

ARTICLE

Received 11 Dec 2015 | Accepted 15 Feb 2016 | Published 18 Mar 2016

DOI: 10.1038/ncomms11046

OPEN

# Versatile protein tagging in cells with split fluorescent protein

Daichi Kamiyama<sup>1,\*</sup>, Sayaka Sekine<sup>1,\*</sup>, Benjamin Barsi-Rhyné<sup>2</sup>, Jeffrey Hu<sup>1</sup>, Baohui Chen<sup>1</sup>, Luke A. Gilbert<sup>3</sup>, Hiroaki Ishikawa<sup>4</sup>, Manuel D. Leonetti<sup>3</sup>, Wallace F. Marshall<sup>4</sup>, Jonathan S. Weissman<sup>3,5</sup> & Bo Huang<sup>1,4</sup>

In addition to the popular method of fluorescent protein fusion, live cell protein imaging has now seen more and more application of epitope tags. The small size of these tags may reduce functional perturbation and enable signal amplification. To address their background issue, we adapt self-complementing split fluorescent proteins as epitope tags for live cell protein labelling. The two tags, GFP11 and sfCherry11 are derived from the eleventh  $\beta$ -strand of super-folder GFP and sfCherry, respectively. The small size of FP11-tags enables a cost-effective and scalable way to insert them into endogenous genomic loci via CRISPR-mediated homology-directed repair. Tandem arrangement FP11-tags allows proportional enhancement of fluorescence signal in tracking intraflagellar transport particles, or reduction of photobleaching for live microtubule imaging. Finally, we show the utility of tandem GFP11-tag in scaffolding protein oligomerization. These experiments illustrate the versatility of FP11-tag as a labelling tool as well as a multimerization-control tool for both imaging and non-imaging applications.

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, USA. <sup>2</sup>Tetrad Graduate Program, University of California, San Francisco, California 94143, USA. <sup>3</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA. <sup>4</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA. <sup>5</sup>Howard Hughes Medical Institute, San Francisco, California 94143, USA. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to B.H. (email: bo.huang@ucsf.edu).

Complementary to direct fluorescent protein fusion, more and more examples of live cell protein labelling and imaging have recently emerged using epitope tags: peptides that do not function by themselves but can be recognized by other intracellularly expressed proteins. An epitope tag fused to the target protein can become an enzyme substrate to be ligated to a small molecule<sup>1</sup>. Alternatively, it can bind an intracellularly expressed antibody/nanobody to bring in functional protein units, such as a fluorescent protein for imaging<sup>2,3</sup> or ubiquitin ligase for protein degradation<sup>4</sup>. This two-component labelling approach has a number of advantages. It is highly modular because the functional units can be easily replaced<sup>2</sup>. A small peptide epitope tag may induce less perturbation than a much larger fluorescent protein. In addition, small epitope tags can be arranged into a multimerization scaffold. For example, the SunTag<sup>2</sup> allows a dramatic enhancement of fluorescence signal by bringing as many as 24 green fluorescent proteins (GFPs) to a single target protein. This arrangement enables it to successfully overcome an intrinsic problem such as the low affinity between the tagged protein and the binding module. Alternatively, the SunTag could have used a higher affinity and a more precise expression stoichiometry between them, so that there is neither incomplete labelling nor background from excessive binding modules.

One approach to relax this affinity and stoichiometry requirement is to have a binding module that becomes fluorescent only when it is on the epitope. In this case, as long as the binding module is expressed in excess of the target protein, the equilibrium can be shifted towards complete labelling. For this purpose, we adopted a split super-folder GFP (sfGFP) that was previously engineered for efficient self-complementation without the assistance of other protein–protein interactions<sup>5</sup>. This split construct breaks the sequence of sfGFP between the tenth and the eleventh  $\beta$ -strand into two parts: GFP1-10 and GFP11, with GFP11 being a short, 16 amino acid peptide (Fig. 1a). The GFP1-10 fragment, which contains the three residues that constitute the GFP chromophore, is nonfluorescent by itself because chromophore maturation requires the conserved E222 residue located on GFP11 (ref. 6). Upon complementation, the reconstituted GFP becomes fluorescent after the chromophore maturation reaction is completed<sup>5,7</sup>. This split GFP system has been previously used for protein quantification<sup>5</sup>, visualization of protein subcellular localization<sup>8–10</sup>, single-molecule imaging<sup>11</sup>, cell-cell contact detection<sup>12</sup>, as well as *in vitro* protein complex assembly<sup>13</sup>. Some of these applications<sup>8–10</sup> fused the small GFP11 fragment to the target protein as an epitope tag. To expand this approach, here we establish 1-10/11 split fluorescent protein constructs (FP11 tags and the corresponding FP1-10 fragments) as a general method for cellular protein visualization including multicolour imaging, especially for the labelling of endogenous protein via genetic knock-in and signal amplification with tandem tags.

## Results

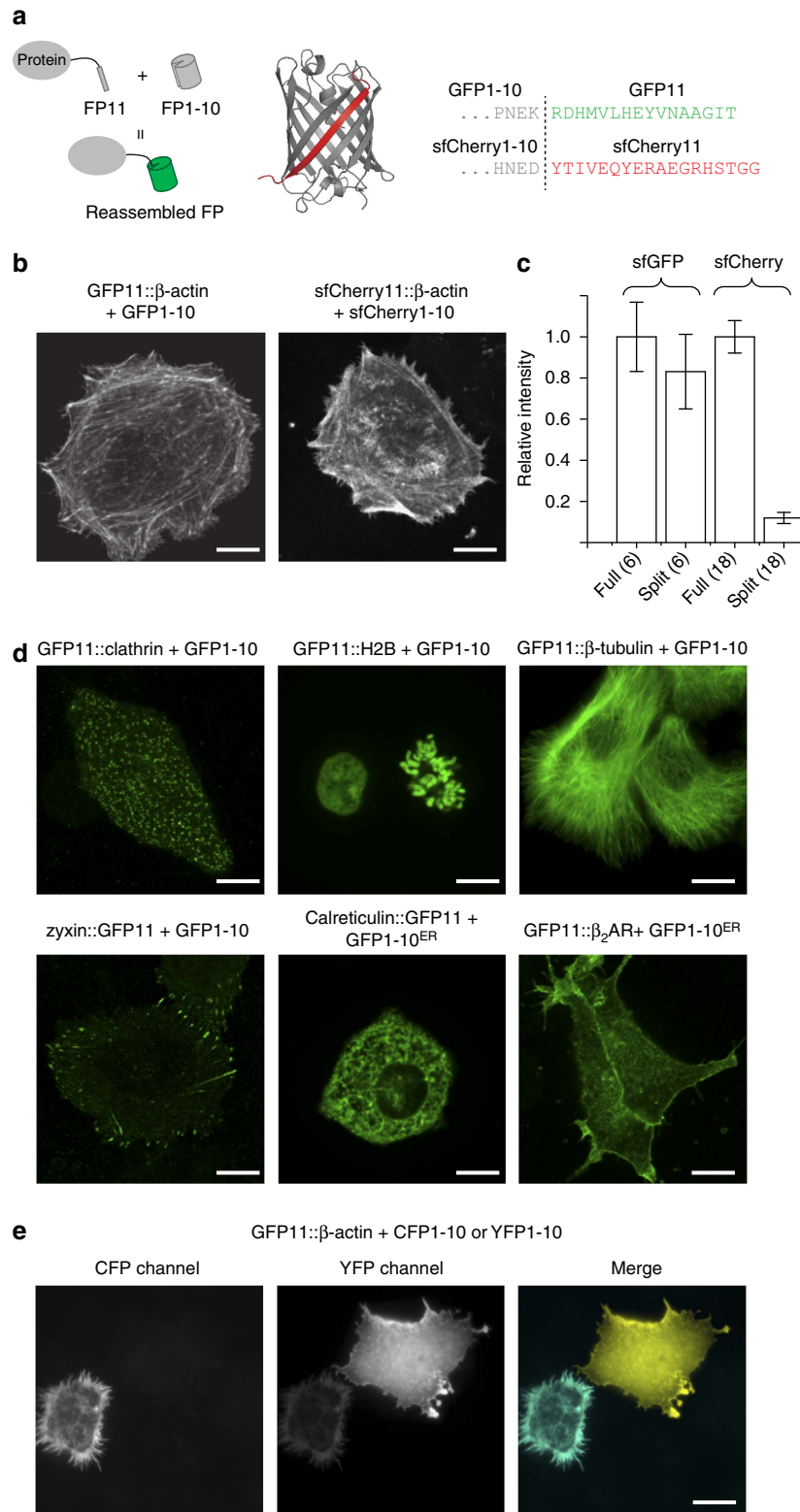
**Live cell protein labelling using GFP11 and sfCherry11.** To validate protein labelling using FP11-tag in mammalian cells, we fused GFP11 to the (amino) N terminus of human  $\beta$ -actin (with an 18 a.a. GSS linker) and co-expressed it with GFP1-10 in HeLa cells. We observed strong actin-like fluorescence signal, whereas GFP11:: $\beta$ -actin or GFP1-10 alone gave no detectable signal (Fig. 1b and Supplementary Fig. 1). We compared the average fluorescence intensity (representing a combination of expression efficiency, complementation efficiency and fluorophore maturation<sup>14</sup>) of  $\beta$ -actin labelled with either full-length sfGFP or GFP11 and expressed using the same vector (Fig. 1c). The difference between the two labelling

approaches was small, indicating that GFP11-tag can label the target protein with a similar efficiency as direct GFP fusion. We then demonstrated that GFP11-tag is compatible with a variety of cellular proteins for live imaging, including cytoskeleton ( $\beta$ -actin and  $\beta$ -tubulin), clathrin-coated pits (clathrin light chain), histone (H2B) and focal adhesion (zyxin) (Fig. 1d). By adding an endoplasmic reticulum (ER) signal peptide to GFP1-10 (GFP1-10<sup>ER</sup>), we were able to specifically label ER luminal protein (calreticulin) or the extracellular domain of plasma membrane protein ( $\beta_2$  adrenergic receptor; Fig. 1d). In addition, we showed that the Y66W mutation for cyan fluorescent protein (CFP) and the T203Y mutation for yellow fluorescent protein (YFP)<sup>15,16</sup> can be successfully introduced into GFP1-10 to label proteins with the corresponding colour (Fig. 1e).

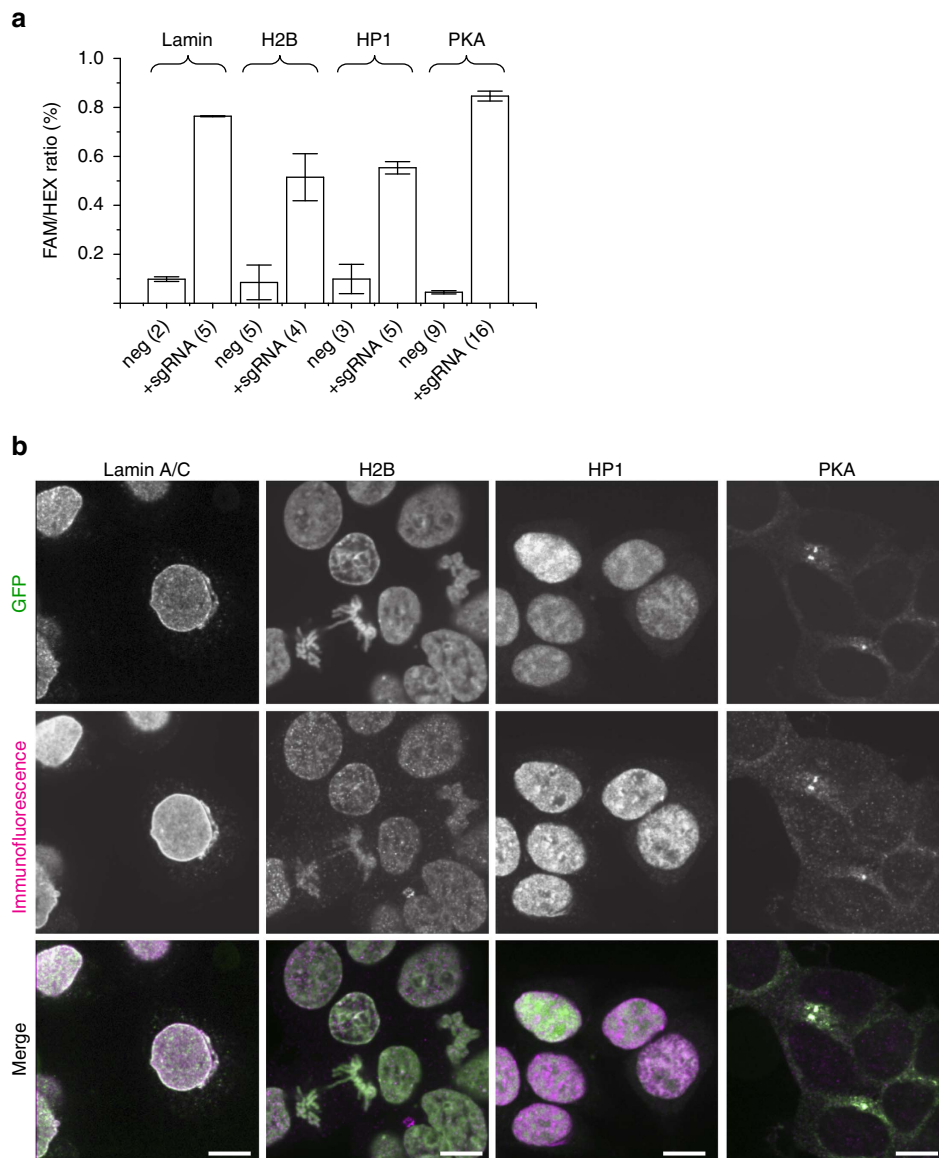
To enable multicolour imaging, we tested splitting mCherry at the same site of the GFP1-10/11 system. However, no fluorescence complementation signal was observed when co-expressing mCherry:: $\beta$ -actin and mCherry1-10 (data not shown), which is consistent with previous characterization of split-mCherry constructs<sup>17</sup>. Instead, we were able to obtain a functional split from superfolder Cherry (sfCherry, derived from mCherry for potentially better folding<sup>18</sup>), resulting in an 18 amino-acid peptide for sfCherry11 (Fig. 1a). We verified that sfCherry11:: $\beta$ -actin can be fluorescently imaged when co-expressing sfCherry1-10 (Fig. 1b). Nevertheless, its overall fluorescence signal level is substantially weaker than that of direct sfCherry fusion (Fig. 1c), potentially caused by the less-efficient self-complementation between sfCherry11 and sfCherry1-10. This problem could be solved by further optimization of the split-construct<sup>1</sup> or by tandem tagging which we will describe later.

**Labelling endogenous proteins with FP11 tag.** A first unique application of FP11 epitope tag is to generate libraries of fluorescently labelled endogenous proteins via genetic knock-in by homology-directed DNA repair. Previously, such a library created for *Saccharomyces cerevisiae* has demonstrated its wide application in the systematic investigation of protein localization, gene expression and cellular behaviour<sup>19,20</sup>. Because the DNA sequence for GFP11 is only 57 nt (including a 3 a.a. linker), we could use very short homology arms (70 nt on either side, plus the start or stop codon for N- or carboxy (C)-terminal labelling, respectively), so that the entire donor DNA can be a 200 nt single-strand oligo-DNA (ssDNA). Unlike the long donor DNA for full-length GFP, which must be made through multiple cloning steps, GFP11 donor ssDNA can be directly synthesized, thus making it cost-effective for library generation. As a demonstration, we used synthetic ssDNA donors to knock GFP11-tag into four genes, *LMNA*, *HIST2H2BE*, *CBX1* and *PRKACA*, which encode lamin A/C, histone H2B (H2B), heterochromatin protein 1 homologue (HP1) and cAMP-dependent protein kinase catalytic subunit  $\alpha$  (PKA), respectively. Using the Cas9/sgRNA plasmid transfection method<sup>21</sup>, we have obtained a knock-in efficiency in the range of 0.5~0.8% (negative controls: 0.01~0.09%) (Fig. 2a and Supplementary Fig. 2). Other established knock-in approaches, such as by direct nucleofection of the Cas9/sgRNA complex<sup>22</sup>, may yield higher knock-in efficiencies. We imaged these cells and validated the specificity by immunofluorescence (Fig. 2b). This approach can be readily extended to other mammalian proteins or even animal systems given the recent advancements in CRISPR-mediate genome engineering technologies.

**Signal amplification using tandem FP11 tags.** With their small sizes, we can arrange FP11-tags into tandem arrays to amplify the fluorescence signal (Fig. 3a), thus addressing the issues of poor



**Figure 1 | Cellular protein labelling with FP11-tag.** (a) Schematic diagram for FP11-Tag, illustrated on the crystal structure of sfCherry, and the split schemes. (b) Images of HeLa cells co-expressing GFP1-10 and GFP11::β-actin or sfCherry 1-10 and sfCherry 11::β-actin. (c) Average fluorescence intensity of whole cells expressing β-actin labelled with full-length sfGFP, sfCherry or the corresponding FP11 tags.  $n = 6-18$ . Error bars are s.e.m. (d) Fluorescence images of *Drosophila* S2 cells expressing GFP11-tagged of β-tubulin and calreticulin, and HeLa cells expressing GFP11-tagged clathrin light chain, histone H2B, zyxin and β<sub>2</sub>AR. (e) An image of mixed S2 cells expressing either CFP1-10 + GFP11::β-actin or YFP1-10 + GFP11::β-actin. The weak fluorescence of the CFP1-10-expressing cell in the YFP channel is due to the bleed-through of CFP emission. All the scale bars indicate 5 μm.



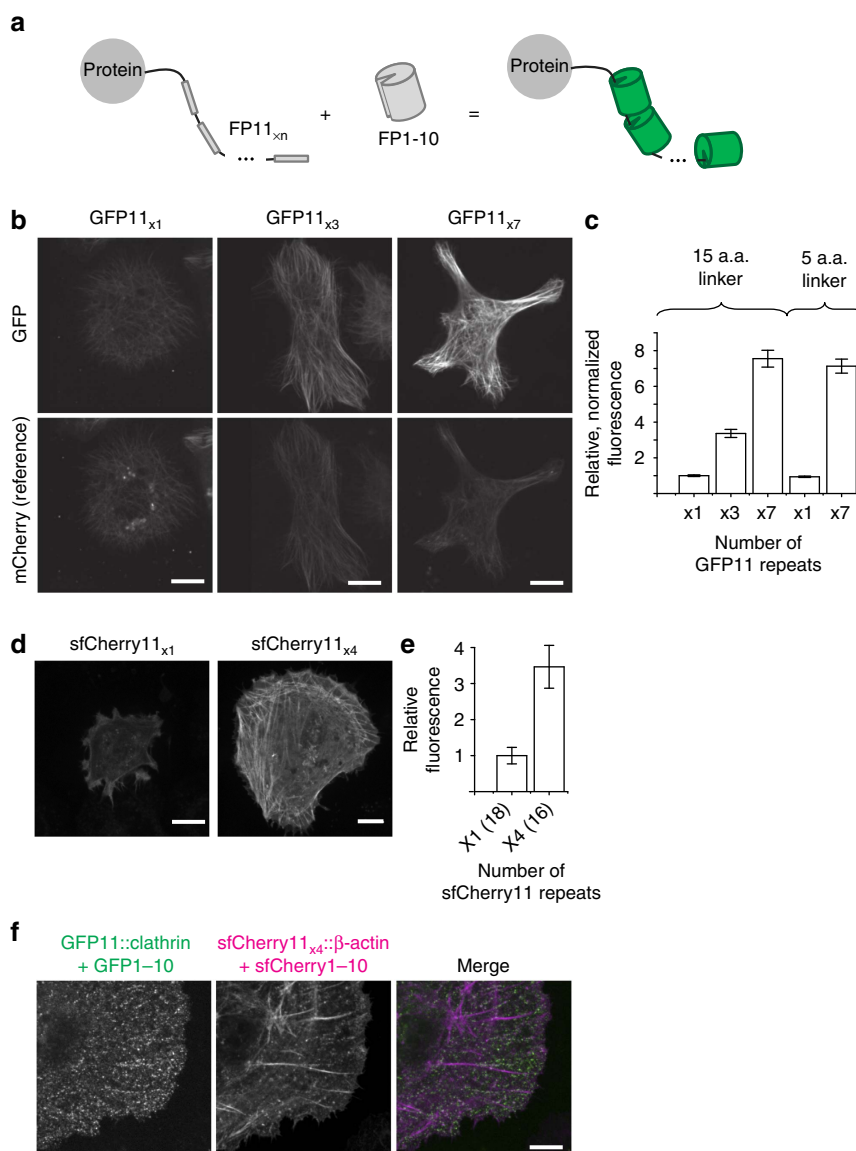
**Figure 2 | Labelling endogenous proteins using GFP11-tag. (a)** GFP11 knock-in efficiencies by co-transfection with Cas9/sgRNA expression plasmids and donor templates, quantified by the combined TaqMan PCR/ddPCR assay (see Supplementary Fig. 2). **(b)** GFP fluorescence and immunofluorescence images of knock-in cells. All the error bars are s.e.m. All the scale bars indicate 5  $\mu$ m. neg, negative.

signal or photobleaching that live cell imaging often suffers from<sup>23</sup>. To design tandem GFP11<sub>x3</sub>, GFP11<sub>x7</sub> and sfCherry11<sub>x4</sub> tags, we purposely used synonymous codons to avoid deleterious recombination during cloning, which can be caused by repetitive nucleic acid sequences. For GFP11<sub>x7</sub>, we tested either 5 a.a. or 15 a.a. linker length between the repeats. Using a fused mCherry as the reference to normalize expression level differences, we observed that the fluorescence signal for GFP11<sub>x3</sub>::mCherry:: $\beta$ -tubulin and GFP11<sub>x7</sub>::mCherry:: $\beta$ -tubulin increased proportionally compared with GFP11::mCherry:: $\beta$ -tubulin without the tandem repeat (Fig. 3b,c). No significant signal difference was recorded for GFP11<sub>x7</sub> using either the 5 a.a. or 15 a.a. linker (Fig. 3c), indicating a negligible interference between reconstituted GFPs in the tandem tag. We also observed a proportional enhancement of overall cellular fluorescence comparing sfCherry11<sub>x4</sub>:: $\beta$ -actin with sfCherry11:: $\beta$ -actin (Fig. 3d,e). Combining sfCherry11<sub>x4</sub>-tagged  $\beta$ -actin with GFP11-tagged clathrin light chain, we showed two colour imaging using FP11 tags (Fig. 3f). Compared with SunTag<sup>2</sup>, tandem FP11 tags have the advantage that FP1-10 can be

overexpressed without causing a fluorescence background, thus allowing a much simpler expression control.

The fluorescence amplification by tandem FP11-tag can greatly benefit imaging experiments that are difficult because of low signal, such as single particle tracking of intraflagellar transport (IFT) in primary cilia<sup>24</sup>. We fused GFP11<sub>x7</sub> to the C terminus of IFT20, which moves along ciliary microtubules. Compared with the image of IFT20::GFP, IFT20::GFP11<sub>x7</sub> generated a much higher signal from IFT particles in the primary cilium of mouse IMCD3 cells (Fig. 4a, Supplementary Fig. 3 and Supplementary Movie 1). Moreover, analysis of the kymograph showed no significant difference between the two labelling methods for either anterograde or retrograde transport speed (Fig. 4a,b), suggesting that IFT20 labelling using tandem GFP11-tag did not perturb its function.

Another advantage of a brighter label is to reduce photobleaching and phototoxicity by lowering excitation intensity while maintaining the same signal level. To demonstrate this application, we imaged *Drosophila* S2 cells expressing

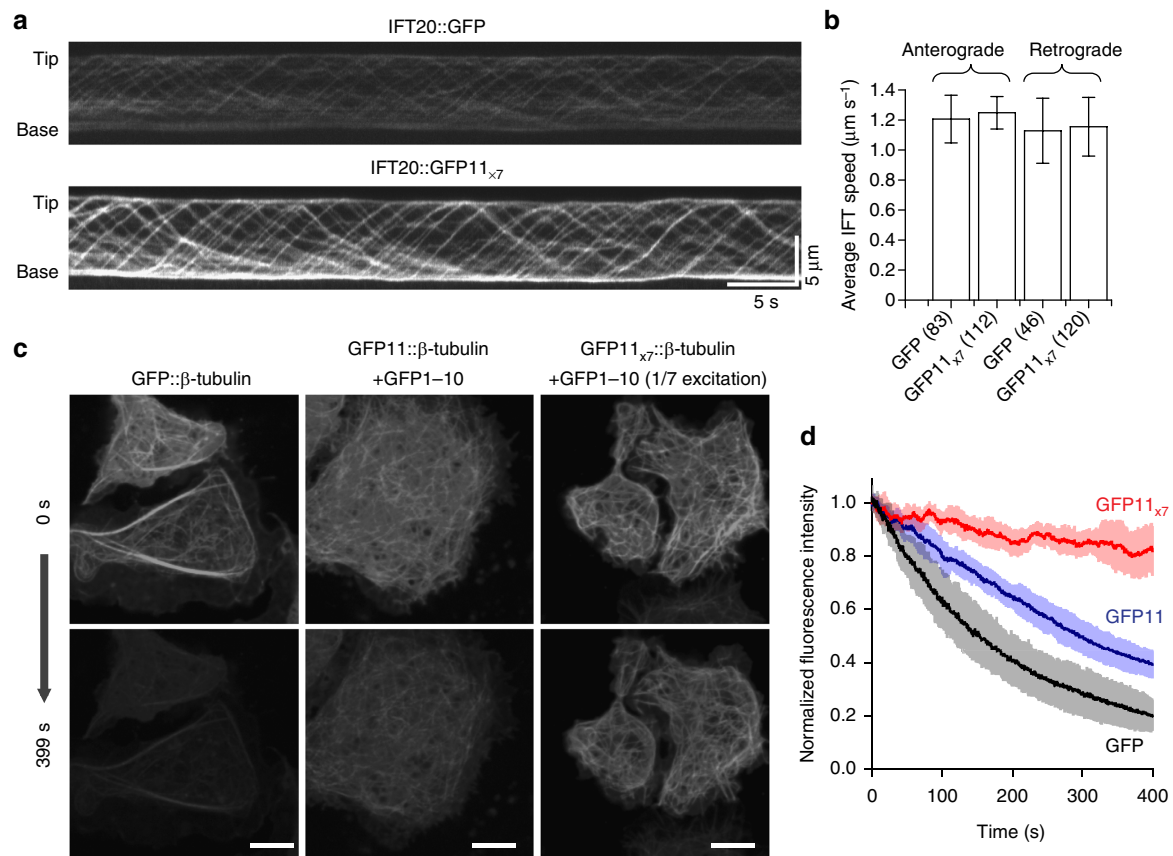


**Figure 3 | Amplification of fluorescence signal by tandem FP11.** (a) Schematic of fluorescence signal amplification by tandem FP11 (b) Images of *Drosophila* S2 cells expressing GFP11<sub>x1</sub>, x3 or x7::mCherry::β-tubulin together with overexpressed GFP1-10. The three panel sets were acquired and displayed with identical settings. (c) Quantification of GFP to mCherry fluorescence intensity ratio with different number of GFP11 repeats and different linker lengths between the repeats. Fifty cells were analysed in each case. Error bars are s.e.m. (d) Images of HeLa cells expressing sfCherry11<sub>x1</sub> or x4::β-actin together with overexpressed sfCherry1-10. (e) Average whole-cell fluorescence intensity with sfCherry11<sub>x1</sub> or x4::β-actin. *n* as indicated in the figure. Error bars are s.e.m. (f) Two-colour imaging of GFP11-labelled clathrin light chain and sfCherry11<sub>x4</sub> labelled β-actin. All the scale bars indicate 5 μm.

GFP::β-tubulin, GFP11::β-tubulin or GFP11<sub>x7</sub>::β-tubulin by spinning disk confocal microscopy (Fig. 4c and Supplementary Movie 2). With one-seventh the excitation laser power, the GFP11<sub>x7</sub> sample generated a similar level of fluorescence signal compared with either GFP or GFP11, but its photobleaching rate was approximately nine times as slow (Fig. 4d). Surprisingly, even the non-repeating GFP11 displayed approximately three times as slow a photobleaching rate as that of direct GFP fusion. A possible explanation for this effect is that the complementation between GFP11 and GFP1-10 can be reversed, allowing photobleached GFP1-10 to be exchanged for an intact one in the cytoplasm, although such exchange may be limited by the maturation process of the GFP chromophore.

**Tandem FP11 tags as multimerization scaffolds.** Tandem FP11-tag can also be used in non-imaging applications<sup>13</sup> as a

scaffold to assemble multiple copies of proteins for synergistic functioning. Previously, we have developed a CRISPR/Cas9-based transcriptional activation system (CRISPRa)<sup>25</sup>, which activates a specific gene by targeting it with a nuclease-deactivated Cas9 (dCas9) fused to the transcription activator VP64. We have then observed that the activation of natively suppressed gene by CRISPRa often require multiple copies of transcription activation VP64 domains at the target<sup>2,26–28</sup>. Here, we showed that tandem GFP11-tag could be used to achieve this goal. We chose to target the C-X-C chemokine receptor type 4 gene *CXCR4*, which is poorly expressed in K265 cells. We fused VP64 to GFP1-10, and then coexpressed it with dCas9::GFP11<sub>x7</sub> and a *CXCR4* targeting sgRNA<sup>26</sup> in K265 cells (Fig. 5a). We immunostained these cells for *CXCR4* and quantified its expression level by flow cytometry (Fig. 5b). A 45 ± 2-fold signal increase was detected compared with the controls, whereas no significant activation can be observed either with a non-targeting sgRNA or with direct



**Figure 4 | Live cell imaging using tandem FP11 tag.** (a) Tracking of IFT particles in mouse IMCD3 cells expressing either IFT20::GFP or IFT20::GFP11<sub>x7</sub> + GFP1-10. Kymographs show both retrograde and anterograde transport (see Supplementary Fig. 3 and Supplementary Movie 1). (b) Comparison of IFT20 movement speed for IFT20::GFP and IFT20::GFP11<sub>x7</sub>. The number of particles that were analysed for each case is indicated in the figure. Error bars are standard deviations. (c) Comparison of the photobleaching rate in imaging  $\beta$ -tubulin labelled with GFP, GFP11 and GFP11<sub>x7</sub>, showing snapshots of S2 cells expressing the three different constructs at the beginning and the end of a 400 s movie. GFP11<sub>x7</sub> samples were imaged with one-seventh the excitation laser power. Scale bars indicate 5  $\mu\text{m}$ . (d) Fluorescence photobleaching time traces. Five cells were averaged in each condition. The error bars are standard deviations.

dCas9::VP64 fusion. Immunofluorescence images confirmed CXCR4 expression in these cells (Fig. 5c). In conjugation with SunTag, tandem GFP11 can serve as an orthogonal multimerization tag multiplexed applications. We also observed more robust expression of GFP1-10::VP64, potentially due to its better solubility than scFv::VP64 in SunTag<sup>2,26</sup>.

## Discussion

The experiments above showcased the versatility of FP11 as an epitope tag for protein labelling, library generation and signal amplification, in both imaging and non-imaging applications. The functionality of the FP11 tags can be further expanded. For example, we have shown that the point mutations converting GFP to CFP and YFP can be successfully introduced to GFP1-10. Similarly, other functional mutants of GFP, such as blue fluorescent protein (BFP)<sup>15</sup>, photoactivatable GFP (PA-GFP)<sup>29</sup> and the pH-sensitive pHluorin<sup>30</sup>, may possibly be adaptable for FP11 labelling, too. Of course, as in any protein labelling method, validating the proper functionality of the labelled protein is always essential, especially when tandem GFP11-tag is used. In our case, GFP11<sub>x7</sub> was well tolerated by multiple proteins including IFT20 and Cas9. Therefore, we expect a broad applicability for tandem GFP11-tag in helping live cell imaging to overcome the issues of poor signal, photobleaching and phototoxicity. We note that when combined with genetic

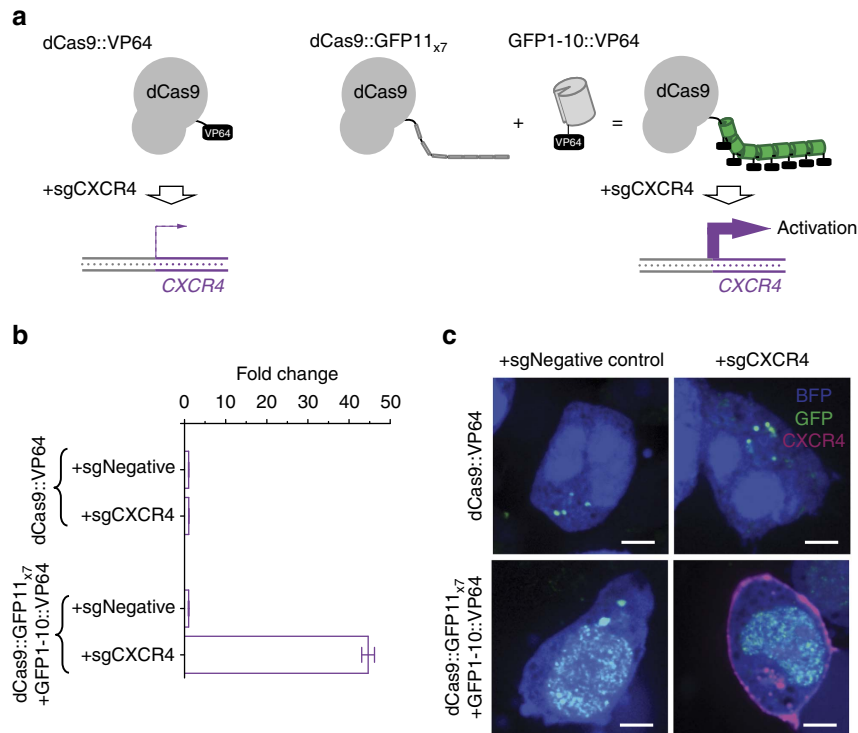
knock-in, the tandem FP11 tag could help the visualization of proteins with low endogenous expression levels. However, knocking in tandem FP11 tags require longer donor ssDNAs that contain the larger insert and potentially longer homology arms. While such long ssDNAs cannot be directly synthesized, they can be produced by reverse transcription methods<sup>31</sup> to avoid the involvement of cloning.

## Methods

**Molecular cloning.** The amino-acid sequence of super-folder mCherry (sfCherry) was obtained from the published literature<sup>18</sup>. sfCherry was split between 208D and 209Y at the middle of the loop between  $\beta$ -strands 10 and 11 (Fig. 1a). For the nucleotide sequence of sfCherry1-10 and sfCherry11, see Supplementary Tables 1 and 2.

The DNAs of GFP11 (ref. 5) and sfCherry11 were directly synthesized. The DNAs of H2B, zyxin, clathrin light chain,  $\beta$ -actin were subcloned from the corresponding sfGFP or mEmerald fusion plasmids (cDNA source: the Michael Davidson Fluorescent Protein Collection at the UCSF Nikon Imaging Center). 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 BglII (sfGFP-H2B C-10); zyxin (6 a.a.): sfGFP sequence between BamHI and NotI (sfGFP-zyxin-6); clathrin light chain (15 a.a.): mEmerald sequence between NheI and BglII (mEmerald-clathrin-15);  $\beta$ -actin (18 a.a.): mEmerald sequence between AgeI and BglII (mEmerald-actin-C-18). PCR-amplified GFP11 as well as sfCherry11 fragments were then inserted into the digested vectors using In-Fusion assembly (Life Technologies).

For the cloning of GFP11-tagged *Drosophila* calreticulin, GFP11 was inserted at the nineteenth a.a. position where the signal peptide ends. GFP11::calreticulin was inserted into *Drosophila* expression pACUH vectors (source: Yuh-Nung Jan) at the EcoRI/XbaI sites. For the cloning of GFP11-tagged human  $\beta_2$ AR, the signalling



**Figure 5 | Controlled protein multimerization using GFP11-tag.** (a) Schematic of CXCR4 gene activation by either dCas9::VP64 or dCas9::GFP11<sub>x7</sub> + GFP1-10::VP64. (b) Plot of CXCR4 fluorescence signals measured by flow cytometry. Only green fluorescence positive cells were analysed, which indicated the expression of both dCas9::GFP11<sub>x7</sub> and GFP1-10::VP64. Error bars are s.e.m. (c) Images of K256 cells infected by lentivirus expressing the two dCas9 constructs with either a targeting or a non-targeting sgRNA. The cells were stained with a fluorescently labelled anti-CXCR4 antibody as an indicator of CXCR4 expression. All the scale bars indicate 5  $\mu\text{m}$ .

peptide (5HT3R: MRLCIPQVLLALFLSMLTGPGECS),  $\beta_2\text{AR}$  and GFP11 were synthesized and cloned into pcDNA3.1 vectors at the BamHI and XhoI sites.

For the expression of GFP1-10, GFP1-10 was synthesized and cloned into pACUH vectors, pcDNA3.1 vectors as well as modified lentiviral pHR-SFFV vector<sup>32</sup>. For GFP1-10 sequence information, see Supplementary Table 1. For the expression of *sfCherry1-10*, *sfCherry1-10* was synthesized and cloned into pcDNA3.1 vectors.

To generate an ER-localized GFP1-10 (GFP1-10<sup>ER</sup>), a signal peptide (5'-ATG ATGTGGTG CAAAACAGTGATAGTGTGCTGGCGACAGTCGCTTTATTA GTGCC-3') and an ER retention sequence (5'-AGCGAACACGACGAATTG-3') were fused to the N-terminal and the C-terminal of GFP1-10, respectively. The SP::GFP1-10::SEHDEL fragment was cloned into pACUH vectors. For mammalian expression, we cloned SP::GFP1-10 into pcDNA3.1 vectors.

For CFP and YFP imaging, we introduced point mutations in GFP1-10. It has been previously reported that the introduced point mutations Y66W and T203Y in GFP alter the GFP spectral properties to CFP and YFP, respectively. We generated the corresponding mutations in GFP1-10 with the Q5 site-Directed Mutagenesis Kit (NEB). The primers were designed using the NEB online primer design software (NEBaseChanger). The used primers were as follows: YFP\_forward (5'-CTACCTCTCAtatCAAACAGTCTGAGCAAAGATC-3'), YFP\_reverse (5'-TGATTATCAGGAAGAAGTACC-3'), CFP\_forward (5'-ACGCTTACGTgg GGAGTTCAGTGC-3'), and CFP\_reverse (5'-TGTTACGAGAGTCGGCCA-3'). We fused these colour variants to the N terminus of *Drosophila*  $\beta$ -actin and cloned them into pACUH vectors.

To prepare GFP11 repeat arrays constructs, we first synthesized GFP11<sub>x1, x3</sub> or x<sub>7</sub> with long (15 a.a.) or short (5 a.a.) linkers (see the sequence information in Supplementary Table 2). Then, these GFP11 fragments were fused to  $\beta$ -tubulin, mouse IFT20 and dCas9. Specifically, for the cloning of GFP11<sub>x1, x3</sub> or x<sub>7</sub>: *mCherry:: $\beta$ -tubulin*, *Drosophila*  $\beta$ -tubulin (cDNA source: R.Vale) was cloned into pACUH vectors. We then inserted *mCherry*-fused GFP11<sub>x1, x3</sub> or x<sub>7</sub> at the EcoRI site. For the cloning of GFP11<sub>x1</sub> or x<sub>7</sub>: *Ift20*, we first fused *Ift20* with GFP11<sub>x1</sub> or x<sub>7</sub>, and then we inserted it into the EcoRI/NotI sites of pEGFP-N1 vectors. For the cloning dCas9::GFP11<sub>x7</sub>, we modified our SunTag vector (pHRdSV40-NLS-dCas9-24xGCN4\_v4-NLS-P2A-BFP-dWPRE (addgene #60910)) (ref. 2). 24xGCN4 was cut out by BamHI and NotI, and GFP11<sub>x7</sub> was inserted into these sites.

For the cloning of *sfCherry:: $\beta$ -actin*, we synthesized full-length *sfCherry* and cloned it into the vector containing  $\beta$ -actin fragment (see above) at the BglIII/NheI site. To construct *sfCherry11<sub>x4</sub>:: $\beta$ -actin*, we also synthesized *sfCherry11<sub>x4</sub>*, and replaced the *sfCherry* fragment to the tandem one at the BglIII/NheI site. For the sequence information of *sfCherry11<sub>x4</sub>*, see Supplementary Table 2.

For our CRISPRa assay in Fig. 5, we fused an NLS sequence to the C-terminal of GFP1-10::VP64 and cloned into modified pHR vectors. For the GFP1-10::VP64::NLS sequence information, see Supplementary Table 3.

**Cell culture.** The cell lines used for imaging were human HeLa (UCSF cell culture facility), HEK293FT (gifted from Bruce Conklin) and K562 cells (UCSF cell culture facility), mouse IMCD 3 cells (UCSF cell culture facility) and *Drosophila* S2 cells (UCSF cell culture facility). HeLa and HEK293FT cells were grown in a Dulbecco's modified Eagle's medium with 10% FCS (UCSF Cell Culture Facility), 100 units ml<sup>-1</sup> streptomycin, 100  $\mu\text{g ml}^{-1}$  penicillin and 2 mM glutamine. K562 cells were grown in RPMI-1640 with 25 mM HEPES and 2.0 g l<sup>-1</sup> NaHCO<sub>3</sub> in 10% fetal bovine serum (FBS), 100 units ml<sup>-1</sup> streptomycin, 100  $\mu\text{g ml}^{-1}$  penicillin and 2 mM glutamine. Mouse IMCD3 cells were maintained in a mixture of DMEM and Ham's F12 medium (1:1 vol/vol) with 10% FBS. *Drosophila* S2 cells were grown in Schneider's *Drosophila* Medium with 10% FBS (Life Technologies).

**Generation of stable cell lines and lentiviral infection.** For viral production, HEK293T were plated in T25 flasks. Twenty-four hours later, the 3  $\mu\text{g}$  of lentiviral vector pHR-SFFV-GFP1-10, 3  $\mu\text{g}$  of lentiviral packaging plasmid pCMV-dR8.91 and 0.3  $\mu\text{g}$  of envelop plasmid pMD2.G were transfected into the cells using FuGENE HD (Promega) following the manufacturer's protocol. Twenty-four hours after transfection, the culture medium was replaced. Fifty-six hours after transfection, the lentiviral supernatant was harvested and kept at -80  $^{\circ}\text{C}$  for later use.

HEK293FT cells stably expressing GFP1-10 were generated by infecting with lentivirus diluted 1:10 in DMEM medium and incubating for 24 h in the medium. We further plated the HEK293FT cells expressing GFP1-10 at one cell per well in a 96-well plate and isolated monoclonal cell lines that showed modest expression.

To construct stable cell lines and measure CRISPRa activity, K562 cells were lentivirally transduced with constructs that express either dCas9::GFP11<sub>x7</sub> and GFP1-10::VP64 or dCas9::VP64::BFP. The cells were additionally transduced with negative control and CXCR4 targeting sgRNA expression constructs<sup>2</sup>.

For IFT imaging, the plasmids encoding IFT20::GFP (ref. 33) and IFT20::GFP11<sub>x7</sub> were transfected into IMCD3 cells using Lipofectamine 2000 (Life Technologies) and then selected with Geneticin (400  $\mu\text{g ml}^{-1}$ ; Life Technologies). After isolating a stable monoclonal cell line, the cells were infected by the GFP1-10 virus-particle-containing medium.

**Transient expression.** Mammalian expression plasmids (200  $\mu\text{g}$  of each construct per well) were transfected using Lipofectamine 2000 (Life Technologies) into cultures

of HeLa or HEK293FT cells grown on a 48-well plate (Eppendorf). *Drosophila* expression plasmids (100 µg of each construct per well) were transfected using Effectene (QIAGEN) into cultures of S2 cells grown on a 12-well plate (Eppendorf). Note that in Fig. 3 and Supplementary Movie 2, *GFP1-10* and *GFP11<sub>x7</sub>::mCherry::β-tubulin* plasmids were transfected into S2 cells at a concentration ratio of 7:1. Forty-eight hours after transfection, the cells were live-imaged or fixed with 4% paraformaldehyde for later imaging. To facilitate cilia generation, IMCD3 cells were cultured in serum-free media 1 day before imaging. For two-colour imaging in Fig. 3f, HeLa cells were infected with *GFP1-10* lentivirus and, 24 h later, 50 ng of *Clathrin light chain::GFP11*, 100 ng of *sfCherry1-10* and 100 ng of *sfCherry11<sub>x4</sub>::β-actin* were transfected using Lipofectamine 2000 (Life Technologies).

**Imaging.** The cells were grown in 96-well glass bottom plates with #1.5 high-performance cover glass (*In Vitro* Scientific) coated with either Poly-L-Lysine (Sigma-Aldrich) or Fibronectine (Roche) and were imaged on an inverted Nikon Ti-E microscope, Yokogawa CSU-22 confocal scanner unit, × 60/1.4 NA oil objective, an Andor EM-CCD camera (iXon DU897) and Micro-Manager software. Images in Fig. 1e were taken using an inverted Nikon Ti-E microscope with wide-field illumination, × 100/1.4 NA oil objective, and a sCMOS camera (Hamamatsu Flash 4.0). For IFT20 time-lapse imaging in Fig. 4, the cells were grown in DMEM and Ham's F12 (1:1 vol:vol) medium without phenol red, and were imaged in a 37 °C chamber on a Nikon Ti-E microscope with Laser TIRF system, an Andor DU897 EMCCD camera and an Apo TIRF × 100/1.49 NA oil objective. All the imaging experiments were performed at UCSF Nikon Image Center.

**Image analysis.** To measure the brightness of full-length FPs and split-FPs in Figs 1c and 3e, we took the images of β-actin-fused full-length FPs or split-FPs. These mCherry, sfCherry or GFP images were taken with × 10 0.3 NA or × 60 1.4 NA objectives, respectively. The acquired images were first background-subtracted; each of an average value of a 10-µm diameter region of interest (ROI) outside a cell was applied for the background subtraction. Then, the mean within a 10-µm diameter ROI inside a cell was calculated for each cell. A total 6~18 cells were measured for each condition.

To characterize the tandem GFP11-tag, we first set the excitation intensity and exposure time using a *GFP11<sub>x1</sub>*-linker-mCherry fusion protein (*GFP11<sub>x1</sub>::mCherry*) so that the camera counts for GFP and mCherry were in the same range. We then imaged the *GFP11<sub>x1</sub>*, *x3* and *x7*-labelled β-tubulin samples using these parameters. To quantify the ratio of signal enhancement, we acquired images on a wide-field microscope using a × 10/0.3 NA objective to achieve a larger field of view. The acquired images were first background-subtracted and then binarized as '0' for outside and '1' for inside the cells. The GFP-to-mCherry signal ratio was calculated for each pixel within the mask, and the mean ratio within a 10-µm diameter ROI was computed for each cell. A total 50 cells were measured for each condition.

To trace the fluorescent intensities for live imaging of microtubules in Fig. 4, we used Plot z-axis Profile function in ImageJ. For each photobleaching experiment, we calculated the average fluorescence intensity in a 3 µm × 3 µm square ROI in one cell over the duration of the experiment, subtracted the background measured from a cell-free region in the same field of view, and normalized the trajectory to the average value of the first 10 time points. Figure 4d reported the mean normalized photobleaching trajectories from five experiments for each construct (*GFP*, *GFP11* and *GFP11<sub>x7</sub>*). For the photobleaching, rates were calculated from the slope of a linear fit to the initial linear segment of the decay curves.

IFT20 moving velocities were measured by kymographs analysis using ImageJ. In brief, we drew a segmented line along a cilium in an image sequence and then created a kymograph using the KymoResliceWide plugin. The velocity of IFP particles were determined by drawing straight lines along IFT trajectories in the kymograph and then measuring the angle of the lines. A detailed particle tracking protocol was described in a previous publication<sup>34</sup>.

**Immunocytochemistry.** Mouse monoclonal antibodies used were anti-GFP ([1:100] Sigma-Aldrich, G6539), acetylated tubulin ([1:1,000] Sigma-Aldrich, T7451) and anti-histone H2B ([1:50] Abcam, ab52484) antibodies. Rabbit polyclonal antibodies used were anti-γ-tubulin ([1:1,000] Sigma-Aldrich, T5192), anti-lamin A/C ([1:20] Santa Cruz Biotechnology, H110), anti-cAMP protein kinase catalytic subunit ([1:1,000] Abcam, ab26322) and anti-CBX/HP1 beta antibodies ([1:100] Abcam, ab10478). TRITC or Alexa Fluor 647 conjugated secondary antibodies (Jackson Immuno Research Laboratories, Inc. or Life Technologies, respectively) were used for indirect immunofluorescence detection. BG-conjugated dyes, including SNAP-Surface Alexa Fluor 647 and 546 (NEB), were used for staining SNAP-tag expressed in cells.

The cells were fixed with 4% formaldehyde (Electron Microscopy Sciences) for 15 min, permeabilized with 0.1% Triton X-100 in phosphate-buffered saline (PBS), stained for primary antibodies and Alexa-Fluor-conjugated secondary antibodies, or BG-conjugated fluorophores for SNAP-tag at 4 °C overnight.

**Design of sgRNAs for CRISPR-mediated knock-in.** We searched the human genome for CRISPR target sites (GN<sub>20</sub>GG). We used the px330 vector provided through AddGene to generate the sgRNA and Cas9 expression construct. For the individual target sequences, see Supplementary Table 4.

**Design of TaqMan PCR primers and probes.** For the design of TaqMan PCR primers and probes, we used the PrimerQuest software (Integrated DNA Technologies) and the Primer-BLAST (NCBI). For the probes, the PrimerQuest default setting gave us optical probe sequences, whereas the primers were identified by the Primer-BLAST. We designed that one of the primers would anneal inside of the donor sequence but not another. The probes and primers were purchased from Integrated DNA Technologies. For sequence of the primers and probes, see Supplementary Table 6.

**Design of homologous recombination templates.** To prepare *EGFP-LMNA* homologous recombination templates (70 or 200 bp homology arms), we synthesized two DNA fragments (*LMNA* 5' regulatory region + one half of a *GFP* coding sequence (1–350 bp), and another half of a *GFP* coding sequence (350–717 bp) + *LMNA* coding region). Subsequently, they were cloned into the pcDNA3.1 vector in which a CMV promoter sequence was removed.

ssDNA donors (200 nt in total length) were manually designed by adding 70 nt homology sequences to both side of the *GFP11* sequence. ssDNA donors were purchased from Integrated DNA Technologies. For the sequence of oligo-nucleotide donors, see Supplementary Table 5.

**CRISPR-mediated knock-in using plasmid transfection.** For the knock-in experiments, 200 ng of Cas9 + sgRNA vector and 400 ng of an oligonucleotide donor DNA were transfected to HEK293FT cells stably expressing *GFP1-10* per 24-well plate (Eppendorf). Genomic DNA was extracted from cells 3 days after transfection with the DNAzol reagent (Molecular Research Center, Inc). We further enriched GFP-positive cells by fluorescence-activated cell sorting.

**Measurement of knock-in efficiency using TaqMan/ddPCR.** The premixtures of TaqMan probes and primers consisted of 5 µM of TaqMan probe, 18 µM of a forward primer and 18 µM of a reverse primer. To optimize PCR annealing temperatures and characterize detection efficiencies, we mixed Bio-Rad ddPCR Supermix for Probes (12.5 µl), FAM- and HEX probe and primer premixture (1.25 µl each), and mixture of a genomic DNA (100 ng) and a knock-in allele plasmid (0.5 pg or series of diluted plasmid) in 25 µl total volume. Droplet generation, PCR reaction and droplets read were performed by a QX100 Droplet Generator, a C1000 Thermal Cycler and a QX100 Droplet Reader (Bio-Rad), respectively according to the instructions from the manufacturer. Droplets were analysed by QuantSoft software (Bio-Rad). The optimal annealing temperatures for *LMNA*, *HIST2H2BE*, *CBX1* and *PRKACA* were 62, 60, 60, 59 and 58 °C, respectively.

To measure the knock-in efficiency, we mixed Bio-Rad ddPCR Supermix for Probes (12.5 µl), FAM- and HEX probe and primer premixture (1.25 µl each), and HEK293FT genomic DNA (100 ng) in 25 µl total volume. After droplet generation and PCR reaction, the knock-in frequencies were calculated by taking concentration ratio of a mutant-allele-specific FAM probe and a wild-type-allele-specific HEX probe (*TUBA1A* genomic locus).

**Transcription activation assay.** For *CXCR4* gene activation, the cells were dissociated in Gibco Cell Dissociation Buffer and then stained in PBS with 10% FBS for 1 h at room temperature using anti-human *CXCR4* antibody (2 µg ml<sup>-1</sup>; Biologend, 12G5) conjugated to allophycocyanine. *CXCR4* activation was measured using LSR II flow cytometer (BD Biosciences) at 72 or 96 h post infection as written above<sup>25</sup>.

## References

1. Fernandez-Suarez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell Biol.* **9**, 929–943 (2008).
2. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
3. Irannejad, R. *et al.* Conformational biosensors reveal GPCR signalling from endosomes. *Nature* **495**, 534–538 (2013).
4. Caussinus, E., Kanca, O. & Affolter, M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat. Struct. Mol. Biol.* **19**, 117–121 (2012).
5. 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).
6. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl Acad. Sci. USA* **100**, 12111–12116 (2003).
7. Kent, K. P., Childs, W. & Boxer, S. G. Deconstructing green fluorescent protein. *J. Am. Chem. Soc.* **130**, 9664–9665 (2008).
8. 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).



9. 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**, 727–728 730, 732 passim (2010).
10. Hyun, S. I., Maruri-Avidal, L. & Moss, B. Topology of endoplasmic reticulum-associated cellular and viral proteins determined with split-GFP. *Traffic* **16**, 787–795 (2015).
11. Pinaud, F. & Dahan, M. Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins. *Proc. Natl Acad. Sci. USA* **108**, E201–E210 (2011).
12. 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).
13. 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).
14. 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).
15. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
16. Ormo, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
17. Fan, J. Y. *et al.* Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein-protein interactions in living cells. *Biochem. Biophys. Res. Commun.* **367**, 47–53 (2008).
18. 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).
19. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
20. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
21. Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat. Methods* **11**, 291–293 (2014).
22. 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).
23. Carlton, P. M. *et al.* Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl Acad. Sci. USA* **107**, 16016–16022 (2010).
24. Ishikawa, H. & Marshall, W. F. Ciliogenesis: building the cell's antenna. *Nat. Rev. Mol. Cell Biol.* **12**, 222–234 (2011).
25. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
26. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
27. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
28. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
29. Patterson, G. H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873–1877 (2002).
30. Miesenbock, G., De Angelis, D. A. & Rothman, J. E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192–195 (1998).
31. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
32. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
33. Ishikawa, H. *et al.* TTC26/DYF13 is an intraflagellar transport protein required for transport of motility-related proteins into flagella. *Elife* **3**, e01566 (2014).
34. Ishikawa, H. & Marshall, W. F. Efficient live fluorescence imaging of intraflagellar transport in mammalian primary cilia. *Methods Cell Biol.* **127**, 189–201 (2015).

## Acknowledgements

We thank Yuichiro Miyaoka, Amanda Chan and Bruce Conklin for helping with TaqMan/ddPCR assay, and Wei Qiang Ong for the CD86-SNAP-tag vector and the  $\beta_2AR$  vector. This work is supported by NIH DP2OD008479 (to B.H.), R21MH101688 (to B.H. and D.K.), R01DA036858 (to L.A.G., M.D.L. and J.S.W.), R01GM097017 (to H.I. and W.F.M.), the W.M. Keck Foundation (to J.H., B.C. and B.H.). S.S. receives support from the Japan Society for the Promotion of Science postdoctoral fellowship for research abroad.

## Author contributions

D.K., S.S. and B.H. conceived and designed the experiments. D.K. and S.S. performed the protein labelling and tandem GFP11 tag imaging experiments. B.B.-R. and D.K. performed the split sfCherry experiments. J.H., S.S., M.D.L. and D.K. performed the CRISPR knock-in experiments using plasmid transfection. B.C. generated and validated GFP11<sub>x7</sub> tagged dCas9. H.I., D.K. and W.F.M. performed the IFT tracking experiment. J.S.W. and B.H. conceived the CRISPRa experiment. L.A.G. performed the CRISPRa experiment. D.K. and B.H. wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

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

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

**How to cite this article:** Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* 7:11046 doi: 10.1038/ncomms11046 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>