DNA Origami offers a versatile method for quantifying protein copy-number in super-resolution

Francesca Cella Zanacchi[1, *], Carlo Manzo[1, 4, $], Angel Sandoval Alvarez[1, $], Nathan D. Derr[2], Maria Garcia Parajo[1, 3], Melike Lakadamyali[1, *, &]

[1] ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain

[2]Smith College, 44 College Lane, Northampton, MA 01063, USA

[3] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

[4]Universitat de Vic – Universitat Central de Catalunya (UVic-UCC), C. de la Laura,13, 08500 Vic, Spain

[&] Present address: Perelman School of Medicine, Department of Physiology, University of Pennsylvania, 764 Clinical Research Building, 415 Curie Boulevard, Philadelphia, PA, 19104

[$] These authors contributed equally to this work

[*] Correspondence should be sent to M.L.: melike.lakadamyali@icfo.es and F.C.Z: francesca.cella.zanacchi@gmail.com

Abstract

Single-molecule-based super-resolution microscopy offers a unique opportunity for quantifying protein copy-number with nanoscale resolution. However, while fluorescent proteins have been extensively characterized for quantitative imaging using calibration standards, similar calibration tools for small organic fluorophores used in conjunction with immunofluorescence are lacking. Here, we show that DNA origami in combination with GFP antibodies is a versatile platform for calibrating fluorophore and antibody labeling efficiency to quantify protein copy-number in cellular contexts through immunofluorescence based super-resolution microscopy.

Main Text

Single molecule localization microscopy has become an important tool for imaging intracellular structures and protein complexes with nanoscale spatial resolution[1]. Recently, an immense effort has been dedicated to the quantification of super-resolution images[2,3]. Among the different quantitative parameters that can be extracted, protein copy-number and stoichiometry have been of particular interest. Single-molecule-based super-resolution methods are uniquely positioned to determine protein copy-numbers, since the single molecule information can be exploited for counting. However, the exact quantification is ultimately impaired by the stochasticity of the labeling method and the complex photophysics of the fluorescent probes. Therefore, it is not surprising that a substantial effort has been dedicated to developing analytical approaches and calibration standards aimed to overcome this challenge[4-12]. For example, the photophysics of photoactivatable and photoconvertible

fluorescent proteins (FPs) have been extensively studied and nano-templates have been developed to calibrate the signal and count FP-tagged proteins[4,5,7,10,11]. Since FPs provide a one to one labeling stoichiometry and have limited blinking or reactivation probability, they are desirable for quantitative imaging. However, a major limitation is imposed by their low photon budget, leading to images with a lower spatial resolution compared to small organic fluorophores, which are the probe of choice for a large number of super-resolution studies. Targeting these bright fluorophores to the protein of interest typically requires immunofluorescent labeling by primary and secondary antibodies. In this case, unfortunately, both the antibody labelling efficiency as well as the number of fluorophores conjugated to the primary or to the secondary antibody are highly stochastic. In addition, fluorophores might undergo repeated blinking or reactivation events. Combined together, these issues pose major challenges for protein-copy quantification. Partial solutions to these challenges have been reported. For example, the fluorophore photophysics can be modelled[8,12] or characterized using single fluorophores conjugated to antibodies or images of sparse spots on the sample[13-17]. In the case of DNA-PAINT approaches - that rely on "on-off" binding of fluorophore-labeled small oligos - the binding kinetics can be modeled and accounted for in the quantification[9]. Nonetheless, in all cases the unknown stoichiometry of antibody-based labeling, resulting from the stochasticity of fluorophore-antibody and antibody-target binding, largely affects the precision of the final quantification. Therefore, there is an urgent need for versatile calibration standards that take into account not only the fluorophore photophysics but also the antibody and fluorophore labeling stoichiometry. Although in other works, ad hoc calibration standards have allowed quantifying complex structures such as nucleosomes[13] there is lack of a general approach toward this problem.

To overcome these challenges and thus develop versatile calibration standards that can be used for quantifying protein copy-number in intracellular contexts, we took advantage of DNA origami. Specifically, we used a previously developed 3D DNA origami chassis comprised of 12 parallel DNA double helices. This chassis serves as a skeleton for attaching additional components via the use of "handle" sequences that project outward from the structure[18]. These handles provide site- and sequence-specific attachment points for single fluorophores as well as proteins of interest and allow testing of several different labeling strategies such as antibody, nanobody and Halo/SNAP tag labeling (**Figure 1a**). We first used this structure to attach complimentary anti-handle sequences labeled with a single AlexaFluor647 to the three handles located at positions 1, 7 and 13 of helix 0 and thus establish a baseline for the efficiency of attaching the anti-handle oligos to the complimentary handle oligos in the DNA origami. This handle/anti-handle labelling efficiency should be independent of the fluorophore used and only depend on the sequence of the oligos. A single TAMRA fluorophore attached at position 14 of each of the other outer helices (h3, h4, h7, h8, h11, **Figure 1a**) was used to identify the DNA origami structures on the glass slide (**Figure 1b**). To determine the number of fluorophores successfully conjugated to the handles, we performed single-step photobleaching experiments and analyzed the intensity-time traces from AlexaFluor647 spots, only taking into account those that showed co-localization with a

TAMRA spot. The traces revealed single, double and triple steps as expected (**Figure 1c**) and the distribution of the number of counted steps fit to a binomial giving a handle/anti-handle labeling probability of 48% (**Supplementary Figure 1a**). Next, we performed STORM imaging and once again analyzed the spots that showed co-localization between the AlexaFluor647 and TAMRA labels (**Figure 1d left panel**). Super-resolution imaging revealed single, double or triple clusters (**Figure 1d inset**), in agreement with the single-step photobleaching experiments. We segmented these clusters using a previously developed cluster identification algorithm[13] (**Figure 1d right panel**) and found that the nearest neighboring clusters were separated by an average of 85±7 nm (Mean±SD) whereas the furthest two clusters were separated by a distance of 157±17 nm (**Supplementary Figure 1b**), matching the expected distance between the individual handles used for the labeling. The number of localizations detected from individual clusters showed a broad distribution (**Figure 1e)** with the median value of the number of localizations obtained for 1, 2 and 3 fluorophores showing a roughly linear increase with the number of fluorophores as expected (**Figure 1f and Supplementary Table 1**).

We next purified a modified dimeric *Saccharomyces cerevisiae* dynein motor[19] (**Methods and Supplementary Figure 2**) whose individual protomers contained both the SNAP-tag and GFP in the motor's tail cargo-binding domain[18]. Because dynein is a homodimer containing two such protomers, each motor thus has two copies of GFP for imaging. We used the SNAP-tag to covalently link the dynein to anti-handle DNA sequences complimentary to the handles at the same 3 positions of the chassis (**Methods and Figure 1a**). We then immunostained the GFP using a combination of primary anti-GFP antibodies and AlexaFluor405-AlexaFluor647 labeled secondary antibodies (**Methods**) and then performed super-resolution microscopy (**Figure 1g left**) and clustering analysis (**Figure 1g right**). Once again, only those clusters that co-localized with the TAMRA signal were selected and additionally filtered to retain only the clusters that were separated by the expected distance between the handles (see **Methods**). The counted number of single, double and triple clusters fit well to a binomial distribution with a labeling efficiency of 38% (**Supplementary Figure 1c**). This efficiency was only slightly lower than the labeling efficiency for attaching single fluorophores (48%, **Supplementary Figure 1a**) suggesting that the main limitation in labeling is due to the attachment of the dynein to the DNA origami, rather than the antibody labeling efficiency. DNA origami immobilized on a Biotin-Streptavidin functionalized glass substrate and on top of cells displayed similar distributions of the number of localizations per cluster (**Supplementary Figure 3**). A calibration curve corresponding to the median number of localizations obtained for 1, 2 and 3 clusters (corresponding to 2, 4 and 6 GFPs) was built (**Figure 1h and Supplementary Table 1**). The calibration curve obtained in this regime (up to 6 copies of GFP) was roughly linear, suggesting that the antibody labeling efficiency was high and the binding of primary antibody to its target GFP did not reach saturation levels. This calibration curve can be used to extract average protein copy-numbers in a given image by comparing the median number of localizations obtained in the cellular context to the curve. The main advantage of using the DNA origami calibration as opposed to single fluorophores or sparse

spots on the sample is the fact that it simultaneously accounts for the stochasticity of the antibody labeling by the fluorophores as well as the binding of the primary and secondary antibodies. Given that most primary and secondary antibodies are polyclonal, this method provides a more precise calibration that properly accounts for the labeling stoichiometry.

To determine whether this method could be used not only to extract average protein copy-numbers but also the percentage of each oligomeric state, we further explored whether we could fit the distribution of the number of localizations per cluster to a functional form. Indeed, the distribution of localizations for single, double or triple clusters (corresponding to 2, 4 and 6 copies of GFP, respectively) could be simultaneously fit using only 2 free parameters (μ and σ) assuming that they correspond to the convolutions of respectively 2, 4 and 6 functions f_1, where f_1 is a log-normal distribution describing the probability distribution of number of localizations obtained by labeling monomeric GFP with A647-conjugated antibodies (see **Methods, Figure 1i**)[20,21]. Therefore, for a general distribution of localizations containing an unknown mixture of oligomeric states, it should be possible to extract both the oligomeric state and the percentage of oligomers corresponding to that particular state by fitting the data to a linear combination of calibration distributions f_n obtained as recursively convoluting f_1 n-times as has been shown before for single fluorophores[16,17,20,21]. We validated this idea in multiple ways. First, we generated a synthetic distribution of localizations by combining a known fraction of single, double and triple clusters from the DNA-origami images, which we then fit to a linear combination of log-normal functions f_n (**Figure 2a**). The fitting was performed at varying the number of functions f_n. The optimal number of functions (N_{max}) was automatically chosen as the one minimizing the fit objective function (**Methods)**. The extracted fraction of single, double and triple motors (2, 4 and 6 GFPs) was in excellent agreement with the expected fractions given sufficient statistics (around 180 datapoints) (**Figure 2b, Supplementary Figure 4a**) and the fit provided the best result for N_{max}=3 functions (**Figure 2c**). Second, to test the range of stoichiometries that can be probed with this method, we combined DNA-origami images to generate synthetic distributions of localizations comprising an equal fraction of 1, 4, 8 and 16 motors (2, 8, 16 and 32 GFPs) (**Figure 2d**). The peaks in the stoichiometry distribution obtained from the fit were in agreement with the chosen oligomeric states (**Figure 2d**) providing a good correlation with the theoretical data (**Supplementary Figure 4b**) and the objective function was minimized at N_{max}=20 functions (**Figure 2f**). Fitting the data to a convolution of more than 20 functions did not change the results (**Supplementary Figure 4c**). The amount of statistics needed for accurate fitting increased with increasing stoichiometry (**Supplementary Figure 4a and b**). Given sufficient statistics, the method performed well at a large range of stoichiometries, albeit with decreasing performance especially at stoichiometries larger than 16 ($0.99<R<0.87$ for stoichiometries $3<N_{stoi}<16$ and $0.76<R<0.5$ for stoichiometries $16<N_{stoi}<32$, **Supplementary Figure 4d**). Finally, we attached dynein to 5 handles on the chassis (thus giving rise to up to 5 dimers and 10 copies of GFP). In this case, due to the short distance between the handle positions (28 nm), we could no longer distinguish clusters corresponding to individual dynein motors (**Figure 2g**). We thus combined together all the localizations coming

from each DNA origami structure (identified by the presence of TAMRA signal) and plotted the distribution of localizations (**Figure 2h**). The fit to a linear combination of calibration functions f_n corresponding to one to five GFP dimers (n=2,4,…,10) revealed a combination of 37% single, 44% two, 14% three, 4% four, 1% five dynein motors, fitting well to a binomial distribution for a labeling efficiency of 33%, and in close agreement with the 38% labeling efficiency obtained for triple handles (**Figure 2i**). The objective function of the fit was minimized at 5 GFP dimers and did not improve by fitting to a larger number of GFPs (**Figure 2l)**.

We finally applied this calibration method to determine copy-numbers of protein complexes imaged in cells. As a first test of a biological structure, we performed immunofluorescence of the nuclear pore complex (NPC) subunit Nup133 fused to GFP, expressed in U2OS cells in the presence of siRNA to knock down the endogenous copy of Nup133 (**Figure 3)**. We chose Nup133 as its stoichiometry has been previously characterized using various methods[22] (**Figure 3a**). Super-resolution images showed ring-like structures as expected (**Figure 3b, c**), albeit with lower than the 8-fold symmetry per NPC. This is likely due to the incomplete siRNA knockdown and potential limitations with antibody access when the protein copy number is much higher than the linear regime we demonstrated for the DNA-origami. To test a scenario in which the expected range of stoichiometries is more narrow and controlled, we manually sorted the NUP133 images taking into account the number of Nup133 clusters that were visible by eye. We could reliably sort up to 5 Nup133 clusters, since the images of individual clusters started significantly merging together within the resolution limit of STORM for higher order structures. We note that the manual sorting is still prone to some errors as multiple clusters in close proximity may be counted as a single cluster. We then extracted the distribution for the number of localizations per NPC ring containing 1-5 Nup133 clusters (**Figure 3d-h**), which showed the expected range of stoichiometries considering 4 copies of GFP per Nup133 cluster (**Figure 3d-h and Supplementary Figure 5 a-e**). Indeed, for single clusters the stoichiometry ranged from 1 to 4, for double clusters from 2-8 and so on. In addition, we performed cluster analysis on the NPC images without manual sorting to combine clusters belonging to the same NPC ring into one cluster (**Figure 3c**) and obtained the corresponding distribution of localizations per NPC ring. After the fitting, we obtained a broad distribution of stoichiometries with a maximum stoichiometry of around 30 and a mean stoichiometry of 12 (**Figure 3i, black bars and Supplementary Figure 5f**). The maximum stoichiometry is consistent with the expected stoichiometry of 32. Given that the majority of the NPCs in the super-resolution images contained less than the 8 Nup133 clusters (**Figure 3i, inset**), we expected to obtain an average stoichiometry lower than 32. In the super-resolution images, the most predominantly observed NPC rings contained around 3 NUP133 clusters (**Figure 3i, inset**), which is consistent with the mean stoichiometry of 12 retrieved from the fit. In addition, weighing the stoichiometries obtained from the sorted data (1-5 rings) (**Figure 3d-h**) with their occurrence in the super-resolution images (**Figure 3i, inset**) gave a distribution that matched remarkably well to the experimentally obtained distribution (**Figure 3i, red line**). Varying the N_{max} around the value for which the objective function was

minimized did not change the results (**Supplementary Figure 6**), while choosing a much smaller N_{max} gave rise to isolated peaks at the tail of the stoichiometry distribution (**Supplementary Figure 6**). For consistency, we recommend choosing the N_{max} for which the objective function is minimum, however the shape of the stoichiometry distribution obtained after the fit can also help guide the choice for N_{max} as its tail should show a smooth decay. Finally, similar results were obtained for another subunit of the NPC, NUP107, belonging to the same sub-complex as NUP133 (**Supplementary Figure 7**).

In conclusion, we developed a versatile calibration standard that can be used to quantify protein copy number from super-resolution images obtained after immunofluorescence labeling. Interestingly, the calibration curve obtained for GFP antibody labeling was mostly linear for a range of up to 10 GFPs suggesting that the antibody labeling is efficient and not affected by crowding and steric hindrance. The use of GFP antibodies provides a particularly versatile strategy for quantifying a large number of proteins of interest using the calibration curve reported here. In order to do so, it is important to point out that same imaging and image analysis conditions should be used as detailed in the **Methods and Supplementary Protocol**. We used standard imaging buffers, laser powers and acquisition settings that are typical for STORM experiments. Finally, the method we developed is not limited to GFP antibodies and is also applicable to the use of antibodies against any endogenous protein of interest. In addition, it can be used to calibrate nanobody labeling, Halo or SNAP-tag fusions and photoactivatable and photoswitchable fluorescent proteins.

Acknowledgements: We thank Pablo Gomez for helpful discussions. The siRNA resistant NUP133-GFP and NUP107-GFP plasmids were a kind gift from Jan Ellenberg, EMBL, Heidelberg. M.L. acknowledges funding from the Fundació Cellex Barcelona, European Union Seventh Framework Programme under the European Research Council grants 337191-MOTORS and Spanish Ministry of Economy and Competitiveness and the Fondo Europeo de Desarrollo Regional (FEDER) grant FIS2015-63550-R (MINECO/FEDER). F.C.Z. acknowledges funding from the "Severo Ochoa" Programme for Centres of Excellence in R&D (SEV-2015-0522). CM acknowledges funding from the Spanish Ministry of Economy and Competitiveness and the European Social Fund (ESF) through the Ramón y Cajal program 2015 (RYC-2015-17896).

References

1 Oddone, A., Vilanova, I. V., Tam, J. & Lakadamyali, M. Super-resolution imaging with stochastic single-molecule localization: concepts, technical developments, and biological applications. *Microscopy research and technique* **77**, 502-509, doi:10.1002/jemt.22346 (2014).

2 Durisic, N., Cuervo, L. L. & Lakadamyali, M. Quantitative super-resolution microscopy: pitfalls and strategies for image analysis. *Current opinion in chemical biology* **20**, 22-28, doi:10.1016/j.cbpa.2014.04.005 (2014).

3 Deschout, H., Shivanandan, A., Annibale, P., Scarselli, M. & Radenovic, A. Progress in quantitative single-molecule localization microscopy. *Histochemistry and cell biology* **142**, 5-17, doi:10.1007/s00418-014-1217-y (2014).

4 Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Identification of clustering artifacts in photoactivated localization microscopy. *Nat Methods* **8**, 527-528, doi:10.1038/nmeth.1627 (2011).
5 Durisic, N., Laparra-Cuervo, L., Sandoval-Alvarez, A., Borbely, J. S. & Lakadamyali, M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nat Methods* **11**, 156-162, doi:10.1038/nmeth.2784 (2014).
6 Finan, K., Raulf, A. & Heilemann, M. A set of homo-oligomeric standards allows accurate protein counting. *Angew Chem Int Ed Engl* **54**, 12049-12052, doi:10.1002/anie.201505664 (2015).
7 Fricke, F., Beaudouin, J., Eils, R. & Heilemann, M. One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localization microscopy. *Sci Rep* **5**, 14072, doi:10.1038/srep14072 (2015).
8 Hummer, G., Fricke, F. & Heilemann, M. Model-independent counting of molecules in single-molecule localization microscopy. *Mol Biol Cell* **27**, 3637-3644, doi:10.1091/mbc.E16-07-0525 (2016).
9 Jungmann, R. *et al.* Quantitative super-resolution imaging with qPAINT. *Nat Methods* **13**, 439-442, doi:10.1038/nmeth.3804 (2016).
10 Lee, S. H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc Natl Acad Sci U S A* **109**, 17436-17441, doi:10.1073/pnas.1215175109 (2012).
11 Puchner, E. M., Walter, J. M., Kasper, R., Huang, B. & Lim, W. A. Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *Proc Natl Acad Sci U S A* **110**, 16015-16020, doi:10.1073/pnas.1309676110 (2013).
12 Rollins, G. C., Shin, J. Y., Bustamante, C. & Presse, S. Stochastic approach to the molecular counting problem in superresolution microscopy. *Proc Natl Acad Sci U S A* **112**, E110-118, doi:10.1073/pnas.1408071112 (2015).
13 Ricci, M. A., Manzo, C., Garcia-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145-1158, doi:10.1016/j.cell.2015.01.054 (2015).
14 Nieuwenhuizen, R. P. *et al.* Measuring image resolution in optical nanoscopy. *Nature methods* **10**, 557-562, doi:10.1038/nmeth.2448 (2013).
15 Ehmann, N. *et al.* Quantitative super-resolution imaging of Bruchpilot distinguishes active zone states. *Nature communications* **5**, 4650, doi:10.1038/ncomms5650 (2014).
16 Bakker, G. J. *et al.* Lateral mobility of individual integrin nanoclusters orchestrates the onset for leukocyte adhesion. *Proc Natl Acad Sci U S A* **109**, 4869-4874, doi:10.1073/pnas.1116425109 (2012).
17 Torreno-Pina, J. A. *et al.* Enhanced receptor-clathrin interactions induced by N-glycan-mediated membrane micropatterning. *Proc Natl Acad Sci U S A* **111**, 11037-11042, doi:10.1073/pnas.1402041111 (2014).
18 Derr, N. D. *et al.* Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science* **338**, 662-665, doi:10.1126/science.1226734 (2012).
19 Reck-Peterson, S. L. *et al.* Single-molecule analysis of dynein processivity and stepping behavior. *Cell* **126**, 335-348, doi:10.1016/j.cell.2006.05.046 (2006).
20 Schmidt, T., Schutz, G. J., Gruber, H. J. & Schindler, H. Local Stoichiometries Determined by Counting Individual Molecules. *Anal. Chem.* **68**, 4397-4401 (1996).
21 Moertelmaier, M., Brameshuber, M., Linimeier, M., Schutz, G. J. & Stockinger, H. Thinning out clusters while conserving stoichiometry of labeling. *Applied Physics Letters* **87** (2005).
22 von Appen, A. *et al.* In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140-143, doi:10.1038/nature15381 (2015).

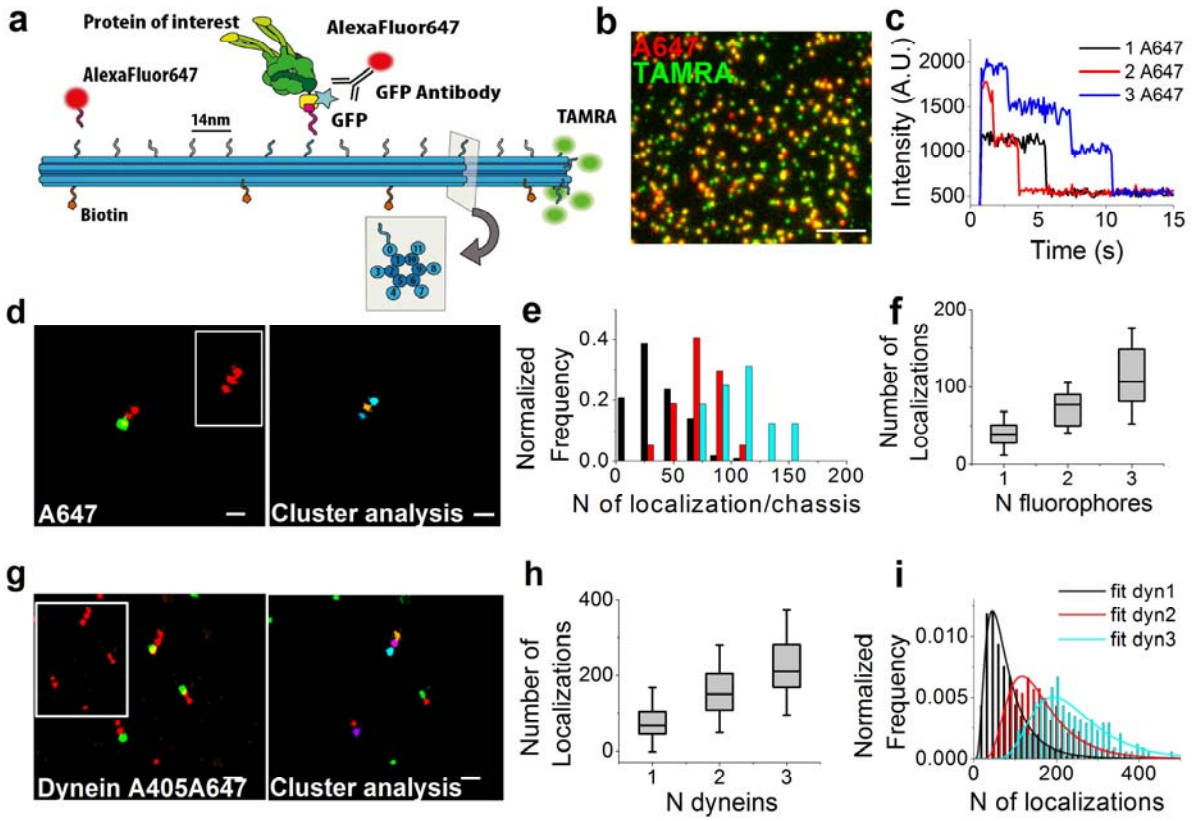

Figure 1. ***DNA origami calibration:*** *(**a**) Schematic representation of the 12 helix DNA origami structure demonstrating different labelling strategies. (**b**) Widefield image showing DNA origami structures functionalized with TAMRA (green) as a reference and with AlexaFluor 647 (red) attached at handle positions 1, 3 and 7 of the helix 0. (**c**) Intensity-time traces corresponding to stepwise photobleaching experiments for structures containing single, double and triple fluorophore occupancy. (**d**) Left: Dual-color STORM image showing DNA origami functionalized with AlexaFluor 647 (red) and TAMRA (green), inset shows the STORM image of AlexaFluor 647 alone, without the TAMRA signal for ease of visualization; (Right) Clustering analysis of the AlexaFluor 647 STORM image corresponding to the inset. (**e**) Distribution for the number of localizations detected for 1 (black), 2 (red) and 3 (cyan) fluorophores (**f**) Number of localizations for 1, 2 and 3 fluorophores (n=3 experiments). The box shows 25/75th percentile, the line is the median value and the whiskers are the standard deviation (see Supplementary Table 1). (**g**) Left: Dual-color STORM image showing DNA origami functionalized with TAMRA (green) and Dynein-GFP (GFP immunostained with Alexa Fluor 405/Alexa Fluor 647, red), inset shows the STORM image of AlexaFluor 647 alone; (Right) Clustering analysis of the AlexaFluor 647 STORM image corresponding to the inset. (**h**) Calibration curve showing the number of localizations for 1, 2 and 3 motors (n=3 experiments). The box shows 25/75th percentile, the line is the median value and the whiskers are the standard deviation (see Supplementary Table 1). (**i**) The localization distribution fits remarkably well to a convolution of 2 (black), 4 (red) and 6 (blue) log-normal distributions f1 corresponding to the distribution of a single GFP. Scale bar 200nm (**d**, **g**), Scale bar 5µm (**b**).*

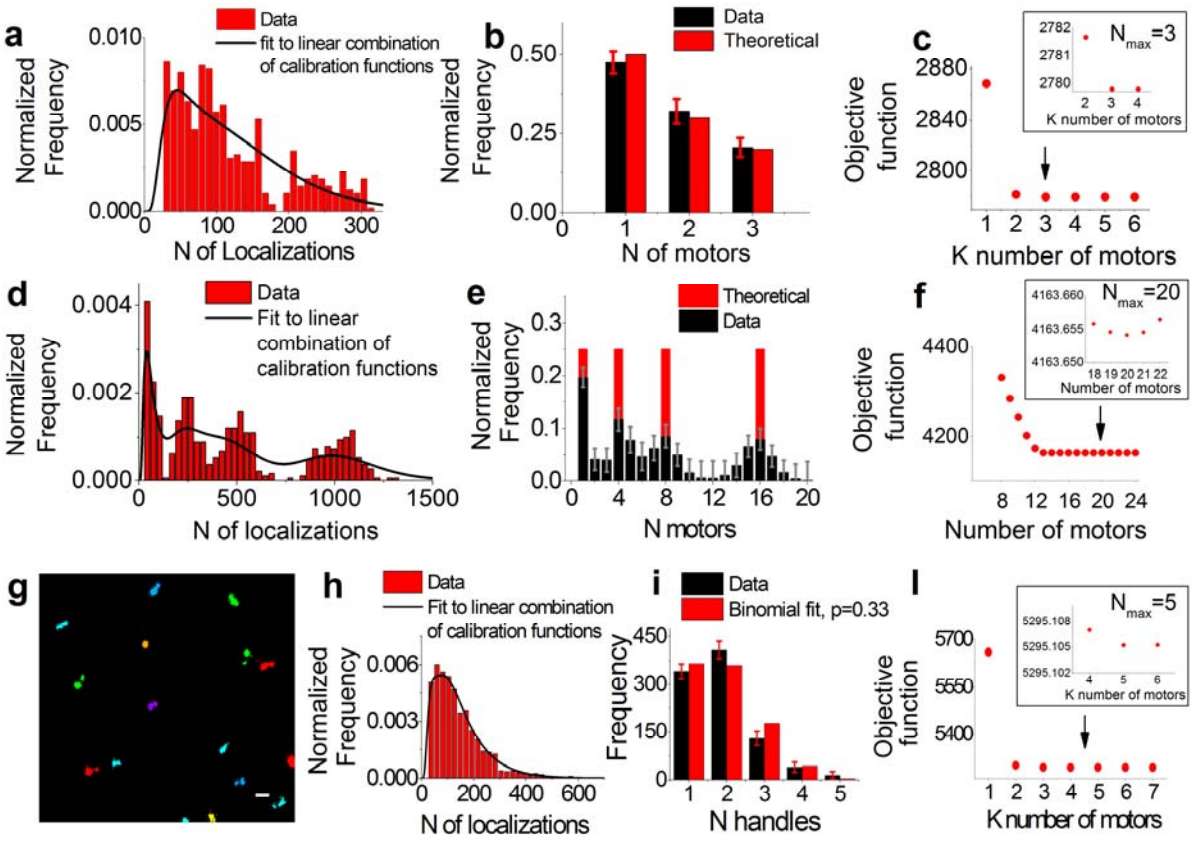

Figure 2 Validation of stoichiometry determination *(**a**) Estimation of the stoichiometry for a synthetic sample with known percentage of single, double and triple motors generated from the DNA-origami images, fit to a linear combination of lognormal distributions (**b**) and the evolution of the objective function F for a number of stoichiometries with a minimum corresponding to a stoichiometry of 3 motors (N_{max}=3) (**c**) (**d**) Estimation of the stoichiometry for a synthetic sample with equal percentage (25%) of 1, 4, 8 and 16 motors (2, 8, 16 and 32 GFPs) generated from the DNA-origami images, fit to a linear combination of lognormal distributions (**e**) and the corresponding objective function F for a number of stoichiometries with a minimum corresponding to a stoichiometry of 20 motors (N_{max}=20) (**f**). (**g**) Clustering analysis of STORM images for DNA chassis functionalized with 5 motors. Scale bar 200nm. (**h**) Distribution showing the total number of localizations per 5 dynein motors (red) and the corresponding fit to a linear combination of log normal functions considering up to 5 dimers (black line) (**i**) The percentage of 1,2,3,4 and 5 motors obtained from the fit in (black) (37% single, 44% two, 14% three, 4% four, 1% five dynein motors) matches to a binomial distribution with a labeling efficiency of p=0.33 (red) (Reduced ChiSquared=0.0009) and (**l**) the corresponding objective function F for a number of stoichiometries with a minimum corresponding to a stoichiometry of 5 motors (N_{max}=5). Errors bars in (b), (e) and (i) refer to the lower bound to the standard errors based on the Fisher Information Matrix.*

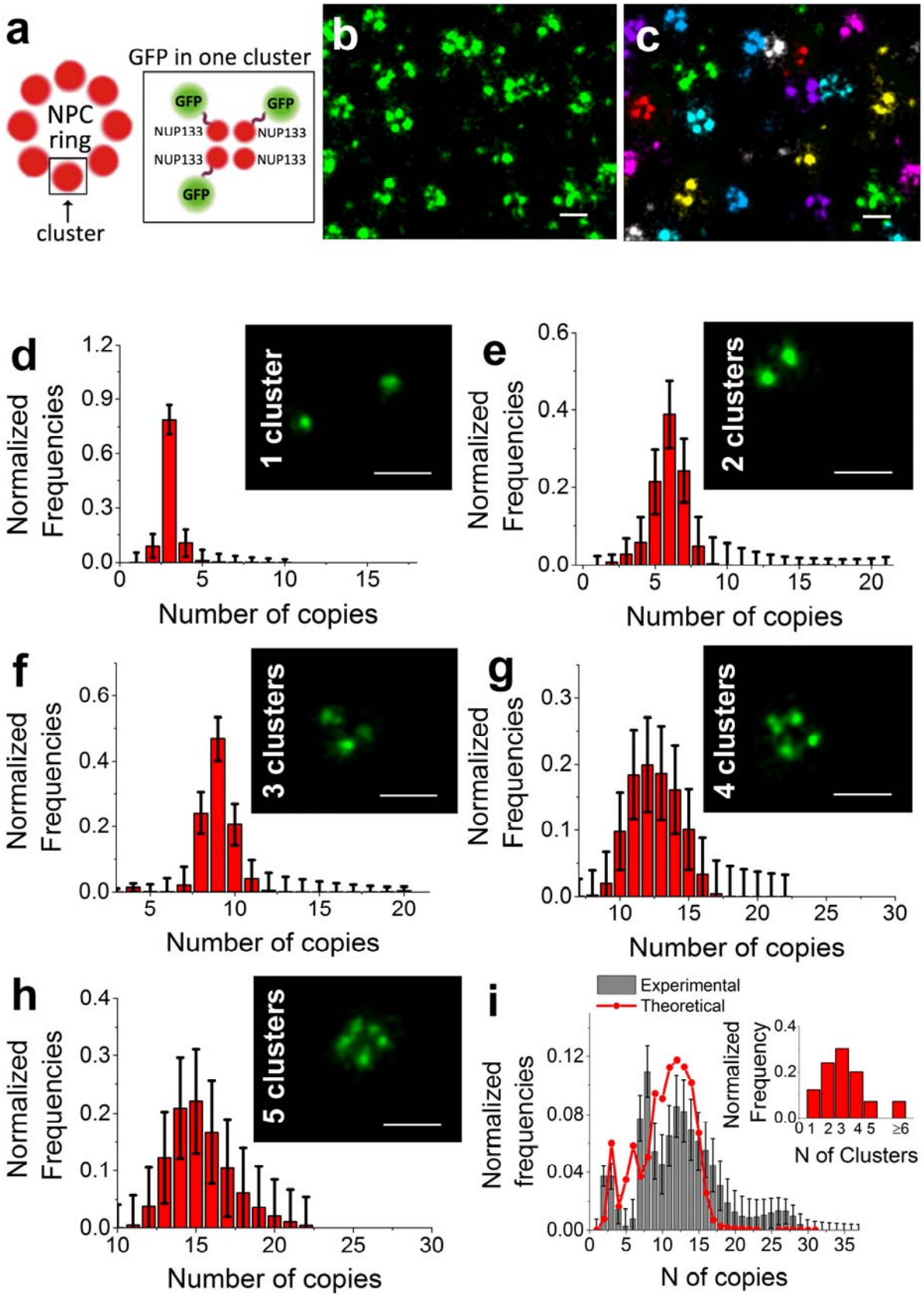

Figure13. Quantification of NUP133 complexes in U2OS cells: (**a)** Schematic representation of the NPC and Nup133 subunit composition reflecting the terminology used in the manuscript **(b)** STORM image showing NUP133 in siRNA resistant NUP133-GFP expressing U20S cell in which the endogenous

copy of NUP133 was knocked down by siRNA. (**c**) Corresponding clustering analysis of the STORM image. GFP copy-number distribution for NUP133 extracted from the fit of manually sorted data (**d-h**) corresponding to 1,2,3,4,5 clusters, respectively (**d-h insets**). GFP copy-number distributions estimated in the whole cell for NUP133 (**i, black bars**) by fitting the distribution of the number of localizations per NPC to a linear combination of calibration functions considering contributions up to 32 monomers. Distribution of stoichiometries obtained by weighing the sorted data (**i, red line**) with their occurrence in the super-resolution images (**i, inset**). Scale bars: 200 nm (**b**-**g)**. Errors bars in (**d-i**) refer to the lower bound to the standard errors based on the Fisher Information Matrix.