max_value ls1_IRGL_sample_mean ls1_IRGL_sample_sd ls1_NCA_avg_value ls1_NCA_max_value ls1_NCA_sample_mean ls1_NCA_sample_sd ls1_TAREA_avg_value ls1_TAREA_max_value ls1_TAREA_sample_mean ls1_TAREA_sample_sd
Gray level features	n2_avg n2_stddev n2_wavg n2_wdist n2_wiavg
Gray level features (min/max normalized)	n_avg n_stddev n_wavg n_wdist n_wiavg

1. Haralick, R., Dinstein & Shanmugam in IEEE Transactions on Systems1973).
2. Walker, R. & Jackway, P. in Pattern Recognition, 1996., Proceedings of the 13th International Conference on, Vol. 21996).

Suppl. Table 2. siRNA oligos used for RNAi screening

Gene HUGO Symbol	Full Gene Name	Entrez Gene ID	RefSeq Accession Number	Ambion siRNA ID	Sense siRNA Sequence	Antisense siRNA Sequence
1-Sep	septin 1	1731	NM_052838,	35578	GGAUGCAGAGAUGAAGGAAtt	UUCCUUCAUCUCUGCAUCCtg
1-Sep	septin 1	1731	NM_052838,	35491	GGAAGAGGAGAUCCACAUCTt	GAUGUGGAUCUCCUCUUCCTt
1-Sep	septin 1	1731	NM_052838,	35399	GGAGCAAUUUGAGCAGUACTt	GUACUGCUCAAAUUGCUCCTc
3-Sep	septin 3	55964	NM_019106,NM_145733,NM_145734,	38214	GGAGCUUGAAGUAAAUGGCTt	GCCAUUUACUUCAAGCUCCTt
3-Sep	septin 3	55964	NM_019106,NM_145733,NM_145734,	38123	GGAAGAAACGCAUCCCUGAtt	UCAGGGAUGCGUUUCUUCCtg
3-Sep	septin 3	55964	NM_019106,NM_145733,NM_145734,	133703	GGGCCAAGCCCUUUUUAGUtt	ACUAAAAAGGGCUUGGCCCtg
6-Sep	septin 6	23157	NM_015129,NM_145799,NM_145800,NM	136765	GGCUAAAGCUCACGAUCGUtt	ACGAUCGUGAGCUUUAGCCtc
6-Sep	septin 6	23157	NM_015129,NM_145799,NM_145800,NM	136764	GCAACGUGAGGCUAAAGCUtt	AGCUUUAGCCUCACGUUGCtc
6-Sep	septin 6	23157	NM_015129,NM_145799,NM_145800,NM	136766	GCUCACGAUCGUUAGCACAtt	UGUGCUAACGAUCGUGAGCtt
8-Sep	septin 8	23176	XM_034872,	264717	CCUCCUUACUCACUAUAGUtt	ACUAUAGUGAGUAAGGAGGtg
8-Sep	septin 8	23176	XM_034872,	264716	CGCCAUACUUUUCCUAUAUtt	AUAUAGGAAAAGUAUGGCGtt
8-Sep	septin 8	23176	XM_034872,	264715	GGCAGAUGUUUGUCAACAAtt	UUGUUGACAAACAUCUGCCtc
10-Sep	septin 10	151011	NM_144710,NM_178584,	215107	CCUUGACAGCAAGGUAAACtt	GUUUACCUUGCUGUCAAGGtt
10-Sep	septin 10	151011	NM_144710,NM_178584,	37634	GGUGGAUGUGAAACAUGAAtt	UUCAUGUUUCACAUCCACCtt
10-Sep	septin 10	151011	NM_144710,NM_178584,	37539	GGCUAUAUGUAUAAGGUGGtt	CCACCUUAUACAUAUAGCCtg
11-Sep	septin 11	55752	NM_018243,	125139	CCUGUACUAAAUGCCUAAUtt	AUUAGGCAUUUAGUACAGGtt
11-Sep	septin 11	55752	NM_018243,	125138	GGUGUUCGGUUAAAAGCCAtt	UGGCUUUUAACCGAACACCtg
11-Sep	septin 11	55752	NM_018243,	125137	CGUUAAUGGACACUUUGUUtt	AACAAAGUGUCCAUUAACGtg
ANLN	anillin, actin binding protein (scraps homolo	54443	NM_018685,	132620	GGAAGCUACAUUCUGUUCCTt	GGAACAGAAUGUAGCUUCCTg
ANLN	anillin, actin binding protein (scraps homolo	54443	NM_018685,	132619	GCCUGGUACCGCUUGUUUAtt	UAAACAAGCGGUACCAGGCtg
ANLN	anillin, actin binding protein (scraps homolo	54443	NM_018685,	132621	GGUUUCACUGAAUGCGAAATt	UUUCGCAUUCAGUGAAACCtt
ARHGAP17	Rho GTPase activating protein 17	55114	NM_018054,NM_001006634,	26221	GGAUCAAGACAAAAAACUUtt	AAGUUUUUUGUCUUGAUCCtg
ARHGAP17	Rho GTPase activating protein 17	55114	NM_018054,NM_001006634,	26127	GGUGGAGAUUCCCAACAUCtt	GAUGUUGGGAAUCUCCACCtc
ARHGAP17	Rho GTPase activating protein 17	55114	NM_018054,NM_001006634,	26031	GGAGACACAAAAAACUGCCtt	GGCAGUUUUUUGUGUCUCCtc
AURKB	aurora kinase B	9212	NM_004217,	495	GGUGAUGGAGAAUAGCAGUtt	ACUGCUAUUCUCCAUCACCtt
AURKB	aurora kinase B	9212	NM_004217,	494	GGAGGAUCUACUUGAUUCUtt	AGAAUCAAGUAGAUCCUCCtc
AURKB	aurora kinase B	9212	NM_004217,	493	GGCAAGUUUGGAAACGUGUtt	ACACGUUUCCAAACUUGCCtt
AURKC	aurora kinase C	6795	NM_003160,NM_001015878,NM_001015	111219	GCGAGAAAUUAGAUGAACAtt	UGUUCAUCUAAUUUCUCGCtt
AURKC	aurora kinase C	6795	NM_003160,NM_001015878,NM_001015	379	GGAAAGCCAUUUCAUUGUGtt	CACAAUGAAAUGGCUUUCCtt
AURKC	aurora kinase C	6795	NM_003160,NM_001015878,NM_001015	378	GGUAGAUGUGAGGUUUCCAtt	UGGAAACCUCACAUCUACCtt
BIRC5	baculoviral IAP repeat-containing 5 (survivin	332	NM_001168,NM_001012270,NM_001012	121296	GGCAGUGGCCUAAAUCCUUtt	AAGGAUUUAGGCCACUGCCtt
BIRC5	baculoviral IAP repeat-containing 5 (survivin	332	NM_001168,NM_001012270,NM_001012	121295	GCCAUUCUAAGUCAUUGGGtt	CCCAAUGACUUAGAAUGGCtt
BIRC5	baculoviral IAP repeat-containing 5 (survivin	332	NM_001168,NM_001012270,NM_001012	121294	CCACUUCCAGGGUUUAUUCtt	GAAUAAACCCUGGAAGUGGtg
BUB1	BUB1 budding uninhibited by benzimidazole	699	NM_004336,	510	GGCAAAAGCUGAAGAAAGUtt	ACUUUCUUCAGCUUUUGCCtt
BUB1	BUB1 budding uninhibited by benzimidazole	699	NM_004336,	509	GGUUAUUUCAGACACGCCUtt	AGGCGUGUCUGAAAUAACCtg
BUB1	BUB1 budding uninhibited by benzimidazole	699	NM_004336,	147346	CGAAGAGUGAUCACGAUUUtt	AAAUCGUGAUCACUCUUCGtt
BUB1B	BUB1 budding uninhibited by benzimidazole	701	NM_001211,	90	GGUGGGAAGGAGAGUAAUAtt	UAUUACUCUCCUUCCCACCtt
BUB1B	BUB1 budding uninhibited by benzimidazole	701	NM_001211,	89	GGCUUCAGAAAUGUAACAAtt	UUGUUACAUUUCUGAAGCCtg
BUB1B	BUB1 budding uninhibited by benzimidazole	701	NM_001211,	88	GGGAUUGGUGUUUCACUUGtt	CAAGUGAAACACCAAUCCCtt
BUB3	BUB3 budding uninhibited by benzimidazole	9184	NM_004725,NM_001007793,	137638	GCAGGGUUAUGUAUUAAGCtt	GCUUAAUACAUAACCCUGCtt
BUB3	BUB3 budding uninhibited by benzimidazole	9184	NM_004725,NM_001007793,	137637	GCCUGAAAAGGUAUAUACCtt	GGUAUAUACCUUUUCAGGCtg
BUB3	BUB3 budding uninhibited by benzimidazole	9184	NM_004725,NM_001007793,	15258	GGUAUAUACCCUCUCAGUGtt	CACUGAGAGGGUAUAUACCtt
CCNB1	cyclin B1	891	NM_031966,	118840	GCUGAUCCAAACCUUUGUAtt	UACAAAGGUUUGGAUCAGCtc
CCNB1	cyclin B1	891	NM_031966,	118839	GCCUAUUUUGGUUGAUACUtt	AGUAUCAACCAAAAUAGGCtc
CCNB1	cyclin B1	891	NM_031966,	118838	GCAAAACCUUCAGCUACUGtt	CAGUAGCUGAAGGUUUUGCtt
CDC10	septin 7	989	NM_001788,NM_001011553,	10504	GGGAAGAUCUUUUAAACUCtt	GAGUUUAAAAGAUCUUCCCtt
CDC10	septin 7	989	NM_001788,NM_001011553,	10417	GGUCCUUCUCAUAGAAUUAtt	UAAUUCUAUGAGAAGGACCtg
CDC10	septin 7	989	NM_001788,NM_001011553,	10323	GGUUUUGAAUUCACGCUUAtt	UAAGCGUGAAUUCAAAACCtc
CDC14A	CDC14 cell division cycle 14 homolog A (S.	8556	NM_033312,NM_033313,NM_003672,	105908	GAAAAUAGUGCACUACACCtt	GGUGUAGUGCACUAUUUUCtt
CDC14A	CDC14 cell division cycle 14 homolog A (S.	8556	NM_033312,NM_033313,NM_003672,	105907	GAUUUUGGACCGCUGAACUtt	AGUUCAGCGGUCCAAAAUCtg
CDC14A	CDC14 cell division cycle 14 homolog A (S.	8556	NM_033312,NM_033313,NM_003672,	105906	GCACAGUAAAUACCCACUAtt	UAGUGGGUAUUUACUGUGCtt
CDC14B	CDC14 cell division cycle 14 homolog B (S.	8555	NM_033331,NM_033332,NM_003671,	45650	GAUUUUGGACCACUCAAUCtt	GAUUGAGUGGUCCAAAAUCtg
CDC14B	CDC14 cell division cycle 14 homolog B (S.	8555	NM_033331,NM_033332,NM_003671,	45559	GAACUUCUACGCAGAUUUUtt	AAAAUCUGCGUAGAAGUUCtc
CDC14B	CDC14 cell division cycle 14 homolog B (S.	8555	NM_033331,NM_033332,NM_003671,	35174	GGUGAUAGACUUCGGGCCUtt	AGGCCCGAAGUCUAUCACCtt
CDC16	CDC16 cell division cycle 16 homolog (S. ce	8881	NM_003903,	137341	GCCUAGUGAAACGGUCAUCtt	GAUGACCGUUUCACUAGGCtt
CDC16	CDC16 cell division cycle 16 homolog (S. ce	8881	NM_003903,	137340	GCGACUGGGAAAUGUCACAtt	UGUGACAUUUCCCAGUCGCtg
CDC16	CDC16 cell division cycle 16 homolog (S. ce	8881	NM_003903,	137342	CCAAUAACUCAAAACUAGCtt	GCUAGUUUUGAGUUAUUGGtc
CDC2	cell division cycle 2, G1 to S and G2 to M	983	NM_033379,NM_001786,	42819	GGAACUUCGUCAUCCAAAUtt	AUUUGGAUGACGAAGUUCCtt
CDC2	cell division cycle 2, G1 to S and G2 to M	983	NM_033379,NM_001786,	1625	GGUUAUAUCUCAUCUUUGAtt	UCAAAGAUGAGAUAUAACCtg
CDC2	cell division cycle 2, G1 to S and G2 to M	983	NM_033379,NM_001786,	1440	GGUCAAGUGGUAGCCAUGAtt	UCAUGGCUACCACUUGACCtg
CDC20	CDC20 cell division cycle 20 homolog (S. ce	991	NM_001255,	215139	CCUUGUGGAUUGGAGUUCUtt	AGAACUCCAAUCCACAAGGtt
CDC20	CDC20 cell division cycle 20 homolog (S. ce	991	NM_001255,	145701	CCAGCUAGUUAUUUGGAAGtt	CUUCCAAAUAACUAGCUGGtt
CDC20	CDC20 cell division cycle 20 homolog (S. ce	991	NM_001255,	145700	CCUGCCGUUACAUUCCUUCtt	GAAGGAAUGUAACGGCAGGtc
CDCA1	cell division cycle associated 1	83540	NM_031423,NM_145697,	131098	GGACCUUUCAGAUAAUAGGtt	CCUAUUAUCUGAAAGGUCCtg
CDCA1	cell division cycle associated 1	83540	NM_031423,NM_145697,	131097	GCAUGCCGUGAAACGUAUAtt	UAUACGUUUCACGGCAUGCtt
CDCA1	cell division cycle associated 1	83540	NM_031423,NM_145697,	131099	CGCACAGUAAUUGAGGAUUtt	AAUCCUCAAUUACUGUGCGtt
CDCA1	kinesin heavy chain member 2	3796	NM_004520,	118425	CGUAGAAAAUCUAAUUGUGtt	CACAAUUAGAUUUUCUACGtg
CDCA8	cell division cycle associated 8	55143	NM_018101,	132285	GGUCAAGCCGUGCUAACACtt	GUGUUAGCACGGCUUGACCtt
CDCA8	cell division cycle associated 8	55143	NM_018101,	132284	GGUAGAUGAAAUGAUAGUGtt	CACUAUCAUUUCAUCUACCtg
CDCA8	cell division cycle associated 8	55143	NM_018101,	132286	GGCUUAUUGUUUGAGUGUGtt	CACACUCAAACAAUAAGCCtg
CENPE	centromere protein E, 312kDa	1062	NM_001813,	121339	GCUACUAAAUCAGGAGAAUtt	AUUCUCCUGAUUUAGUAGCtt
CENPE	centromere protein E, 312kDa	1062	NM_001813,	121337	CCAAUCAUCGAUUCUGCCAtt	UGGCAGAAUCGAUGAUUGGtg
CENPE	centromere protein E, 312kDa	1062	NM_001813,	10706	GGAAUUAAAGGCUAAAAGAtt	UCUUUUAGCCUUUAAUUCCtg
CENPF	centromere protein F, 350/400ka (mitosin)	1063	NM_016343,	146738	GGUGACUCCAAGUCGAUCAtt	UGAUCGACUUGGAGUCACCtc
CENPF	centromere protein F, 350/400ka (mitosin)	1063	NM_016343,	146737	CCAAGUCAAUAUUAUAGUGtt	CACUAUAAUAUUGACUUGGtg
CENPF	centromere protein F, 350/400ka (mitosin)	1063	NM_016343,	146739	GGUCGAUGAAUUAACAACUtt	AGUUGUUAAUUCAUCGACCtt
CEP1	centrosomal protein 1	11064	NM_007018,	136173	GCUAUAAUCUAAUAGGGAAtt	UUCCCUAUUAGAUUAUAGCtg
CEP1	centrosomal protein 1	11064	NM_007018,	136172	GGAGUUAGAUAUAUUACAGtt	CUGUAAUAUAUCUAACUCCtg
CEP1	centrosomal protein 1	11064	NM_007018,	136174	GCCACUAAAUUAUUAUCCAtt	UGGAUAAUAAUUUAGUGGCtc
ch-TOG	cytoskeleton associated protein 5	9793	NM_014756,NM_001008938,	122705	GGUGUUCAAUCAACCUAAAtt	UUUAGGUUGAUUGAACACCtt
ch-TOG	cytoskeleton associated protein 5	9793	NM_014756,NM_001008938,	122704	GGUGUUGUAAGUAAGGUGUtt	ACACCUUACUUACAACACCtg
ch-TOG	cytoskeleton associated protein 5	9793	NM_014756,NM_001008938,	122703	GCAAGGUUAAGUGGGUAUGtt	CAUACCCACUUAACCUUGCtt
CHC1	chromosome condensation 1	1104	NM_001269,	145718	GGAAACGACCACUUGGUGAtt	UCACCAAGUGGUCGUUUCCtg
CHC1	chromosome condensation 1	1104	NM_001269,	145717	CCGUGUGUCUAAGCAAAAGtt	CUUUUGCUUAGACACACGGtg
CHC1	chromosome condensation 1	1104	NM_001269,	145719	GCAUACAGUCUUAUUAGUCtt	GACUAAUAAGACUGUAUGCtg
CIT	citron (rho-interacting, serine/threonine kina	11113	NM_007174,	103737	GGAUAAAUUAAGGGUCAUUtt	AAUGACCCUUAAUUUAUCCtg
CIT	citron (rho-interacting, serine/threonine kina	11113	NM_007174,	103729	GGGAUAUUAGAUGCCCUCUtt	AGAGGGCAUCUAAUAUCCCtt
CIT	citron (rho-interacting, serine/threonine kina	11113	NM_007174,	103721	GGCUGAAUCUGUUCUUCCAtt	UGGAAGAACAGAUUCAGCCtg
CLASP1	cytoplasmic linker associated protein 1	23332	NM_015282,	136867	GCACAGACUUUAACACUAAtt	UUAGUGUUAAAGUCUGUGCtc

CLASP1	cytoplasmic linker associated protein 1	23332	NM_015282,	136866	CCAUGUUAGAUAAACUUGUtt	ACAAGUUUAUCUAACAUGGtc
CLASP1	cytoplasmic linker associated protein 1	23332	NM_015282,	136868	CGACACAUAUCAGUAUUAGtt	CUAAUACUGAUAUGUGUCGtt
CSNK2B	casein kinase 2, beta polypeptide	1460	NM_001320,	9896	GGAACCCUGUAUGGUUUUUtt	AAAAACCAUACAGGGUUCCtg
CSNK2B	casein kinase 2, beta polypeptide	1460	NM_001320,	9806	GGAGACUUUGGUUACUGUCtt	GACAGUAACCAAAGUCUCCtt
CSNK2B	casein kinase 2, beta polypeptide	1460	NM_001320,	9710	GGCAGCCGAGAUGCUUUAUtt	AUAAAGCAUCUCGGCUGCCtg
DCTN1	dynactin 1 (p150, glued homolog, Drosophil	1639	NM_023019,NM_004082,	242562	CCACAUUAAGUUCACGCAGtt	CUGCGUGAACUUAAUGUGGtc
DCTN1	dynactin 1 (p150, glued homolog, Drosophil	1639	NM_023019,NM_004082,	242561	GGAGAAAGAGUUUGAGGAGtt	CUCCUCAAACUCUUUCUCCtt
DCTN1	dynactin 1 (p150, glued homolog, Drosophil	1639	NM_023019,NM_004082,	242560	GGCAGAGAGCACCAUUGAUtt	AUCAAUGGUGCUCUCUGCCtg
DCTN2	dynactin 2 (p50)	10540	NM_006400,	135759	GGUGCACCAGCUAUAUGAAtt	UUCAUAUAGCUGGUGCACCtt
DCTN2	dynactin 2 (p50)	10540	NM_006400,	135758	GGACAGGAUAUGAAUCUGGtt	CCAGAUUCAUAUCCUGUCCtc
DCTN2	dynactin 2 (p50)	10540	NM_006400,	135757	GCGGAGUUCGAUGCGUUUGtt	CAAACGCAUCGAACUCCGCtt
DLG7	discs, large homolog 7 (Drosophila)	9787	NM_014750,	138400	GCCAAAAAAGGCUAUUCCAtt	UGGAAUAGCCUUUUUUGGCtc
DLG7	discs, large homolog 7 (Drosophila)	9787	NM_014750,	138399	CGAAAUAGACACUUUGGUUtt	AACCAAAGUGUCUAUUUCGtt
DLG7	discs, large homolog 7 (Drosophila)	9787	NM_014750,	138401	CGAGGAAUAUUUAAAGUGGtt	CCACUUUAAAUAUUCCUCGtt
DNCH1	dynein, cytoplasmic, heavy polypeptide 1	1778	NM_001376,	118311	GCCAAAAGUUACAGACUUUtt	AAAGUCUGUAACUUUUGGCtt
DNCH1	dynein, cytoplasmic, heavy polypeptide 1	1778	NM_001376,	118310	GCAAAAUAUUGAAAUUCCGtt	CGGAAUUUCAAUAUUUUGCtg
DNCH1	dynein, cytoplasmic, heavy polypeptide 1	1778	NM_001376,	118309	CGUACUCCCGUGAUUGAUGtt	CAUCAAUCACGGGAGUACGtt
ECT2	epithelial cell transforming sequence 2 oncc	1894	NM_018098,	26257	GGCCAAUAAUUUAAGUUGCtt	GCAACUUAAAUUAUUGGCCtt
ECT2	epithelial cell transforming sequence 2 oncc	1894	NM_018098,	26165	GGUUUGGAUUCUCCGGAAUtt	AUUCCGGAGAAUCCAAACCtt
ECT2	epithelial cell transforming sequence 2 oncc	1894	NM_018098,	26070	GGACAUUAAAGUGGGCUUUtt	AAAGCCCACUUUAAUGUCCtt
ESPL1	extra spindle poles like 1 (S. cerevisiae)	9700	NM_012291,	121653	CCAUUAAUAAAAAGUGUCCtt	GGACACUUUUUAUUAAUGGtg
ESPL1	extra spindle poles like 1 (S. cerevisiae)	9700	NM_012291,	121652	GCAGCUGACUGCUAAGCUAtt	UAGCUUAGCAGUCAGCUGCtg
ESPL1	extra spindle poles like 1 (S. cerevisiae)	9700	NM_012291,	121651	GCUUGUGAUGCCAUCCUGAtt	UCAGGAUGGCAUCACAAGCtt
FZR1	fizzy/cell division cycle 20 related 1 (Drosop	51343	NM_016263,	241641	UUAAAUGCCUGAUUGUGAAtt	UUCACAAUCAGGCAUUUAAtg
FZR1	fizzy/cell division cycle 20 related 1 (Drosop	51343	NM_016263,	241640	GCAAAACCCGUUCGACAAAtt	UUUGUCGAACGGGUUUUGCta
FZR1	fizzy/cell division cycle 20 related 1 (Drosop	51343	NM_016263,	241639	GUCAGAACCGGAAAGCCAAtt	UUGGCUUUCCGGUUCUGACtg
INCENP	inner centromere protein antigens 135/155k	3619	NM_020238,	28431	GGAGAAGAAGAAGCAGAUUtt	AAUCUGCUUCUUCUUCUCCtc
INCENP	inner centromere protein antigens 135/155k	3619	NM_020238,	145370	CGGAAGAAGAGACGGAUUUtt	AAAUCCGUCUCUUCUUCCGtc
INCENP	inner centromere protein antigens 135/155k	3619	NM_020238,	145369	GCGCAUGUUCACCAGAGAAtt	UUCUCUGGUGAACAUGCGCtc
INCENP	inner centromere protein antigens 135/155k	3619	NM_020238,	28244	GGACUUGGUGUGGCUUGAGtt	CUCAAGCCACACCAAGUCCtt
KEAP1	kelch-like ECH-associated protein 1	9817	NM_012289,NM_203500,	138235	CGAGUGGCGAAUGAUCACAtt	UGUGAUCAUUCGCCACUCGtt
KEAP1	kelch-like ECH-associated protein 1	9817	NM_012289,NM_203500,	138234	GGAACGAGUGGCGAAUGAUtt	AUCAUUCGCCACUCGUUCCtc
KEAP1	kelch-like ECH-associated protein 1	9817	NM_012289,NM_203500,	138233	CGGGACAAACCGCCUUAAUtt	AUUAAGGCGGUUUGUCCCGtc
KIF11	kinesin family member 11	3832	NM_004523,	118431	GGAGUGAUAAUUAAAGGUUtt	AACCUUUAAUUAUCACUCCtc
KIF11	kinesin family member 11	3832	NM_004523,	118430	GCUCAAGGAAAACAUACACtt	GUGUAUGUUUUCCUUGAGCtc
KIF11	kinesin family member 11	3832	NM_004523,	118429	CCAUUUAAUUUGGCAGAGCtt	GCUCUGCCAAAUUAAAUGGtc
KIF2	kinesin heavy chain member 2	3796	NM_004520,	118424	GCCAAAGUAAACAAAAUUGtt	CAAUUUUGUUUACUUUGGCtg
KIF2	kinesin heavy chain member 2	3796	NM_004520,	118423	CCUGGAGAGCAUCUUUUCAtt	UGAAAAGAUGCUCUCCAGGtc
KIF20A	kinesin family member 20A	10112	NM_005733,	118443	GGUUAAAGCUAAAUUACAGtt	CUGUAAUUUAGCUUUAACCtc
KIF20A	kinesin family member 20A	10112	NM_005733,	118442	GCAGCAGGUUCCAUCUGAGtt	CUCAGAUGGAACCUGCUGCtt
KIF20A	kinesin family member 20A	10112	NM_005733,	118441	CCUGCUAUCAGACUGCUCUtt	AGAGCAGUCUGAUAGCAGGtt
KIF23	kinesin family member 23	9493	NM_138555,NM_004856,	118503	CCAUAACAUGUAUGUUGCAtt	UGCAACAUACAUGUUAUGGtt
KIF23	kinesin family member 23	9493	NM_138555,NM_004856,	118502	GGUUGAUGCCUUAUUAGAAtt	UUCUAAUAAGGCAUCAACCtc
KIF23	kinesin family member 23	9493	NM_138555,NM_004856,	118501	CCGAAAUGGAGACUAUAAGtt	CUUAUAGUCUCCAUUUCGGtt
KIF2C	kinesin family member 2C	11004	NM_006845,	118446	GCAACUUGUUUUGCAUAUGtt	CAUAUGCAAAACAAGUUGCtt
KIF2C	kinesin family member 2C	11004	NM_006845,	118445	GCUUCUUCCCUUACAUCCGtt	CGGAUGUAAGGGAAGAAGCtg
KIF2C	kinesin family member 2C	11004	NM_006845,	214569	GCAGGCUAGCAGACAAAUAtt	UAUUUGUCUGCUAGCCUGCtc
KIF4A	kinesin family member 4A	24137	NM_012310,	118455	GCGAAUGAAAAAAUGAACGtt	CGUUCAUUUUUUCAUUCGCtt
KIF4A	kinesin family member 4A	24137	NM_012310,	118454	GCAAGCGAAUGAAAAAAUGtt	CAUUUUUUCAUUCGCUUGCtc
KIF4A	kinesin family member 4A	24137	NM_012310,	118453	GGUAAUAGCCAUACUCUUAtt	UAAGAGUAUGGCUAUUACCtc
KIF5B	kinesin family member 5B	3799	NM_004521,	118428	GCUGAGUGGAAAACUUUAUtt	AUAAAGUUUUCCACUCAGCtt
KIF5B	kinesin family member 5B	3799	NM_004521,	118427	GCACAUCUCAAGAGCAAGUtt	ACUUGCUCUUGAGAUGUGCtt
KIF5B	kinesin family member 5B	3799	NM_004521,	118426	GCCUUAUGCAUUUGAUCGGtt	CCGAUCAAAUGCAUAAGGCtt
KIFC1	kinesin family member C1	3833	NM_002263,	118527	CCUAAAUGCAGAACUAAAAtt	UUUUAGUUCUGCAUUUAGGtc
KIFC1	kinesin family member C1	3833	NM_002263,	118526	GGCCAGACCACAGCUCAAAtt	UUUGAGCUGUGGUCUGGCCtt
KIFC1	kinesin family member C1	3833	NM_002263,	118525	CGACCAAAAUUACCACAUCtt	GAUGUGGUAAUUUUGGUCGtt
KNS2	kinesin 2 60/70kDa	3831	NM_005552,NM_182923,	118512	GCAUCUGGAGUUUAUGAAUtt	AUUCAUAAACUCCAGAUGCtt
KNS2	kinesin 2 60/70kDa	3831	NM_005552,NM_182923,	118511	GCACAAUUCCAUUUUACAAtt	UUGUAAAAUGGAAUUGUGCtc
KNS2	kinesin 2 60/70kDa	3831	NM_005552,NM_182923,	118510	GCUUUGAAGAAUGAGCACAtt	UGUGCUCAUUCUUCAAAGCtt
LATS1	LATS, large tumor suppressor, homolog 1 (D	9113	NM_004690,	567	GGAGUGUUACUCCUCCACCtt	GGUGGAGGAGUAACACUCCtt
LATS1	LATS, large tumor suppressor, homolog 1 (D	9113	NM_004690,	566	GGUUCUGAGAGUAAAAUUAtt	UAAUUUUACUCUCAGAACCtc
LATS1	LATS, large tumor suppressor, homolog 1 (D	9113	NM_004690,	565	GGACAGAGAGGCAUUAGUUtt	AACUAAGUCCUCUCUGUCCtt
LIMK1	LIM domain kinase 1	3984	NM_002314,NM_016735,	1413	GGACAAGAGGCUCAACUUCtt	GAAGUUGAGCCUCUUGUCCtt
LIMK1	LIM domain kinase 1	3984	NM_002314,NM_016735,	1318	GGUGACACACCGUGAGACAtt	UGUCUCACGGUGUGUCACCtt
LIMK1	LIM domain kinase 1	3984	NM_002314,NM_016735,	1223	GGAUCUAUGAUGGCCAGUAtt	UACUGGCCAUCAUAGAUCCtc
LOC285643	LOC285643	285643	XM_209695,	118521	GCAAGUGAAUGAAAAACUGtt	CAGUUUUUCAUUCACUUGCtc
LOC285643	LOC285643	285643	XM_209695,	118520	GGUAACAGCCACACUCUUAtt	UAAGAGUGUGGCUGUUACCtc
LOC285643	LOC285643	285643	XM_209695,	118519	CCAACAGUUGGCAUUAUUCtt	GAAUAAUGCCAACUGUUGGtt
MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast)	8379	NM_003550,NM_001013836,NM_001013	121449	CCAAAGUGCUGCACAUGAGtt	CUCAUGUGCAGCACUUUGGtc
MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast)	8379	NM_003550,NM_001013836,NM_001013	121448	GCGAUUGUGAAGAACAUGAtt	UCAUGUUCUUCACAAUCGCtg
MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast)	8379	NM_003550,NM_001013836,NM_001013	121447	GGAUGCAGCGAUUGUGAAGtt	CUUCACAAUCGCUGCAUCCtg
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	4085	NM_002358,	143483	GCGUGGCAUAUAUCCAUCUtt	AGAUGGAUAUAUGCCACGCtg
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	4085	NM_002358,	11455	GGAUGACAUGAGGAAAAUAtt	UAUUUUCCUCAUGUCAUCCtc
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	4085	NM_002358,	11361	GGAUGAAAUCCGUUCAGUGtt	CACUGAACGGAUUUCAUCCtg
MAP1B	microtubule-associated protein 1B	4131	NM_032010,NM_005909,	144102	GCUCAAACAUCUAGACUUUtt	AAAGUCUAGAUGUUUGAGCtt
MAP1B	microtubule-associated protein 1B	4131	NM_032010,NM_005909,	144101	GCCAGCUUAACCCUGUUCUtt	AGAACAGGGUUAAGCUGGCtt
MAP1B	microtubule-associated protein 1B	4131	NM_032010,NM_005909,	144100	CCCUUCUGAUGAAGCAGUCtt	GACUGCUUCAUCAGAAGGGtt
MAPRE1	microtubule-associated protein, RP/EB fami	22919	NM_012325,	136500	GCAGGUCAACGUAUUGAAAtt	UUUCAAUACGUUGACCUGCtg
MAPRE1	microtubule-associated protein, RP/EB fami	22919	NM_012325,	136499	GCUAAGCUAGAACACGAGUtt	ACUCGUGUUCUAGCUUAGCtt
MAPRE1	microtubule-associated protein, RP/EB fami	22919	NM_012325,	136501	GGUCAACGUAUUGAAACUUtt	AAGUUUCAAUACGUUGACCtg
MOBK1B	MOB1, Mps One Binder kinase activator-like	55233	NM_018221,	132366	GGCACAACAAGUAUUAUACtt	GUAUAAUACUUGUUGUGCCtc
MOBK1B	MOB1, Mps One Binder kinase activator-like	55233	NM_018221,	26356	GGAUCUCAUCAGUAUGAACtt	GUUCAUACUGAUGAGAUCCtt
MOBK1B	MOB1, Mps One Binder kinase activator-like	55233	NM_018221,	26450	GGGAGAGGAUCUCAAUGAAtt	UUCAUUGAGAUCCUCUCCCtc
MOBKL1A	MOB1, Mps One Binder kinase activator-like	92597	NM_173468,	148184	GGUUUUGGAGGUUAAUUUAtt	UAAAUUAACCUCCAAAACCtc
MOBKL1A	MOB1, Mps One Binder kinase activator-like	92597	NM_173468,	148183	CCAUAUUCUAUUGCUAGGGtt	CCCUAGCAAUAGAAUAUGGtt
MOBKL1A	MOB1, Mps One Binder kinase activator-like	92597	NM_173468,	148185	GGAUGGAUAAAACACUACAtt	UGUAGUGUUUUAUCCAUCCtg
MPP1	membrane protein, palmitoylated 1, 55kDa	4354	NM_002436,	121379	CCGAGGACAUGUACACCAAtt	UUGGUGUACAUGUCCUCGGtc
MPP1	membrane protein, palmitoylated 1, 55kDa	4354	NM_002436,	121381	GCACAGCUCGAUUUUUGAUtt	AUCAAAAAUCGAGCUGUGCtt
MPP1	membrane protein, palmitoylated 1, 55kDa	4354	NM_002436,	121380	GCCGUCUUCCUGCACUACAtt	UGUAGUGCAGGAAGACGGCtt
MSF	septin 9	10801	NM_006640,	135963	CGCACGAUAUUGAGGAGAAtt	UUCUCCUCAAUAUCGUGCGtg
MSF	septin 9	10801	NM_006640,	18321	GGAGGAGGUCAACAUCAACtt	GUUGAUGUUGACCUCCUCCtg
MSF	septin 9	10801	NM_006640,	18228	GGGCUUCGAGUUCAACAUCtt	GAUGUUGAACUCGAAGCCCtg
NEDD4	neural precursor cell expressed, developme	4734	NM_198400,NM_006154,	120779	GGGAUUCUUUGAACUAAUAtt	UAUUAGUUCAAAGAAUCCCtc
NEDD4	neural precursor cell expressed, developme	4734	NM_198400,NM_006154,	120778	GGACCUAUUAUGUAAACCAtt	UGGUUUACAUAAUAGGUCCtt
NEDD4	neural precursor cell expressed, developme	4734	NM_198400,NM_006154,	120777	GGAUAUCCUUGGAAGGACCtt	GGUCCUUCCAAGGAUAUCCtg

NEDD5	septin 2	4735	NM_004404,NM_001008491,NM_001008	14802	GGCUGUCUCAUUUAUUUUUtt	AAAAAUAAAUGAGACAGCCtg
NEDD5	septin 2	4735	NM_004404,NM_001008491,NM_001008	14709	GGCAAUACACAACAAGGUGtt	CACCUUGUUGUGUAUUGCCtt
NEDD5	septin 2	4735	NM_004404,NM_001008491,NM_001008	14614	GGCGGCACACAUCAUUGAUAAtt	UUAUCAAUGAUGUGCCGCCtg
PLCB2	phospholipase C, beta 2	5330	NM_004573,	15072	GGUUGAAGAGAGAGAUUAAtt	UUAAUCUCUCUCUUCAACCtc
PLCB2	phospholipase C, beta 2	5330	NM_004573,	14977	GGAGAUGGAGUUUCUGGAUtt	AUCCAGAAACUCCAUCUCCtt
PLCB2	phospholipase C, beta 2	5330	NM_004573,	14881	GGGCUACUACUUAUACUGGtt	CCAGUAUAAGUAGUAGCCCtt
PLK1	polo-like kinase 1 (Drosophila)	5347	NM_005030,	42856	GGAGGUGUUCGCGGGCAAGtt	CUUGCCCGCGAACACCUCCtt
PLK1	polo-like kinase 1 (Drosophila)	5347	NM_005030,	1341	GGUUUUCGAUUGCUCCCAGtt	CUGGGAGCAAUCGAAAACCtt
PLK1	polo-like kinase 1 (Drosophila)	5347	NM_005030,	103548	GGUGGAUGUGUGGUCCAUUtt	AAUGGACCACACAUCCACCtc
PMF1	polyamine-modulated factor 1	11243	NM_007221,	253762	CCUCUGAGAACGGCUGAAAtt	UUUCAGCCGUUCUCAGAGGta
PMF1	polyamine-modulated factor 1	11243	NM_007221,	253761	GUGAGGAGACCGCCAGCCCtt	GGGCUGGCGGUCUCCUCACtc
PMF1	polyamine-modulated factor 1	11243	NM_007221,	253760	UUUCGAGGGUGAAGCUCCUtt	AGGAGCUUCACCCUCGAAAtg
PNUTL1	septin 5	5413	NM_002688,NM_001009939,	12081	GGCAUUGCAUGAGAAGGUCtt	GACCUUCUCAUGCAAUGCCtt
PNUTL1	septin 5	5413	NM_002688,NM_001009939,	11993	GGGAGUCAAGCUGAAGCUCtt	GAGCUUCAGCUUGACUCCCtt
PNUTL1	septin 5	5413	NM_002688,NM_001009939,	11898	GGCUUUGACUUCACACUCAtt	UGAGUGUGAAGUCAAAGCCtt
PNUTL2	septin 4	5414	NM_080415,NM_080416,NM_080417,NN	142770	GCUUGAUCCCUAUGAUUCCtt	GGAAUCAUAGGGAUCAAGCtt
PNUTL2	septin 4	5414	NM_080415,NM_080416,NM_080417,NN	14882	GGAUUUCUCAGGAAAUGCGtt	CGCAUUUCCUGAGAAAUCCtt
PNUTL2	septin 4	5414	NM_080415,NM_080416,NM_080417,NN	15073	GGAGAUGCUACACAAAAUAtt	UAUUUUGUGUAGCAUCUCCtg
PRC1	protein regulator of cytokinesis 1	9055	NM_003981,NM_199413,NM_199414,	137453	GCGGUUACAAAGAACUGAGtt	CUCAGUUCUUUGUAACCGCtg
PRC1	protein regulator of cytokinesis 1	9055	NM_003981,NM_199413,NM_199414,	137452	CCAGCGGUUACAAAGAACUtt	AGUUCUUUGUAACCGCUGGtc
PRC1	protein regulator of cytokinesis 1	9055	NM_003981,NM_199413,NM_199414,	137454	CCAUUAUGUCUGGGUCAAAtt	UUUGACCCAGACAUAAUGGtg
PTTG1	pituitary tumor-transforming 1	9232	NM_004219,	42068	GAGUUUGUGUGUAUUUGUAtt	UACAAAUACACACAAACUCtg
PTTG1	pituitary tumor-transforming 1	9232	NM_004219,	41990	GUCUGUAAAGACCAAGGGAtt	UCCCUUGGUCUUUACAGACtt
PTTG1	pituitary tumor-transforming 1	9232	NM_004219,	41900	GAUCUCAAGUUUCAACACCtt	GGUGUUGAAACUUGAGAUCtc
RAB8A	RAB8A, member RAS oncogene family	4218	NM_005370,	3022	GGAAAGCACAAAUGAAGGAtt	UCCUUCAUUUGUGCUUUCCtt
RAB8A	RAB8A, member RAS oncogene family	4218	NM_005370,	2930	GGGAGUCAAAAUCACACCGtt	CGGUGUGAUUUUGACUCCCtg
RAB8A	RAB8A, member RAS oncogene family	4218	NM_005370,	2836	GGCCAACAUCAAUGUGGAAtt	UUCCACAUUGAUGUUGGCCtt
RABGAP1	RAB GTPase activating protein 1	23637	NM_012197,	147469	GGAAUUGUGAGACUCUUAGtt	CUAAGAGUCUCACAAUUCCtt
RABGAP1	RAB GTPase activating protein 1	23637	NM_012197,	147468	CCCUGUGCCAUUAGUAGGGtt	CCCUACUAAUGGCACAGGGtt
RABGAP1	RAB GTPase activating protein 1	23637	NM_012197,	147470	GCUGUAAGCCGGAUACUUUtt	AAAGUAUCCGGCUUACAGCtt
RACGAP1	Rac GTPase activating protein 1	29127	NM_013277,	20374	GGCAACUUUUUUGCUUCUCtt	GAGAAGCAAAAAAGUUGCCtt
RACGAP1	Rac GTPase activating protein 1	29127	NM_013277,	20283	GGAUUUCCGUAAAAAGUGGtt	CCACUUUUUACGGAAAUCCtc
RACGAP1	Rac GTPase activating protein 1	29127	NM_013277,	20189	GGACUUUGAGGAUUUCCGUtt	ACGGAAAUCCUCAAAGUCCtt
RASSF1	Ras association (RalGDS/AF-6) domain fan	11186	NM_007182,NM_170712,NM_170713,NN	137076	GGAUAUCCUUAUCAGAGCUtt	AGCUCUGAUAAGGAUAUCCtg
RASSF1	Ras association (RalGDS/AF-6) domain fan	11186	NM_007182,NM_170712,NM_170713,NN	137075	CGGUUCUUACACAGGCUUCtt	GAAGCCUGUGUAAGAACCGtc
RASSF1	Ras association (RalGDS/AF-6) domain fan	11186	NM_007182,NM_170712,NM_170713,NN	40417	GGUUCAGCUGAAGCUGGUGtt	CACCAGCUUCAGCUGAACCtt
Rif1	RAP1 interacting factor homolog (yeast)	55183	NM_018151,	223389	CCACAUUUUUUCAGAACAGtt	CUGUUCUGAAAAAAUGUGGtt
Rif1	RAP1 interacting factor homolog (yeast)	55183	NM_018151,	223388	CCUGAGUUCGAAUUUAGGUtt	ACCUAAAUUCGAACUCAGGtt
Rif1	RAP1 interacting factor homolog (yeast)	55183	NM_018151,	223387	GCACUUUGGGUGAUAUCUAtt	UAGAUAUCACCCAAAGUGCtc
ROCK2	Rho-associated, coiled-coil containing prote	9475	NM_004850,	596	GGAAAGAACUUUAAAACAGtt	CUGUUUUAAAGUUCUUUCCtc
ROCK2	Rho-associated, coiled-coil containing prote	9475	NM_004850,	595	GGUGCUUUUGGUGAAGUGCtt	GCACUUCACCAAAAGCACCtc
ROCK2	Rho-associated, coiled-coil containing prote	9475	NM_004850,	110867	GGCCACAAAGGCACGACUAtt	UAGUCGUGCCUUUGUGGCCtt
RSN	restin (Reed-Steinberg cell-expressed interr	6249	NM_002956,NM_198240,	241427	AUCAAUUACCAAAGGUGAUtt	AUCACCUUUGGUAAUUGAUtc
RSN	restin (Reed-Steinberg cell-expressed interr	6249	NM_002956,NM_198240,	241426	CCUUCAGUUCCGGGUUGAAtt	UUCAACCCGGAACUGAAGGtc
RSN	restin (Reed-Steinberg cell-expressed interr	6249	NM_002956,NM_198240,	142518	GGUAAAUCGGAAAUGAAGAtt	UCUUCAUUUCCGAUUUACCtt
RSU1	Ras suppressor protein 1	6251	NM_152724,NM_012425,	143078	CCGAGAUAUGCUAAUUUAAtt	UUAAAUUAGCAUAUCUCGGtg
RSU1	Ras suppressor protein 1	6251	NM_152724,NM_012425,	108352	GGCCGUAGCAGUUUGACGAtt	UCGUCAAACUGCUACGGCCtt
RSU1	Ras suppressor protein 1	6251	NM_152724,NM_012425,	3713	GGUGCUCAACUUUUUUAAUtt	AUUAAAAAAGUUGAGCACCtc
SEC5L1	SEC5-like 1 (S. cerevisiae)	55770	NM_018303,	147719	GGUGCAAGUUUUCAAGAAAtt	UUUCUUGAAAACUUGCACCtc
SEC5L1	SEC5-like 1 (S. cerevisiae)	55770	NM_018303,	147718	GGACUUAGAAAUGCUAUUCtt	GAAUAGCAUUUCUAAGUCCtt
SEC5L1	SEC5-like 1 (S. cerevisiae)	55770	NM_018303,	147717	GGAACAAAGGAAUUCCGCCtt	GGCGGAAUUCCUUUGUUCCtg
SEC8L1	SEC8-like 1 (S. cerevisiae)	60412	NM_021807,NM_001037126,	147808	GCAUAAGCAUGUCCUGAACtt	GUUCAGGACAUGCUUAUGCtc
SEC8L1	SEC8-like 1 (S. cerevisiae)	60412	NM_021807,NM_001037126,	147807	GCCUACGAGAAAUGUGACCtt	GGUCACAUUUCUCGUAGGCtt
SEC8L1	SEC8-like 1 (S. cerevisiae)	60412	NM_021807,NM_001037126,	29348	GGACUCUGUCUACUAGUGAtt	UCACUAGUAGACAGAGUCCtg
SMC2L1	SMC2 structural maintenance of chromosor	10592	NM_006444,	135817	GGGCCGAAUUACAAAAGUAtt	UACUUUUGUAAUUCGGCCCtg
SMC2L1	SMC2 structural maintenance of chromosor	10592	NM_006444,	135816	GGUGGUUAUUGGUGGUAGAtt	UCUACCACCAAUAACCACCtg
SMC2L1	SMC2 structural maintenance of chromosor	10592	NM_006444,	135818	GCUCUAUUAGCCAGAUUUCtt	GAAAUCUGGCUAAUAGAGCtt
SMC4L1	SMC4 structural maintenance of chromosor	10051	NM_005496,NM_001002799,NM_001002	137983	GCAAUUAGAUGAAUGUGCUtt	AGCACAUUCAUCUAAUUGCtt
SMC4L1	SMC4 structural maintenance of chromosor	10051	NM_005496,NM_001002799,NM_001002	137982	GGCCAGACUGAACACGAUGtt	CAUCGUGUUCAGUCUGGCCtt
SMC4L1	SMC4 structural maintenance of chromosor	10051	NM_005496,NM_001002799,NM_001002	137981	GCUCCUCGGCUUAUGAUAAtt	UUAUCAUAAGCCGAGGAGCtc
SPAG5	sperm associated antigen 5	10615	NM_006461,	135834	GGCCCGUUUAGAUACCAUGtt	CAUGGUAUCUAAACGGGCCtc
SPAG5	sperm associated antigen 5	10615	NM_006461,	135833	CCAUCGUCCUUGUACCAUCtt	GAUGGUACAAGGACGAUGGtt
SPAG5	sperm associated antigen 5	10615	NM_006461,	135835	GCAGUAGAAGAUGUUGGUAtt	UACCAACAUCUUCUACUGCtg
STK3	serine/threonine kinase 3 (STE20 homolog,	6788	NM_006281,	792	GGAUAGUUUUUCAAAUAGGtt	CCUAUUUGAAAAACUAUCCtg
STK3	serine/threonine kinase 3 (STE20 homolog,	6788	NM_006281,	791	GGACAUGCAAAAUUGGCAGtt	CUGCCAAUUUUGCAUGUCCtt
STK3	serine/threonine kinase 3 (STE20 homolog,	6788	NM_006281,	793	GGGUCUUAUGGAAGUGUAUtt	AUACACUUCCAUAAGACCCtt
STK38	serine/threonine kinase 38	11329	NM_007271,	865	GGAUUUUCUUCCAUAAUGCtt	GCAUUAUGGAAGAAAAUCCtc
STK38	serine/threonine kinase 38	11329	NM_007271,	864	GGAGAAACGACUCCGGAGAtt	UCUCCGGAGUCGUUUCUCCtc
STK38	serine/threonine kinase 38	11329	NM_007271,	863	GGCCUAAAAGAUGAGGAGAtt	UCUCCUCAUCUUUUAGGCCtt
STK4	serine/threonine kinase 4	6789	NM_006282,	795	GGAACUAUGAAAAGAAGGGtt	CCCUUCUUUUCAUAGUUCCtt
STK4	serine/threonine kinase 4	6789	NM_006282,	794	GGGACUUGAAUACCUUCAUtt	AUGAAGGUAUUCAAGUCCCtt
STK4	serine/threonine kinase 4	6789	NM_006282,	796	GGGUCCUAUGGCAGCGUAUtt	AUACGCUGCCAUAGGACCCtt
STK6	serine/threonine kinase 6	6790	NM_003600,NM_198433,NM_198434,NN	427	GGCAACCAGUGUACCUCAUtt	AUGAGGUACACUGGUUGCCtg
STK6	serine/threonine kinase 6	6790	NM_003600,NM_198433,NM_198434,NN	426	GGAACUGGCAUCAAAACAGtt	CUGUUUUGAUGCCAGUUCCtc
STK6	serine/threonine kinase 6	6790	NM_003600,NM_198433,NM_198434,NN	425	GGUCCAAAACGUGUUCUCGtt	CGAGAACACGUUUUGGACCtc
STMN1	stathmin 1/oncoprotein 18	3925	NM_005563,NM_203399,NM_203401,	144044	GCGUGUUUCUAGAGAACAGtt	CUGUUCUCUAGAAACACGCtt
STMN1	stathmin 1/oncoprotein 18	3925	NM_005563,NM_203399,NM_203401,	144043	GCUGACUAAUUUGUUCUGAtt	UCAGAACAAAUUAGUCAGCtt
STMN1	stathmin 1/oncoprotein 18	3925	NM_005563,NM_203399,NM_203401,	144042	GCCCUCGGUCAAAAGAAUCtt	GAUUCUUUUGACCGAGGGCtg
STMN3	stathmin-like 3	50861	NM_015894,	134696	CCUGUGUGUUUAAGACAUGtt	CAUGUCUUAAACACACAGGtg
STMN3	stathmin-like 3	50861	NM_015894,	134695	GGGCUGGGAUAUUCCUCAUtt	AUGAGGAAUAUCCCAGCCCtc
STMN3	stathmin-like 3	50861	NM_015894,	134694	CGUUCGGGUUUUGGUUUUGtt	CAAAACCAAAACCCGAACGtg
TD-60	regulator of chromosome condensation 2	55920	NM_018715,	123152	GCUUUGCUAAACUAACCUAtt	UAGGUUAGUUUAGCAAAGCtc
TD-60	regulator of chromosome condensation 2	55920	NM_018715,	123151	GGGUCAAAGUGCAAAGGGCtt	GCCCUUUGCACUUUGACCCtt
TD-60	regulator of chromosome condensation 2	55920	NM_018715,	123150	GGAGCGCGUCAAACUUGAAtt	UUCAAGUUUGACGCGCUCCtt
TPX2	TPX2, microtubule-associated protein homo	22974	NM_012112,	136426	GCCAAGGGUAGACAUACUGtt	CAGUAUGUCUACCCUUGGCtt
TPX2	TPX2, microtubule-associated protein homo	22974	NM_012112,	136425	GGGCAAAACUCCUUUGAGAtt	UCUCAAAGGAGUUUUGCCCtg
TPX2	TPX2, microtubule-associated protein homo	22974	NM_012112,	136427	CCUUGCCCUACUAAGAUUUtt	AAAUCUUAGUAGGGCAAGGtc
TUBG1	tubulin, gamma 1	7283	NM_001070,	120784	GAACCUGUCGCCAGUAUGAtt	UCAUACUGGCGACAGGUUCtc
TUBG1	tubulin, gamma 1	7283	NM_001070,	120193	CCUGUCGCCAGUAUGACAAtt	UUGUCAUACUGGCGACAGGtt
TUBG1	tubulin, gamma 1	7283	NM_001070,	9227	GGGAGAAAAGAUCCAUGAGtt	CUCAUGGAUCUUUUCUCCCtg
TUBG2	tubulin, gamma 2	27175	NM_016437,	120877	GCAGAUGGAAGUGACAGUUtt	AACUGUCACUUCCAUCUGCtt
TUBG2	tubulin, gamma 2	27175	NM_016437,	120570	GCUUCACCUCAUGGACAACtt	GUUGUCCAUGAGGUGAAGCtg
TUBG2	tubulin, gamma 2	27175	NM_016437,	120876	CCGAGAAGCAGAUGGAAGUtt	ACUUCCAUCUGCUUCUCGGtc
ZW10	ZW10 homolog, centromere/kinetochore pro	9183	NM_004724,	137635	GCAGUUGGAAAGAGACUCAtt	UGAGUCUCUUUCCAACUGCtg
ZW10	ZW10 homolog, centromere/kinetochore pro	9183	NM_004724,	137634	GGUGUGCAAUAUGAUUAGCtt	GCUAAUCAUAUUGCACACCtc

ZW10	ZW10 homolog, centromere/kinetochore pro	9183	NM_004724,	137636	GCAAAUCGGAGAUAUUUUAtt	UAAAAUAUCUCCGAUUUGCtc
ZYX	zyxin	7791	NM_003461,NM_001010972,	115435	CCUCCCAGCUUCACCUAUGtt	CAUAGGUGAAGCUGGGAGGtt
ZYX	zyxin	7791	NM_003461,NM_001010972,	115434	GCAGUAUUGAUUUGGAGAUtt	AUCUCCAAAUCAAUACUGCtc
ZYX	zyxin	7791	NM_003461,NM_001010972,	139068	CCCAACAUGGUCUAGGGAUtt	AUCCCUAGACCAUGUUGGGtc

Supplementary Table 3

	Inter	*Pro*	*Meta*	*Ana*
Precision (positive predictive value)	*97.9%*	*95.3%*	*98.7%*	*98.4%*
Sensitivity (recall)	*97.9%*	*96.5%*	*97.4%*	*98.4%*

Performance of support vector machine prediction on four classes, similar to an analysis by Wang M., et al., *Context-based mixture model for cell phase identification in automated fluorescence microscopy*, BMC Bioinformatics (2007). Manually annotated objects for interphase, prophase, metaphase, and anaphase class of the same data shown in Fig. 1d were used for training. Precision and sensitivity of class predictions were calculated for each class individually, considering the respective class as positive, and the respective other classes as negative. The calculations were based on the amount of true-positive (tp), true-negative (tn), false-positive (fp), and false-negative (fn) predictions. Precision is defined as tp / (tp + fp), which sometimes is also referred to as *positive predictive value*. Sensitivity is defined as tp / (tp + fn), sometimes referred to as *recall*.

Supplementary Table 4

	Inter	**Pro**	**Prometa**	**Meta**	**Early ana**	**Late ana**	**Telo**
Precision (w/o HMM)	99.8%	80.2%	84.6%	99.8%	62.7%	94.2%	82.9%
Precision (HMM)	100.0%	100.0%	98.0%	99.8%	93.1%	100.0%	96.4%
Sensitivity (w/o HMM)	95.5%	99.5%	97.0%	86.0%	75.0%	96.1%	97.5%
Sensitivity (HMM)	100.0%	100.0%	100.0%	95.7%	96.4%	99.3%	100.0%

Per class prediction performance compared with manual annotation of data without error correction (Fig. 2a) and with HMM error correction (Fig. 2f).

Supplementary Movie legends

Movie 1. Time-lapse imaging of HeLa cells stably expressing the fluorescent chromatin marker H2B-mCherry (imaged with widefield epifluorescence 20x dry objective). The movie shows a region of interest of 512*512*30 (x*y*t; overall movie dimensions: 1392*1040*206 (x*y*t); time-lapse: 4.6 min.

Movie 2. Object detection and supervised classification of morphologies. The contours were derived by the automated segmentation, and the color code for different morphology classes is as indicated in legend of Fig. 1B. Original data is shown in Suppl. Movie 1.

Movie 3. Automated extraction of mitotic events. The movie displays 100 randomly selected examples for cells progressing through mitosis (same as in Fig. 2A). The cells were *in silico* synchronized to the prophase - prometaphase transition and sorted based on total prometaphase and metaphase duration. The morphology classes annotated as in Fig. 2A are indicated by color-coding as in the legend of Fig. 1B.

Movie 4. Classification error correction based on free hidden Markov model. The same cells as shown in Fig. 2A and Suppl. Movie 3 were classified based on morphological features as well as the temporal context.

Movie 5. Time-lapse imaging of HeLa cells stably expressing the fluorescent chromatin marker H2B-mCherry (red) and mEGFP-α-tubulin (green) with widefield epifluorescence 20x dry objective. The movie shows a region of interest of 512*512*30 (x*y*t). The overall movie dimensions were 1392*1040*206 (x*y*t); time-lapse: 4.6 min.

Movie 6. Annotation of spindle dynamics in movies of cells expressing H2B-mCherry and mEGFP-α-tubulin. The movie displays 100 randomly selected examples for automatically annotated cells progressing through mitosis (same as in Fig. 3D). The cells were *in silico* synchronized to the prophase - prometaphase transition in the H2B-mCherry channel and sorted by total prometaphase and metaphase duration. The morphology classes are indicated by color-coding as indicated in the legend of Fig. 3A.

Movie 7. Time-lapse imaging of HeLa cells stably expressing the fluorescent chromatin marker H2B-mCherry (red) and GalT-EGFP (green) with widefield epifluorescence 10x dry objective. The movie shows a region of interest of 512*512*30 (x*y*t). The overall movie dimensions were 1392*1040*482 (x*y*t); time-lapse: 2.8 min.

Movie 8. Annotation of Golgi dynamics in movies of cells expressing H2B-mCherry and GalT-EGFP. The movie displays 100 randomly selected examples for automatically annotated cells progressing through mitosis (same as in Fig. 3E). The cells were *in silico* synchronized to the prophase - prometaphase transition in the H2B-mCherry channel and sorted by total prometaphase and metaphase duration. The morphology classes are indicated by color-coding as indicated in the legend of Fig. 3B.

Movie 9. Time-lapse imaging of HeLa cells stably expressing the fluorescent chromatin marker H2B-mCherry (red) and DNA replication factory marker EGFP-PCNA (green) with widefield epifluorescence 10x dry objective. The movie shows a region of interest of 350*350*54 (x*y*t; every 2nd time point shown). The overall movie dimensions were 1392*1040*482 (x*y*t); time-lapse: 5.9 min.

Movie 10. Annotation of S-phase progression in movies of cells expressing H2B-mCherry and EGFP-PCNA. The movie displays 100 randomly selected examples for automatically annotated cells progressing through the cell cycle (same as in Fig. 3F). The cells were *in silico* synchronized to the G1 – early S transition in the EGFP-PCNA channel and sorted by total S-phase duration. Every 2nd time point of original data is shown. The morphology classes are indicated by color-coding as indicated in the legend of Fig. 3C.

Movie 11. Time-lapse imaging of untreated control HeLa cells stably expressing H2B-mCherry and Securin-mEGFP with widefield epifluorescence 20x dry objective. The movie shows a region of interest of 400*400*100 (x*y*t). The overall movie dimensions were 1392*1040*500 (x*y*t); time-lapse: 2.7 min.

Movie 12. Time-lapse imaging of Mad2 siRNA transfected HeLa cells stably expressing H2B-mCherry and Securin-mEGFP with widefield epifluorescence 20x dry objective. The movie shows a region of interest of 400*400*100 (x*y*t). The overall movie dimensions were 1392*1040*500 (x*y*t; time-lapse: 2.7 min).

Movie 13. Time-lapse imaging of HeLa cells stably expressing H2B-mCherry and Securin-mEGFP with widefield epifluorescence 20x dry objective, treated with 50 ng/ml Nocodazol immediately before starting the imaging. The movie shows a region of interest of 400*400*100 (x*y*t). The overall movie dimensions were 1392*1040*500 (x*y*t); time-lapse: 2.7 min.

Movie 14. Time-lapse imaging of control HeLa cells stably expressing H2B-mCherry and IBB-EGFP transfected with non-silencing siRNA, using widefield epifluorescence 10x dry objective. The movie shows 80 time frames of the entire imaging field downsampled in x/y by a factor of 2 for display. Original movie dimensions: 1392*1040*744 (x*y*t); time-lapse: 3.7 min. 108 movies of different RNAi conditions were captured simultaneously in this experiment by multi-location time-lapse imaging.

Movie 15. Time-lapse confocal imaging of HeLa cells stably expressing H2B-mCherry and mEGFP-α-tubulin (63x oil immersion objective). Cells were transfected with non-silencing siRNA. Movie dimensions are 512*512*132 (x*y*t); time-lapse: 7.1 min.

Movie 16. Time-lapse confocal imaging of HeLa cells stably expressing H2B-mCherry and mEGFP-α-tubulin (63x oil immersion objective). Cells were transfected with siRNA targeting Cdc20. Movie dimensions are 512*512*132 (x*y*t); time-lapse: 7.1 min.