

# UCSF

## UC San Francisco Previously Published Works

### Title

Improved split fluorescent proteins for endogenous protein labeling

### Permalink

<https://escholarship.org/uc/item/7dw83335>

### Journal

Nature Communications, 8(1)

### ISSN

2041-1723

### Authors

Feng, Siyu

Sekine, Sayaka

Pessino, Veronica

et al.

### Publication Date

2017

### DOI

10.1038/s41467-017-00494-8

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>


Peer reviewed

ARTICLE

DOI: 10.1038/s41467-017-00494-8

OPEN

# Improved split fluorescent proteins for endogenous protein labeling

Siyu Feng<sup>1</sup>, Sayaka Sekine<sup>2</sup>, Veronica Pessino<sup>3</sup>, Han Li<sup>4</sup>, Manuel D. Leonetti<sup>4</sup> & Bo Huang<sup>1,2,5,6</sup> 

Self-complementing split fluorescent proteins (FPs) have been widely used for protein labeling, visualization of subcellular protein localization, and detection of cell–cell contact. To expand this toolset, we have developed a screening strategy for the direct engineering of self-complementing split FPs. Via this strategy, we have generated a yellow–green split-mNeonGreen<sub>2<sub>1–10</sub>/11</sub> that improves the ratio of complemented signal to the background of FP<sub>1–10</sub>-expressing cells compared to the commonly used split GFP<sub>1–10/11</sub>; as well as a 10-fold brighter red-colored split-sfCherry<sub>2<sub>1–10</sub>/11</sub>. Based on split sfCherry<sub>2</sub>, we have engineered a photoactivatable variant that enables single-molecule localization-based super-resolution microscopy. We have demonstrated dual-color endogenous protein tagging with sfCherry<sub>2<sub>11</sub></sub> and GFP<sub>11</sub>, revealing that endoplasmic reticulum translocon complex Sec61B has reduced abundance in certain peripheral tubules. These new split FPs not only offer multiple colors for imaging interaction networks of endogenous proteins, but also hold the potential to provide orthogonal handles for biochemical isolation of native protein complexes.

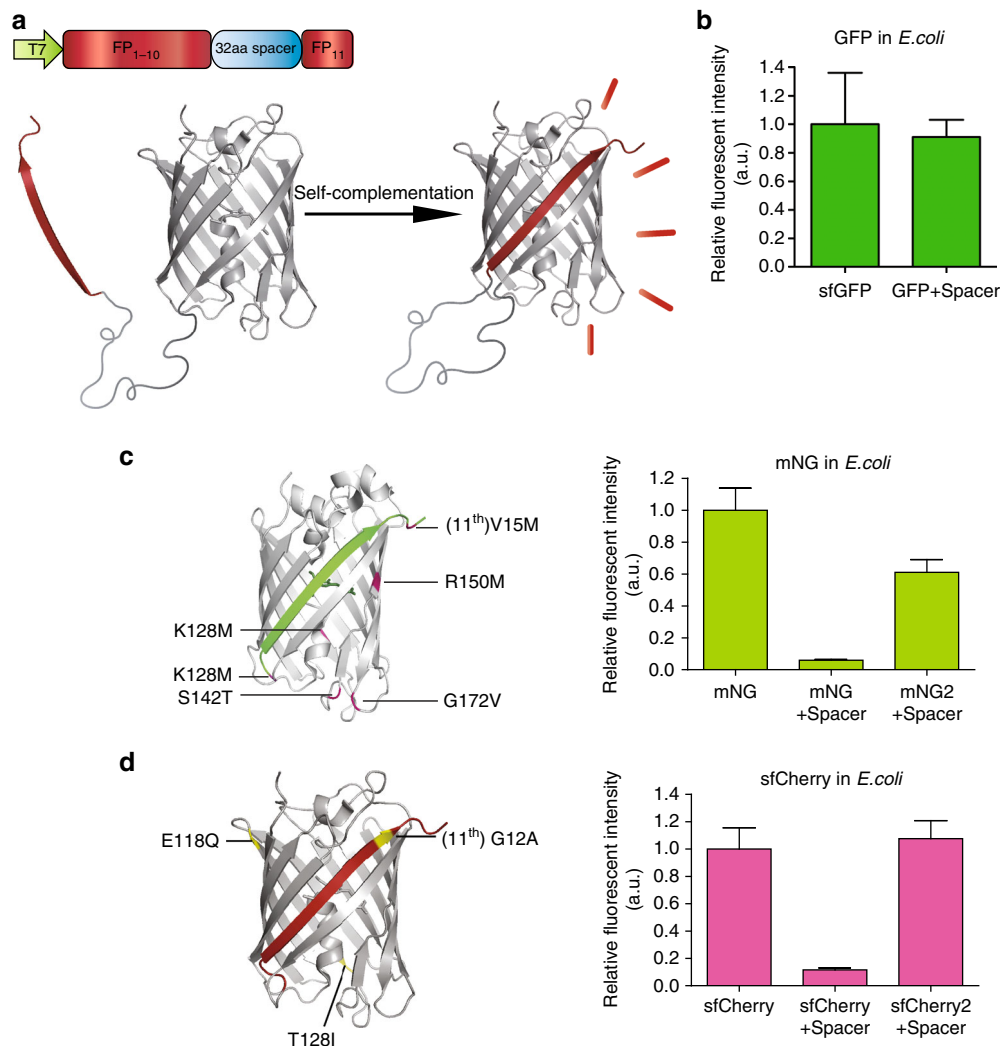
<sup>1</sup>The UC Berkeley–UCSF Graduate Program in Bioengineering, San Francisco, CA 94143, USA. <sup>2</sup>Department of Pharmaceutical Chemistry, University of California in San Francisco, San Francisco, CA 94143, USA. <sup>3</sup>Graduate Program of Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>4</sup>Department of Cellular and Molecular Pharmacology, and Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>5</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>6</sup>Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. Correspondence and requests for materials should be addressed to B.H. (email: [bo.huang@ucsf.edu](mailto:bo.huang@ucsf.edu))

Self-complementing split fluorescent proteins (FPs) are split FP constructs in which the two fragments can associate by themselves to form a fully functional FP without the assistance of other protein–protein interactions. By fusing one fragment on a target protein and detecting its association with the other fragment, these constructs have demonstrated powerful applications in the visualization of subcellular protein localization<sup>1–3</sup>, quantification of protein aggregation<sup>4</sup>, detection of cytosolic peptide delivery<sup>5, 6</sup>, identification of cell contacts and synapses<sup>7, 8</sup>, as well as scaffolding protein assembly<sup>3, 9, 10</sup>. Recently, they have also enabled the generation of large-scale human cell line libraries with fluorescently tagged endogenous proteins through CRISPR/Cas9-based gene editing<sup>11</sup>.

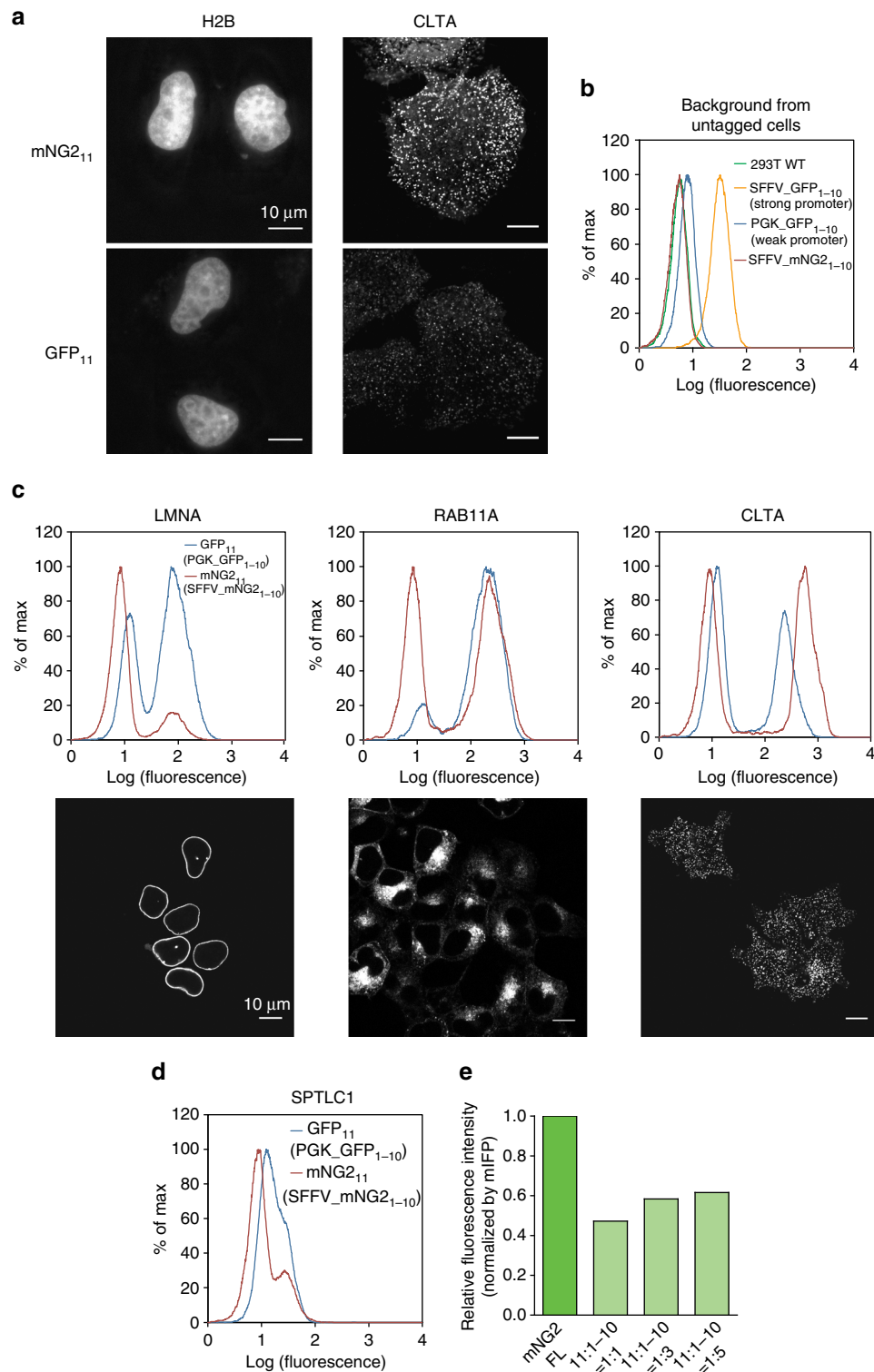
So far, the most commonly used self-complementing split FP was GFP<sub>1–10</sub>D7/11M3<sub>OPT</sub> (which we refer to as GFP<sub>1–10/11</sub>), engineered from super-folder GFP (sfGFP)<sup>12</sup>. With the splitting point between the tenth and eleventh  $\beta$ -strands, the resulting GFP<sub>11</sub> fragment is a 16-amino acid (a.a.) short peptide. The corresponding GFP<sub>1–10</sub> fragment remains almost non-fluorescent until complementation, making GFP<sub>1–10/11</sub> well suited for protein

labeling by fusing GFP<sub>11</sub> to the target protein and over-expressing GFP<sub>1–10</sub> in the corresponding subcellular compartments. However, there lacks a second, orthogonal split FP system with comparable complementation performance for multicolor imaging and multiplexed scaffolding of protein assembly. Previously, a sfCherry<sub>1–10/11</sub> system<sup>3</sup> was derived from super-folder Cherry, an mCherry variant optimized for folding efficiency<sup>13</sup>. However, its overall fluorescent brightness is substantially weaker than an intact sfCherry fusion, potentially due to its limited complementation efficiency<sup>3</sup>. Although two-color imaging with sfCherry<sub>1–10/11</sub> and GFP<sub>1–10/11</sub> has been done using tandem sfCherry<sub>11</sub> to amplify the sfCherry signal for over-expressed targets, it is still too dim to detect most endogenous proteins.

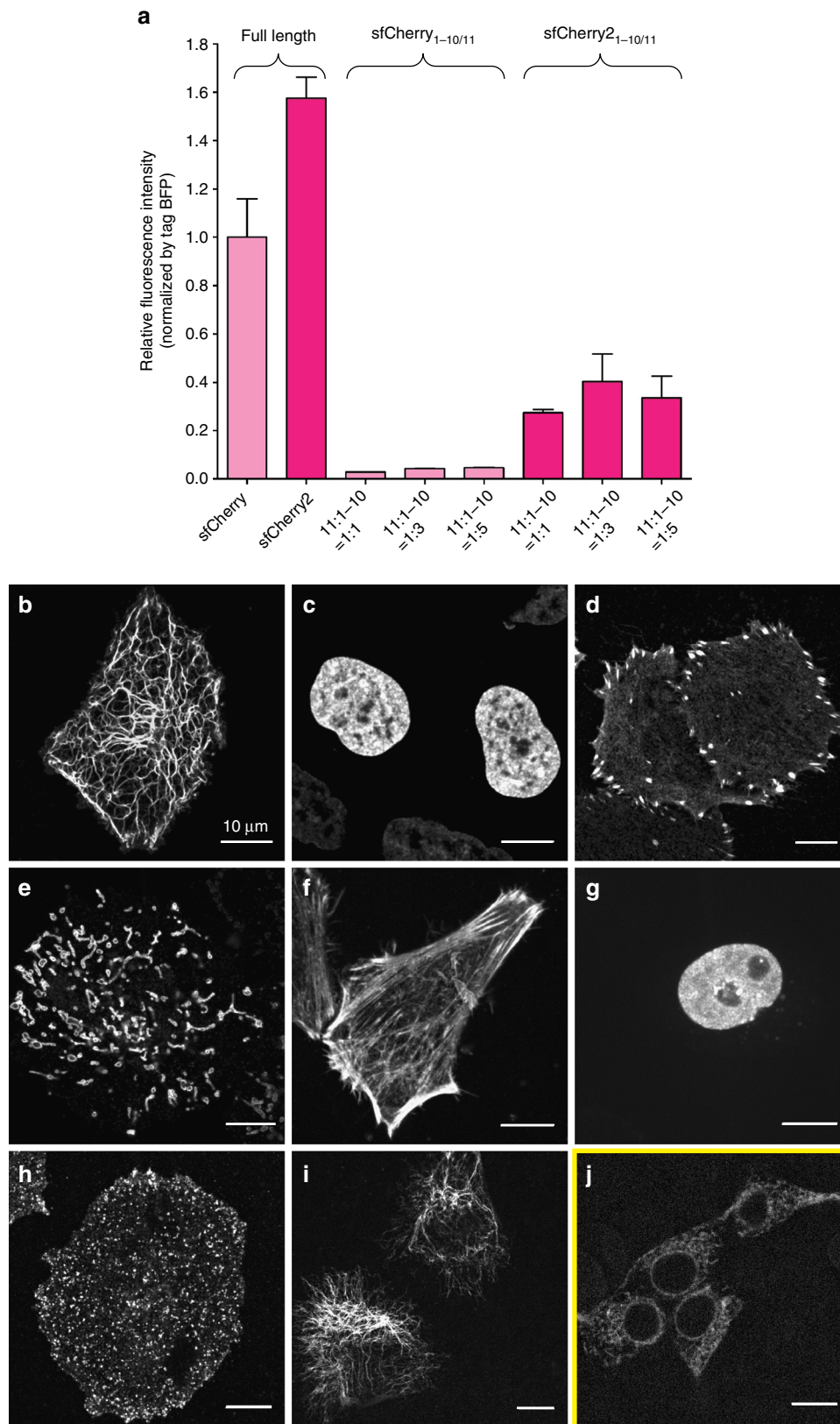
In this paper, we report a screening strategy for the direct engineering of self-complementing split FPs. Using this strategy, we have generated a yellow–green-colored mNeonGreen<sub>2</sub><sub>1–10/11</sub> (mNG2) that has an improved ratio of complemented signal to the background of FP<sub>1–10</sub>-expressing cells as compared to GFP<sub>1–10/11</sub>, as well as a red-colored sfCherry<sub>2</sub><sub>1–10/11</sub> that is about



**Fig. 1** Engineering split FP constructs using the spacer-insertion strategy. **a** Construct design and complementation scheme of the spacer-insertion screening system. **b** Relative fluorescence intensities of sfGFP and GFP + spacer in *E. coli* colonies. **c** Left: schematic diagram of split mNeonGreen2 with 6 mutations highlighted in pink, illustrated on the crystal structure of lanGFP. Right: relative fluorescence intensities of mNG, mNG + spacer, mNG2 + spacer in *E. coli* colonies. **d** Left: schematic diagram of split sfCherry2 with 3 mutations highlighted in yellow, illustrated on the crystal structure of sfCherry. Right: relative fluorescence intensities of sfCherry, sfCherry2 + spacer, sfCherry + spacer in *E. coli* colonies. For each bar in all bar graphs, Number of colonies > 400 and error bars are standard deviations



**Fig. 2** Protein labeling through transient expression or endogenous knock-in using mNG2<sub>11</sub>. **a** Fluorescence images of HeLa cells co-expressing either mNG2<sub>11</sub> or GFP<sub>11</sub> labeled H2B or CLTA with the corresponding FP<sub>1-10</sub>. Scale bars are 10 μm. **b** FACS background from wild-type HEK 293T cells, or HEK 293T cells stably expressing GFP<sub>1-10</sub> (using the strong SFFV promoter or the weak PGK promoter) or mNG2<sub>1-10</sub> (using SFFV). **c** Comparison of tagging endogenous LMNA, RAB11A, and CLTA using mNG2<sub>11</sub> or GFP<sub>11</sub> knock-in in SFFV-mNG2<sub>1-10</sub> and PGK-GFP<sub>1-10</sub> cells, respectively, (flow cytometry histograms) and the corresponding confocal images of mNG2<sub>11</sub> knock-in cells. Scale bars are 10 μm. **d** Comparison of tagging a low-abundance endogenous protein SPTLC1 through mNG2<sub>11</sub> or GFP<sub>11</sub> knock-in (flow cytometry histograms). **e** Whole cell fluorescence intensity of full length mNG2 and mNG2<sub>11</sub>-CLTA/mNG2<sub>1-10</sub>, measured by flow cytometry and normalized for expression level by mIFP fluorescence signal. Number of cells > 6000



**Fig. 3** Protein labeling and imaging using sfCherry<sub>211</sub>. **a** Whole cell fluorescence intensity of HEK 293T cells expressing full length sfCherry, full length sfCherry<sub>2</sub>, sfCherry<sub>1-10/11</sub> or sfCherry<sub>21-10/11</sub>, measured by flow cytometry and normalized for expression level by TagBFP fluorescence signal. Number of cells > 2500. Error bars are s.e.m. **b-i** Confocal images of sfCherry<sub>211</sub> labeling in HeLa cells cotransfected with sfCherry<sub>21-10</sub>: keratin **b**, intermediate filament); histone H2B **c**, nuclear); zyxin **d**, focal adhesion); TOMM20 **e**, mitochondria outer membrane);  $\beta$ -actin **f**, actin cytoskeleton); heterochromatin protein 1 **g**, nuclear); clathrin light chain A **h**, clathrin-coated pits); vimentin **i**, intermediate filament). **j** Confocal image of endogenous Sec61B (ER) labeled by sfCherry<sub>211</sub> in HEK 293T cells stably expressing sfCherry<sub>21-10</sub>. Scale bars are 10  $\mu$ m



10 times as bright as the original sfCherry<sub>1–10/11</sub>. Further, we have engineered a photoactivatable PAsfCherry<sub>2<sub>1–10/11</sub></sub> for single-molecule switching-based super-resolution microscopy. Using these split FPs, we have demonstrated dual-color endogenous protein tagging, which has revealed the reduced abundance of the endoplasmic reticulum (ER) translocon component Sec61B from certain peripheral ER tubules.

## Results

**Engineering split FPs with the spacer-insertion strategy.** Inspired by assays previously used to optimize a protease reporter<sup>9</sup>, we devised a general strategy for the engineering of self-complementing split FPs. Specifically, we inserted a 32 a.a. spacer (DVGGGGSEGGGSGGPGSGGEGSAGGGSAGGGS) between the tenth and eleventh  $\beta$ -strands of a fluorescent protein (Fig. 1a). This long spacer hinders the folding of the FP, which results in a fluorescence level much lower than its full length counterpart without the spacer. To improve the fluorescence, we then subjected the spacer-inserted FP to multiple rounds of directed evolution in *Escherichia coli*. In each round, the coding sequence was randomly mutagenized or shuffled. Then, the brightest 1 or 2 colonies from each plate were selected for the next round.

We first aimed to produce a green-colored split FP that has improved brightness compared to GFP. A recent quantitative assessment of FPs<sup>14</sup> reported that the brightness of mNeonGreen (mNG)<sup>15</sup>, a yellow-green fluorescent protein derived from *Branchiostoma lanceolatum*, is more than 2 times higher than sfGFP. mNG also demonstrates good photostability, acid tolerance and monomeric quality. Guided by the crystal structure of the closely related lanGFP (PDB: 4HVF), we chose to split between tenth and eleventh  $\beta$ -strands of mNG and removed the additional GFP-like C terminus (GMDELYK), resulting in a 213-amino acid fragment which we called mNG<sub>1–10</sub> and a 16-amino acid, mNG<sub>11</sub>. Unlike the highly optimized<sup>12</sup> split GFP<sub>1–10/11</sub> system, whose fluorescence signal is only slightly reduced with the spacer insertion (Fig. 1b), inserting the spacer between mNG<sub>1–10</sub> and mNG<sub>11</sub> drastically reduced its fluorescence signal (Fig. 1c). Using our spacer-assisted screening system, after three rounds of random mutagenesis, we identified five substitutions in the 1–10 fragment (K128M, S142T, R150M, G172V, and K213M) and one substitution in eleventh strand (V15M) (Fig. 1c). We named this improved mNG, mNG2. In *E. coli* colonies grown on LB-agar plates, spacer-inserted mNG2 demonstrated a 10-fold improvement in brightness after directed evolution, which is ~60% as bright as a full length mNG (Fig. 1c).

To improve the complementation efficiency of split sfCherry, we subjected the spacer-inserted sfCherry to three rounds of random mutagenesis and one round of DNA shuffling. We identified a new variant, named sfCherry2, which contains two mutations on the 1–10 fragment (E118Q and T128I) and one on the eleventh strand (G12A) (Fig. 1d). In *E. coli* colonies, spacer-inserted sfCherry2 is ~9 times as bright as the spacer-inserted original sfCherry (Fig. 1d). We have also used this strategy to split FusionRed, a red fluorescent protein with minimal cell toxicity and dimerization tendencies<sup>16</sup>. Unfortunately, we have not been able to obtain a brightly fluorescent, spacer-inserted variant even after four rounds of random mutagenesis.

**Protein labeling by mNG2<sub>1–10/11</sub> in mammalian cells.** To test protein labeling using mNG2<sub>11</sub>, we expressed two proteins, histone H2B (H2B) or clathrin light chain A (CLTA) fused to mNG2<sub>11</sub> in HeLa cells. With the co-expression of mNG2<sub>1–10</sub>, we could correctly image the localization of these proteins, similar to that using GFP<sub>11</sub> (Fig. 2a). Interestingly, when GFP<sub>1–10</sub> is expressed by itself, we observed a weak but non-negligible

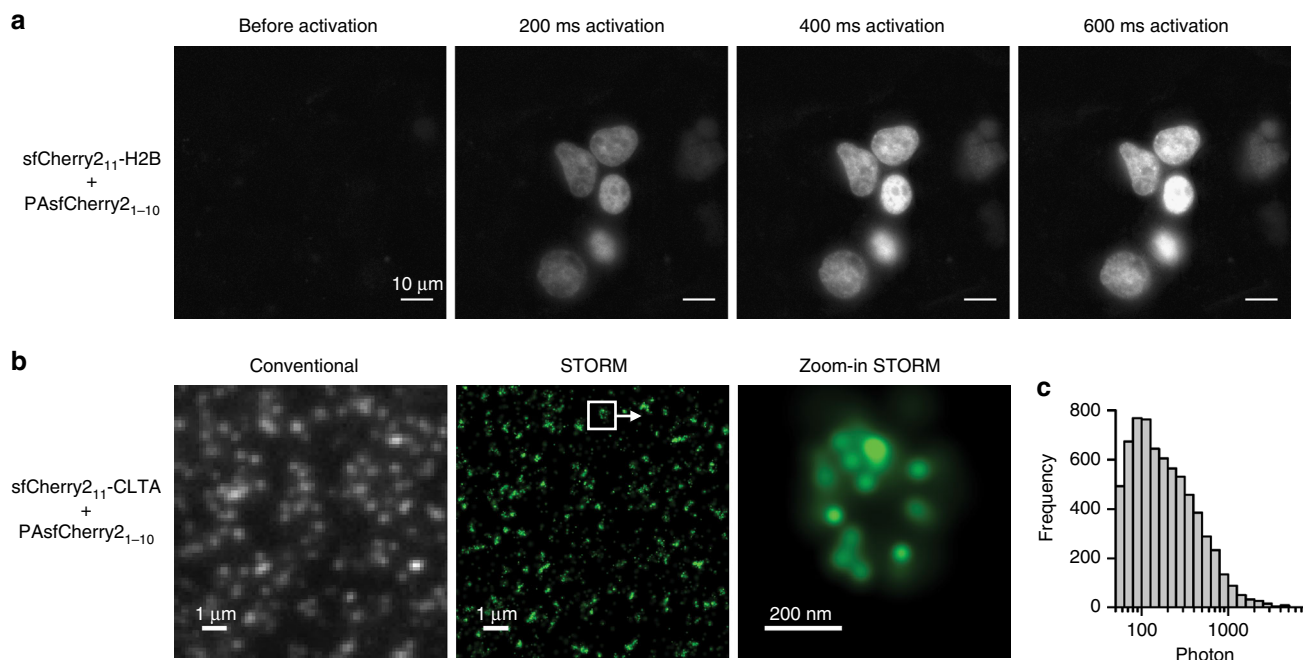
fluorescence background even without the expression of GFP<sub>11</sub> fragment. Comparing HEK 293T cell lines stably expressing GFP<sub>1–10</sub> from the strong SFFV promoter and the weak PGK promoter, we found that the background is positively correlated with the GFP<sub>1–10</sub> expression level (Fig. 2b). This background might be attributed to either weak fluorescence from GFP<sub>1–10</sub> or elevated cell autofluorescence caused by GFP<sub>1–10</sub> expression. In contrast, even when using the SFFV promoter, the signal from mNG2<sub>1–10</sub>-expressing cells is indistinguishable from wild-type cell autofluorescence (Fig. 2b).

This low background fluorescence from mNG2<sub>1–10</sub> is important for the labeling of endogenous proteins using FP<sub>11</sub> knock-in. Previously, utilizing the small GFP<sub>11</sub> tag we have developed a scalable scheme to fluorescently tag endogenous proteins in human cell lines by gene editing using electroporation of Cas9/sgRNA ribonucleoprotein (RNP) and a single-stranded donor DNA, followed by enrichment of integrated cells by fluorescence activated cell sorting (FACS)<sup>11</sup>. Similar to GFP<sub>11</sub>, mNG2<sub>11</sub> is also a 16 a.a. peptide, allowing its DNA and the homology arms to fit in a 200 nucleotide (nt) donor DNA that can be directly obtained from commercial synthesis, which is a key contributor to the efficiency and cost-effectiveness of our method. We compared genetic knock-in using GFP<sub>11</sub> or mNG2<sub>11</sub> for three targets: Lamin A/C (LMNA, inner nuclear membrane), RAB11A and CLTA. In all three cases, we observed similar or stronger fluorescence from mNG2<sub>11</sub> knock-in cells than GFP<sub>11</sub> knock-in cells. The background from non-integrated cells is substantially lower from mNG2<sub>1–10</sub>, despite the use of the low expression PGK GFP<sub>1–10</sub> and high expression SFFV mNG2<sub>1–10</sub>. This better separation between FP<sub>11</sub>-integrated and non-integrated cells is especially advantageous for labeling low-abundance proteins, for example, SPTLC1 (Serine Palmitoyl-transferase Long Chain Base Subunit 1) (Fig. 2d).

Finally, to compare the brightness of mNG2<sub>11</sub> with that of full length mNG2 on proteins, we constructed plasmids encoding full length mNG2 or mNG2<sub>11</sub>-CLTA fused to mIFP<sup>17</sup> through a self-cleaving P2A site, so that expression level differences can be normalized by mIFP signal level. We cotransfected HEK 293T cells with either of the two plasmids and a plasmid expressing mNG2<sub>1–10</sub>. With 11:1–10 transfected DNA ratio increased from 1:1 to 1:3 and 1:5, the normalized whole cell fluorescence by flow cytometry from mNG2<sub>11</sub> is approximately 50–60% of that from full length mNG2 (Fig. 2e). This relative brightness likely indicates the overall effect of complementation efficiency, chromophore maturation and potential perturbation to chromophore environment by the protein split. For reference, we performed the same measurement on GFP<sub>11</sub> and observed a brightness close to that of the full length sfGFP (Supplementary Fig. 1).

**Protein labeling using sfCherry2<sub>1–10/11</sub> in mammalian cells.** To quantify the improvement of sfCherry2<sub>1–10/11</sub> over sfCherry<sub>1–10/11</sub> for mammalian protein labeling and compare their performances with that of full length sfCherry and sfCherry2, we performed a similar measurement as in the previous section. We used TagBFP instead of mIFP for expression level normalization because sfCherry fluorescence bleeds into the mIFP detection channel (Fig. 3a). We found that sfCherry2<sub>1–10/11</sub> is ~10 times as bright as sfCherry<sub>1–10/11</sub> in mammalian cells. We have also observed that full length sfCherry2 is ~50% brighter than sfCherry<sup>13</sup>, suggesting a better overall folding efficiency. The normalized fluorescence signal of sfCherry2<sub>1–10/11</sub> reached ~30% of full length sfCherry and ~20% of full length sfCherry2.

To test sfCherry2<sub>11</sub> as a fluorescent tag for live-cell imaging, we constructed mammalian expression vectors encoding target



**Fig. 4** Photoactivatable sfCherry<sub>2<sub>1-10/11</sub></sub> and its use in super-resolution microscopy. **a** Photoactivation of H2B labeling in HEK 293 T cells cotransfected with sfCherry<sub>2<sub>11</sub></sub>-H2B and PAsfCherry<sub>2<sub>1-10</sub></sub>. Scale bars: 10 μm. **b** Conventional wide-field fluorescence image (left) and super-resolution image (middle/right) of endogenous CLTA. **c** Histogram of photon number detected per photoactivation event in **b**

proteins tagged with sfCherry<sub>2<sub>11</sub></sub> at either the N or C terminus and co-expressed each with cytoplasmic sfCherry<sub>2<sub>1-10</sub></sub> in HeLa cells (Fig. 3b–i). For the diverse array of target proteins tested, including nuclear proteins histone H2B and heterochromatin protein 1 (HP1), cytoskeletal proteins β-actin, keratin and vimentin, focal adhesion protein zyxin, CLTA, and mitochondrial outer membrane protein TOMM20, we observed their correct localization. No fluorescent signal was detected with the sfCherry<sub>2<sub>1-10</sub></sub> fragment alone.

We also demonstrated sfCherry<sub>2<sub>11</sub></sub> labeling of endogenous proteins by knocking it into the ER translocon complex protein Sec61B in HEK 293T cells stably expressing both GFP<sub>1-10</sub> and sfCherry<sub>2<sub>1-10</sub></sub> (293 T<sup>double1-10</sup>). After sorting for red-fluorescence-positive cells by FACS, confocal imaging (Fig. 3j) confirmed the ER specificity of the sfCherry<sub>2<sub>11</sub></sub> signal, which is substantially weaker than the other over-expression cases (Fig. 3b–i). We noticed that many cells also display additional punctate structures, which we have verified to be lysosomes (see the last result section). This artifact is common for mCherry-derived FPs when expressed for a prolonged period of time, especially when targeting proteins in secretory pathways<sup>18</sup>. It likely results from the fact that mCherry has a β-barrel structure resisting lysosomal proteolysis and a low pKa, such that it remains fluorescent in the acidic lysosome lumen<sup>19</sup>. We have not observed lysosome labeling in any of the non-ER targets that we have imaged in this work with sfCherry<sub>2<sub>11</sub></sub> labeling.

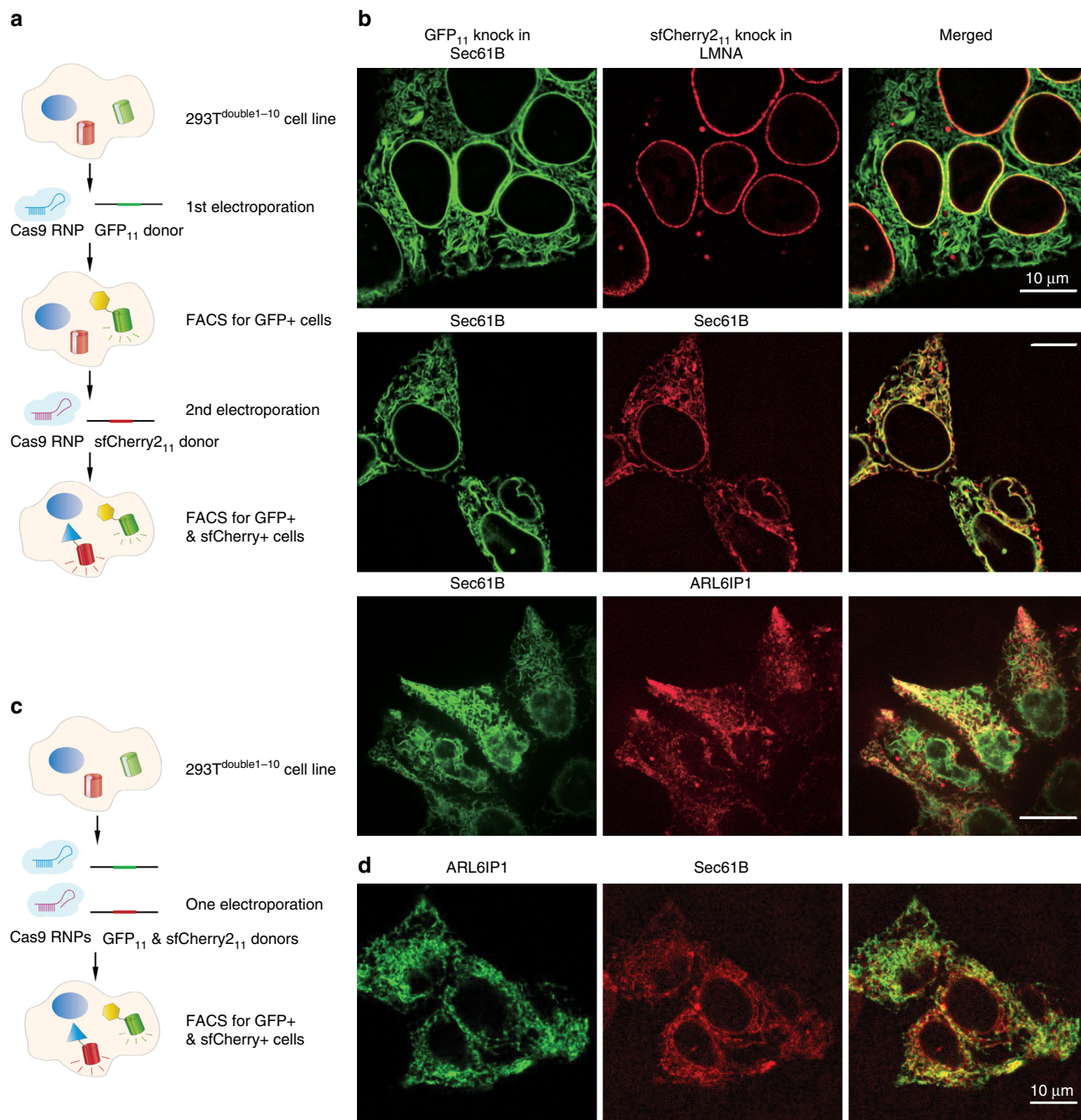
**Super-resolution microscopy using PAsfCherry<sub>2<sub>1-10/11</sub></sub>.** With their capability to change their fluorescence properties upon light (usually ultraviolet) irradiation, photoactivatable (PA) FPs<sup>20</sup> enable tracking of protein trafficking and imaging of labeled proteins using single-molecule switching-based super-resolution microscopy<sup>21</sup> (more commonly known as stochastic optical reconstruction microscopy, STORM, or photoactivated localization microscopy, PALM). Previously, mCherry has been engineered into a PA FP (PAmCherry1) after introducing 10 a.a.

substitutions<sup>22</sup>. Because these substitutions are located on the 1–10 fragment, we merged them with the sequence of sfCherry<sub>2<sub>1-10</sub></sub> (designated as PAsfCherry<sub>2<sub>1-10</sub></sub>). When imaging HEK 293T cells co-expressing sfCherry<sub>2<sub>11</sub></sub>-H2B and PAsfCherry<sub>2<sub>1-10</sub></sub>, we observed little initial fluorescence in red channel and a large fluorescence increase after 405 nm light irradiation (Fig. 4a), confirming the photoactivation property of complemented sfCherry<sub>2<sub>11</sub></sub> and PAsfCherry<sub>2<sub>1-10</sub></sub>.

Combining PAsfCherry<sub>2<sub>1-10</sub></sub> with our FP<sub>11</sub> tag knock-in method, we can easily perform STORM imaging of endogenous proteins tagged by sfCherry<sub>11</sub>, which could avoid potential overexpression artifacts. For demonstration, we imaged endogenous CLTA in HEK 293 T cells. Because PAsfCherry<sub>2</sub> is non-fluorescent before activation, we knocked-in a tandem GFP<sub>11</sub>-sfCherry<sub>2<sub>11</sub></sub> tag in HEK 293T cells stably expressing GFP<sub>1-10</sub>, so that integrated cells can be isolated using GFP fluorescence signal. We then transfected these cells with PAsfCherry<sub>2<sub>1-10</sub></sub> plasmid for STORM imaging. Compared to conventional wide-field images, clathrin-coated pits can be clearly seen as sub-diffraction-limit objects in the STORM images (Fig. 4b). We were able to collect on average 260 photons in each photoactivation event (Fig. 4c). We note that the labeling of most pits is incomplete, potentially because we labeled only one of the two clathrin light chain genes that have nearly identical functionalities<sup>23</sup>. In addition, non-homozygous knock-in and inefficient complementation between the two fragments can further contribute to the incomplete labeling. Selecting for homozygous knock-in cells and/or the use of tandem sfCherry<sub>2<sub>11</sub></sub> tags<sup>3</sup> may solve the latter two problems.

**Dual-color endogenous protein tagging in human cells.** The orthogonal sfCherry<sub>2<sub>11</sub></sub> now enables two-color imaging of endogenous proteins in order to visualize their differential spatial distribution and interactions. We tested double knock-in of GFP<sub>11</sub> and sfCherry<sub>2<sub>11</sub></sub> either sequentially or simultaneously in HEK 293T cells stably expressing both GFP<sub>1-10</sub> and sfCherry<sub>2<sub>1-10</sub></sub> (293 T<sup>double1-10</sup>). For sequential knock-in, we first performed



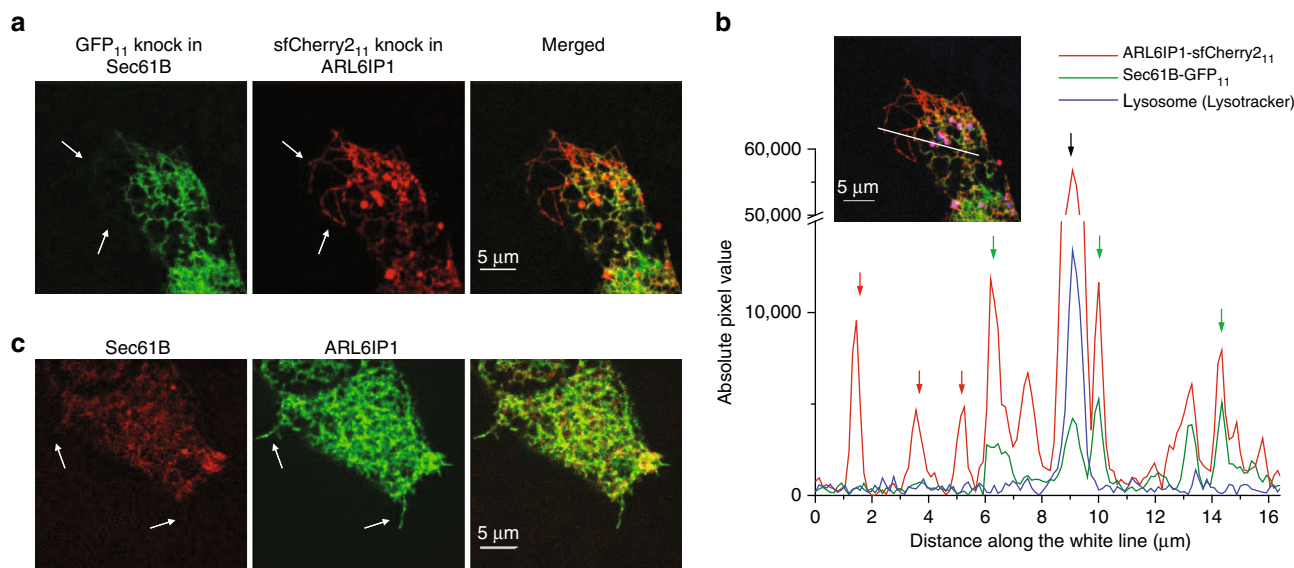


**Fig. 5** Double labeling and dual-color imaging of endogenous proteins using GFP<sub>11</sub> and sfCherry<sub>211</sub>. **a** Schematic diagram for double labeling endogenous proteins through a sequential knock-in strategy. **b** Dual-color fluorescence images of GFP<sub>11</sub>/sfCherry<sub>211</sub> sequential knocked-in cells. **c** Schematic diagram double labeling endogenous proteins through a simultaneous knock-in strategy. **d** Dual-color fluorescence images of GFP<sub>11</sub>/sfCherry<sub>211</sub> simultaneous knocked-in cells. Scale bars are 10 μm

electroporation of Cas9 RNP and single-strand donor DNA for GFP<sub>11</sub>, sorted GFP positive cells by FACS, and then knocked-in sfCherry<sub>211</sub>, followed by a second round of sorting (Fig. 5a). We targeted Sec61B using GFP<sub>11</sub> and then three proteins with distinctive subcellular localizations using sfCherry<sub>211</sub>: LMNA, ARL6IP1 (tubular ER), and Sec61B (to verify co-localization). When imaging the double knock-in cells by confocal microscopy, we observed co-labeling of the nuclear envelope by Sec61B-GFP<sub>11</sub> and LMNA-sfCherry<sub>211</sub>, almost complete co-localization of Sec61B-GFP<sub>11</sub> and Sec61B-sfCherry<sub>211</sub>, and the exclusion of ARL6IP1-sfCherry<sub>211</sub> from the nuclear envelope (Fig. 5b).

For simultaneous knock-in, we mixed Cas9 RNPs and donor DNAs for both targets in one electroporation reaction and used FACS to enrich cells positive for both green and red signal (Fig. 5c). We chose to tag ARL6IP1 and Sec61B with GFP<sub>11</sub> and sfCherry<sub>211</sub>, respectively. We obtained 0.4% double-positive cells, consistent with the multiplication of the 28 and 1.2% efficiency for GFP<sub>11</sub> and sfCherry<sub>211</sub> single knock-in into the ARL6IP1 and Sec61B gene, respectively (Supplementary Fig. 2). The lower efficiency for sfCherry<sub>211</sub> knock-in could be attributed to its overall lower fluorescence signal level compared to GFP<sub>11</sub>, making it more difficult to distinguish from the background by FACS. Confocal microscopy of sorted cells showed similar





**Fig. 6** Reduced abundance of Sec61B in certain peripheral ER tubules. **a** Dual-color image of a Sec61B-GFP<sub>11</sub> and ARL6IP1-sfCherry2<sub>11</sub> knock-in HEK 293 T cell. Arrows indicate ER tubules marked by ARL6IP1 but contain much reduced Sec61B signal. **b** A cross-section in **a**, with red arrows indicating peripheral ER tubules with Sec61B signal at the background level, green arrows indicating ER tubules positive for both Sec61B and ARL6IP1, and the black arrow indicating a lysosome marked by LysoTracker Blue. **c** Dual-color image of a Sec61B-sfCherry2<sub>11</sub> and ARL6IP1-GFP<sub>11</sub> knock-in HEK 263 T cell. Scale bars are 5 μm

arrangement of the two proteins (Fig. 5d) as in Sec61B-GFP<sub>11</sub>, ARL6IP1-sfCherry2<sub>11</sub> sequential knock-in.

#### Dual-color images reveal peripheral ER with reduced Sec61B.

The endoplasmic reticulum (ER) is a large organelle that spreads throughout the cytoplasm as a continuous membrane network of tubules and sheets with a single lumen<sup>24</sup>. The Sec61 complex, which is composed of alpha, beta, and gamma subunits, is the central component of the protein translocation apparatus of the ER membrane<sup>25</sup>. Researchers have traditionally used Sec61B as an ER marker for imaging because it is thought to be distributed ubiquitously throughout the ER membrane, including nuclear envelope, sheet-like cisternae, and a polygonal array of tubules<sup>26, 27</sup>. However, our dual-color images of endogenous Sec61B and ARL6IP1 in HEK 293 T cells using GFP<sub>11</sub> and sfCherry2<sub>11</sub>, showed that certain peripheral ER tubules marked by ARL6IP1 contain very weak to non-detectable Sec61B signal (Fig. 6a). This large reduction of Sec61B signal in certain ER tubules is clearly visible in cross-section, where the Sec61B signal is at the background level despite Sec61B having the brighter GFP<sub>11</sub> label than sfCherry2<sub>11</sub> on ARL6IP1 (Fig. 6b). We have also ruled out the presence of sfCherry2-containing lysosomes in these areas by staining lysosomes with lysotracker (Supplementary Fig. 3). By visual inspection in z maximum projects of 9 confocal images containing a total of 108 cells, we identified that 29 of them contain such peripheral ER tubules lacking strong Sec61B presence. Furthermore, we confirmed this differential distribution of ER membrane proteins by swapping the tags on the two proteins (Fig. 6c).

#### Discussion

In summary, we have devised a simple platform for the engineering of self-complementing split fluorescence proteins. Using this platform, we have developed a bright yellow–green-colored mNG2<sub>1–10/11</sub> system and substantially increased the performance of the red-colored sfCherry2<sub>1–10/11</sub>. These split constructs have allowed us to obtain two-color or super-resolution images of

endogenous proteins and have revealed ER tubules with greatly reduced abundance of the translocon component Sec61B.

Our platform can be easily extended to the engineering of other self-complementing split FPs with distinct colors (e.g., mTurquoise2, mTagBFP2)<sup>28, 29</sup>, good photoactivation performance (e.g., mMaple3, PATagRFP)<sup>30, 31</sup>, or new functionalities (e.g., pH sensitivity)<sup>32</sup>. We note that for non-self-complementing split FPs, which are used to detect protein–protein interactions in bimolecular fluorescence complementation (BiFC) assays<sup>33</sup>, a different engineering platform is needed to ensure minimum affinity between the two FP fragments by themselves. On the other hand, although our optimized spacer-inserted sfCherry2 has already reached the brightness level of sfCherry without spacer, split sfCherry2<sub>1–10/11</sub> is still much dimmer compared to full length sfCherry. Further improvement could be possible using a longer or more rigid spacer, which adds more spatial hindrance to complementation, or a spacer containing the self-cleaving P2A site, which better mimics the actual split system.

With its higher ratio of complemented signal to the background of FP<sub>1–10</sub>-expressing cells compared to GFP<sub>1–10/11</sub>, our mNG2<sub>1–10/11</sub> system will be advantageous for tagging low-abundance endogenous proteins<sup>3</sup>. More importantly, it provides an orthogonal handle for scaffolding protein oligomerization<sup>3</sup> and biochemical isolation of native protein complexes<sup>11</sup>. For the sfCherry2<sub>1–10/11</sub> system, the fact that we have only observed lysosome puncta when labeling ER proteins suggests that this problem could be potentially resolved by increasing its pK<sub>a</sub> with rational designs<sup>32</sup>. Moreover, our engineering platform can be easily adapted to generating other red split FPs based on novel bright FPs such as TagRFP-T<sup>34</sup>, mRuby3<sup>35</sup>, or mScarlet<sup>36</sup>.

Given the wide use of Sec61B as a marker to label the entire ER, it is surprising that Sec61B has a substantially reduced abundance in certain peripheral tubular structures labeled by ARL6IP1. It is possible that these tubules, often containing a closed end pointing toward the edge of the cell, serve distinct functions from other ER tubules. Further analysis of their protein compositions and contacts with plasma membrane, cytoskeleton and other

organelles may help clarify their functional roles. This observation also suggests that the base of these ER tubules may contain a diffusion barrier that hinders the crossing of Sec61B. Because over-expressing exogenous ER shaping proteins can drastically alter ER morphology<sup>37</sup>, this finding highlights the opportunity of visualizing endogenous proteins to study native interaction networks.

## Methods

**Mutagenesis and screening.** The amino acid sequence of mNG was obtained from the published literature<sup>15</sup>. Because the crystal structure of mNG has not been determined, we used the lanGFP (PDB: 4HVF) structure as a guide to decide splitting between K213 and T214 at the middle of the loop between tenth and eleventh  $\beta$ -strands. And we also removed the additional seven-residue GFP-like C terminal (GMDELYK) to minimize the size of mNG11 tag. The amino acid sequence of sfCherry and the split site were from our previous published literature<sup>3</sup>. The spacer-inserted split mNG and split sfCherry were subjected to random mutagenesis using a GeneMorph II Random Mutagenesis Kit (Agilent). The cDNA library pool was transformed into *E. coli* BL21 (DE3) electrocompetent cells (Lucigen) by electroporation using the Gene Pulser Xcell Electroporation Systems (BioRad). The expression library was plated on nitrocellulose membrane (Whatman, 0.45  $\mu$ m pore size), which was sitting on an LB-agar plate with 30 mg/ml kanamycin. After overnight growth at 37 °C, the nitrocellulose membrane was transferred to a new LB-agar plate containing 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) and 30 mg/ml kanamycin and cultured for 3–6 h at 37 °C to induce the protein expression. Clone screening was performed by imaging the second LB-agar plate using a BioSpectrum Imaging System (UVP). The brightest candidates in each library were pooled (typically 20–30 selected from ~10,000 colonies) and used as templates for the next round of evolution. For DNA shuffling, we used the method described by Yu et al.<sup>38</sup>. Specifically, we digested plasmids containing spacer-inserted sfCherry<sub>1–10/11</sub> variants with BamHI and XhoI. Fragments of ~800 bp were purified from 1% agarose gels using zymo clean gel DNA gel recovery kit (Zymo Research). The DNA concentrations were measured and the fragments were mixed at equal amounts for a total of ~2  $\mu$ g. The mixture was then digested with 0.5 unit DNase I (New England Biolabs) for 13 min and terminated by heating at 95 °C for 10 min. The DNase I digests were run on a 2% agarose gel, and the size of 50–100 bp fragments were cut out and purified. Ten microliter of purified fragments was added to 10  $\mu$ l of Phusion High-Fidelity PCR Master Mix and reassembled with a PCR program of 30 cycles, with each cycle consisting of 95 °C for 60 s, 50 °C for 60 s, and 72 °C for 30 s. After gene reassembly, 1  $\mu$ l of this reaction was amplified by PCR. The shuffled library was expressed and screened as described above. After the directed evolution was saturated, the brightest clone was selected and the DNA sequences of the constructs were confirmed by sequencing (Quintara Biosciences).

**Molecular cloning.** The screening construct was expressed from a pET28a bacterial expression vector and contained a 32-residue spacer (DVGGGGSE GGGGCGPGSGGEGSAGGGSAGGGS) between tenth and eleventh  $\beta$ -strands of FPs. The DNA sequences of mNG1–10\_32aaspacer\_mNG11 and sfCherry1–10\_32aaspacer\_sfCherry11 were directly synthesized (Integrated DNA Technologies, IDT). For the nucleotide sequence of identified final mutants of mNG2<sub>1–10</sub>/mNG2<sub>11</sub> and sfCherry2<sub>1–10</sub>/sfCherry2<sub>11</sub>, Supplementary Table 1.

The DNAs of mNG2<sub>11</sub> and sfCherry2<sub>11</sub> were PCR amplified from identified pET28a constructs (final mutants) using Phusion High-Fidelity DNA Polymerase (Thermo Scientific). The DNAs of histone H2B, clathrin light chain A, keratin,  $\beta$ -actin, zyxin, heterochromatin protein 1, TOMM20, vimentin, laminB, mIFP were subcloned from mEmerald, sfGFP or mIFP fusion plasmids (cDNA source: the Michael Davidson Fluorescent Protein Collection at the UCSF Nikon Imaging Center). The P2A sequence used for all mIFP constructs are GCTACTAA CTTCAGCCTGCTGAAGCAGGCTGGAGCGTGGAGGAGGAAACCCTG GACCT. We performed the following restriction enzyme digestion (amino acid linker length shown in parentheses for each): histone H2B (10 a.a.): sfGFP sequence between AgeI and BglIII (sfGFP-H2B-C-10); clathrin light chain A (15 a.a.): mEmerald sequence between AgeI and BglII (mEmerald-Clathrin-15); keratin (18 a.a.): mEmerald sequence between BamHI and NotI (mEmerald-Keratin14-N-18);  $\beta$ -actin (18 a.a.): mEmerald sequence between AgeI and BglII (mEmerald-Actin-C-18); zyxin (6 a.a.): sfGFP sequence between BamHI and NotI (sfGFP-Zyxin-6); TOMM20 (20 a.a.): mEmerald sequence between BamHI and NotI (mEmerald-TOMM20-N-10); vimentin (9 a.a.): mEmerald sequence between BamHI and NotI (mEmerald-Vimentin-N-18); laminB (10 a.a.): mEmerald sequence between AgeI and BglII (mEmerald-LaminB1-10). PCR amplified mNG2<sub>11</sub> and sfCherry2<sub>11</sub> were then ligated with the digested vectors using In-Fusion HD Cloning kit (Clontech). For the mammalian expression and lentiviral production, DNAs of mNG2<sub>1–10</sub> or sfCherry2<sub>1–10</sub> were directly PCR amplified from identified pET28a constructs (final mutants) and cloned into pcDNA3.1 vectors (HindIII/BamHI) as well as lentiviral pHR-SFFV vector (BamHI/NotI).

To generate the split PAsfCherry2, we introduced 10 photoactivation point mutations (based on PAmCherry1 sequence from literature<sup>22</sup>) to the sfCherry2<sub>1–10</sub> sequence and cloned the synthesized PAsfCherry2<sub>1–10</sub> fragment into a pcDNA3.1 vector. Those substitutions are: E26V/A58T/K69N/L84F/N99K/S148L/I165V/Q167P/L169V/I203R. For the complete nucleotide sequence of PAsfCherry2<sub>1–10</sub>, Supplementary Table 1.

**Cell culture and generation of stable cell lines.** Human HEK 293 T and human HeLa cells (obtained from UCSF Cell Culture Facility) were maintained in Dulbecco's Modified Eagle Medium (DMEM) with high glucose (UCSF Cell Culture Facility), supplemented with 10% (vol/vol) FBS and 100  $\mu$ g/ml penicillin/streptomycin (UCSF Cell Culture Facility). All cells were grown at 37 °C and 5% CO<sub>2</sub> in a humidified incubator. Both cell lines have been tested to be free of mycoplasma contamination. For the lentiviral production, HEK 293T cells were plated into 6-well plates 1 day prior to transfection. A quantity of 110 ng of pMD2.G plasmid, 890 ng of pCMV-dR8.91 plasmid and 1000 ng of the lentiviral plasmid (pSFFV-mNG2<sub>1–10</sub>, pSFFV-GFP<sub>1–10</sub> or pSFFV-sfCherry2<sub>1–10</sub>) were cotransfected into HEK 293T in each well using FuGENE HD (Promega) following the manufacturer's recommended protocol. Lentivirus was collected 48 h after transfection and were stored immediately in –80 °C freezer for future use. To generate the HEK 293T cells stably expressing (1) mNG2<sub>1–10</sub> or (2) both GFP<sub>1–10</sub> and sfCherry2<sub>1–10</sub> (293T<sup>double1–10</sup>), HEK 293T cells were infected with (1) pSFFV-mNG2<sub>1–10</sub> or (2) both pSFFV-GFP<sub>1–10</sub> and pSFFV-sfCherry2<sub>1–10</sub> lentiviruses. Forty-eight hours after infection, we transiently transfected them with (1) mNG2<sub>11</sub>-H2B plasmid or (2) GFP<sub>11</sub>-H2B and sfCherry2<sub>11</sub>-H2B plasmids in order to isolate cells with successful lentiviral integration of (1) mNG2<sub>1–10</sub> or (2) GFP<sub>1–10</sub> and sfCherry2<sub>1–10</sub> by FACS.

**Sample preparation and data analysis in flow cytometry.** To characterize the complementation efficiency of split mNG2 and split GFP, we made pSFFV-mIFP\_P2A\_FP[full length] and pSFFV-mIFP\_P2A\_FP<sub>11</sub>-CLTA constructs. Corresponding to each bar in Fig. 2e and Supplementary Fig. 1a, HEK 293 T cells grown on 48-well plate (Eppendorf) were cotransfected with (1) 80 ng pSFFV-mIFP\_P2A\_mNG2[full length] with 80 ng pSFFV-mNG2<sub>1–10</sub>, (2) 80 ng pSFFV-mIFP\_P2A\_mNG2<sub>11</sub>-CLTA with 80 ng pSFFV-mNG2<sub>1–10</sub>, (3) 80 ng pSFFV-mIFP\_P2A\_mNG2<sub>11</sub>-CLTA with 160 ng pSFFV-mNG2<sub>1–10</sub>, (4) 80 ng pSFFV-mIFP\_P2A\_mNG2<sub>11</sub>-CLTA with 240 ng pSFFV-mNG2<sub>1–10</sub>, (5) 80 ng pSFFV-mIFP\_P2A\_sfGFP[full length] with 80 ng pSFFV-GFP<sub>1–10</sub>, (6) 80 ng pSFFV-mIFP\_P2A\_GFP<sub>11</sub>-CLTA with 80 ng pSFFV-GFP<sub>1–10</sub>, (7) 80 ng pSFFV-mIFP\_P2A\_GFP<sub>11</sub>-CLTA with 160 ng pSFFV-GFP<sub>1–10</sub>, (8) 80 ng pSFFV-mIFP\_P2A\_GFP<sub>11</sub>-CLTA with 240 ng pSFFV-GFP<sub>1–10</sub>. We varied the ratio of FP<sub>11</sub> fragment to FP<sub>1–10</sub> fragment from 1:1 to 1:3 to 1:5, to verify the saturation of complementation.

To characterize the increased fluorescence intensity of split sfCherry2 vs. original split sfCherry, we made pSFFV-FP<sub>11</sub>-TagBFP and pSFFV-FP<sub>11</sub>-TagBFP constructs. Corresponding to each bars in Fig. 3a, HEK 293 T cells grown on 24-well plate (Eppendorf) were transfected using with (1) 500 ng pSFFV-sfCherry<sub>11</sub>-TagBFP, (2) 500 ng pSFFV-sfCherry2<sub>11</sub>-TagBFP, (3) 500 ng pSFFV-sfCherry<sub>11</sub>-TagBFP with 500 ng pSFFV-sfCherry<sub>1–10</sub>, (4) 500 ng pSFFV-sfCherry<sub>11</sub>-TagBFP with 1000 ng pSFFV-sfCherry<sub>1–10</sub>, (5) 500 ng pSFFV-sfCherry<sub>11</sub>-TagBFP with 1500 ng pSFFV-sfCherry<sub>1–10</sub>, (6) 500 ng pSFFV-sfCherry<sub>211</sub>-TagBFP with 500 ng pSFFV-sfCherry<sub>21–10</sub>, (7) 500 ng pSFFV-sfCherry<sub>211</sub>-TagBFP with 1000 ng pSFFV-sfCherry<sub>21–10</sub>, (8) 500 ng pSFFV-sfCherry<sub>211</sub>-TagBFP with 1500 ng pSFFV-sfCherry<sub>21–10</sub>.

Analytical flow cytometry was carried out on a LSR II instrument (BD Biosciences) and cell sorting on a FACSAria II (BD Biosciences) in Laboratory for Cell Analysis at UCSF. Flow cytometry data analysis (gating and normalization) was done using the FlowJo software (FlowJo LLC) and plotted in GraphPad Prism.

**Fluorescence microscopy.** We transfected human HeLa cells grown on an 8-well glass bottom chamber (Thermo Fisher Scientific) using FuGene HD (Promega). In order to achieve better cell attachment, 8-well chamber was coated with Fibronectin (Sigma-Aldrich) for 1 h before seeding cells. Total plasmid amount of 120 ng per well with the FP<sub>11</sub> to FP<sub>1–10</sub> ratio in 1:2 was used to achieve optimal labeling. Thirty-six to forty-eight hours after transfection, live cells were imaged and then fixed with 4% paraformaldehyde. For lysosome staining, LysoTracker Blue DND-22 (Thermo Fisher Scientific) was added directly to the culture medium (50 nM final concentration) and incubate for 30 min before imaging.

Most of live-cell imaging was acquired on an inverted Nikon Ti-E microscope (UCSF Nikon Imaging Center), a Yokogawa CSU-W1 confocal scanner unit, a PlanApo VC 100x/1.4NA oil immersion objective, a stage incubator, an Andor Zyla 4.2 sCMOS or an Andor iXon Ultra DU888 EMCCD camera and MicroManager software. PAsfCherry2 photoactivation in H2B labeling and split mNG2 vs. split GFP comparison in H2B labeling were imaged on a Nikon Ti-E inverted wide-field fluorescence microscope equipped with an LED light source (Excelitas X-Cite XLED1), a 100X NA 1.40 PlanApo oil immersion objective, a motorized stage (ASI) and an sCMOS camera (Hamamatsu Flash 4.0). Microscopy images were subjected to background subtraction using a rolling ball radius of 100 pixels in ImageJ Fiji<sup>39</sup> software. Analysis of conventional fluorescence microscopy images were performed in ImageJ.

**STORM image acquisition and analysis.** Super-resolution images were collected using a TIRF-STORM microscope, home-built from a Nikon Eclipse Ti-E inverted microscope. A 405 nm activation laser (OBIS 405, Coherent), 488 nm imaging laser (OBIS 488, Coherent) and a 561 nm imaging laser (Sapphire 561, Coherent) were aligned, expanded, and focused at the back focal plane of the UPlanSapo 1.4NA 100X oil immersion objective (Olympus). Images were recorded with an electron multiplying CCD camera (iXon + DU897E-C20-BV, Andor), and processed with a home-written software. The OBIS lasers were controlled directly by the computer whereas the Sapphire 561 nm laser was shuttered using an acoustic optical modular (Crystal Technology). A quadband dichroic mirror (ZT405/488/561/640rpc, Chroma) and a band-pass filter (ET525/50 m, Chroma for 488 nm and ET595/50 nm, Chroma for 561 nm) separated the fluorescence emission from the excitation light. Because PAsfCherry2 is initially in a dark state, knock-in cells were carefully located through GFP, using low green intensity (11  $\mu$ W) and a wide-field setting (5 Hz), so as to minimize potential bleaching of PAsfCherry2. Cells were then subjected to STORM imaging, and if fluorophores appeared in the red channel, it was clear that they were indeed expressing both fluorescent proteins. Maximum laser power used during STORM measured before the objective was 16  $\mu$ W for 405 nm and 25 mW for 561 nm. By increasing the power on the 405 nm laser (zero to 16  $\mu$ W) during imaging, PAsfCherry2 was photoconverted from a dark to a red state, and could be observed as single fluorophores. These images were recorded at a frame rate of 30 Hz, with an EMCCD camera gain of 60. During image acquisition, the axial drift of the microscope stage was stabilized by a home-built focus stabilization system utilizing the reflection of an IR laser off the sample. Frames were collected until sample was bleached. Analysis of the STORM images was performed on the Insight3 software<sup>40</sup>. Cells were imaged in PBS buffer.

**Preparation and electroporation of Cas9/sgrRNA RNP.** All synthetic nucleic acid reagents were purchased from Integrated DNA Technologies (IDT). sgrRNAs and Cas9/sgrRNA RNP complexes were prepared using the following procedure<sup>11</sup>. sgrRNAs were obtained by in vitro transcribing DNA templates containing a T7 promoter (TAATACGACTCACTATAG), the gene-specific 20-nt sgrRNA sequence and a common sgrRNA scaffold region. DNA templates were generated by overlapping PCR using a set of 4 primers: 3 primers common to all reactions (forward primer T25: 5'-TAA TAC GAC TCA CTA TAG-3'; reverse primer BS7: 5'-AAA AAA AGC ACC GAC TCG GTG C-3' and reverse primer ML611: 5'-AAA AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT AAA CTT GCT ATG CTG TTT CCA GCA TAG CTC TTA AAC-3') and one gene-specific primer (forward primer 5'-TAA TAC GAC TCA CTA TAG NNN NNN NNN NNN NNN NNG TTT AAG AGC TAT GCT GGA A-3'). For each template, a 100- $\mu$ L PCR was performed with iProof High-Fidelity Master Mix (BioRad) reagents with the addition of 1  $\mu$ M T25, 1  $\mu$ M BS7, 20 nM ML611, and 20 nM gene-specific primer. The PCR product was purified and eluted in 12  $\mu$ L of RNase-free DNA buffer (2 mM Tris pH 8.0 in DEPC-treated water). Next, a 100- $\mu$ L in vitro transcription reaction was performed with 300 ng DNA template and 1000 U of T7 RNA polymerase in buffer containing 40 mM Tris pH 7.9, 20 mM MgCl<sub>2</sub>, 5 mM DTT, 2 mM spermidine and 2 mM of each NTP (New England BioLabs). Following a 4 h incubation at 37 °C, the sgrRNA product was purified and eluted in 15  $\mu$ L of RNase-free RNA buffer (10 mM Tris pH 7.0 in DEPC-treated water). The sgrRNA was quality-checked by running 5 pg of the product on a 10% polyacrylamide gel containing 7 M urea (Novex TBE-Urea gels, Thermo Fisher Scientific).

For the knock-in of mNG2<sub>11</sub>, sfCherry2<sub>11</sub>, or GFP<sub>11</sub>, 200-nt homology-directed recombination (HDR) templates were ordered in single-stranded DNA (ssDNA) form as ultramer oligos (IDT). For knock-in of GFP<sub>11</sub>-sfCherry2<sub>11</sub> in tandem, HDR template was ordered in double-stranded (dsDNA) form as gBlock fragments (IDT) and processed to ssDNA as described below. For the complete set of DNA sequence used for sgrRNA in vitro transcription or HDR templates, see Supplementary Tables 2 and 3.

Cas9 protein (pMJ915 construct, containing two nuclear localization sequences) was expressed in *E. coli* and purified by the University of California Berkeley Macrolab<sup>41</sup>. 293T mNG2<sub>1-10</sub> stable cells or 293T<sup>double1-10</sup> cells were treated with 200 ng/mL nocodazole (Sigma) for ~15 h before electroporation to increase HDR efficiency<sup>42</sup>. Cas9/sgrRNA RNP complexes were assembled with 100 pmol Cas9 protein and 130 pmol sgrRNA just prior to electroporation and combined with HDR template in a final volume of 10  $\mu$ L. Electroporation was carried out in Amaxa 96-well shuttle Nucleofector device (Lonza) using SF-cell line reagents (Lonza). Nocodazole-treated cells were resuspended to 10<sup>4</sup> cells/ $\mu$ L in SF solution immediately prior to electroporation. For each sample, 20  $\mu$ L of cells was added to the 10  $\mu$ L RNP/template mixture. Cells were immediately electroporated using the CM130 program and transferred to 24-well plate with pre-warmed medium. Electroporated cells were cultured for 5–10 days prior to FACS selection of integrated cells.

**Preparation of sfCherry2<sub>11</sub>-GFP<sub>11</sub>-CLTA ssDNA Template.** sfCherry2<sub>11</sub>-GFP<sub>11</sub>-CLTA ssDNA template was prepared from a commercial dsDNA fragment (gBlock, IDT) containing the template sequence preceded by a T7 promoter<sup>11</sup>. The dsDNA fragment was amplified by PCR using Kapa HiFi reagents (Kapa Biosystems) and purified using SPRI beads (AMPure XP resin, Beckman Coulter) at a 1:1 DNA:resin volume ratio (following manufacturer's instructions) and eluted

in 25  $\mu$ L RNase-free water. Next, RNA was produced by in vitro transcription using T7 HiScribe reagents (New England BioLabs). Following a 4 h reaction at 37 °C, the mixture was treated with 4U TURBO DNase (Thermo Fisher Scientific) and incubated for another 15 min at 37 °C. The RNA product was then purified using SPRI beads and eluted in 60  $\mu$ L RNase-free water. DNA:RNA hybrid was then synthesized by reverse transcription using Maxima H RT reagents (Thermo Fisher Scientific). Finally, ssDNA was made by hydrolyzing the RNA strand through the addition of 24  $\mu$ L NaOH solution (0.5 M NaOH + 0.25 M EDTA, in water) and incubation at 95 °C for 10 min. The final ssDNA product was purified using SPRI beads at a 1:1.2 DNA:resin volume ratio and eluted in 15  $\mu$ L water.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon reasonable request. All relevant DNA sequences are listed in the Supplementary Information.

Received: 9 November 2016 Accepted: 30 June 2017

Published online: 29 August 2017

## References

- Kaddoum, L., Magdeleine, E., Waldo, G. S., Joly, E. & Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49**, 727–736 (2010).
- Van Engelenburg, S. B. & Palmer, A. E. Imaging type-III secretion reveals dynamics and spatial segregation of Salmonella effectors. *Nat. Methods* **7**, 325–330 (2010).
- Kamiyama, D. et al. Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* **7**, 11046 (2016).
- Chun, W., Waldo, G. S. & Johnson, G. V. Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3beta activation and caspase 3 cleavage. *J. Neurochem.* **103**, 2529–2539 (2007).
- Schmidt, S. et al. Detecting cytosolic peptide delivery with the GFP complementation assay in the low micromolar range. *Angew. Chem. Int. Ed. Engl.* **54**, 15105–15108 (2015).
- Milech, N. et al. GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci. Rep.* **5**, 18329 (2015).
- Feinberg, E. H. et al. GFP reconstitution across synaptic partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **57**, 353–363 (2008).
- Macpherson, L. J. et al. Dynamic labelling of neural connections in multiple colours by trans-synaptic fluorescence complementation. *Nat. Commun.* **6**, 10024 (2015).
- To, T. L. et al. Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. *Proc. Natl Acad. Sci. USA* **112**, 3338–3343 (2015).
- Kim, Y. E., Kim, Y. N., Kim, J. A., Kim, H. M. & Jung, Y. Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. *Nat. Commun.* **6**, 7134 (2015).
- Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl Acad. Sci. USA* **113**, E3501–3508 (2016).
- Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
- Nguyen, H. B., Hung, L. W., Yeates, T. O., Terwilliger, T. C. & Waldo, G. S. Split green fluorescent protein as a modular binding partner for protein crystallization. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 2513–2523 (2013).
- Cranfill, P. J. et al. Quantitative assessment of fluorescent proteins. *Nat. Methods* **13**, 557–562 (2016).
- Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013).
- Shemiakina, I. I. et al. A monomeric red fluorescent protein with low cytotoxicity. *Nat. Commun.* **3**, 1204 (2012).
- Yu, D. et al. A naturally monomeric infrared fluorescent protein for protein labeling in vivo. *Nat. Methods* **12**, 763–765 (2015).
- Costantini, L. M. et al. A palette of fluorescent proteins optimized for diverse cellular environments. *Nat. Commun.* **6**, 7670 (2015).
- Huang, L., Pike, D., Sleat, D. E., Nanda, V. & Lobel, P. Potential pitfalls and solutions for use of fluorescent fusion proteins to study the lysosome. *PLoS ONE* **9**, e88893 (2014).
- Lippincott-Schwartz, J. & Patterson, G. H. Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol.* **19**, 555–565 (2009).
- Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010).



22. Subach, F. V. et al. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods* **6**, 153–159 (2009).
23. Brodsky, F. M. Diversity of clathrin function: new tricks for an old protein. *Annu. Rev. Cell Dev. Biol.* **28**, 309–336 (2012).
24. English, A. R., Zurek, N. & Voeltz, G. K. Peripheral ER structure and function. *Curr. Opin. Cell Biol.* **21**, 596–602 (2009).
25. Greenfield, J. J. & High, S. The Sec61 complex is located in both the ER and the ER-Golgi intermediate compartment. *J. Cell Sci.* **112**, 1477–1486 (1999).
26. Friedman, J. R., Webster, B. M., Mastronarde, D. N., Verhey, K. J. & Voeltz, G. K. ER sliding dynamics and ER-mitochondrial contacts occur on acetylated microtubules. *J. Cell Biol.* **190**, 363–375 (2010).
27. Nixon-Abell, J. et al. Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. *Science* **354**, aaf3928 (2016).
28. Goedhart, J. et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, 751 (2012).
29. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS ONE* **6**, e28674 (2011).
30. Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl Acad. Sci. USA* **111**, 8452–8457 (2014).
31. Subach, F. V., Patterson, G. H., Renz, M., Lippincott-Schwartz, J. & Verkhusha, V. V. Bright monomeric photoactivatable red fluorescent protein for two-color super-resolution sptPALM of live cells. *J. Am. Chem. Soc.* **132**, 6481–6491 (2010).
32. Shen, Y., Rosendale, M., Campbell, R. E. & Perraiss, D. pHuji, a pH-sensitive red fluorescent protein for imaging of exo- and endocytosis. *J. Cell Biol.* **207**, 419–432 (2014).
33. Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002).
34. Shaner, N. C. et al. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545–551 (2008).
35. Bajar, B. T. et al. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, 20889 (2016).
36. Bindels, D. S. et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).
37. Yamamoto, Y., Yoshida, A., Miyazaki, N., Iwasaki, K. & Sakisaka, T. Arl6IP1 has the ability to shape the mammalian ER membrane in a reticulon-like fashion. *Biochem. J.* **458**, 69–79 (2014).
38. Yu, D. et al. An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat. Commun.* **5**, 3626 (2014).
39. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
40. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
41. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
42. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).

## Acknowledgements

We thank Daichi Kamiyama and David A. Brown in the Bo Huang laboratory (UCSF) for advice on molecular cloning and other extensive discussion, Dan Yu in the Xiaokun Shu laboratory (UCSF) for advice on mutagenesis and *E. coli* library screening, and Joseph DeRisi (UCSF) for generously sharing the nucleofector device with us. This work is supported by the National Institutes of Health Director's New Innovator Award DP2OD008479 (to B.H.) and R21EB022798 (to B.H., S.F., and V.P.), and by the W. M. Keck Foundation Medical Research Grant (to B.H. and S.S.). V.P. is supported by a predoctoral fellowship from the American Heart Association. B.H. is a Chan Zuckerberg Biohub Investigator.

## Author contributions

S.F. and B.H. conceived and designed the research; S.F. performed the screening and protein labeling experiments; S.F., H.L., and M.L. performed the mNG2<sub>11</sub> knock-in and imaging experiment; S.F. and S.S. performed the dual-color knock-in and imaging experiments; S.F. and V.P. performed the STORM experiments and analysis; S.F. analyzed the data; S.F. and B.H. wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-00494-8.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017