1   **Precise measurement of nanoscopic septin ring structures in deep learning-**

2   **assisted quantitative superresolution microscopy**

3

4   Amin Zehtabian*#, Paul Markus Müller*#, Maximilian Goisser, Leon Obendorf, Lea Jänisch,

5   Nadja Hümpfer, Jakob Rentsch and Helge Ewers

6

7   Institut für Chemie und Biochemie, Freie Universität Berlin, Thielallee 63, 14195, Berlin, Germany

8   *** Corresponding authors**

9   A. Zehtabian: amin.zehtabian@fu-berlin.de

10   P.M. Müller: paulmarkus.mueller@fu-berlin.de

11   **# equal contribution**

12

13   ## Abstract

14   The combination of image analysis and fluorescence superresolution microscopy methods

15   allows for unprecedented insight into the organization of macromolecular assemblies in cells.

16   Advances in deep learning-based object recognition enables the automated processing of

17   large amounts of data, resulting in high accuracy through averaging. However, while the

18   analysis of highly symmetric structures of constant size allows for a resolution approaching

19   the dimensions of structural biology, deep learning methods are prone to different forms of

20   bias. A biased recognition of structures may prohibit the development of readouts for

21   processes that involve significant changes in size or shape of amorphous macromolecular

22   complexes. What is required to overcome this problem is a detailed investigation of potential

23   sources of bias and the rigorous testing of trained models using real or simulated data

24   covering a wide dynamic range of possible results. Here we combine single molecule

25   localization-based superresolution microscopy of septin ring structures with the training of

26   several different deep learning models for a quantitative investigation of bias resulting from

27   different training approaches and finally quantitative changes in septin ring structures. We

28   find that trade-off exists between measurement accuracy and the dynamic range of

29   recognized phenotypes. Using our trained models, we furthermore find that septin ring size

30   can be explained by the number of subunits they are assembled from alone. Our work

31   provides a new experimental system for the investigation of septin polymerization.

1 Introduction

2

3 Light microscopy is a technique of fundamental importance in cell biology and with the

4 development of superresolution techniques has allowed unprecedented insight into cellular

5 processes on the brink of structural biology[1–7]. Machine learning-based computational image

6 processing furthermore allows for high-throughput analysis of even complex phenotypes in

7 cells[8–12] and has accelerated progress in cell biological discovery. In recent years a number of

8 open source deep learning platforms have been made available[13–22] that require minimal

9 know-how on the user side. At the same time, methods have been developed to reduce

10 background[23,24] and to accelerate image processing in superresolution microscopy[25–27].

11 Especially in superresolution microscopy, image processing allows for ultrahigh resolution of

12 multiprotein complexes such as nuclear pores[3,28] or centrioles[29,30]. However, these are highly

13 symmetric structures of constant size that bear several symmetry axes. As a result, they

14 contain intrinsic means to allow for accurate averaging. For superresolution microscopy and

15 averaging to harness its full power with the help of deep learning-based image processing to

16 improve resolution of more amorphous structures and to develop quantitative readouts for

17 structural changes it remains to be determined to what extent deep learning methods can

18 handle changes in organization or even size of target structures. Overfitting may make it

19 impossible to detect even small changes in morphology or size, on the other hand poor fitting

20 may not allow for accurate measurements.

21 Here we combine deep learning-assisted image analysis based on an open-source platform

22 with single molecule localization-based superresolution microscopy (SMLM) of subresolution

23 septin ring structures. Septins are the fourth cytoskeleton[31,32] and assemble from nonpolar

24 heteromultimeric rod-shaped complexes into ring-like structures of subresolution size when

25 not associated to tubulin or actin[33]. These rings are free of actin and are thought to represent

26 *bona fide* septin filamentous polymers[34]. The composition of septin complexes plays an

27 important role in septin function[35,36] and septin ring size depends on the composition of

28 septin complexes[37]. Septin ring size may thus be a long sought-after readout for complex

29 assembly and composition. Here we ask what factors in supervised deep learning affect ring

30 recognition and finally, accuracy of ring size measurements in native and perturbed cells with

31 differing ring size. To do so, we train six different models on data annotated according to

1   several different parameters and by three different experts and validate them on cellular and

2   synthetic SMLM datasets. We find that while most models readily recognize ring-structures

3   in the size range expected in native cells and allow for highly accurate determination of septin

4   ring diameter from SMLM data, only few models allow for the accurate recognition and

5   measurement of rings of differing size. We furthermore show that septin ring size can be

6   explained by septin complex composition alone, and we thus provide a new experimental

7   paradigm for the investigation of septin complex assembly. Our results demonstrate that the

8   combination of deep learning with SMLM can provide accurate readouts of changes in the

9   size of amorphous multiprotein structures and that experimental design and careful

10   validation are essential for the generation of reliable experimental pipelines.

11

12

1 Results

2 We chose septin rings as a model to test the robustness and accuracy of deep learning-

3 assisted superresolution microscopy for measurements of subresolution cellular structures.

4 The septins are a family of GTP-binding proteins that comprise the fourth cytoskeleton.

5 Septins can assemble into filaments themselves or attach to actin stress fibers, however,

6 when actin stress fibers are collapsed via the use of cytochalasin D, they form ring structures

7 of around 350 nm diameter[33,37] (Figure 1A). When imaged in SMLM, rings appear as

8 homogenously sized, symmetric rings with a thin perimeter (Figure 1B), but also incomplete

9 rings, arcs, small filamentous structures and aggregates can be found. When we acquired

10 many cells in SMLM imaging after Cytochalasin D treatment and selected hundreds of rings

11 from the stained septin structures, we could average them for accurate determination of their

12 diameter of 343.9 ± 5.0 nm (n = 181, Figure 1C). We concluded that we could accurately

13 determine septin ring diameter using this method.
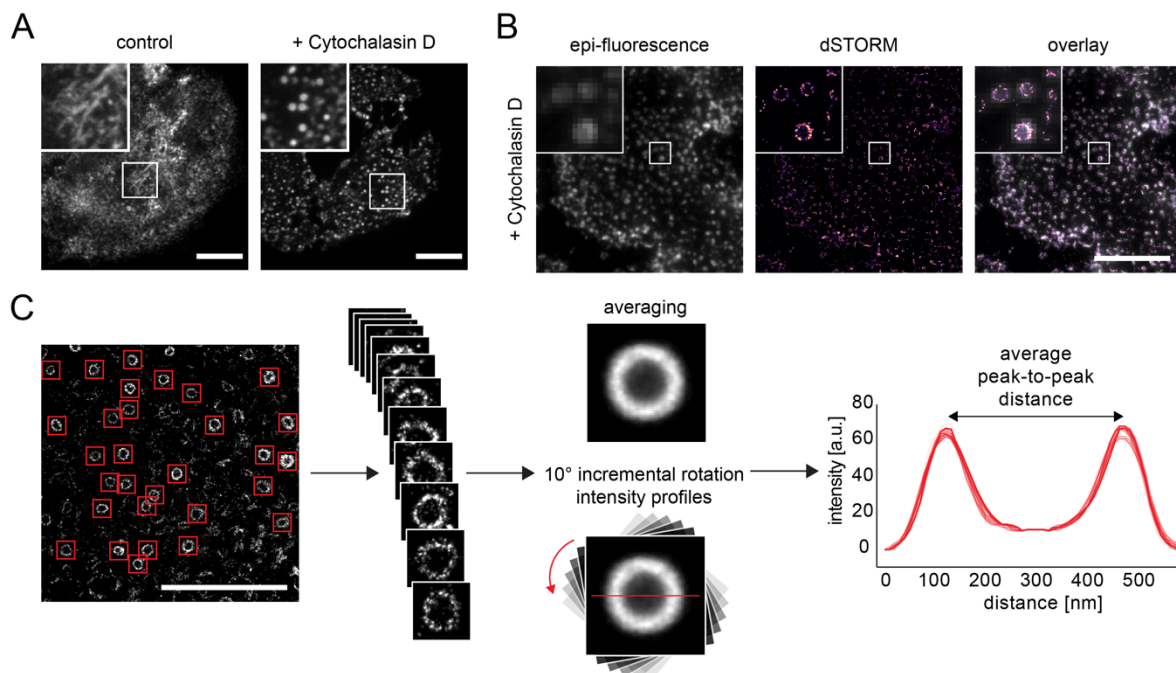
14
15 **FIGURE 1: Superresolution microscopy assay for quantification of septin ring diameter.**
16 **(A)** Genome-edited NRK52E cells expressing mEGFP-Septin-2 from both alleles were treated with 5 µM
17 Cytochalasin D for 30 min (or DMSO as control) prior to fixation, resulting in the formation of ring-like
18 structures. Images of mEGFP fluorescence were taken by spinning disc confocal microscopy. Scale bar: 10
19 µm. **(B)** Septin rings can be readily resolved by single molecule localization-based superresolution
20 microscopy. NRK52E cells were treated with 5 µM Cytochalasin D for 30 min and immunostained against
21 Septin-7. Images were taken by conventional epi-fluorescence microscopy and by dSTORM. Scale bar: 10
22 µm. **(C)** Scheme of the procedure for the measurement of septin ring diameter. Left to right: Septin rings
23 from superresolution microscopy images were cropped and averaged. The intensity profile of the averaged
24 rings was measured at 18 successive 10° incremental rotations and the peak-to-peak distance of all of these
25 profiles was averaged to return the ring diameter. Scale bar: 5 µm.

4

1   We next aimed to use a widely available and versatile deep learning network to train it to
2   recognize septin rings. We decided to use the ZeroCostDL4Mic[13] platform which provides free
3   cloud-based computation and the StarDist[38] network that has been widely used for detection
4   of cellular structures. However, since it had not been used for ring detection we decided to
5   test, if it could be trained from scratch to recognize septin rings in SMLM reconstructions. To
6   do so, we used an annotated dataset of 496 images for training and adjusted parameters of
7   the StarDist network until we could reliably reach convergence of our model.

8   We next aimed to ask, how annotation bias may affect model learning and the capability of
9   the model to detect phenotypes strongly different from the wildtype (*wt*) situation in our
10   images. In cellular SMLM images, septin rings are of variable circularity and completeness in
11   labeling and occasionally touch. We hence decided to test if higher or lower stringency in
12   determining a ring would be suitable for training and if masks precisely following the ring
13   outline would lead to more accurate results. After training, we furthermore asked, whether
14   the deep learning models would be able to accurately recognize and measure also
15   significantly smaller or bigger rings.

16   For training, we generated a large amount of SMLM images of cytochalasin D-treated cells,
17   which we separated into a training set of 496 images and a cellular holdout test dataset of 32
18   images. We then designed a framework to train deep learning models using six different
19   annotation strategies and to test the models on the cellular holdout data, the synthetic data
20   and a biological test case, that is rings devoid of Septin-9 (Figure 2). We asked one human
21   expert to annotate in three different ways in terms of accuracy of the masks (how does it fit
22   to the ring) and labeling stringency (how circular and/or continuous do rings have to be) and
23   after that asked two more human experts to individually annotate with custom masks and
24   low stringency. A final sixth model was trained using the consensus masks. All trained models
25   resulted in robust recognition of rings in the cellular test dataset of 32 SMLM images and we
26   benchmarked the models in terms of precision (fraction of recognized objects that are
27   present in ground truth), recall (fraction of objects in ground truth that are recognized) and
28   measured average ring size compared to ground truth (Figure 2). We then went on to run the
29   models on synthetic SMLM data generated using FluoSim[39] in which ring sizes from 125 to
30   500 nm were simulated. Finally, we tested the two best performing models on a cellular
31   example of ring size change, cells in which Septin-9 had been removed from the complex,
32   leading to significantly smaller rings[37].
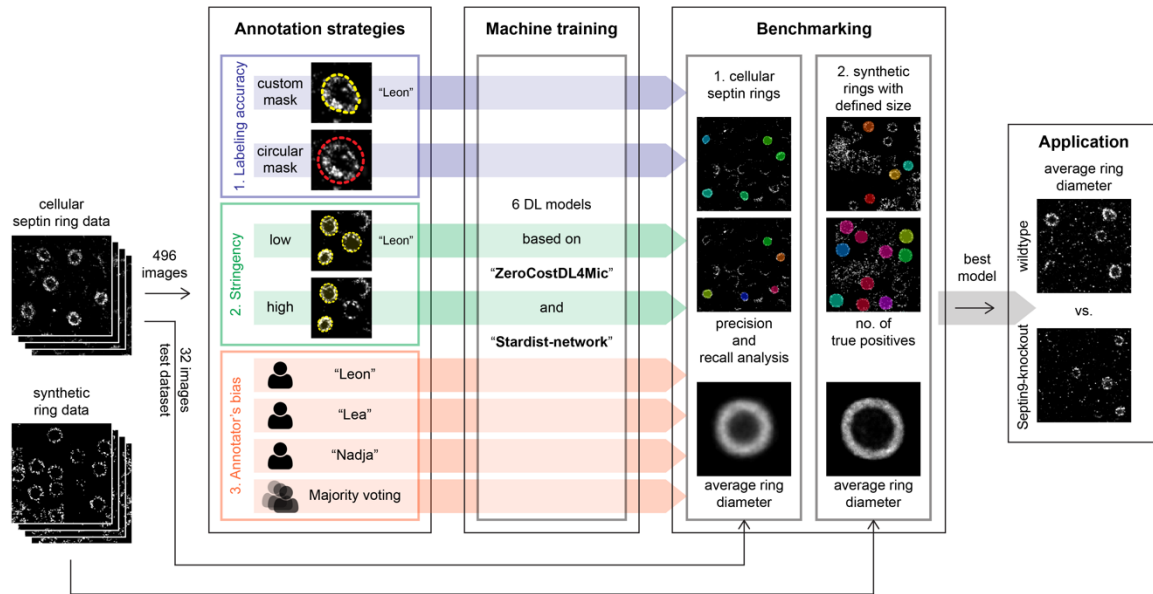
1
2 **FIGURE 2: General annotation, training and benchmarking strategy for deep learning-assisted septin ring**
3 **recognition and measurement.**
4 528 dSTORM images of septin rings in NRK52E cells were generated and used for annotation and training
5 (496 images) or as holdout test dataset (32 ground truth images). To benchmark annotation strategies for
6 deep learning, images were annotated with different labeling accuracy, annotation stringency, and by
7 three independent human experts as well as the majority voting thereof. For each annotation strategy we
8 generated a deep learning models based on the ZeroCostDL4Mic platform using the StarDist network. For
9 benchmarking, we compared the ability of the models to recognize the rings and to return the correct ring
10 diameter, both in cellular data and synthetic data with rings of known diameter. The best performing
11 models were then applied to measure the diameter of septin rings in knockout cells.

12

13 To evaluate the influence of labeling strategies on recognition by the deep learning models,

14 a septin biologist recognized septin rings in the cellular test dataset of SMLM data. These we

15 accepted as *bona fide* septin rings and the resulting dataset and average diameter are in the

16 following termed ground truth. We then asked a human expert, "Leon", to annotate the

17 training set according to three different strategies. The first strategy used low stringency in

18 recognition as a septin ring and required the expert to draw a custom mask around each ring

19 (Figure 2). The second strategy required the expert to use low stringency in septin ring

20 recognition, but to use a circular mask or, to be more precise, a convex circular-shape mask

21 with flexible size that covered the target ring for annotation. The third strategy required the

22 expert to apply high stringency in selecting rings and to use a custom drawn mask. When the

23 annotation-specific models were trained and applied to the cellular test dataset, we found

24 that all models recognized a large fraction of rings reliably when compared to ground truth

25 albeit with varying precision and recall (Figure 3A). Specifically, the low stringency, custom

26 mask trained model led to the highest recall with 89.5% of rings found that were present in

1   the ground truth. On the other hand, the high stringency, custom mask trained model led to

2   the highest precision with 89% of all recognized rings being true positives, but recall was low.

3   The circular mask, low stringency model was of intermediate precision and recall and no

4   model showed precision nor recall below 50%. Overall, we concluded that our models could

5   reliably recognize septin rings in our cells.
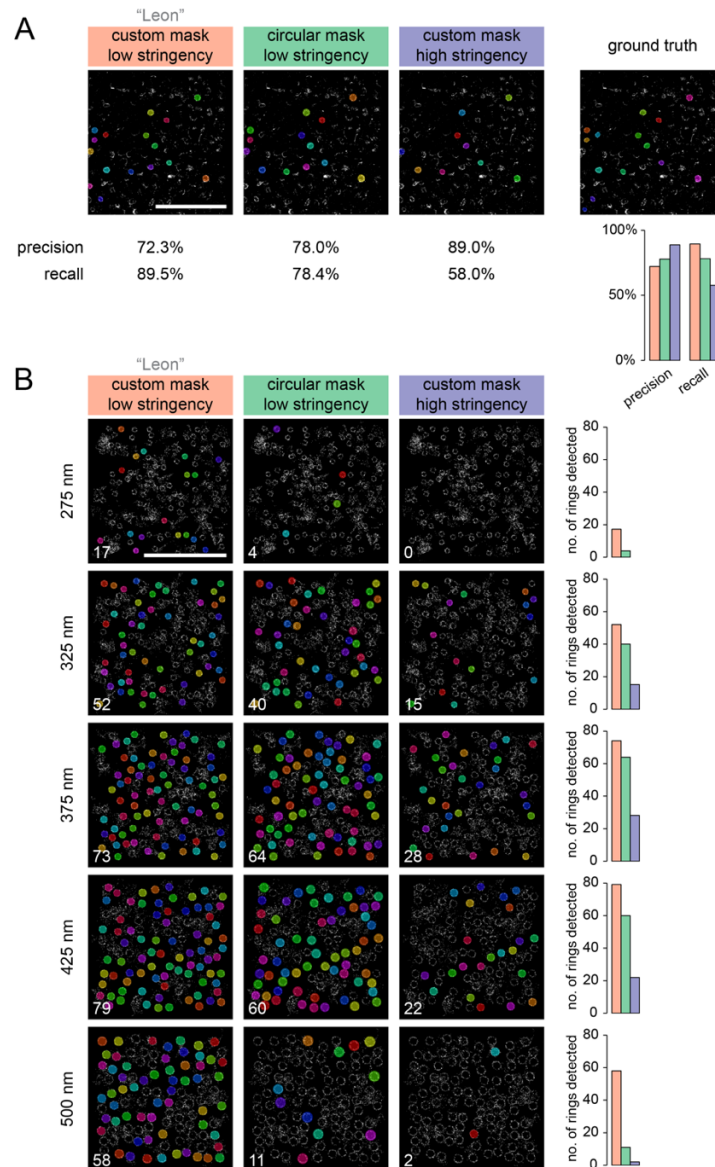
6
7   **FIGURE 3: Effect of annotation accuracy and stringency on the ability of deep learning models to detect**
8   **rings of different size.**
9   **(A)** Overlay of randomly colored masks of recognized septin ring structures in dSTORM images by 3
10   different deep learning models: custom mask and low stringency model (aka "Leon"); circular mask and
11   low stringency model; custom mask and high stringency model. The rings detected in a representative
12   dSTORM reconstruction from the holdout data are shown together with the respective ground truth.
13   Precision and recall calculations are shown below. Scale bar: 5 µm. **(B)** Overlay of randomly colored masks
14   of recognized ring structures in synthetic dSTORM images with predetermined ring diameters from 275 to
15   500 nm as recognized by the deep learning models as in (A). Number of detected rings in each condition is
16   shown on the bottom left of each image and plotted on the right. Scale bar: 5 µm.

7

When the composition of septin complexes is changed, the diameter of septin rings changes as well[37], suggesting that ring size may provide a readout for septin complex assembly and composition. However, to accurately measure the diameter of rings with different sizes, our models must recognize rings of even strongly differing diameters for an accurate readout of ring-size changes. We thus aimed to test whether one of our training strategies would lead to a bias in the recognition of differently sized rings and only recognize rings in the size-range found in our untreated cells. To do so, we analyzed a panel of synthetic SMLM test datasets in which rings of a predefined diameter from 275 to 500 nm, respectively, were present. When we then applied our models to recognize rings in these synthetic test datasets, we found striking differences in ring recognition (Figure 3B). While the low stringency, custom mask model (hereafter called "Leon") reliably recognized rings in all test cases, the other two models failed to reliably recognize rings for cases significantly below or above the ring sizes present in our cells. We concluded that the custom mask, high stringency model (hereafter called "high stringency") and the circular mask, low stringency model (hereafter called "circular mask") were biased for rings in the size range found in cells and that the training strategy significantly impacts recognition of septin rings towards superresolution measurements.

Next, we aimed to investigate how three different experts using the same annotation strategy would influence ring recognition by the resulting models. We asked two additional experts, "Lea" and "Nadja" to annotate the cellular SMLM training dataset according to the low stringency, custom mask strategy. From the annotations of the three experts "Leon", "Lea" and "Nadja" we furthermore created a "majority voting" annotation. When we then trained additional models according to the "Lea", "Nadja" and "majority voting" annotations on our cellular test dataset, we found that all models exhibited a precision above 60% and a recall above 80% when compared to ground truth (Figure 4A). The "Nadja" model exhibited an exceptional recall of 92.8%, but only 60.4% precision, meaning that the "Nadja" model recognized many objects that were not found in the ground truth. The performance of the "majority voting" model was very similar to that of the "Leon" model. We concluded that indeed the specific annotation style of different experts influences ring recognition significantly. When we next applied the "Lea", "Nadja" and "majority voting" models to the panel of synthetic test datasets and compared them to the "Leon" model, we found significant discrepancies in performance (Figure 4B).
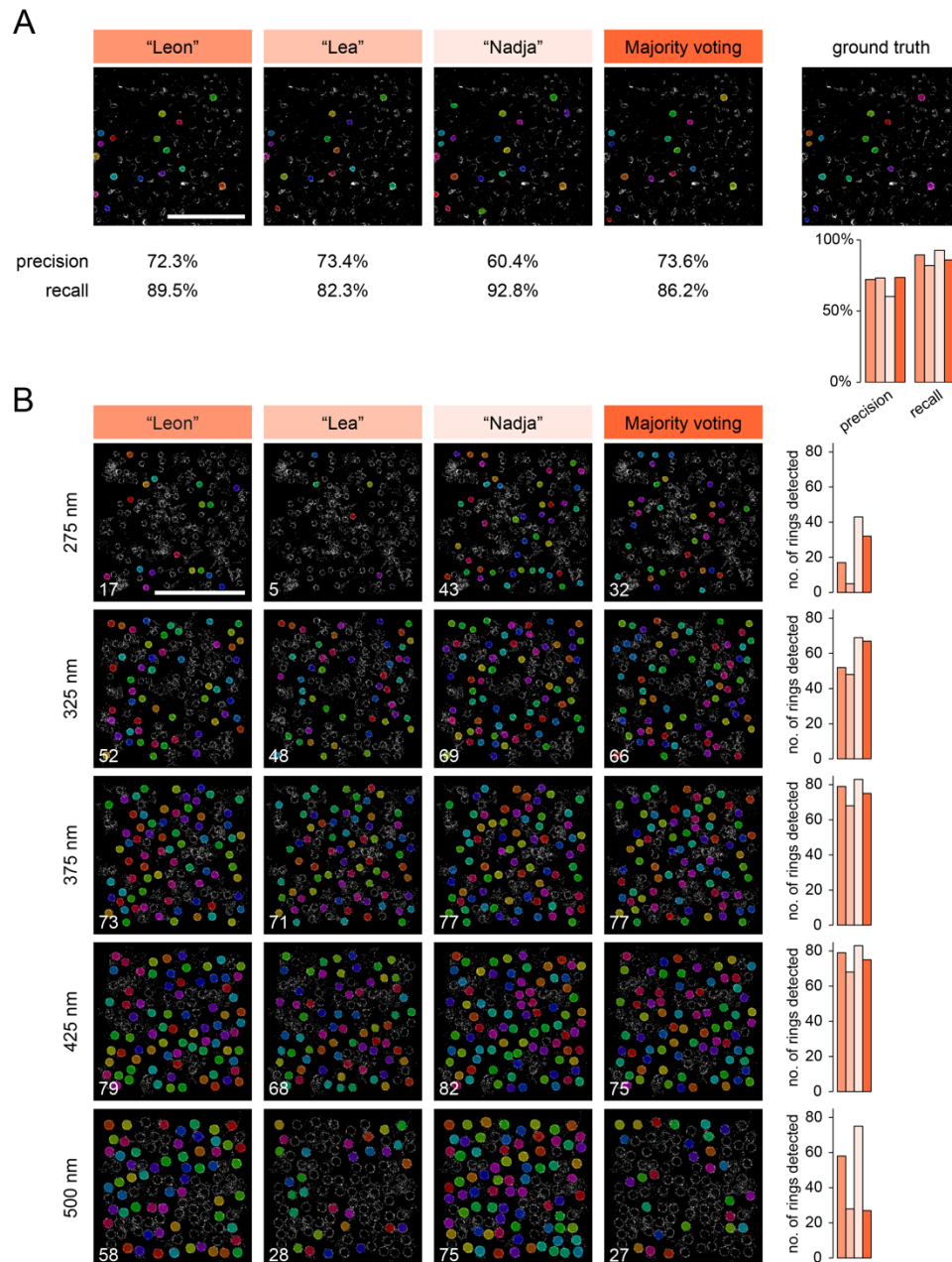
FIGURE 4: Effect of annotator's bias on the ability of deep learning models to recognize rings of different size.
(A) Overlay of randomly colored masks of recognized septin ring structures in dSTORM images by 4 different deep learning models: custom mask and low stringency model each for individual annotators "Leon", "Lea", "Nadja" and for majority voting model. The rings detected in a representative dSTORM reconstruction from the holdout data are shown together with the respective ground truth. Precision and recall calculations are shown below. Scale bar: 5 µm. **(B)** Overlay of randomly colored masks of recognized ring structures in synthetic dSTORM images with predetermined ring diameters from 275 to 500 nm as recognized by the deep learning models as in (A). Number of detected rings in each condition is shown on the bottom left of each image and plotted on the right. Scale bar: 5 µm.

Both the "Nadja" and the "majority voting" models reliably outperformed the other models for smaller rings, while for rings in the range of those found in cells, all models performed

1   equally well. We also tested rings below 275 nm in size (data not shown). For the smallest size

2   of 125 nm, no model could recognize any rings in synthetic SMLM test datasets. Only the

3   "Nadja" model could recognize rings reliably from a diameter of 245 nm upward, making it

4   clearly stand out from the other models. Large (500 nm) rings on the other hand were readily

5   recognized by the "Nadja" and "Leon" models, while the "Lea" and "majority voting" models

6   recognized less than half of the rings detected by either the "Nadja" or "Leon" models. We

7   concluded that the bias introduced by individual annotators has significant consequences on

8   the recognition of septin rings in SMLM data.

9   We next aimed to ask, to what extent these biases in septin ring annotation would impact the

10 accurate measurement of septin ring diameter in datasets analyzed by our models. To do so,

11 we averaged the recognized septin rings for every model and condition and measured the

12 diameter in nm as described in Figure 1C. When we compared the results of our models on

13 the cellular SMLM test dataset, we found that all models resulted in a measurement with an

14 accuracy within 3% of the 343.9 nm found in the ground truth (Figure 5A). We next aimed to

15 ask, if we could measure differences as small as 5 nm in ring diameter. To do so, we generated

16 synthetic SMLM test datasets with diameters in 5 nm steps from 350 nm, to 380 nm. When

17 we ran our models on these datasets and measured the average diameters of the respectively

18 recognized rings, we found that all models resulted in very accurate measurements of the

19 respective ring sizes, with errors maximally 3 - 4 nm (Figure 5B). Overall, the highest spread

20 was found for the "circular mask" and "high stringency" models (Figure 5C), while the

21 residuals were lowest for the "majority voting" model (Figure 5D). We concluded that all

22 models were suitable to automatically extract and measure ring size in native cellular SMLM

23 datasets. When we plotted the measured ring diameter against ground truth ring diameter in

24 synthetic SMLM test data, we found that all models exhibited excellent linearity over the

25 range from 275 to 500 nm with R^2-values above 0.94 (Figure 6A). While thus all models were

26 very precise in their measurement of ring diameter, a clear dependence of measurement

27 accuracy on the number of recognized rings was found (Figure 6B), suggesting that the

28 generation of higher amounts of data could further ameliorate measurement accuracy for the

29 less well performing models. In Figure 6C, we summarize the data from Figures 3-5 in an

30 accessible overview according to the z-scores for ring recognition (top) and ring diameter

31 measurement (bottom). Found values are horizontally z-scored, thus emphasizing differences

1 between models even if they may be small in absolute size and describing relative

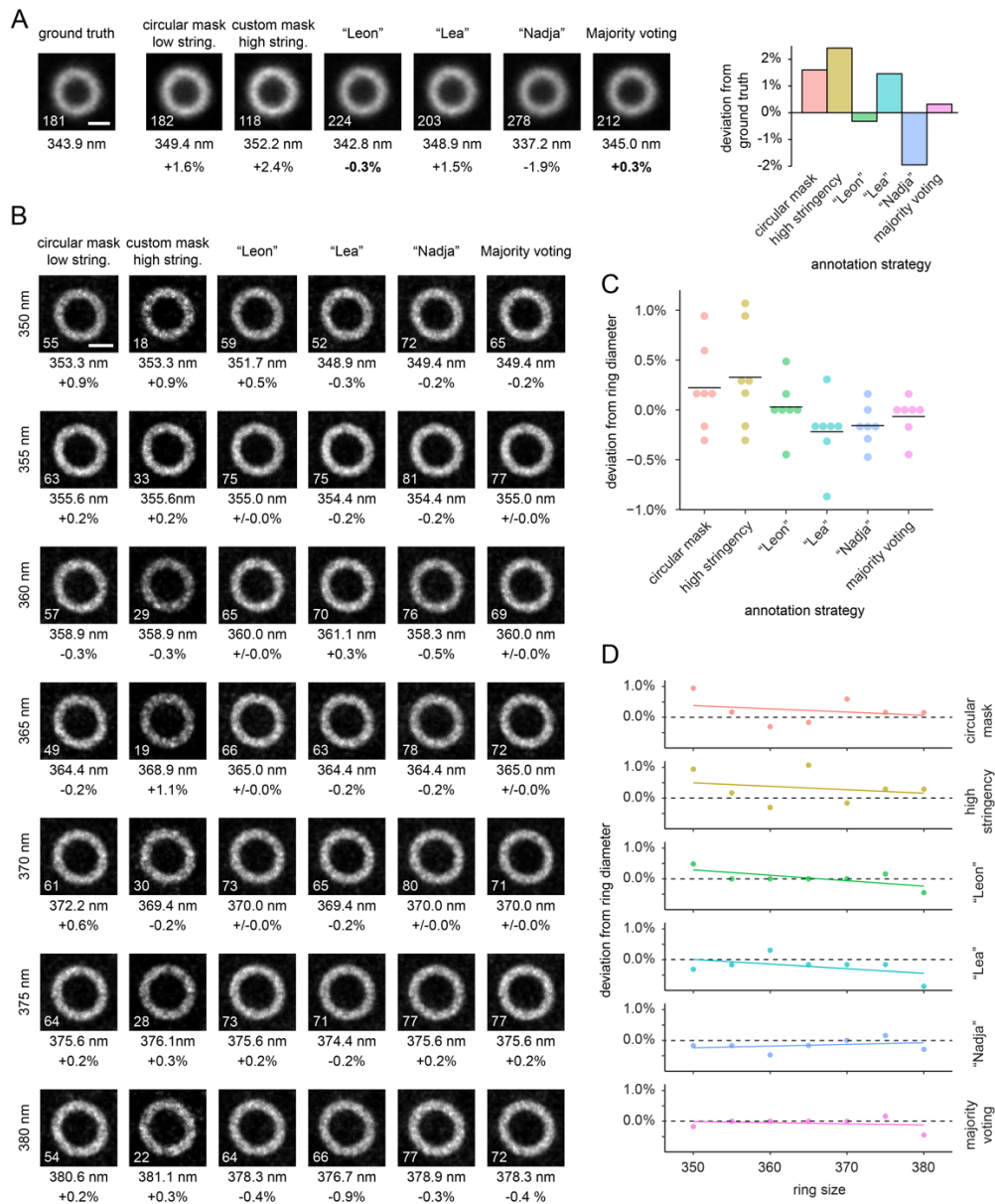2 performance. However, this overview allows for general observations.

3
4 **FIGURE 5: Quantification of the precision of deep learning-based ring measurements.**

5 **(A)** Results for septin ring diameter measurements in cellular dSTORM holdout data by the respective deep
6 learning models. Averaged recognized ring images are shown for each model and average ring diameter
7 and deviation from ground truth are shown below the images. Number of rings averaged is shown on the
8 bottom left of each image. The plot on the right shows the deviation of measured septin ring diameter
9 from ground truth for each model in %. Scale bar: 200 nm. **(B)** Results for septin ring diameter
10 measurements in synthetic dSTORM data of known ring size by respective deep learning models.
11 Underlying ground truth ring diameter is given on the left. Averaged recognized ring images are shown for
12 each model and diameter, average ring diameter and deviation from ground truth are shown below the
13 images. Number of rings averaged is shown on the bottom left of each image. Scale bar: 200 nm. **(C)**
14 Deviation from respective ground truth ring diameter in percentage points measured for all models and
15 for all 7 ground truth ring diameters shown in (B). Black lines indicate mean deviation. **(D)** Deviation from
16 respective ground truth ring diameter in percentage plotted against ground truth diameter for all models.
17 Dashed line marks 0% deviation, colored lines show a linear fit.
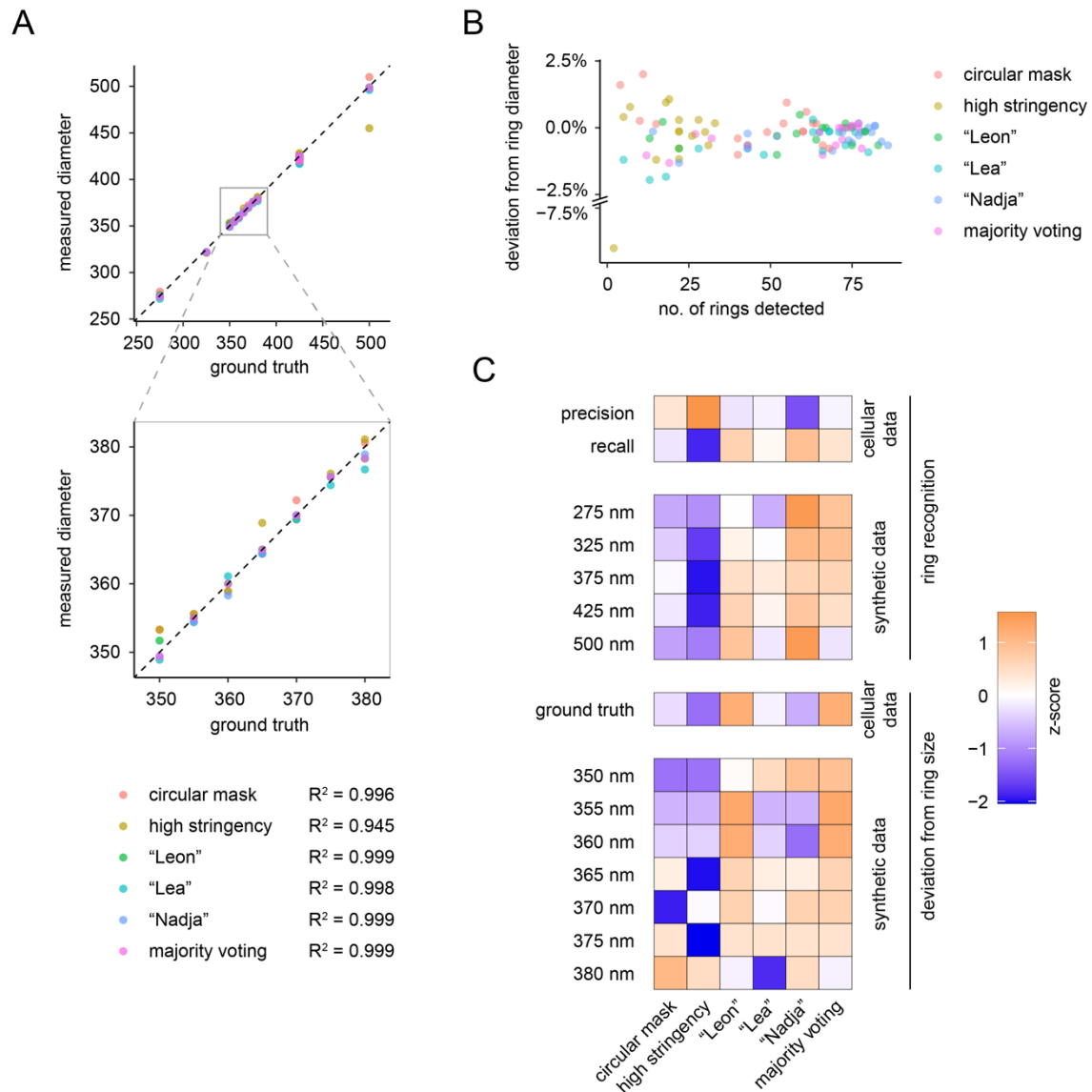
1
2 **FIGURE 6: Deep learning models accurately and robustly measure septin ring size in dSTORM data.**
3 **(A)** Measured ring diameter in synthetic dSTORM data plotted against ground truth ring diameter for all
4 models used in this study. The magnified inset below shows the ring size range exemplified in Figure 5B.
5 **(B)** Deviation from respective ground truth ring diameter for synthetic rings from 275 nm to 500 nm in
6 percentage points measured for all models plotted against the respective number of detected rings in each
7 case. Note that deviation from ground truth decreases with larger number of rings detected. **(C)** Global
8 analysis of deep learning model performance. From top to bottom: Precision and recall values in cellular
9 data, number of rings recognized in synthetic data, ring size deviation from ground truth in cellular data
10 and deviation from ground truth ring size in synthetic data. Heat map shows horizontally z-scored values.
11

12 It seems that the z-scores for recall and precision, respectively, are correlated with specific

13 outcomes. The model with the lowest precision, "Nadja", performed best in recognizing rings

14 of sizes significantly different from those measured in native cells, suggesting that a less

15 stringent selection of cellular structures as rings allows the model to recognize rings of

16 different sizes more easily. On the other hand, the model with the highest precision, "high

12

1   stringency", did not perform well in synthetic data of ring sizes. For recall on the other hand,

2   the models that most reliably recognized and most accurately measured rings in the expected

3   size in cells, showed the highest values ("Leon" and "majority voting"), whereas the models

4   with low recall ("high stringency" and "circular mask") exhibited the worst performance. We

5   concluded that precision and recall are important indicators of model performance, but they

6   do not suffice to predict experimental outcome.

7   Having found models that can recognize septin rings of different sizes and accurately measure

8   their diameter, we decided to test our models on an experimental system, where septin ring

9   size is known to vary. Septin complexes in mammalian cells are non-polar, palindromic rod-

10   shaped heterooctamers of around 34 nm length[40–44] that bear two Septin-9 subunits in their

11   middle. When Septin-9 is removed from cells through RNA interference, septin rings are

12   known to be smaller[37]. This suggests that ring size is connected to complex length as the

13   septin complexes created in absence of Septin-9 are known to be hexamers[45] of around 26

14   nm length. If that is the case, septin ring size should reflect complex size and ring size should

15   be dependent on the number of complexes in the ring. To ask, whether we can thus estimate

16   the number of septin complexes in rings from SMLM measurements, we performed SMLM on

17   embryonic fibroblast cells from *wt* and Septin-9$^{-/-}$ mice[46]. We found septin rings to be in the

18   size range of 300 - 400 nm and decided to analyze ring diameter using the "Leon" and the

19   "majority voting" models as they most accurately measured the average ring diameter in that

20   range. Indeed, both models readily recognized many rings in measured SMLM data (Figure

21   7A). In *wt* cells, we measured an average ring diameter of 413.3 nm ("majority voting") or

22   425.5 nm ("Leon"). Geometrically, this would translate into 39 septin octamers with a 9.2°

23   association angle ("Leon") or 38 octamers with a 9.4° association angle ("majority voting").

24   On the other hand, in Septin-9$^{-/-}$ cells, we measured ring diameters of 334.4 nm ("Leon") or

25   335.0 nm ("majority voting"), which would translate into 40 septin hexamers with an 8.9°

26   association angle for both models (Figure 7B). Strikingly, we could thus determine that septin

27   ring size in our cells must be determined by the propagation angle of about 9° between

28   polymerizing septin complexes that would then lead to 40 septin complexes required to form

29   a ring. We concluded that machine learning supported SMLM superresolution microscopy

30   could reveal information on the ultrastructural arrangement of septin complexes in rings.

31
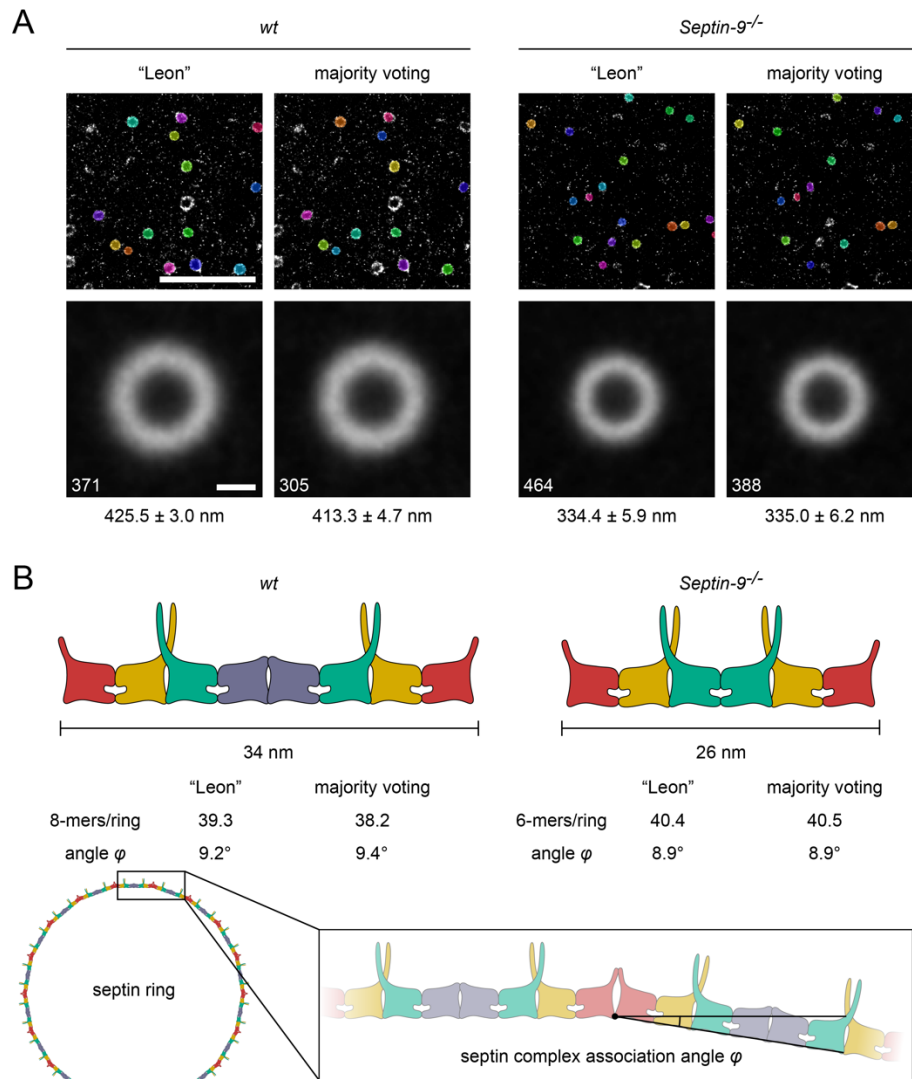
FIGURE 7: Accurate measurement of cellular septin ring size reveals the geometric architecture of septin rings.

(A) Septin ring recognition in wild-type (*wt*) and Septin-9$^{-/-}$ MEFs. Septin ring recognition was performed with the two deep learning models "Leon" and "majority voting". Representative overlay of randomly colored masks of recognized septin ring structures in dSTORM images and averaged ring images are shown for each case. Average ring diameter is shown below. Number of rings averaged is shown on the bottom left of each image. Scale bars: upper row 5 µm, bottom row 200 nm. **(B)** Model of septin ring assembly. Shown are schematic representations of septin complexes. In mouse fibroblasts the knockout of Septin-9 is thought to remove the central Septin-9 dimer (blue) from the palindromic, heterooctameric complex, resulting in a heterohexamer. Since the polymerization interface for septin filament assembly is the same for hexamer and octamer, the rules for assembly of rings should be the same for both complexes and resulting rings thus smaller. Septin rings measured via deep learning-assisted superresolution microscopy data analysis in Septin-9$^{-/-}$ cells are smaller than wildtype rings, consistent with a constant mode of assembly of rings from ~ 40 complexes with an interface angle φ of ~ 9°.

14

1 **Discussion**

2 We here developed a deep learning-based assay for the recognition and measurement of

3 septin ring size in superresolution images of mammalian cells. We trained deep learning

4 models in several different ways to investigate possible bias. We demonstrate that the outline

5 of the annotation mask, the stringency in annotation, and the generation of a consensus

6 between experts influence the outcome of both the accuracy of measurements as well as

7 robustness in detection of target structures. For example, we found that some of the models

8 were more sensitive to changes in ring size than others. Especially if the models are supposed

9 to recognize structures with a different phenotype in cells, such bias can be decisive. This

10 suggests that benchmarking on simulated data covering the whole phenotypic range provides

11 an excellent means to control the reliability of the deep learning models.

12 So far, deep learning models for the recognition of fluorescence labeled structures in cells

13 have mostly been used to recognize structures of low or highly predictable variability such as

14 cellular nuclei[17,38,47]. The correct recognition of different phenotypes is, however, an essential

15 property of data analysis models in experimental biology. To ensure this is possible for our

16 models, we use a wide array of synthetic SMLM data to test our models on the capability to

17 recognize phenotypes that are different from what is observed in *wt* cells. Our data suggest

18 that very stringent criteria in the selection of native rings may lead to overfitting and thus be

19 detrimental to the recognition of rings of different size in cells, as observed for the "high

20 stringency" and likely also the "Lea" model. On the other hand, less stringent criteria lead to

21 a slightly lower accuracy in the measurement of ring diameter, as observed for the "Nadja"

22 model. From all of our models, the "majority voting" model belonged to the best performing

23 models both for recognizing *wt* cellular rings as well as for recognizing rings of significantly

24 different size. It also resulted in very low error in ring size determination. Our data agree with

25 previous observations that consensus or majority voting models often perform better than

26 models based on individual annotators[48].

27 Our data provide insight into the requirements in experimental design of deep learning-based

28 data analysis pipelines of superresolution microscopy specifically and into the generation of

29 robust deep learning models in general. We furthermore provide a robust quantitative assay

30 for the investigation of subresolution structures assembled from an essential cytoskeletal

31 component, the septins.

1   Our finding that the septin ring diameter both in Septin-9 *wt* and in Septin-9$^{-/-}$ cells can be

2   explained by 40 complexes adding an angle of 9° each towards ring closure suggests that

3   superresolution microscopy of septin complexes may be useful as a readout to investigate

4   septin complex assembly in the future. That the angle in Septin-9 containing complexes is the

5   same as in Septin-9-free complexes suggests that the 9° propagation angle is not introduced

6   at the center of the hexamer, but rather at the polymerization interface between terminal

7   Septin-2-group septins in adjacent complexes. In baker's yeast *Saccharomyces cerevisiae,* it is

8   known that the terminal septin in the heterooctamer, here *cdc11* or *shs1*, determines the

9   degree to which septin filaments bend. Straight filaments form with *cdc11* and ring-like

10   filaments form with *shs1* at the termini of the complex[49]. It is tempting to speculate that the

11   Septin-2-group septins, which occupy the termini of the mammalian octamer may play a

12   similar role. It is known that the Septin-2 group members (Septin-1, Septin-2, Septin-4, and

13   Septin-5) can replace one another at position X in a Septin-X/6/7/7/6/X heterohexamer

14   biochemically[50] and that they exhibit non-overlapping functions[36,51–53] that may be a result of

15   their mode of filament assembly. In the future, our assay will allow for investigations into

16   potential influences of complex-terminal septin subunits on ring-size and thus septin

17   polymerization properties.

1   Materials and methods

2

3   Cell culture and immunostaining.

4   The NRK52E-Septin-2-GFP[EN/EN] genome edited cell line expressing mEGFP-Septin2 from both

5   alleles was generated as described[54]. All steps in the following were carried out at room

6   temperature unless otherwise stated. NRK52E-Septin-2-GFP[EN/EN], Septin-9[c/c] (*wt*) or Septin-9[-]

7   [/-] (knockout) mouse embryonic fibroblasts[46] were seeded on round coverslips (thickness no.

8   1.5) coated with 0.01% (w/v) poly-L-Lysine solution (Sigma) and grown in DMEM (Gibco,

9   Thermo Fisher Scientific) supplemented with 10% (v/v) fetal calf serum, 2 mM L-Glutamine

10   (Gibco, Thermo Fisher Scientific) and 100 U/ml penicillin–streptomycin (Sigma) at 37 °C and

11   5% CO_2 in a humidified incubator. The next day, cells were treated with 5 µM Cytochalasin D

12   for 30 min at 37 °C in DMEM without supplements to induce septin ring formation. Cells were

13   fixed in 4% (w/v) paraformaldehyde solution in PHEM buffer (60 mM PIPES-KOH pH 6.9, 25

14   mM HEPES, 10 mM EGTA, 1 mM $MgCl_2$) for 15 min at 37°C. 1.5. Fixation was quenched with

15   50 mM NH_4Cl in PHEM buffer for 7 min. Cells were permeabilized using 0.25% (v/v) TritonX-

16   100 in PHEM buffer for 5 min and then blocked with Image-iT FX signal enhancer (Invitrogen,

17   Thermo Fisher Scientific) for 30 min before blocking for 1h with 4% (v/v) horse serum, 1%

18   (w/v) BSA and 0.1% (v/v) TritonX-100 in PHEM buffer. The sample was then stained with 10

19   µg/ml anti-GFP nanobody coupled to Alexa Fluor-647 at 4°C overnight or with rabbit anti-

20   human-Septin-7 IgG (1:500, IBL America) overnight and subsequently with goat anti-rabbit

21   Alexa Fluor-647 conjugate IgG-Fab (1:500, Thermo Fisher Scientific) for 45 min in PHEM buffer

22   supplemented with 1% (w/v) BSA and 0.1% (v/v) TritonX-100. Samples were post-fixed in 4%

23   (w/v) paraformaldehyde in PHEM buffer for 10 minutes and quenched as described above.

24   For the cellular septin ring dataset NRK52E 13 samples were prepared out of which 6 were

25   stained with anti-GFP nanobody and 7 were stained with anti-Septin-7 antibody resulting in

26   288 and 240 images, respectively. Out of these 528 images 32 were randomly selected to be

27   used as test dataset while the remaining 496 images were used for training of the deep

28   learning models. For the Septin-9[c/c] or Septin-9[-/-] mouse embryonic fibroblasts 96 images

29   were generated from each sample.

30

31

Spinning disc confocal microscopy.

Images were acquired on an inverted IX71 microscope (Olympus) equipped with a CSU-X1 spinning disk unit (Yokogawa) and an iLas laser illumination system (Gataca Systems) with a 491 nm laser for illumination. A 60x NA 1.42 oil objective (Olympus) was used, and images were taken with an ORCA Flash 4.0LT sCMOS camera (Hamamatsu). The System was operated using the software MetaMorph.

dSTORM of septin rings.

Coverslips with immunostained samples were mounted in a mounting chamber and filled with GLOX + BME buffer (4% (w/v) glucose, 140 mM β-mercaptoethanol, 10 mM NaCl, 200 mM Tris-HCl pH 8, 500 µg/ml glucose oxidase, 40 µg/ml glucose catalase and 10% (v/v) glycerol). dSTORM images were acquired on a Vutara 352 superresolution microscope (Bruker) equipped with an ORCA Flash4.0 sCMOS camera (Hamamatsu) and a 60x NA 1.49 ApoN oil immersion total internal reflection fluorescence (TIRF) objective (Olympus), yielding a pixel size of 98 nm. Data were acquired in TIRF illumination at a laser power density of ~35 kW/cm^2 using a 639 nm laser. dSTORM images were reconstructed from 10,000 images taken with an exposure of 20 ms per image. Reference epi-fluorescence images were taken with an exposure of 500 ms and low laser power density. Images were reconstructed with a pixel size of 10 nm using the software Vutara SRX v.6.04.14.

Generation of synthetic dSTORM ring data.

Synthetic storm data was generated using FluoSim[39]. Geometry files were created using a python script, which generates non-overlapping randomly spaced rings with a thickness of 150 nm. Each geometry file contains 100 rings of the same radius. To generate background geometry files a python script was used, which creates 100 overlapping irregular icosahedrons per geometry with random side lengths of up to 500 nm. The geometry files were imported into FluoSim. To mimic the immunostaining process 10000 dyes were distributed randomly onto the geometries. Single-molecule localization reconstructions were then generated using parameters which resemble the experimental STORM imaging parameters (pixel size 10 nm, switch on rate 0.027 Hz, switch off rate 10 Hz, simulation time step 20 ms, localization precision 10 nm, 10000 frames).

Data preparation for deep learning.

The deep learning trainings have been carried out on 496 cellular SMLM reconstructed images of size 1024 by 1024 pixels rendered at a pixel size of 10 nm, where the number of septin rings per image was ranging from 0 to ~30 rings. Each training image has been manually annotated using each of the six different annotation strategies resulting in six different paired sets of raw and corresponding mask images. The manual annotation was performed using the Fiji plugin Labkit (https://imagej.net/plugins/labkit/). This way, our annotators manually drew masks around each potential septin ring resulting in mask images in which all pixels within an annotated ring were labeled with a distinct integer, while background pixels were labeled with zero. Based on the model as well as the number of septin-like rings in the image, a few to tens of seconds are spent on manual annotation of a single image. Moreover, the "majority voting" annotation - where the certainty level of the labels is high - is created from the annotations made by the three expert annotators. A given pixel in the "majority voting" label image is considered as a part of a mask only if it has been annotated by at least two individual annotators, otherwise it is considered background. The unseen test dataset comprising 32 cellular SMLM images was manually annotated by a septin biologist and the result was used as ground truth for further benchmarking of the models.

Deep learning model training and parameter settings.

For septin ring recognition, we trained each of the six models from scratch with the same cellular SMLM image data paired with the corresponding set of model specific mask images. To do so, we selected the 2D variant of the StarDist[38] machine learning network with the U-Net structure as its core. We trained the StarDist network on the ZeroCostDL4Mic platform [13] as it provided simple access to a range of popular deep learning networks including StarDist. The hyperparameters of the StarDist structure, originally designed for cell and nuclei segmentation problems, were tuned in our implementations for the recognition of rings in SMLM image data. To do so, we adjusted the ZeroCostDL4Mic StarDist notebook to enable a long-running grid search for hyperparameter optimization as well as offline data analysis. The grid search yielded different sets of hyperparameters when optimizing for recall and precision, of which we chose the one that ensured a reasonable balance between the two criteria in all the six trained models. Table 1 summarizes the values of the StarDist network's hyperparameters used in our experiments. In order to avoid adding another cause for bias,

19

1 no augmentation was used for training. Furthermore, transfer learning is not considered in

2 our work. The batch size was set to 8 and a mae function was chosen as the loss function. Of

3 the training dataset, 10% were used as a validation set to monitor possible overfitting during

4 training.

5 To guarantee a fair comparison of the six different annotation strategies, we trained each pair

6 of raw images and corresponding mask images on the same StarDist network with fixed

7 hyperparameters. In order to train the StarDist network, the key adopted python packages

8 included tensorflow (v 0.1.12), Keras (v 2.3.1), csbdeep (v 0.6.3), numpy (v 1.19.5), and cuda

9 (v 11.1.105). The training process has been accelerated using a Tesla P100 GPU and depending

10 on the model it took between 123 to 161 minutes.

11

12

Table 1. Hyperparameter configuration of the Stardist network used in this research

Parameter	Value
Number of epochs	200
Patch size	1024x1024
Batch size	8
Percentage validation	10
n rays	32
Grid parameter	2
Initial learning rate	0.0003
Augmentation	No
Transfer learning	No

13

14

15 **Ring recognition by means of trained models.**

16 For benchmarking, we ran the six trained models on the cellular test dataset and the synthetic

17 ring dataset using the ZeroCostDL4Mic platform. For the ring recognition in the Septin9$^{c/c}$ and

18 Septin9$^{-/-}$ cell data we exported the models from the ZeroCostDL4Mic platform and ran them

19 using the StarDist Fiji plugin. The outputs generated by the models were mask images of the

20 same size of the corresponding test image.

1 **Ring averaging and diameter measurement.**

2 To generate average ring images, the model output mask images were converted into binary

3 images using Fiji. The centers of the individual ring masks were calculated and fixed-size

4 bounding boxes were created around the center points. Bounding boxes touching the image

5 borders were excluded from further measurements. The bounding box ROIs were then

6 together with the corresponding images fed to CellProfiler[47] to extract per-ring crops from

7 the corresponding raw images. Using a custom written Fiji macro, we measured the average

8 ring diameter as follows: the ring crops per condition were averaged and the diameter was

9 determined by measuring horizontal line profiles with a line thickness of 1 pixel on 18

10 successive 10° rotations of the image (see Figure 1C). In each of these 18 intensity plots the

11 peak-to-peak distance was measured and then averaged to calculate the ring diameter.

12

13 **Calculation of septin ring geometry**

14 The septin ring circumference C was calculated based on the formula $C = \pi*d$ where d is the

15 measured ring diameter. The circumference was then divided by the approximate

16 periodicity length of heterooctameric or heterohexameric complexes as can be found in

17 filaments including the Septin-2-NC interface gap in the open conformation as predicted by

18 pdb structure data (pdb codes: 2QAG, 7M6J, 6UQQ, 5CYO). This was in line with the observed

19 periodicity length of the heterooctameric complex as found in filamentous septin

20 assemblies[40–44]. The septin complex association angle was calculated by dividing 360° by the

21 number of septin complexes found in the ring.

22

23 **Statistics**

24 We used the precision metric to show how relevant the recognized objects are in comparison

25 to the septin rings in the ground truth. It was calculated as *TP / (TP + FP)*, where *TP* stands for

26 the number of true positives and *FP* is the number of falsely recognized objects. Recall, on the

27 other hand, was used to show how well the recognized objects match the septin rings in the

28 ground truth. Thus, it was calculated as *TP / (TP + FN)*, where *FN* is the number of relevant

29 septin rings in the ground truth that were not recognized by the models. Both measures fall

30 in the [0%, 100%] interval, with higher values corresponding to better septin ring recognition

31 performance. R^2 values were calculated in R (R Project) using the "lm" function. *z*-score values

32 were calculated across the six models using the formula $z = (x-\mu)/\sigma$, where x is the variable

1 (recall, precision, ring diameter, no of rings detected), μ is the mean over the group of

2 variables of all six models and σ is the standard deviation from the mean.

3

Acknowledgements

12

References

14 1. Sochacki, K. A., Dickey, A. M., Strub, M.-P. & Taraska, J. W. Endocytic proteins are

15   partitioned at the edge of the clathrin lattice in mammalian cells. *Nat Cell Biol* (2017)

16   doi:10.1038/ncb3498.

17 2. Mund, M. *et al.* Systematic Nanoscale Analysis of Endocytosis Links Efficient Vesicle

18   Formation to Patterned Actin Nucleation. *Cell* **174**, 884-896.e17 (2018).

19 3. Szymborska, A. *et al.* Nuclear Pore Scaffold Structure Analyzed by Super-Resolution

20   Microscopy and Particle Averaging. *Science* (2013) doi:10.1126/science.1240672.

21 4. Gunzenhäuser, J., Olivier, N., Pengo, T. & Manley, S. Quantitative Super-Resolution

22   Imaging Reveals Protein Stoichiometry and Nanoscale Morphology of Assembling HIV-

23   Gag Virions. *Nano Lett* (2012) doi:10.1021/nl3021076.

24 5. Gambarotto, D. *et al.* Imaging cellular ultrastructures using expansion microscopy (U-

25   ExM). *Nat Methods* 1–11 (2018) doi:10.1038/s41592-018-0238-1.

6. Willig, K. I., OO, R. S., Westphal, V., Jahn, R. & Hell, S. W. STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 935–939 (2006).

7. Balzarotti, F. *et al.* Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* (2016) doi:10.1126/science.aak9913.

8. Sommer, C. & Gerlich, D. W. Machine learning in cell biology - teaching computers to recognize phenotypes. *J Cell Sci* **126**, 5529–5539 (2013).

9. Piccinini, F. *et al.* Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. *Cell Syst* **4**, 651-655.e5 (2017).

10. Jones, T. R. *et al.* Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci USA* **106**, 1826–1831 (2009).

11. Held, M. *et al.* CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Methods* **7**, 747–754 (2010).

12. Morone, D., Marazza, A., Bergmann, T. J. & Molinari, M. Deep learning approach for quantification of organelles and misfolded polypeptide delivery within degradative compartments. *MBoC* **31**, 1512–1524 (2020).

13. von Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat Commun* **12**, 2276 (2021).

14. Lucas, A. M. *et al.* Open-source deep-learning software for bioimage segmentation. *MBoC* **32**, 823–829 (2021).

15. Gómez-de-Mariscal, E. *et al.* DeepImageJ: A user-friendly environment to run deep learning models in ImageJ. *Nat Methods* **18**, 1192–1195 (2021).

1   16. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts

2   with deep learning. *Nat Neurosci* 1–12 (2018) doi:10.1038/s41593-018-0209-y.

3   17. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat*

4   *Methods* 1–10 (2019) doi:10.1038/s41592-018-0261-2.

5   18. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat Methods* 1–14 (2019)

6   doi:10.1038/s41592-019-0403-1.

7   19. Sommer, C., Hoefler, R., Samwer, M. & Gerlich, D. W. A deep learning and novelty

8   detection framework for rapid phenotyping in high-content screening. *Mol Biol Cell* **28**,

9   3428–3436 (2017).

10 20. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for

11   cellular segmentation. *Nat Methods* 1–21 (2020) doi:10.1038/s41592-020-01018-x.

12 21. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-

13   configuring method for deep learning-based biomedical image segmentation. *Nat*

14   *Methods* **18**, 203–211 (2021).

15 22. Zhang, J. *et al.* PIXER: an automated particle-selection method based on segmentation

16   using a deep neural network. *BMC Bioinformatics* **20**, 41 (2019).

17 23. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence

18   microscopy. *Nat Methods* **15**, 1090–1097 (2018).

19 24. Fang, L. *et al.* Deep learning-based point-scanning super-resolution imaging. *Nat*

20   *Methods* **18**, 406–416 (2021).

21 25. Speiser, A. *et al.* Deep learning enables fast and dense single-molecule localization with

22   high accuracy. *Nat Methods* **18**, 1082–1090 (2021).

23 26. Nehme, E. *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by

24   deep learning. *Nat Methods* 1–10 (2020) doi:10.1038/s41592-020-0853-5.

27. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat Biotechnol* **36**, 460–468 (2018).

28. Heydarian, H. *et al.* 3D particle averaging and detection of macromolecular symmetry in localization microscopy. *Nat Commun* **12**, 2847 (2021).

29. Sieben, C., Banterle, N., Douglass, K. M., Gönczy, P. & Manley, S. Multicolor single-particle reconstruction of protein complexes. *Nat Methods* 1–9 (2018) doi:10.1038/s41592-018-0140-x.

30. Mennella, V. *et al.* Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. *Nat Cell Biol* **14**, 1159–1168 (2012).

31. Mostowy, S. & Cossart, P. Septins: the fourth component of the cytoskeleton. *Nat Rev Mol Cell Biol* **13**, 183–194 (2012).

32. Weirich, C. S., Erzberger, J. P. & Barral, Y. The septin family of GTPases: architecture and dynamics. *Nat Rev Mol Cell Biol* **9**, 478–489 (2008).

33. Kinoshita, M., Field, C. M., Coughlin, M. L., Straight, A. F. & Mitchison, T. J. Self- and actin-templated assembly of Mammalian septins. *Dev Cell* **3**, 791–802 (2002).

34. Sellin, M. E., Holmfeldt, P., Stenmark, S. & Gullberg, M. Microtubules support a disc-like septin arrangement at the plasma membrane of mammalian cells. *Mol Biol Cell* (2011) doi:10.1091/mbc.E11-09-0754.

35. Hu, J. *et al.* Septin-driven coordination of actin and microtubule remodeling regulates the collateral branching of axons. *Curr Biol* **22**, 1109–1115 (2012).

36. Kaplan, C., Steinmann, M., Zapiorkowska, N. A. & Ewers, H. Functional Redundancy of Septin Homologs in Dendritic Branching. *Front Cell Dev Biol* **5**, 11 (2017).

1   37. Kim, M. S., Froese, C. D., Estey, M. P. & Trimble, W. S. SEPT9 occupies the terminal

2   positions in septin octamers and mediates polymerization-dependent functions in

3   abscission. *J Cell Biol* **195**, 815–826 (2011).

4   38. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-convex

5   Polygons. *arXiv:1806.03535 [cs]* **11071**, 265–273 (2018).

6   39. Lagardère, M., Chamma, I., Bouilhol, E., Nikolski, M. & Thoumine, O. FluoSim: simulator

7   of single molecule dynamics for fluorescence live-cell and super-resolution imaging of

8   membrane proteins. *Sci Rep* 1–14 (2020) doi:10.1038/s41598-020-75814-y.

9   40. Bertin, A. *et al.* Saccharomyces cerevisiae septins: supramolecular organization of

10   heterooligomers and the mechanism of filament assembly. *Proc Natl Acad Sci USA* **105**,

11   8274–8279 (2008).

12   41. DeMay, B. S. *et al.* Septin filaments exhibit a dynamic, paired organization that is

13   conserved from yeast to mammals. *J Cell Biol* **193**, 1065–1081 (2011).

14   42. Frazier, J. A. *et al.* Polymerization of purified yeast septins: evidence that organized

15   filament arrays may not be required for septin function. *J Cell Biol* **143**, 737–749 (1998).

16   43. Kaplan, C. *et al.* The absolute arrangement of subunits in cytoskeletal septin filaments in

17   cells measured by fluorescence microscopy. *Nano Lett* (2015)

18   doi:10.1021/acs.nanolett.5b00693.

19   44. Bertin, A. *et al.* Phosphatidylinositol-4,5-bisphosphate promotes budding yeast septin

20   filament assembly and organization. *Journal of Molecular Biology* **404**, 711–731 (2010).

21   45. Sellin, M. E., Stenmark, S. & Gullberg, M. Mammalian SEPT9 isoforms direct

22   microtubule-dependent arrangements of septin core heteromers. *Mol Biol Cell* **23**,

23   4242–4255 (2012).

1   46. Füchtbauer, A. *et al.* Septin9 is involved in septin filament formation and cellular

2   stability. *Biol Chem* **392**, 769–777 (2011).

3   47. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and

4   quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).

5   48. Segebarth, D. *et al.* On the objectivity, reliability, and validity of deep learning enabled

6   bioimage analyses. *eLife* **9**, e59780 (2020).

7   49. Garcia, G. *et al.* Subunit-dependent modulation of septin assembly: Budding yeast septin

8   Shs1 promotes ring and gauze formation. *J Cell Biol* (2011) doi:10.1083/jcb.201107123.

9   50. Kinoshita, M. Assembly of mammalian septins. *J. Biochem.* **134**, 491–496 (2003).

10   51. Song, K. *et al.* A SEPT1-based scaffold is required for Golgi integrity and function. *J Cell*

11   *Sci* **132**, jcs225557 (2019).

12   52. Ihara, M. *et al.* Cortical organization by the septin cytoskeleton is essential for structural

13   and mechanical integrity of mammalian spermatozoa. *Dev Cell* **8**, 343–352 (2005).

14   53. Asada, A. *et al.* Neuronal expression of two isoforms of mouse Septin 5. *J Neurosci Res*

15   **88**, 1309–1316 (2010).

16   54. Banko, M., Mucha Kruczynska, I., Weise, C., Heyd, F. & Ewers, H. A homozygous

17   genome-edited Sept2-EGFP fibroblast cell line. *Cytoskeleton (Hoboken)* **76**, 73–82

18   (2019).

19