An experimental and computational framework to build a dynamic protein atlas of human cell division

Yin Cai[1,3]*, M. Julius Hossain[1]*, Jean-Karim Hériché[1], Antonio Z. Politi[1], Nike Walther[1], Birgit Koch[1,4], Malte Wachsmuth[1,5], Bianca Nijmeijer[1], Moritz Kueblbeck[1], Marina Martinic[2], Rene Ladurner[2,6], Jan-Michael Peters[2] & Jan Ellenberg[1]

1 European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. Correspondence should be addressed to J.E. (jan.ellenberg@embl.de)

2 Research Institute of Molecular Pathology (IMP), Vienna, Austria.

3 Current address: Roche Diagnostics, Waiblingen, Germany

4 Current address: Max Planck Institute for Medical Research, Heidelberg, Germany

5 Current address: Luxendo GmbH, Heidelberg, Germany

6 Current address: Stanford School of Medicine, Stanford (CA), USA

* These authors contributed equally to this work

Essential biological functions of human cells, such as division, require the tight coordination of the activity of hundreds of proteins in space and time. While live cell imaging is a powerful tool to study the distribution and dynamics of individual proteins after fluorescence tagging, it has not yet been used to map protein networks due to the lack of systematic and quantitative experimental and computational approaches. Using the cell and nuclear boundaries as landmarks, we generated a 4D image data-driven, canonical, computational model for the morphological changes during mitotic progression of human cells. We show that this model can be used to integrate the dynamic distribution of 3D concentration data for many mitotic proteins imaged by absolutely calibrated fluorescence microscopy in a large number of dividing cells. Analysis of a pilot data set, containing 28 proteins of interest imaged in dividing cells, allowed us to automatically identify sub-cellular structures and quantify the timing and magnitude of protein fluxes between them, as well as predicting dynamic multi-molecular biological processes such as organelle dis/assembly. Our integrated experimental and computational method enables building a 4D protein atlas of the dividing human cell. As part of the MitoSys consortium, we provide here an approach that is generic and allows the systematic mapping and mining of dynamic protein localization networks that drive cellular functions.

INTRODUCTION

The specific localization, the timing of particular interactions and the favored binding partners are all crucial to ensure correct function and regulation of proteins and their complexes[1–3]. Moreover, changes in cellular shape and internal structure, which are determined by the activities of proteins, can often feedback on their functions inside the cell[4]. To understand how cells function at the molecular level, it is therefore essential to understand how networks of proteins behave within dynamic cellular boundaries. Mitosis as a fundamental process of life is a good example of a dynamic process orchestrated by a large network of proteins. Within an hour, cells undergo dramatic changes in their morphology and sub-cellular organization. The MitoCheck consortium identified hundreds of essential proteins[5] which need to act in a coordinated fashion in both space and time to drive these changes and respond to them to ensure correct division.

Despite a considerable amount of data on individual mitotic proteins and complexes, our comprehensive understanding of human cell division as a system is limited. To overcome this limitation, the integration of the current knowledge into a single model is necessary. This model can then be used to predict and analyze the dynamic interactions between the molecular parts of the mitotic machinery. Building such a model would require standardized and quantitative data on many different proteins observed in live dividing human cells, as well as the computational tools to integrate and mine such large data sets. Due to the advent of endogenous gene tagging and automated fluorescence correlation spectroscopy (FCS)[6], quantitative 4D protein localization data can now be generated not only in yeast[7], but also in human cells. Although previous computational studies on multiple subcellular localizations have achieved automatic recognition[8,9] and modeling of individual sub-cellular structures either with descriptive models[10,11] or with generative models[12,13], we are still lacking computational methods that can deal with the complexity of dynamic morphological changes of cells and the simultaneous localization of proteins to multiple structures.

Within the MitoSys consortium, we developed a new integrated experimental and computational framework for time-resolved, 3D protein localization analysis within a dynamically changing cellular context. It integrates the large amount of 3D+time image data acquired by live imaging of fluorescently tagged mitotic proteins by automatic FCS-calibrated confocal microscopy, into a single canonical model of the human dividing cell. This allows both the estimation of subcellular fluxes and the prediction of interactions between all proteins in the model.

RESULTS

Automatic generation of a quantitative 4D protein localization data set

To provide standardized, quantitative data on the dynamic 3D localization of mitotic proteins, we imaged HeLa cell lines, in which such proteins were knocked-in fluorescently (see Online Methods). For each protein, the cell surface and chromosome volume were recorded in separate channels as spatio-temporal landmarks (using either histone H2B-mCherry or SiR-DNA[14] and extracellular dextran-Dy481XL, see Online Methods). To effectively record high-resolution data during mitosis, cells beginning cell division (prophase), were automatically detected in low resolution images using the chromosome signal. They were then imaged by high resolution, confocal microscopy in 3D for one hour until division was completed (**Fig. 1a**). At the end of each 3D mitosis movie, FCS[15] was performed in the mitotic protein channel at six automatically selected positions in the nucleus and cytoplasm (**Fig. 1a**). This yielded the number of fluorescent protein molecules within the effective illumination volume. By correlating molecule numbers with fluorescence intensities (**Fig. 1b**), we calibrated the microscope detector and converted protein fluorescence movies to time-resolved protein concentration distribution maps (**Fig. 1c**). Using this automatic experimental pipeline, we acquired a pilot dataset for 28 proteins. Most of these proteins were tagged with EGFP at all alleles of the endogenous locus by targeted recombination using zinc finger nucleases[16] or CRISPR-Cas9 nickase[17], while for some genes stable integration of cDNAs or bacterial artificial chromosomes[18] had to be used (**Supplementary Table S1,** see Online Methods). Time-resolved 3D protein distribution was recorded for 18 dividing cells per protein on average (499 cells in total, ranging from 10 to 35 cells per protein, Fig. 1d), giving us a sufficiently large dataset to develop and test our computational framework.

Definition of mitotic standard time

Traditionally, mitosis is divided into five stages: prophase, prometaphase, metaphase, anaphase and telophase[19]. However, cell division is a continuous process. Except for nuclear envelope breakdown and chromosome segregation that mark the onset of prometaphase and anaphase respectively, the other stages are not separated by sharp kinetic boundaries. Moreover, mitotic progression varies significantly between individual cells (**Fig. 1d** and **Supplementary Fig. S1**). To align the kinetic data from different cells, we therefore first defined a "mitotic standard time" based on objectively measured changes in chromosome morphology. To this end, cell and

chromosome boundaries of all imaged mitoses were automatically segmented in 3D using the landmark channels (**Fig. 2a** and **Supplementary Fig. S2**; see Online Methods). Three geometric features were extracted from the segmented data: the distance between the two segregated chromosome masses, the total chromosome volume and the third eigenvalue of the chromosome mass (**Fig. 2b**). Each mitosis movie could thus be represented as a six-dimensional vector sequence of these parameters and their first derivative indicating kinetic transitions. To align the vector sequences, we used a modified Barton-Sternberg algorithm with customized multidimensional dynamic time warping (see Online Methods and **Supplementary Fig. S3a-c**) that locally stretches or compresses the sequences such that sections with similar feature values are assigned to the same standard time point (**Supplementary Fig. S3a, b**). After a few iterations, the alignment converged and all vector sequences displayed very similar kinetic signatures (**Fig. 2c** and **Supplementary Fig. S3d**), allowing us to use the final alignment as a mitotic standard time reference (**Supplementary Fig. S4**). We identified reproducible kinetic transitions by the local maximums in the second derivative of the average feature sequences (**Fig. 2d** and **Supplementary Fig. S5**, see Online Methods), which objectively defined a set of 20 clearly distinguishable mitotic steps. To check that these mitotic standard stages represented real biological transitions, we automatically selected the 3D image stack closest to the average feature values of each stage (**Supplementary Fig. S6**). Although the images picked in this way are from different cells, their automatically assigned sequential order reconstitutes a virtual mitosis with an error-free chronology (**Fig. 2e**), in which all classically known mitotic transitions such as nuclear envelope breakdown (between stage 2 and 3), and anaphase onset (between stage 11 and 12), were correctly identified. Moreover, the method could identify previously hard-to-define stages such as the first formation of the metaphase plate in late prometaphase (between stage 7 and 8), and could differentiate between the different anaphase and telophase stages (stage 12 to 18). Thus, our mitotic standard time provides a quantitative and objective way to define mitotic stages.

A canonical model of the mitotic HeLa cell for interactive exploration of dynamic protein distributions

The mitotic standard time provides an objective way to map all cell images to a standard time reference or bin them to 20 standard stages for averaging. In order to enable visualization, interactive navigation and analysis of all imaged mitotic protein distributions, we next computed the canonical geometry from late prometaphase (stage 7) to cytokinesis (stage 20), where little

deviation from rotational symmetry around the division axis occurs. As illustrated with three example cells in late metaphase 1 (stage 10; **Supplementary Fig. S7a**), after translational and rotational registrations based on their chromatin masses (**Supplementary Fig. S7b**), the mean geometry was computed from the cylindrical coordinates of all segmented surfaces of chromatin and cell masses (**Supplementary Fig. S7c-f**; see Online Methods). The canonical geometry model was then reconstructed from the average geometry of several hundred cells for each analyzed mitotic standard stage (**Supplementary Fig. S7g**), preserving the information on cell-to-cell variability (**Supplementary Fig. S7h**).

Evolution of the mitotic standard geometry over the mitotic standard time defines the 4D canonical mitotic cell model (**Supplementary Fig. S7g**), enabling us to register all cell divisions for the 28 proteins in space and time based on the landmark channels (see Online Methods). For each protein, we mapped each 3D stack to the corresponding mitotic standard stage and then computed the mean and variance of the concentration for each standard voxel and time point of the canonical cell (**Fig. 3a,** see Online Methods). This resulted in 4D concentration maps representing the average behavior and biological variability of each mitotic protein (**Supplementary Fig. S8**). Maps of many proteins can then be freely combined (**Fig. 3b**), to compare their localization patterns, dynamics and abundance interactively and provide intuitive navigation of the integrated data set (www.mitocheck.org/mitotic_cell_atlas)**.**

To illustrate the power of integrated data exploration for multiple proteins in the canonical model, we investigated the nuclear/chromosome localization of eight mitotic chromosome structure proteins throughout mitosis (**Fig. 3c** and **Supplementary Fig. S9a-b**). Five out of eight showed a clear chromatin localization from NEBD to anaphase (**Fig. 3c**). Plotting the total number of proteins bound to mitotic chromosomes and in the daughter nuclei against the mitotic standard time provided a quantitative comparison of protein dynamics and revealed (**Fig. 3d-e,** see Online Methods) that the amount of most chromosomal proteins bound to the genome in metaphase is within the same order of magnitude (3 000 – 26 000 per cell and genome), except for TOP2A which shows a 25 times higher abundance, potentially suggesting a structural rather than a purely enzymatic role (**Fig. 3d-e,** see Online Methods). We found cohesins to slowly and progressively dissociate from chromatin (RAD21, STAG1, and STAG2), with a final more rapid release of approximately 3 000 remaining cohesin complexes before anaphase onset, indicating that no more than 100 cohesins are sufficient to connect the sister chromatids on an average human chromosome, mostly at the kinetochore (see Online Methods and [20]). Interestingly, chromatin bound cohesins consisted of very similar amounts of the two isoforms containing the

HEAT repeat subunits STAG1/2, which summed up to the concentration of the shared kleisin subunit RAD21 (Fig. 3e). This is in contrast to the situation in interphase nuclei, where STAG2-containing complexes dominate[21]. Furthermore, we observed that a significant amount of STAG2 ($p < 0.025$, paired Wilcoxon signed rank-test), but not RAD21 appears to rebind chromosomes in anaphase, suggesting a potential non-cohesive function on chromosomes at anaphase onset **(Fig. 3d-e, Supplementary Fig. S9a)**. In contrast to the complete dissociation of cohesin, about 17 000 molecules of the chromatin organizer CTCF remain associated with the genome throughout mitosis[22], consistent with a "bookkeeping" mechanism of interphase chromosome architecture. Once chromosome segregation was initiated, KIF4A, TOP2A and CTCF further accumulated on chromatin in anaphase, which suggests a role for these proteins in further DNA compaction that has been observed in anaphase[23] (Fig. 3d-e). Interestingly, the cytoplasmic pool of mitotic chromosome proteins shows a stepwise order of re-entry into the reformed nucleus and different rates of import. CTCF is reimported first with the highest rate (391 proteins/sec), followed by simultaneous import of the cohesin subunits STAG1 and RAD21 (352 and 239 proteins/sec, respectively), while STAG2 and WAPL enter the nucleus later and at a lower rate (69 and 89 proteins/sec, respectively). This shows that mitotic decondensation can proceed in the presence of CTCF and subsequently STAG1-containing cohesin complexes, while WAPL and STAG2 are not present.

In addition to the chromosomal proteins, we also explored the assembly of the machinery that imports them into the nucleus, the nuclear pore complex during anaphase. Consistent with previous observations[24,25] we found that the nuclear and cytoplasmic ring components (NUP107, NUP214) assemble early, but surprisingly found that nuclear basket as well as cytoplasmic filaments (TPR and NUP358) form only much later, at a time when import of CTCF is already completed. This suggests that nuclear and cytoplasmic filaments of the NPC are not required for the rapid import of nuclear proteins needed for re-establishing the interphase genome architecture (**Fig. 3d, Supplementary Fig. S9c**).

Prediction of dynamic co-localization networks of mitotic proteins

To understand which proteins work together where and when inside the cell, it is essential to measure the similarity of their dynamic subcellular distributions. Although the visualization of all mitotic proteins in the canonical mitotic cell allows interactive visual inspection and quantitative exploration of the data (**Fig. 3b, Supplementary Fig. S9** and see www.mitocheck.org/mitotic_cell_atlas), visually analyzing similarity becomes difficult for more

than three proteins. In order to compute the similarity between all mitotic proteins, we therefore transformed their spatial distribution into numerical features. To this end, we used a segmentation-free approach based on a speeded-up robust features (SURF) detector[26], to extract a series of so-called interest point clusters from the maximum intensity projection of each 3D mitotic protein image stack (**Fig. 4a** and **Supplementary Fig. S10**; see Online Methods). Each 3D movie was thus transformed into a sequence of 100-dimensional feature vectors, with each feature value representing the fraction of the total amount of the detected cellular protein assigned to the respective interest point cluster (**Fig. 4a**). By averaging the feature vectors of all images of the same protein and mitotic standard stage, the dynamic distribution of all proteins in the dataset could then be represented as a third order tensor of size 28 × 100 × 20 (proteins × features × stages, **Fig. 4b**). Soft clustering with non-negative tensor factorization (NTF) could detect seven clusters of dynamic quantitative protein distribution patterns across all 20 stages (**Fig. 4b**). The identity of the proteins in each pattern revealed a striking correspondence between the statistically defined clusters and major mitotic organelles and structures (e.g. CENPA identifies centromeres/kinetochores, RACGAP1 identifies the midplane in late mitotic stages, **Supplementary Fig. S11**), thus validating our unsupervised approach. Since NTF assigns the fraction of a mitotic protein to each pattern over time (see Online Methods), it reliably deals with promiscuous proteins present in multiple sub-cellular structures. Linking proteins with similar patterns at each time point allowed us to derive a dynamic multigraph (see Online Methods), which showed the dynamic protein co-localization network highlighting the activities of different compartments over time and allowed us to predict where and when proteins may interact (**Fig. 4c**). Results from the above clustering can be used to generate hypotheses that can be visualized in the canonical cell model. Looking for instance at the temporal evolution of AURKB and CDCA8 (aka Borealin), two members of the chromosome passenger complex (CPC), predicts that the two proteins relocate to the midplane (**Fig. 4b**, purple in the right panel) in different proportions and with different kinetics. Since the two proteins are present in a 1:1 ratio in the CPC[27], this observation suggested that a fraction of AURKB is not part of the CPC in the midplane. Exploring the 4D localization of CDCA8 and AURKB in the canonical cell model (**Supplementary Fig. S12**) indeed revealed that while these two proteins partially colocalize at the midbody, AURKB but not CDCA8 shows an additional striking localization in an equatorial cortical ring that contracts as mitosis progresses. This localization of the key mitotic regulator AURKB suggests that it is an integral part of the contractile cytokinetic ring. While unexpected, this observation is consistent with detection of AURKB[28] and its interaction partner INCENP[29] at the lateral edge of the

midzone and with its function in cytokinesis[30–32]. This raises the very interesting possibility that the midplane and cytokinetic ring pools of AURKB have different functions for central spindle architecture and cytokinesis during mitotic exit, which will be very interesting to investigate further in the future.

Automatic measurement of subcellular fluxes predicts the (dis)assembly of mitotic protein machines

Since our clustering of common 4D localization patterns does not a priori yield pure sub-cellular compartments as defined ultrastructurally or by fractionation, we developed a supervised machine learning approach to define subcellular structures using known resident proteins of six compartments/organelles relevant for mitosis: chromosome, nuclear envelope, kinetochore, spindle, centrosome and midbody (see Online Methods). Using the same interest point cluster feature data described above (**Fig. 4a**) as input, we trained a multivariate linear regression model that could assign protein amounts to the six reference compartments (**Fig. 5a**). This allowed us to quantitatively compare the subcellular fluxes between these compartments for all proteins (**Supplementary Fig. S13**).

Mining this data is uniquely powerful to dissect dynamic molecular events inside living cells, including for example the assembly or disassembly of organelles. As an example, we calculated the number of all imaged proteins localized to kinetochores, to investigate the disassembly of this large supramolecular complex essential for cell division. The data allowed us to determine that kinetochore disassembly starts in early metaphase with dissociation of BUB1B and PLK1 followed in late metaphase by BUB1, AURKB, MIS12 and CDCA8 (**Fig. 5b**). In addition, this analysis showed that the stoichiometry of these proteins prior to disassembly differs up to six-fold and that their maximal dissociation rates span over an order of magnitude ranging from 17-173 molecules per second (**Table 1**). The predicted number of ~420 CENPA molecules per kinetochore was consistent with data from biochemical methods[33,34] and the predicted disassembly order was consistent with reports of late dissociation of MIS12[35]. Predictions for the other proteins can be further confirmed by inspection of underlying quantitative live cell images (**Supplementary Fig. S14**), or by exploring the average distribution data of the most interesting proteins in the canonical cell model (www.mitocheck.org/mitotic_cell_atlas).

Our automatic assignment of protein amounts to cellular organelles, thus allowed us for the first time to determine the exact timing, stoichiometry and dissociation rates for multiple mitotic proteins that reside dynamically on kinetochores.

DISCUSSION

With this study we provide an integrated experimental and computational framework to build a comprehensive and quantitative 4D model of the mitotic protein localization network in a dividing human cell. Our model provides a standardized yet dynamic spatio-temporal reference system for the mitotic cell that can be used to integrate quantitative information on any number of protein distributions sampled in thousands of different experiments. Using a pilot data set of 28 proteins of interest, imaged in hundreds of dividing cells, we illustrate the power of this model by mining the data to automatically define dynamic localization patterns to subcellular structures as well as predicting the order, stoichiometry and rates of assembly and disassembly of sub-cellular organelles. Our computational model allows building an interactive 4D atlas of the human mitotic cell, which we have made available via the MitoCheck web site (http://www.mitocheck.org/mitotic_cell_atlas), to provide the pilot data as a resource and host data on additional mitotic proteins as it becomes available. The digital atlas allows the navigation of multiple protein dynamics with a spatial and temporal resolution and continuity that are currently impossible to reach experimentally. We illustrate with mitotic chromosome formation, kinetochore disassembly, NPC assembly and cytokinesis how the knowledge gained through the exploration and mining of the atlas data can be used to formulate new mechanistic hypotheses about the function of proteins inside the cell. The concept of standardizing the spatio-temporal cellular context for analyzing dynamic protein distributions in order to understand cellular processes molecularly as presented here is generic and we envision its adaptation to other essential biological functions such as for example cell migration or cell differentiation.

ACKNOWLEDGEMENTS

We thank T. Hyman and I. Poser for donating multiple mouse BAC protein-GFP cell lines and Toru Hirota for giving us the EGFP-CENPA cell line. The automatic imaging would not have been possible without the generosity of R. Höfler and D. W. Gerlich who developed the Micronaut

software. We thank all members of the Ellenberg laboratory for support, especially M. Isokane, M. J. Roberti, J. Mergenthaler, S. Otsuka, for helping us with generating cell lines, reagents and constructs. We thank Stephanie Alexander for carefully reading the manuscript. We also thank W. Huber, B. Fischer, B Klaus and L. P. Coelho for discussions, the EMBL mechanical and electronic workshop, the EMBL advanced light microscopy facility and the EMBL flow cytometry core facility for their excellent support. This study has hugely benefited from the collaboration with Carl ZEISS Jena, especially with Thomas Ohrt. The work was supported by grants from the EU-FP7-SystemsMicroscopy NoE (Grant Agreement 258068) to J.E. and J.M.P., from EU-FP7-MitoSys (Grant Agreement 241548) to J.E. and J.M.P., from EU-H2020-iNEXT (Grant Agreement 653706), as well as by the European Molecular Biology Laboratory (Y.C., M.J.H., J.K.H., A.Z.P., N.W., B.K., M.W., B.N., M.K., J.E.). Y. C. and N.W. were also supported by the EMBL International PhD Programme (EIPP).

AUTHOR CONTRIBUTIONS

B.K, B.N., M.K., R.L., M.M. and N.W constructed and validated GFP knock-in cell lines. A.Z.P. with the help of Y.C and M.W. developed the imaging pipeline. Y.C., A.Z.P., and N.W. performed the calibrated live cell imaging. Y.C., M.J.H., J.K.H., and A.Z.P. developed the computational analysis pipeline. Y.C., J.K.H., M.J.H., A.Z.P., and J.E. wrote the manuscript. J.M.P. coordinated the Mitosys consortium and supervised part of the work. J.E. supervised the work overall and originally conceived the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Rodriguez-Bravo, V. *et al.* Nuclear pores protect genome integrity by assembling a premitotic and mad1-dependent anaphase inhibitor. *Cell* **156,** 1017–1031 (2014).
2. Dick, A. E. & Gerlich, D. W. Kinetic framework of spindle assembly checkpoint signalling. *Nat. Cell Biol.* **15,** 1370–1377 (2013).
3. Ghongane, P., Kapanidou, M., Asghar, A., Elowe, S. & Bolanos-Garcia, V. M. The dynamic protein Knl1 - a kinetochore rendezvous. *J. Cell Sci.* **127,** 3415–3423 (2014).
4. Fink, J. *et al.* External forces control mitotic spindle positioning. *Nat. Cell Biol.* **13,** 771–778 (2011).
5. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).
6. Wachsmuth, M. *et al.* High-throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells. *Nat. Biotechnol.* **33,** 384–389 (2015).
7. Wu, J.-Q. & Pollard, T. D. Counting Cytokinesis Proteins Globally and Locally in Fission Yeast. *Science* **310,** 310–314 (2005).
8. Murphy, R. F. Communicating subcellular distributions. *Cytom. Part A* **77,** 686–692 (2010).
9. Coelho, L. P. *et al.* Determining the subcellular location of new proteins from microscope images using local features. *Bioinformatics* **29,** 2343–2349 (2013).
10. Shariff, A., Murphy, R. F. & Rohde, G. K. Automated estimation of microtubule model parameters from 3-D live cell microscopy images. in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* 1330–1333 (IEEE, 2011). doi:10.1109/ISBI.2011.5872646
11. Schauer, K. *et al.* Probabilistic density maps to study global endomembrane organization. *Nat. Methods* **7,** 560–566 (2010).
12. Zhao, T. & Murphy, R. F. Automated learning of generative models for subcellular location: Building blocks for systems biology. *Cytom. Part A* **71,** 978–990 (2007).
13. Johnson, G. R., Buck, T. E., Sullivan, D. P., Rohde, G. K. & Murphy, R. F. Joint modeling of cell and nuclear shape variation. *Mol. Biol. Cell* **26,** 4046–56 (2015).
14. Lukinavičius, G. *et al.* SiR-Hoechst is a far-red DNA stain for live-cell nanoscopy. *Nat. Commun.* **6,** 1–7 (2015).
15. Magde, D., Elson, E. & Webb, W. W. Thermodynamic fluctuations in a reacting system measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* **29,** 705–708 (1972).
16. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27,** 851–

857 (2009).
17. Cong, L., Ran, F., Cox, D., Lin, S. & Barretto, R. Multiplex Genome Engineering Using CRISPR / Cas Systems. *Science* **339,** 819–823 (2013).
18. Poser, I. *et al.* BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5,** 409–415 (2008).
19. Alberts, B. *et al. Molecular Biology of the Cell. 4th Edition, New York* (2002). doi:10.1091/mbc.E14-10-1437
20. Landry, J. J. M. *et al.* The Genomic and Transcriptomic Landscape of a HeLa Cell Line. *G3 Genes|Genomes|Genetics* **3,** 1213–1224 (2013).
21. Holzmann, J., Fuchs, J., Pichler, P., Peters, J. M. & Mechtler, K. Lesson from the stoichiometry determination of the cohesin complex: A short protease mediated elution increases the recovery from cross-linked antibody-conjugated beads. *J. Proteome Res.* **10,** 780–789 (2011).
22. Burke, L. J. *et al.* CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. *EMBO J.* **24,** 3291–3300 (2005).
23. Mora-Bermúdez, F., Gerlich, D. & Ellenberg, J. Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. *Nat. Cell Biol.* **9,** 822–831 (2007).
24. Dultz, E. & Ellenberg, J. Live imaging of single nuclear pores reveals unique assembly kinetics and mechanism in interphase. *J. Cell Biol.* **191,** 15–22 (2010).
25. Otsuka, S. *et al.* Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *Elife* **5,** (2016).
26. Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **110,** 346–359 (2008).
27. Carmena, M., Wheelock, M., Funabiki, H. & Earnshaw, W. C. The chromosomal passenger complex (CPC): From easy rider to the godfather of mitosis. *Nat. Rev. Mol. Cell Biol.* **13,** 789–803 (2012).
28. Murata-Hori, M. & Wang, Y. L. Both midzone and astral microtubules are involved in the delivery of cytokinesis signals: Insights from the mobility of aurora B. *J. Cell Biol.* **159,** 45–53 (2002).
29. Earnshaw, W. C. & Cooke, C. a. Analysis of the distribution of the INCENPs throughout mitosis reveals the existence of a pathway of structural changes in the chromosomes during metaphase and early events in cleavage furrow formation. *J. Cell Sci.* **98 (Pt 4),** 443–461 (1991).
30. Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161,** 281–294 (2003).
31. Steigemann, P. *et al.* Aurora B-Mediated Abscission Checkpoint Protects against Tetraploidization. *Cell* **136,** 473–484 (2009).
32. Isokane, M. *et al.* ARH GEF17 is an essential spindle assembly checkpoint factor that targets Mps1 to kinetochores. *J. Cell Biol.* **212,** 647–659 (2016).
33. Hasson, D. *et al.* The octamer is the major form of CENP-A nucleosomes at

human centromeres. *Nat. Struct. Mol. Biol.* **20,** 687–695 (2013).
34. Scott, K. C. & Bloom, K. S. Lessons learned from counting molecules: how to lure CENP-A into the kinetochore. *Open Biol.* **4,** (2014).
35. Gascoigne, K. E. & Cheeseman, I. M. CDK-dependent phosphorylation and nuclear exclusion coordinately control kinetochore assembly state. *J. Cell Biol.* **201,** 23–32 (2013).
36. Mahen, R. *et al.* Comparative assessment of fluorescent transgene methods for quantitative imaging in human cells. *Mol. Biol. Cell* **25,** 3610–3618 (2014).
37. Koch, B. *et al.* Generation and validation of homozygous fluorescent knock-in cells using genome editing. *bioRxiv* (2017).
38. Bancaud, A. *et al.* Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J.* **28,** 3785–98 (2009).
39. Li, C. H. & Lee, C. K. Minimum cross entropy thresholding. *Pattern Recognit.* **26,** 617–625 (1993).
40. Heriche, J.-K. *et al.* Integration of biological data by kernels on graph nodes allows prediction of new genes involved in mitotic chromosome condensation. *Mol. Biol. Cell* **25,** 2522–2536 (2014).
41. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **9,** 62–66 (1979).
42. Meyer, F. Topographic distance and watershed lines. *Signal Processing* **38,** 113–125 (1994).
43. Politi, A. Z. *et al.* Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four dimensional imaging. *bioRxiv* 188862 (2017). doi:10.1101/188862
44. Barton, G. J. & Sternberg, M. J. E. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J. Mol. Biol.* **198,** 327–337 (1987).
45. ten Holt, G., Reinders, M. & Hendriks, E. Multi-Dimensional Dynamic Time Warping for Gesture Recognition. in *Thirteenth annual conference of the Advanced School for Computing and Imaging* (2007).
46. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust.* **29,** 1153–1160 (1981).
47. Lazebnik, S., Schmid, C. & Ponce, J. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27,** 1265–1278 (2005).
48. Heikkilä, M. & Pietikäinen, M. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **28,** 657–662 (2006).
49. Tran, T. N., Drab, K. & Daszykowski, M. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemom. Intell. Lab. Syst.* **120,** 92–96 (2013).

50. Jolliffe, I. T. Principal Component Analysis. *Principal Component Analysis* (2002).
51. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401,** 788–791 (1999).
52. Cichocki, A., Zdunek, R., Phan, A. H. & Amari, S.-I. *Nonnegative Matrix and Tensor Factorizations*. (John Wiley & Sons, Ltd, 2009). doi:10.1002/9780470747278
53. Tucker, L. R. A method for synthesis of factor analysis studies. *Pers. Res. Sect. Rep.* **984,** (1951).
54. Aitchison, J. *The statistical analysis of compositional data*. (Blackburn Press, 2003).
55. Maliga, Z. *et al.* A genomic toolkit to investigate kinesin and myosin motor function in cells. *Nat. Cell Biol.* **15,** 325–334 (2013).
56. Hutchins, J. R. *et al.* Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328,** 593–599 (2010).
57. Ladurner, R. *et al.* Sororin actively maintains sister chromatid cohesion. *EMBO J.* **35,** 635–653 (2016).

FIGURE LEGENDS

Figure 1 | Quantitative imaging of mitotic proteins. (a) Automatic calibrated 3D live confocal imaging pipeline. Cells in prophase were identified by online classification, imaged through mitosis in the landmarks and protein of interest channels, and measured by FCS at selected positions. (b) The local protein concentrations determined by FCS fitting linearly correlate with the background subtracted image intensities at the corresponding positions (shown are data acquired on the same day). (c) Example cell showing concentration map resulting from FCS-based intensity calibration (mean Z-projection). (d) Individual cells have different mitotic spatio-temporal dynamics. Scale bar: 10 µm.

Figure 2 | Modeling of mitotic standard time. (a) Cellular and chromosomal volumes were segmented from the landmarks channel. Scale bar: 10 µm. (b) Three morphological features (in red) were extracted from the chromosomal volume. (c) Mitotic standard time was generated in the feature space by multiple sequence alignment visualized here in the feature dimension describing chromosomal volume. (d) Detection of mitotic transitions. Peaks (red circles) above threshold (grey dashed line) were detected on the second derivatives in the feature dimension of chromosomal volume (left). For the second derivatives of all six feature dimensions, a total of 19 peaks defining 19 transitions (grey circles) for 20 mitotic stages were detected (colored). The solid curve shows the maximum projection through the feature dimensions. (e) The standard mitotic cell was represented by the cell closest to the average of each stage. Each mitotic stage was assigned duration (colored line), its duration standard deviation (dashed line) and a biological annotation.

Figure 3 | Visualization of 4D protein distribution maps. (a) Through averaging of a large number of cells, models were generated for all mitotic standard stages with symmetrical geometries. Example image sequences were registered to the standard space of the corresponding mitotic standard stage. A distribution map over time was then generated for each protein by averaging through multiple cells. (b) Average distributions of four proteins are displayed in different mitotic stages. Colored lines indicate mitotic stages according to figure 2e. (c) Maximal intensity projection of the concentration map for chromatin binding proteins. Scale bar:10 µm. (d) Amount of chromatin-bound and nuclear molecules for eight chromatin remodelers. (e) Fraction of chromatin bound proteins relative to NCAPH2. Shown is the mean and standard deviation. In (d) and (e) TOP2A has been scaled down by a factor 10 for visualization.

Figure 4 | Identification of dynamic protein clusters. (a) SURF interest points were detected (top panel) and assigned to one of 100 clusters of similar interest points. Each image was represented as the distribution of intensity in each of the 100 interest point clusters (bottom panel). Scale bar: 10 µm. (b) Non-negative factorization of the data tensor of 28 proteins × features × mitotic stages (left panel) produced a non-negative tensor of reduced dimension (middle top panel) whose entries can be interpreted as the fraction of protein belonging to each cluster over time (right panel). Scale bar: 10 µm. (c) Dynamic multi-graph of protein co-localization, shown for 5 stages. Each edge colour corresponds to a localization cluster as in (b) and the edge thickness corresponds to the product of the linked genes fractions in the corresponding cluster and can be loosely interpreted as a probability of interactions.

Figure 5 | Analyzing sub-cellular (dis)assembly kinetics using a supervised approach. (a) Example of maximally Z-projected images of marker proteins for the selected subcellular compartments. Scale bar: 10 µm. (b) Kinetics of kinetochore disassembly. The predicted number of molecules localized on kinetochore/centromeres are plotted for eight proteins in the mitotic standard time (left panel) and zoomed in for anaphase (right panel). (c) Model showing the order of protein removal from the kinetochore in anaphase.

TABLES

Table 1 Kinetics of kinetochore disassembly

	Late metaphase	Anaphase kinetochore dissociation	
Gene	Protein number [10^3]	Mitotic standard time [min]	Rate [proteins/s]
BUB1	11.9	31.5	35
AURKB	55.6	33.4	173
CDCA8	49.8	35.0	70
MIS12	7.6	35.3	17

ONLINE METHODS

Cell culture | Cells were cultured in high glucose Dulbecco's modified Eagle's medium (DMEM; Life Technologies) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 units/ml penicillin, 0.1 mg/ml streptomycin, 2 mM Glutamine and 1 mM (v/v) Sodium pyruvate at 37 °C and 5% CO_2. Depending on the genetic modification, one or more of the following antibiotics were supplied to the culture at the stated final concentration: Geneticin (Life Technologies) 500 µg/ml, Hygromycin B (Invitrogen) 200 µg/ml or Puromycin (Invitrogen or Calbiochem) 0.5 µg/ml. Once the cells reached 80-90% confluence, they were passaged and only a fraction of the cells were cultured in a fresh dish. All cells tested negative for mycoplasma contamination.

Cell modification | HeLa Kyoto cells were used for genetic modifications and imaging. HeLa Kyoto are hypotriploid with on average 64 chromosomes, thus during mitosis the cells have on average 64*2 = 128 kinetochores[20]. All cell lines used for this project are listed in **Supplementary Table S1** with their providers indicated. Several cell lines were generated within this project as follows: The cell lines expressing PLK1-mEGFP, CEP192-mEGFP and NUP107-mEGFP were generated using the Zinc finger nuclease (ZFN) pipeline as in[36]. Zinc finger nucleases were purchased from Sigma-Aldrich with DNA-binding sequences listed in **Supplementary Table S2**. The other genome-edited cell lines were generated using the CRISPR/Cas9 system[17] with double nicking. For these cell lines, both gRNAs (**Supplementary Table S2**) and the donor plasmid were designed based on ENSEMBL release 75 and transfected together into HeLa Kyoto cells with jetPrime (Polyplus) according to the manufacturer's instruction. A single clone was selected using our previously developed validation pipeline[36,37]. For 4 out of 20 genome-edited cell lines (BUB1B-EGFP, TPR-mEGFP, mEGFP-NUP107 and CEP192-mEGFP) we detected in the Western-Blots (anti-GFP) a band of the size of free GFP. Therefore, for these cell lines, the total free amount may be overestimated. In order to label the chromosomal volume, an H2B-mCherry[5] cDNA was transfected into some genome-edited cell lines with Fugene6 (Promega) according to the manufacturer's instructions. The pmEGFP2-N1-NES construct was generated by sub-cloning two tandem repeats of mEGFP (mEGFP2)[38] and the NES of MAPKK (NLVDLQKKLEELELDEQQ) into the pEGFP-N1 vector (Clontech Laboratories). The pmEGFP2-N1-NES construct was transfected into HeLa Kyoto cells and cells with stable expression were selected by culturing with the appropriate antibiotics. Single cells or a cell population with the desired expression level were harvested for imaging by fluorescence activated cell sorting (FACS, performed by the EMBL Heidelberg facility).

Calibrated fluorescence confocal microscopy | Confocal microscopy was performed on a Zeiss LSM780, Confocor 3, laser scanning microscopes using 40x, NA 1.2 water DIC Plan-Apochromat objectives and the GaAsP detectors equipped with an incubation chamber (EMBL workshops). Cells were imaged at 37 °C in a CO_2-independent medium (Life Technologies) fluorescently colored with 500 kDa Dextran (Life Technologies) -DY481XL (Dyomics) produced in house. Time-lapse imaging was performed using the ZEN 2012 software as well as in-house software applications. The acquisition was supported by an in-house developed objective cup and a water pump, such that water drops were regularly supplied to the objective-sample interface. Before starting imaging, a number of positions were selected manually. During live cell imaging, the microscope determined the focus automatically by performing line-scan imaging of the reflection signal of the 633 nm laser. The vertical position of the glass bottom was determined as the position with the maximum reflection intensity, and used as reference for acquiring a volume of the specimen at a particular depth.

HeLa Kyoto cell lines with H2B-mCherry were imaged live using an excitation laser at 561 nm every 5 min for about 16 hours on the Zeiss 780 microscopy system. Three confocal planes were acquired at a resolution of 0.32 µm in x and y and 2.5 µm in z. Images were projected in z by taking the maximum intensity value. Images of the H2B signal were analyzed on the fly by the Micronaut software (Gerlich lab, IMBA, Vienna) using a support vector machine classifier that was trained beforehand (with the software CellCognition, http://www.cellcognition.org/) to distinguish between cells in interphase, prophase, mitosis (prometaphase till telophase) and artefacts (apoptosis, on the border of the imaging field, out of focus, too low expression). The classification score for the prophase, interpreted as the probability of a cell being in the class of interest, was output, and a pre-defined threshold was used to make a decision on whether imaging setups for mitotic cell acquisition should be activated. According to how different the sample's H2B-mCherry expression levels were from the training set, the threshold on class probability was set between 0.85 and 0.96. Once a prophase cell was found, it was then imaged using a different imaging setup. For our purpose, mitotic cells were imaged live every 90 seconds for 31 z-planes with a spatial resolution of 0.25 µm in x-y and 0.75 µm in z with a 488 nm laser (high expression of H2B-mCherry allowed it to be excited at 488 nm and produce adequate signal). For cells that do not express H2B-mCherry we used SiR-DNA to stain the chromatin (Spirochrome, final dilution 50 nM added 2 hours before imaging). Cells were imaged with the

633nm laser (3 confocal planes every 5 min at the same resolution as for H2B-mCherry) and processed as for H2B-mCherry.

The signal from the GaASP detector was separated into three channels: GFP, varied from 490 to 552 nm depending on the expression level to avoid detection saturation; mCherry, 587 – 621 nm (**Supplementary Fig. S2a**), and Dy-481XL, 622 – 695 nm (**Supplementary Fig. S2b**). For SiR-DNA we used 622 – 695nm. Once the mitosis was recorded for 40 frames, a single-plane image was then acquired at 2.5 µm above the cover glass surface. Using an in-house-developed Fiji (http://fiji.sc/Fiji) macro, the image was thresholded using the method developed by[39] and the object closest to the image center with a proper size was selected as one of the two daughter nuclei in the cell of interest. The segmented nuclear boundary was fitted with an ellipse. FCS measurements were performed with the 488 nm laser and the APD detector at 505 – 540 nm at two positions in and four around the nucleus with a distance of 2 µm to the ellipse boundary for 30 seconds each.

A manual quality control was then performed. Videos of cells with no expression of the protein of interest, with wrongly selected FCS positions (e.g. outside of the cell) or without anaphase onset were excluded from further processing.

Segmentation of the landmarks | A fully automated computational pipeline was implemented in MATLAB (MATLAB R2017a, The MathWorks, Inc., Natick, Massachusetts, United States) to segment and track cells of interest and reconstruct chromosomal and cellular surfaces. The pipeline was composed of three major steps: segmentation of chromosomal volume, segmentation of cell volume and extraction of parameters out of the landmarks geometry. Chromosomal regions were segmented from the mCherry channel which had high H2B–mCherry signal and very low Dextran-Dy481XL intensity (**Supplementary Fig. S2a**) or from the SiR-DNA channel which had no crosstalk from other channels. In order to perform isotropic 3D image processing, adjacent x-y planes were linearly interpolated along the z direction. A 3D Gaussian filter was applied to reduce the effects of noise. To detect chromosomal regions, the filtered image stack was binarized first using a multi-level thresholding method as described in[40]. In this approach, a global Otsu threshold[41] was determined for the entire stack and the threshold was then adapted for each 2D slice, validated by the connectivity of binary components in 3D. Tiny connected components were removed from the binary image leaving only chromosomal components from all cells in the imaging field. All components were used as seeds for the detection of the cell boundary in a later stage. The connected chromosomal volume in the x-y

center of the first frame identified the cell of interest due to the centering step in the imaging pipeline. The cell of interest was tracked sequentially through the entire image sequence using a nearest neighbor approach. At each time point, an event of chromosome segregation was also probed by analyzing the chromosomal volume around the tracked location. Once segregation was detected, both daughter nuclei were tracked in the subsequent frames (**Supplementary Fig. S2c**).

The cell region was segmented from the Dy481XL channel showing the cell-free regions in high intensities and histone signal with low intensities (**Supplementary Fig. S2b**). Upon interpolation and filtering as for the chromosomal segmentation, a ratio image was created by dividing the filtered image stack of the mCherry channel by that of the Dy481XL channel in order to diminish bleed-through signal from the H2B-mCherry channel. The ratio image stack was then binarized as described above. When using SiR-DNA there is no bleed-through in the Dy481XL channel when excited at 488. In this case the Dy488XL has been directly binarized. To separate individual cell regions, the nuclear seeds previously detected were used, considering the fact that each cell region can have only one or two chromosomal volumes. This was implemented by applying a marker-controlled watershed algorithm[42]. To obtain a better separation between touching cells, the algorithm was applied on the distance transformed image that made use of the geometric properties of the cell surface. The cell region of interest was defined by taking the connected region(s) containing the detected chromosomal volume(s) of interest (**Supplementary Fig. S2d**).

The chromosomal mass at each time point was represented by its three orthogonal eigenvectors and associated eigenvalues where the eigenvector with the largest eigenvalue represented the longest elongated axis of the chromosomal volume. Metaphase frames were automatically detected based on the low value of the smallest eigenvalue of the chromosomal volume. Division axis for metaphase cell was then predicted by taking the eigenvector having the minimum eigenvalue. By definition, this vector is always orthogonal to the metaphase plate. Using the predicted axis in the first and last metaphase frame, axes for the remaining frames were propagated backwards and forwards for stages before and after metaphase, respectively, where the eigenvector with the smallest discrepancy in angle to the axis predicted for the adjacent frame was used (**Supplementary Fig. S2e**). For further analysis, the plane orthogonal to the division axis going through the centroid of both daughter nuclei was predicted as the midplane.

Image processing and calibration | Image processing and calibration were performed according to[43]. Before each calibrated live cell confocal microscopy experiment, the focal volume was calibrated using a 20 nM solution of Alexa488 (Life Technologies), and single mEGFP brightness was calibrated by performing FCS measurements on the HeLa Kyoto expressing mEGFP[36]. All FCS measurements were processed using Fluctuation Analyzer[6]. Autocorrelation functions of dye solutions were fitted using a one-component diffusion model with triplet-like blinking, and measurements of fluorophore-fused proteins were fitted using a two-component anomalous diffusion model with fluorescent protein-like blinking. The effective confocal volume was calculated from

$$V_{eff} = (4 \cdot \pi \cdot D \cdot \tau)^{\frac{3}{2}} \cdot \kappa$$

where D was the diffusion coefficient of the Alexa488 which is 464 $\mu m^2/s$ at 37 °C and κ the structural parameter. The averaged time passing through the confocal volume τ and the structural parameter (typically between 4-7) was fitted to the ACF of Alexa488. The number of fluorescent molecules within a confocal volume was calculated by multiplying the fitted number of molecules (N) with correction factors for background and photobleaching[6]. As proteins might exist in complexes with multiple molecules, a count per molecule (CPM) value was used to correct the number of molecules. As reference, the CPM value of mEGFP was used as measured in the HeLa cells expressing mEGFP where the mean value of all mEGFP measurements were taken. If the CPM of a measurement of a fusion POI within a cell was larger than that of the mEGFP, the fitted number of molecules was corrected by multiplication with the ratio between the two. Finally, the local concentration of the measured protein was determined as the corrected number of molecules divided by the effective confocal volume. As quality control of the FCS measurements, we pre-defined thresholds and deleted data points with too low coefficient of variation R^2 or APD counts or too high fitting χ^2 or bleaching or outlier CPM values.

The calibration of the image acquired with the GaAsP detector was based on the assumption of linear correlation between the local protein concentration and the EGFP intensity which we could verify (**Fig. 1b**). The averaged intensity of the GFP channel in cell-free areas was considered as background. For all measurement points the coefficient ρ between local protein concentration and background-corrected imaging intensity, mean filtered with a 9 × 9 pixels window to avoid noise, was calculated by performing a linear regression. The 3D protein concentration map was generated by multiplying the pixels intensities with the linear coefficient ρ. The protein number in each voxel was obtained by multiplying the concentration with the voxel volume. The absolute protein abundance could be calculated by summing up the map over the

cell volume. After NEBD and before the nuclear envelope reforms, we estimated the number of proteins bound to chromatin by subtracting the cytoplasmic average concentration (representing the background concentration of proteins that freely diffuse between the cytoplasmic and chromosome volume) from the average protein concentration on the chromosome mask (**Fig. 3d-e** and **Supplementary Fig. S9**). Finally, to obtain the number of proteins, the concentration difference is multiplied by the number of voxels of the chromosome mask and the voxel volume.

Modeling of the mitotic standard time | The mitotic standard time was modeled in a six-dimensional feature space using three morphological features of the chromosomal volume: the distance between the two daughter nuclei, the total volume and the third eigenvalue (**Fig. 2b**) and their first derivatives. The model was generated by aligning 132 mitotic image sequences using the Barton-Sternberg multiple sequence alignment algorithm (**Supplementary Fig. S3c**)[44]. The two sequences with the smallest distance to the average of all sequences were selected to initiate the alignment and each of the remaining sequences was then aligned to the average among all aligned sequences. The alignment was implemented as a modified multidimensional dynamic time warping[45] where the total Euclidean distance over time between the pair of sequences was used as the objective of the optimization (**Supplementary Fig. S3a**). The timeline of the averaged sequences was calculated as the mean of the alignment matrix as shown in **Supplementary Figure S3b**. The Barton-Sternberg algorithm was terminated after four rounds as the standard deviation over time remained stable after 2 rounds (**Supplementary Fig. S3d**). The mitotic standard time was defined at a temporal resolution of 15 seconds by subsampling the averaged timeline. In order to find transitions in the mitotic standard time, the second derivative of the model at each time point for each feature dimension was calculated from

$$x_t'' = \left|\left(x_{t^-} - x_t\right) - \left(x_t - x_{t+}\right)\right|,$$

$$\text{where } x_{t^-} = \sum_{i=t-18}^{t-1} x_i/18, \quad x_{t+} = \sum_{i=t+1}^{t+18} x_i/18.$$

Peaks above a pre-defined threshold were selected across all dimensions as transitions. In the later part of the model where the values of the second derivatives were generally low, small peaks were selected as additional transitions such that no stage between two transitions lasted longer than 12 minutes. Furthermore, transitions with lower values were deleted to ensure a minimum duration of 1.5 minutes for each stage (**Supplementary Fig. S5**).

Modeling of the canonical cell | To support spatial averaging, all cells assigned to the same standard mitotic stage were registered into a common reference coordinate system to give them

the same location and orientation. To this end, a virtual coordinate system was defined with its origin at the center of a volume chosen large enough to accommodate all cells after registration. Landmarks (i.e. the cell boundary and chromosomal volumes) were then registered to the virtual coordinate system by applying a transformation function involving translation and rotation in 3D. This transformation function was estimated such that the predicted cell axis was aligned with the x axis in the virtual coordinate system. This transformation was applied to both landmarks to preserve their interrelationship in the registered image stacks as shown in **Supplementary Fig. S7a** and **b**. Bicubic interpolation was used when applying the transformation[46].

Registered landmarks were subsequently represented using a cylindrical coordinate system that transforms 3D coordinates into radial distances providing greater flexibility in shape analysis (**Supplementary Fig. S7c**). To this end, we converted landmarks in each plane along the z axis and along the predicted cell axis to polar coordinates in which object boundaries are represented by their radial distances from the object centroid (**Supplementary Fig. S7c, d**). As the centroids were aligned on the z axis, the cylindrical representation was formed by concatenating into a vector the polar representations for all planes (**Supplementary Fig. S7d**). After chromosome segregation, two separate cylindrical representations were used to encode each of the two daughter nuclei. In this case, the cylindrical axis of each chromosome passes through the centroid of that chromosomal volume.

The standard mitotic space represented by the averaged landmarks was computed in three steps. In the first step, the cylindrical coordinate vectors were averaged separately for each landmark across all cells within each standard mitotic stage (**Supplementary Fig. S7e**). The average vectors were then transformed back to a Cartesian coordinate system from which binary image stacks were generated. In a second step, to reconstruct the landmarks, the average volume of each landmark was obtained by combining two binary image stacks: one obtained using the z axis as the cylinder axis and the other using the cell axis as cylinder axis (**Supplementary Fig. S7f**). This combination involved first taking the intersection between the two binary images and then extending it until the average volume of all the cells belonging to the mitotic stage being processed was reached (**Supplementary Fig. S7f**). Because multiple frames of a cell could be assigned to the same mitotic standard stage, cells could have unequal contributions to each stage with some cells represented more than others at a given stage. To ensure uniform contribution from each cell towards the average mitotic space, in the third step, for each given mitotic stage and for each cell, the frame that was most similar to the average shape obtained in step 2 was selected. These selected cells were then used to re-compute the average shape of the

corresponding mitotic stage as described above. This final average shape was also used to calculate the standard deviation of all cells in the same mitotic stage. Average mitotic space and standard deviation were generated for all the stages (7-20) for each of the landmarks (examples in **Supplementary Fig. S7g** and **h**).

Generation of the protein density map | Standard mitotic spaces were used as reference to register and integrate protein distributions from many different cells to generate protein density maps (**Fig. 3a**). All calibrated protein concentration maps having the same protein in a given mitotic stage were registered first to the corresponding standard mitotic space using the predicted cell division axis. This transformed all individual protein image stacks to the same coordinate system. Bicubic interpolation[46] was used during the rotation. Registered image stacks were then accumulated in the standard mitotic space. Pixels outside the segmented cell region and mapped outside the standard mitotic space were discarded. A protein density map was then created by averaging the accumulated intensities in the standard mitotic space (**Fig. 3a**). Density maps of all proteins for mitotic stages 7-20 were estimated in the same way and can be explored on http://www.mitocheck.org/mitotic_cell_atlas.

Feature extraction of images | The protein z-stack concentration map was processed using a Gaussian filter (Matlab smoth3 function with a kernel of size [3 3 1] and standard deviation 0.65) followed by a maximum projection along the z-axis and normalization to the theoretical saturation intensity. SURF interest points[26] were then detected on the image resized to a 0.063 µm resolution using three octaves each including four Haar wavelet filters at different sizes from 9-by-9 till 99-by-99 pixels ranging from about half a micrometer to more than six micrometers. Interest points were further selected such that most of the protein signals were counted in one of the interest points. Each of these interest points was then described by a numerical vector quantifying features in the following four categories: locations relative to the landmarks (four features), correlation to the H2B signal or the predicted midplane/midbody volume according to the localization of the interest point (one feature), flattened soft spin image features[47] describing the intensity distribution within an interest point (30 features), and summarized uniform Local Binary Patterns (uLBP)[48] describing the orientation of the signal (4 features).

5% of the cells were randomly selected and their interest points were used to construct a training set for identifying clusters of interest points with similar features (**Supplementary Fig. S10**). All training interest points were first separated into 16 clusters by their localization feature

and the ¼-level of their metric values. Interest points in clusters with a sufficient size were then further clustered based on the correlation features separated by pre-defined thresholds followed by a dbscan[49] clustering for each sub-cluster in the reduced feature space covering 85% of the variance according to a principal component analysis[50] on the uLBP and spin image features[47]. The final clustering step was performed only for the clusters with the highest contrast value in their location category based on the spin image features where the interest points were separated into homogeneous bright, structured bright and dim clusters by pre-defined threshold. The total number of clusters was not deterministic since the training set was randomly generated but eight rounds of clustering yielded between 87 and 100 clusters and a run with 100 clusters was used for further analysis. Interest points in the same cluster share similar textures (**Fig. 4a**). All interest points in the remaining images in the data set were then each assigned to one cluster. The total intensity within each interest point was then counted and the fraction of intensities recorded in each interest point cluster was calculated for each cell so that each image was represented by a 100-dimensional feature vector with a sum of one.

Non-negative tensor factorization | Each protein was represented at each mitotic stage by the average of all its vectors present at that stage. Due to binning of consecutive imaging time points, a cell can be represented by several vectors at a given mitotic stage. These duplicates were replaced by their average resulting in each cell being represented by only one vector per mitotic stage. The resulting dataset is a three-dimensional tensor $\underline{\mathbf{X}}$ of 28 proteins × 100 features × 20 mitotic stages. We view canonical subcellular localizations as latent features of the data, that is, we assume that, at any time point, the observed vector for a protein was generated by a combination of the different canonical subcellular localizations the protein occupied at this stage. A protein vector x can then be expressed as the product of a subcellular localization membership vector z and a matrix $\mathbf{A}$ of canonical subcellular localization features: $x = z\mathbf{A}$. Therefore, we wish to model our data tensor $\underline{\mathbf{X}}$ such that for each frontal (temporal) slice $\mathbf{X}_t$,

$$X_t = Z_t A + E_t \, (t = 1,2, \dots ,20)$$

where $\mathbf{Z}_t$ is a matrix whose rows are localization membership vectors and $\mathbf{E}_t$ is a matrix containing the errors.

Given that all feature values are non-negative, a possible solution for each time point can be found by non-negative matrix factorization (NMF) of individual matrices $\mathbf{X}_t$[51]. However, processing time points independently results in loss of information with the undesirable effect that different canonical localizations are learned for different time points. Simultaneous non-

negative factorization of a set of matrices is a special form of non-negative tensor factorization (NTF) which can be reduced to a standard NMF using column-wise unfolding of the data tensor **X**[52]:

$$X = ZA + E$$

where **X** is formed by vertically stacking the $\mathbf{X}_t$ matrices and **Z** is formed by the correspondingly stacked $\mathbf{Z}_t$ matrices and **E** contains the errors. **Z** and **A** are then found using multiplicative updates[51] to minimize the objective function ||**X** - **ZA**|| where ||.|| indicates the Frobenius norm. As a final step, the rows of **Z** are normalized to sum one. Values in **Z** can be interpreted as fractions of the amount of protein (captured by the features) present at each canonical localization.

The method requires choosing the number *k* of canonical subcellular localizations we want to represent our data with. There is no good strategy for finding this number a priori because increasing *k* corresponds to a higher resolution of the localization description e.g. a low *k* results in lumping all chromatin proteins together while a higher *k* resolves kinetochore proteins from other chromatin proteins. Thus the optimal number of subcellular localizations is partly subjective, depending on the level of granularity desired. However, we can use heuristics to help guide the choice of *k*. If the number of selected canonical localizations is too low, many proteins will share the same temporal profile, i.e. their corresponding vectors in $\mathbf{Z}_t$ will be highly similar for all time points. As more canonical localizations are added, we can expect more proteins to resolve into distinct profiles, i.e. the similarity between their corresponding vectors will decrease until eventually adding more canonical localizations will not improve resolution and similarity will stop decreasing. Similarity between vectors across time points can be measured using Tucker's congruence coefficient (TCC)[53]. Therefore, for each value of *k* from 2 to 25, we plot the fraction of TCC values above 0.6. The value of *k* for which the fraction of highly similar proteins reaches a low value plateau indicates that there are enough canonical localizations to describe each protein individually and therefore this value of *k* represents an upper bound on the number of canonical localizations. Following this procedure, *k* was set to seven for the current data. Because the NMF algorithm can converge to a local minimum of the objective function, ten runs with random initialization of the matrices were performed and the run with lowest value of the objective function was kept. A flattened representation of the resulting tensor can be obtained by assigning a different color to each cluster and plotting each protein distribution as a bar chart in which the height of each color band at each time point is proportional to the fraction of the protein amount in the corresponding cluster (**Supplementary Fig. S11**). A dynamic multigraph can be derived from the cluster memberships as follows: First an edge type is defined for each cluster. If two

genes share a cluster at a given time point, then an edge of that type is added between them at that time point. The edge weight is set to the product of the linked genes fractions in the corresponding cluster and can be loosely interpreted as a probability of interaction. For visualization, only edges with a weight greater than an arbitrary threshold (here set to 0.3) were kept (**Fig. 4c**).

Analysis of proteins' localization kinetics using supervised annotation | A multivariate linear regression model with a multivariate Gaussian response was trained with an elastic net regularization and non-negativity constraints on the coefficients with the feature vectors described above as predictors and localization vectors as response. The response vectors were defined using cells with tagged proteins known to be specific markers of unique subcellular compartments (**Supplementary Table S3**) as follows: For each of the marker proteins, the fraction of total intensity in the foreground was determined by Otsu thresholding of the 3D image stack and the corresponding protein amount assigned to the compartment with the complement assigned to cytoplasm. Each cell is thus represented by a 7-dimensional response vector containing the fraction of the tagged protein in the following compartments: chromatin, kinetochore, centrosome, spindle, midbody, nuclear envelope and cytoplasm. To deal with the compositional nature of this data, all features and response vectors are transformed using the additive log-ratio transformation[54] with the inverse hyperbolic sine function as a generalized logarithm to handle occurrences of 0. The model with the best fit using 5-fold cross-validation was selected.

The predictions from the model were transformed back to proportions using the inverse of the log-ratio transformation then multiplied by the total number of proteins to predict the absolute number of molecules in each mitotic subcellular structure for each image. Predictions were then smoothed by local polynomial regression fitting.

To compute the anaphase dissociation kinetics for each kinetochore protein (**Table 1**), we fitted each curve between 30 min and 42 min mitotic standard time (late metaphase to telophase) with a 4 parameter sigmoidal decay function:

$$y = \frac{a - d}{(1 + bc^{(-time)})} + d$$

whose first and second derivatives were analytically calculated. The time of disassembly was defined as the point at which the second derivative is equal to 0 (inflection point of the curve). The disassembly rate was computed as the minimum value of the first derivative in the time interval.

Code availability | The code will be made accessible upon acceptance.

Figure 1

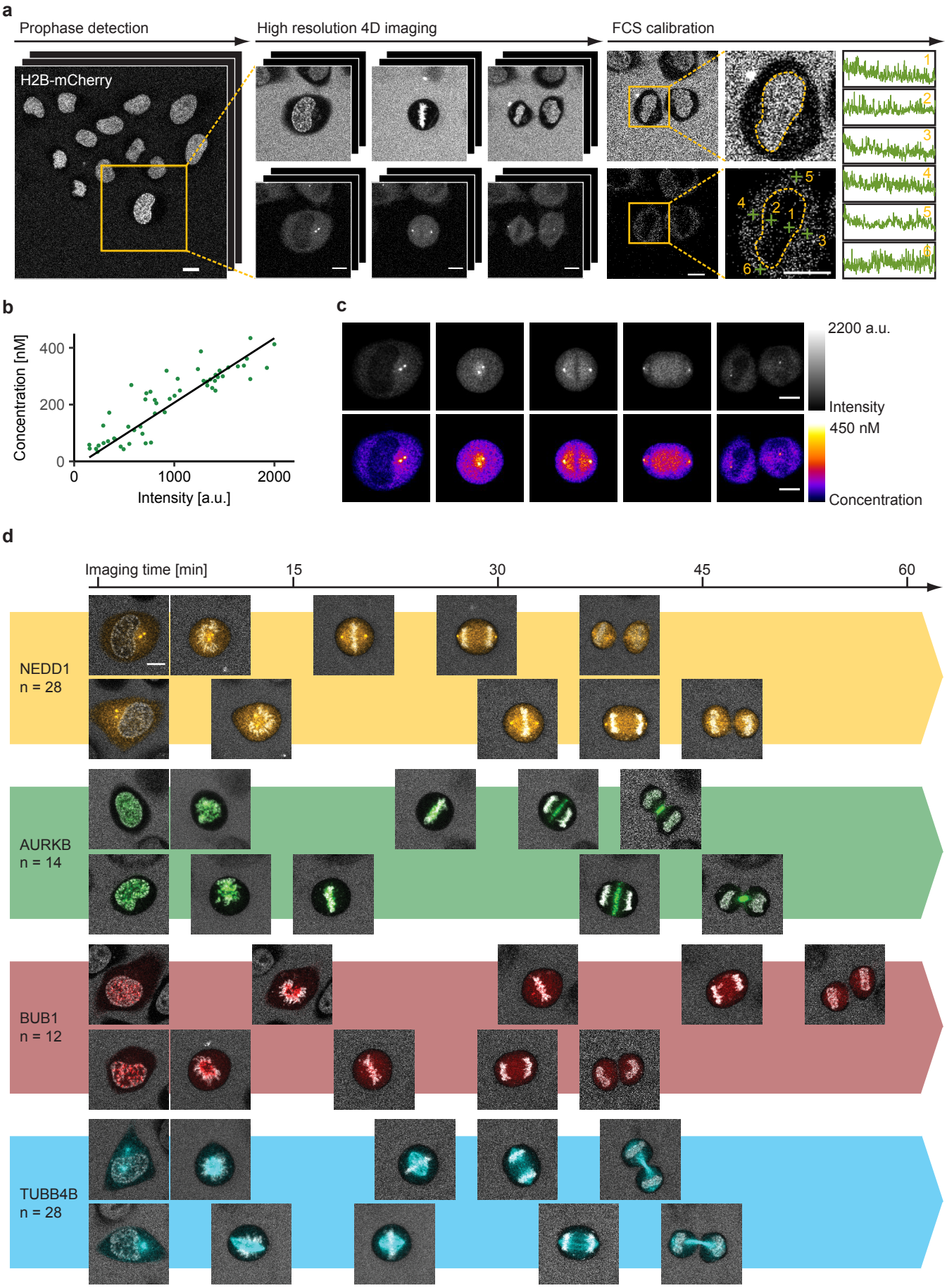

a
Prophase detection
High resolution 4D imaging
FCS calibration
H2B-mCherry
b
Concentration [nM]
400
200
0
0
1000
2000
Intensity [a.u.]
c
2200 a.u.
Intensity
450 nM
Concentration
d
Imaging time [min]
15
30
45
60
NEDD1
n = 28
AURKB
n = 14
BUB1
n = 12
TUBB4B
n = 28

Figure 2

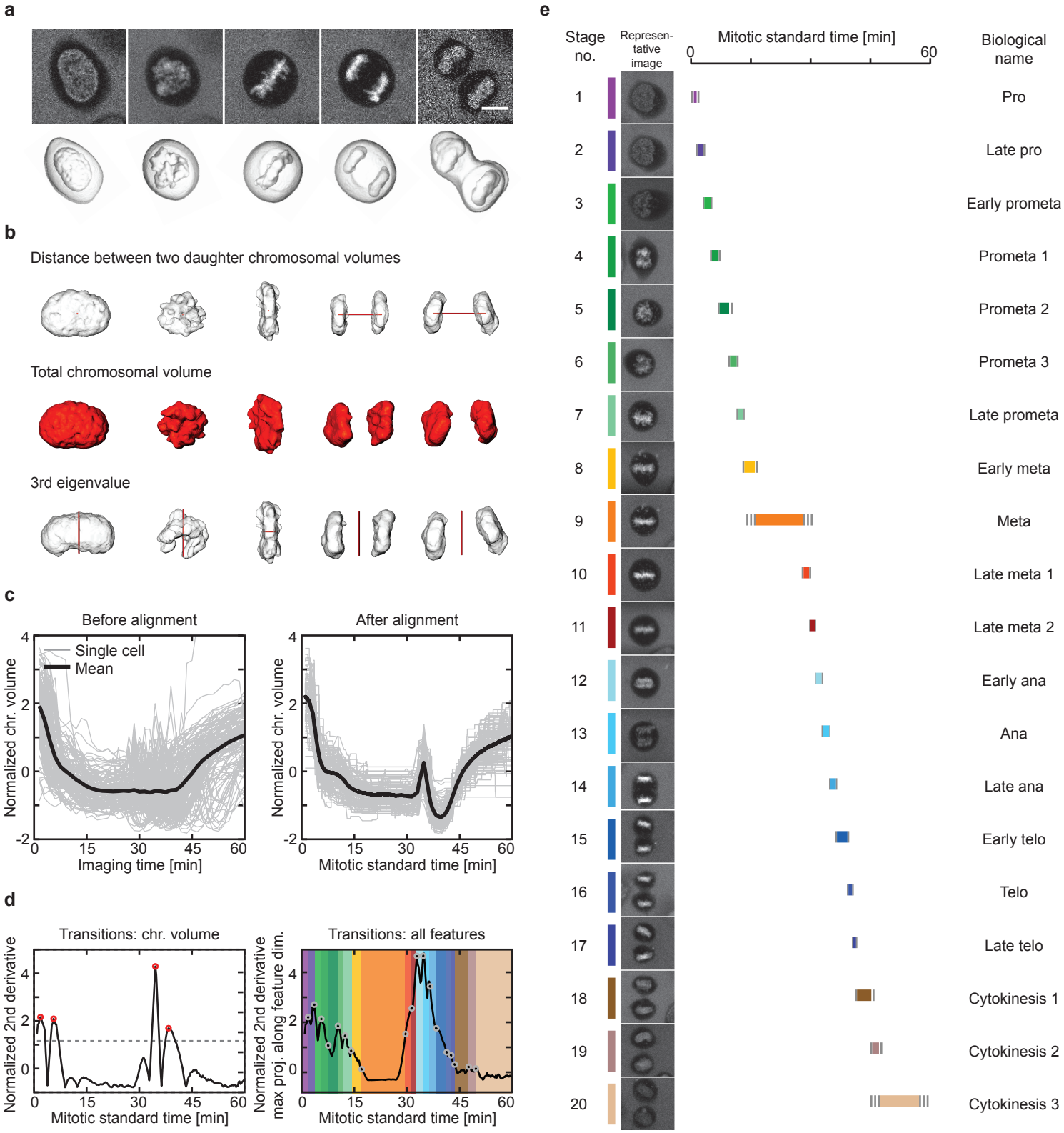

a
b
Distance between two daughter chromosomal volumes
Total chromosomal volume
3rd eigenvalue
c
Before alignment
After alignment
Single cell
Mean
Normalized chr. volume
Imaging time [min]
Mitotic standard time [min]
d
Transitions: chr. volume
Transitions: all features
Normalized 2nd derivative
Normalized 2nd derivative max proj. along feature dim.
Mitotic standard time [min]
e
Stage no.
Represen-tative image
Mitotic standard time [min]
Biological name
1 Pro
2 Late pro
3 Early prometa
4 Prometa 1
5 Prometa 2
6 Prometa 3
7 Late prometa
8 Early meta
9 Meta
10 Late meta 1
11 Late meta 2
12 Early ana
13 Ana
14 Late ana
15 Early telo
16 Telo
17 Late telo
18 Cytokinesis 1
19 Cytokinesis 2
20 Cytokinesis 3

Figure 3

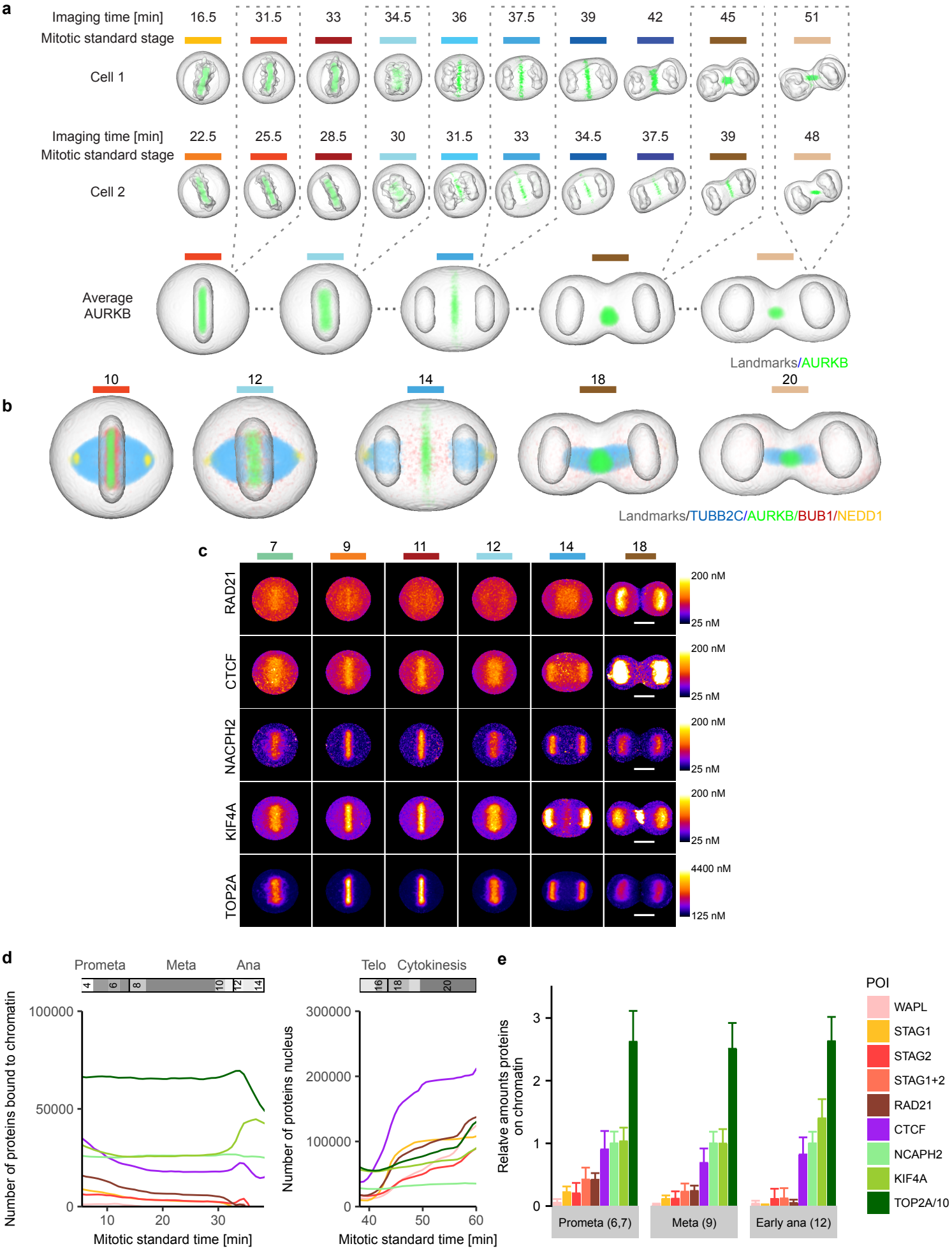

a
Imaging time [min]
16.5
31.5
33
34.5
36
37.5
39
42
45
51
Mitotic standard stage
Cell 1
Imaging time [min]
22.5
25.5
28.5
30
31.5
33
34.5
37.5
39
48
Mitotic standard stage
Cell 2
Average AURKB
Landmarks/AURKB
b
10
12
14
18
20
Landmarks/TUBB2C/AURKB/BUB1/NEDD1
c
7
9
11
12
14
18
RAD21
CTCF
NACPH2
KIF4A
TOP2A
200 nM
25 nM
4400 nM
125 nM
d
Prometa
Meta
Ana
Telo
Cytokinesis
Number of proteins bound to chromatin
Number of proteins nucleus
Mitotic standard time [min]
e
Relatve amounts proteins on chromatin
Prometa (6,7)
Meta (9)
Early ana (12)
POI
WAPL
STAG1
STAG2
STAG1+2
RAD21
CTCF
NCAPH2
KIF4A
TOP2A/10

Figure 4

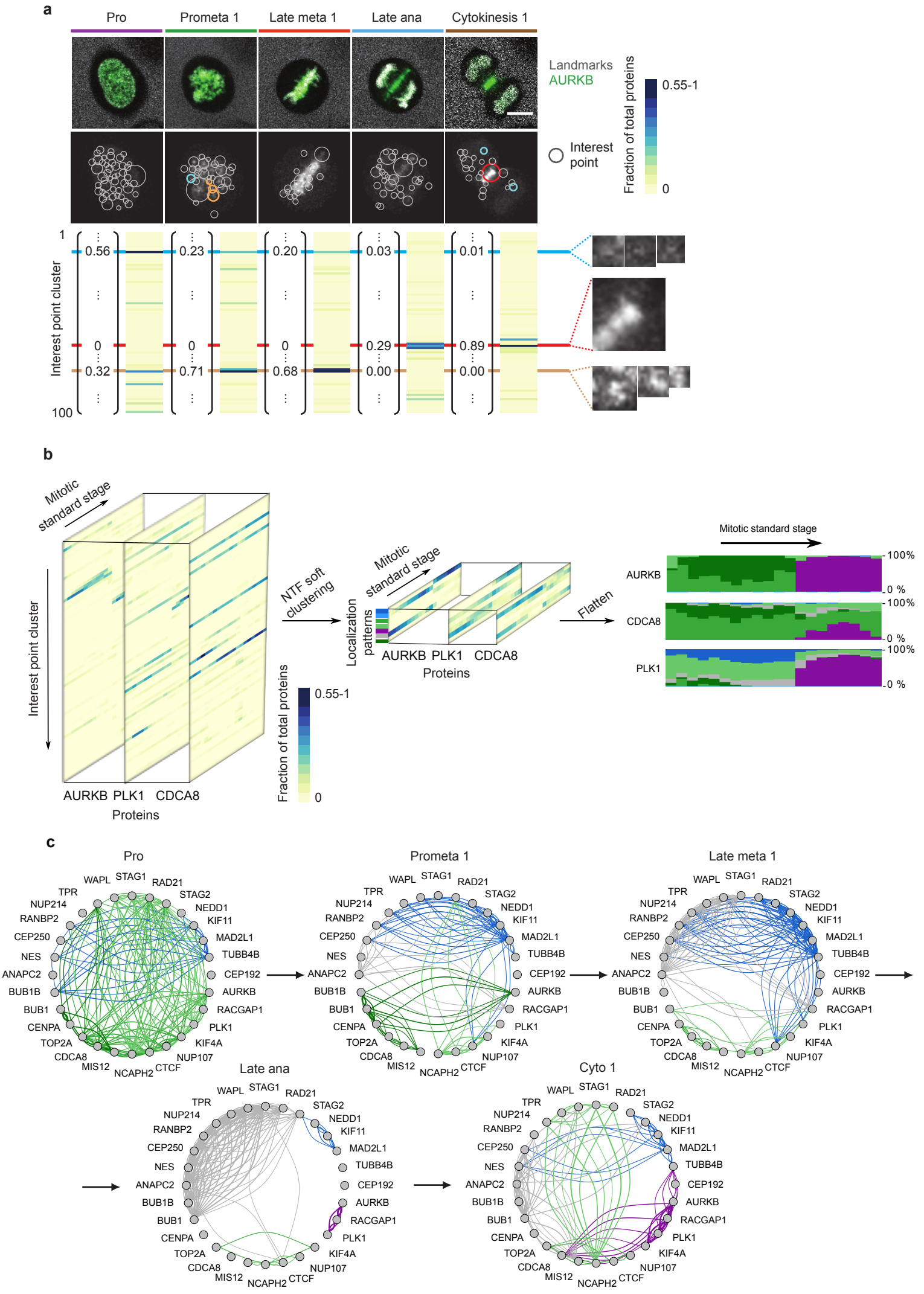
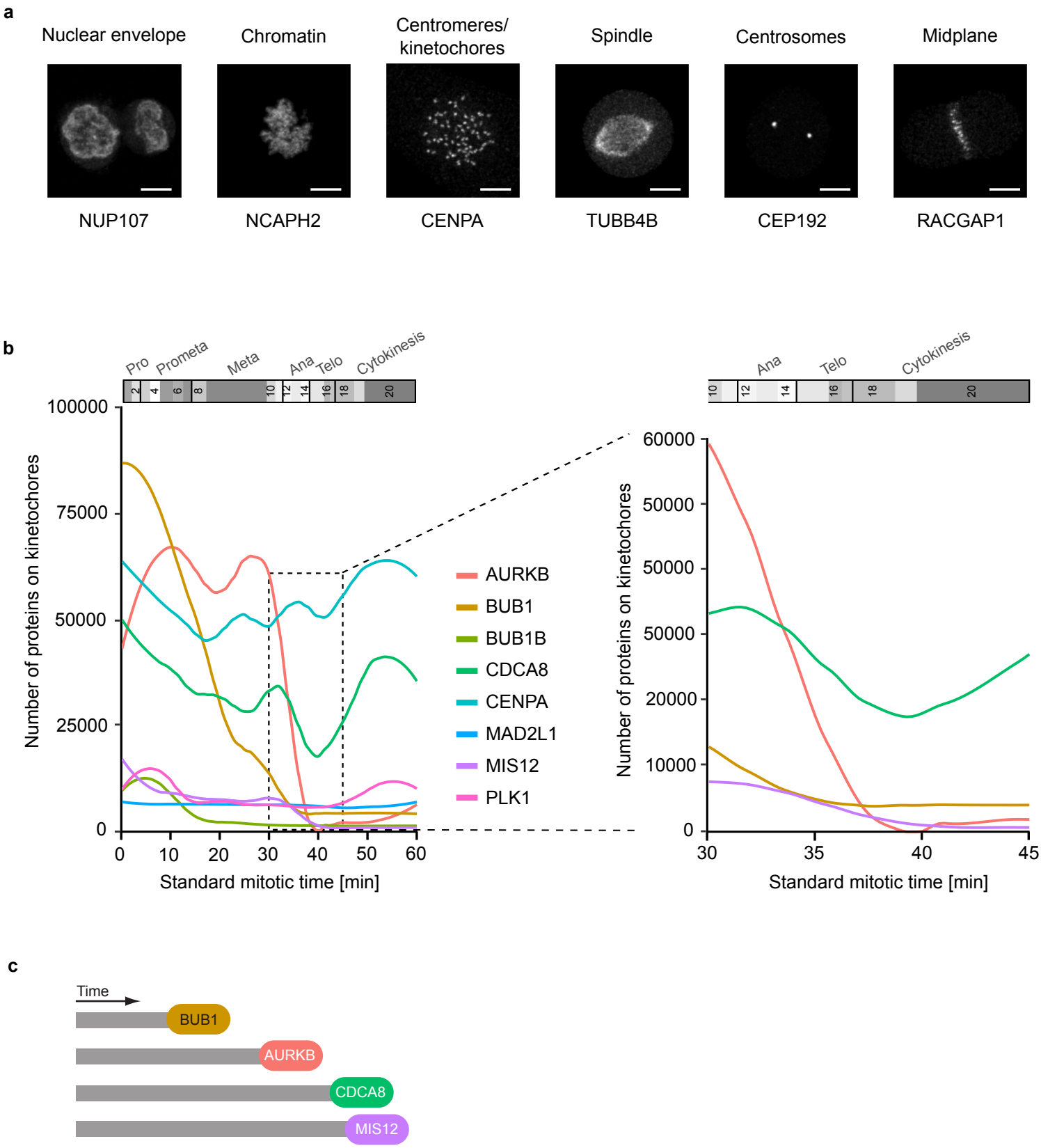

a
Pro
Prometa 1
Late meta 1
Late ana
Cytokinesis 1
Landmarks
AURKB
Interest point
Fraction of total proteins
0.55-1
0
Interest point cluster
1
100
0.56
0.23
0.20
0.03
0.01
0
0.29
0.89
0.32
0.71
0.68
0.00
0.00
b
Mitotic standard stage
Interest point cluster
AURKB PLK1 CDCA8
Proteins
NTF soft clustering
Localization patterns
Flatten
Mitotic standard stage
AURKB
CDCA8
PLK1
100%
0 %
c
Pro
Prometa 1
Late meta 1
Late ana
Cyto 1
WAPL STAG1 RAD21 STAG2 NEDD1 KIF11 MAD2L1 TUBB4B CEP192 AURKB RACGAP1 PLK1 KIF4A NUP107 CTCF NCAPH2 MIS12 CDCA8 TOP2A CENPA BUB1 BUB1B ANAPC2 NES CEP250 RANBP2 NUP214 TPR

Figure 5


a
Nuclear envelope
Chromatin
Centromeres/ kinetochores
Spindle
Centrosomes
Midplane
NUP107
NCAPH2
CENPA
TUBB4B
CEP192
RACGAP1
b
Pro
Prometa
Meta
Ana
Telo
Cytokinesis
Number of proteins on kinetochores
Standard mitotic time [min]
AURKB
BUB1
BUB1B
CDCA8
CENPA
MAD2L1
MIS12
PLK1
c
Time
BUB1
AURKB
CDCA8
MIS12