Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses

Siyu Feng 1, Aruna Varshney2, Doris Coto Villa2, Cyrus Modavi3, John Kohler4, Fatima Farah2, Shuqin Zhou5,6, Nebat Ali2, Joachim D. Müller4, Miri K. Van Hoven2 & Bo Huang 6,7,8

Self-associating split fluorescent proteins (FPs) are split FPs whose two fragments spontaneously associate to form a functional FP. They have been widely used for labeling proteins, scaffolding protein assembly and detecting cell-cell contacts. Recently developments have expanded the palette of self-associating split FPs beyond the original split GFP1-10/11. However, these new ones have suffered from suboptimal fluorescence signal after complementation. Here, by investigating the complementation process, we have demonstrated two approaches to improve split FPs: assistance through SpyTag/SpyCatcher interaction and directed evolution. The latter has yielded two split sfCherry3 variants with substantially enhanced overall brightness, facilitating the tagging of endogenous proteins by gene editing. Based on sfCherry3, we have further developed a new red-colored trans-synaptic marker called Neuroligin-1 sfCherry3 Linker Across Synaptic Partners (NLG-1 CLASP) for multiplexed visualization of neuronal synapses in living C. elegans, demonstrating its broad applications.
Self-associating split fluorescent proteins (SASFPs) are a powerful tool for protein labeling and live-cell imaging. In this system, the eleventh β-strand of FP (FP11, 16 amino acids) is separated out from the remainder of FP (FP1–10) and is genetically fused to the protein of interest (POI). Specific fluorescence signal is detected when FP1–10 reconstitutes with the on-target FP11 to generate a functional fluorescent protein. Since the initial development of self-associating split GFP1–10/11, this technology has been modified and adapted for an extensive range of applications including protein labeling and visualization1–3, scaffolding protein assembly4, protein solubility and aggregation assays5,6, and monitoring the membrane fusion process7. One prominent application is generating a library of human cells with fluorescently tagged endogenous proteins via CRISPR/Cas9-mediated homology-directed repair8. The small size of the GFP11 tag markedly improves the knock-in efficiency and simplifies the donor DNA preparation. In another application, the split GFP1–10/11 system has also been utilized to visualize synapses in living nervous systems by Neuroligin-1 GFP Reconstitution Across Synaptic Partners (NLG-1 GRASP)8.

While all these applications have focused on a single, green-colored channel, expanding the color palette will greatly benefit the investigation of more complex biological systems by enabling multicolor imaging. For this purpose, we have recently developed a red-colored split sfCherry1–10/11 and then a brighter split sfCherry2–10/11, enabling dual-color endogenous labeling in human cells using orthogonal FP1 tags9 and visualization of Listeria protein secretion in infection10. However, unlike split GFP1–10/11, which is as bright as its full-length, split sfCherry2–10/11 produces substantially lower overall fluorescence signal than its full-length counterpart.

Here, we have characterized the complementation mechanism of split FP systems by examining their overall and single-molecule fluorescence brightness. The results suggest a two-step complementation model in which the affinity between the FP1–10 and FP11 fragment is the major limitation to the overall fluorescence signal. Based on this model, we have devised a SpyTag/SpyCatcher-assisted approach to improve the complementation efficiency of sfCherry2–10/11. Furthermore, we have engineered two split sfCherry3 variants with much-enhanced complementation efficiency through a combination of cycles of directed evolution and structure-based site-directed mutagenesis. For tagging endogenous proteins by gene editing, sfCherry3 improves the sorting efficiency for successfully knocked-in cells by 5–10-fold in six tested targets, as compared to sfCherry2. Moreover, we have also developed a new red-colored trans-synaptic marker called Neuroligin-1 sfCherry3 Linker Across Synaptic Partners (NLG-1 CLASP). We established that like NLG-1 GRASP, NLG-1 CLASP labels connections between correct synaptic partners and has a spatial pattern similar to that predicted by electron micrograph reconstruction11,12. As a validation, NLG-1 CLASP labeling is disrupted by loss of the clr-1 gene, which is required for synaptic partner recognition13. To demonstrate the utility of this new biomarker, we utilize it in combination with the NLG-1 GRASP marker to differentially label subsets of synapses made between the same interneurons and different neuronal partners. We also demonstrate that in this system the fluorescence of both biomarkers are severely reduced in clr-1 mutants, as clr-1 is required for both sets of synapses13. These experiments indicate that background fluorescence is minimal, and that these markers correctly visualize synaptic connections.

Results
The complementation process of SASFPs. Previously, we have shown that split GFP1–10/11 has nearly identical overall brightness as its full-length counterpart, whereas both split mNeonGreen2–10/11 (mNG2–10/11) and split sfCherry2–10/11 are substantially dimmer1. This suboptimal performance of mNG2–10/11 and sfCherry2–10/11 could be attributed to either (a) the lower molecular brightness of complemented split FPs or (b) incomplete complementation between the FP1–10 and FP11 fragments. To test (a), we measured the single-molecule brightness of these three split FPs and their full-length counterparts in living cells using fluorescence fluctuation spectroscopy14. We observed no significant difference in single-molecule brightness between the split and the full-length FPs in all three cases (Fig. 1a). Therefore, incomplete complementation should be the cause of the reduced overall fluorescence signal. The ratio of overall fluorescence between split and full-length FPs then reflects the complementation efficiency.

For mNG2–10/11 and sfCherry2–10/11, to determine their complementation efficiency and compare them to GFP1–10/11, we took a similar approach as previously done2. We transiently expressed in HEK 293T cells the full-length FP or the two fragments: FP1–10 and FP11 on a well-folded carrier protein. We quantified whole cell fluorescence by flow cytometry when using a co-expressed infrared fluorescent protein miFP to measure the expression level. miFP was linked to the full-length FP or the FP11 fragment through a 2PA self-cleavage site to ensure equimolar expression. Fluorescence intensities in infrared and green/red channels of each single cell events were displayed in log–log-scale scatter plots (Fig. 1b–d).

In the cases of full-length FPs, as expected, single cell fluorescence intensities followed the trend line with a slope of 1 because miFP was expressed equimorally (Fig. 1e). In the split cases, split GFP1–10/11 almost completely followed the same diagonal line as its full-length counterpart (except for a subtle deviation at the lower-expression end) (Fig. 1b), suggesting a complementation efficiency of almost 100% across a wide range of expression levels. In contrast, both split mNG2–10/11 and split sfCherry2–10/11 deviated from the trend lines of their full-length counterparts (Fig. 1c, d). Considering that the complementation process is ultimately irreversible, as we have shown by both fluorescence recovery after photobleaching (FRAP) and photo-activation measurements (Supplementary Figs. 1 and 2), this observation prompted us to consider a two-step complementation process for self-associating split FPs: the two fragments undergo a reversible association/dissociation equilibrium before entering an irreversible process of folding and/or chromophore maturation:

$$FP_{1 \rightarrow 10} + FP_{11} \Leftrightarrow FP_{1 \rightarrow 10/11}^{\text{folding/maturation}} \rightarrow FP_{1 \rightarrow 10/11} \rightarrow \text{degradation}.$$  

We have verified that in our co-transfection scheme, the expression levels (concentrations) of FP1–10 and FP11 fragments are proportional through an independent experiment in cells co-expressing miFP-P2A-sfCherry2–10/11-Carrier and TagBFP-P2A-sfCherry2–10 (Fig. 1e). Then, at a steady state, this two-step model predicts the following relationship between the two channels of flow cytometry (see Supplementary Notes):

$$[\text{miFP}] \propto \sqrt{K_D[FP_{1 \rightarrow 10/11}]} + [FP_{1 \rightarrow 10/11}],$$  

where $[\text{miFP}]$ represents the sum of concentrations of uncomplemented and complemented $FP_{1 \rightarrow 10/11}$, $[FP_{1 \rightarrow 10/11}]$ is the concentration of matured $FP_{1 \rightarrow 10/11}$, and $K_D$ is the effective dissociation constant of the overall complementation/maturation process. When $K_D$ is much lower than the expression level of $FP_{11}$ (the case of $FP_{1 \rightarrow 10/11}$), the second term dominates, leading to a near proportional relationship between the two flow cytometry channels (slope of 1 in log–log plot, Fig. 1b). In the opposite case of high $K_D$ (the cases of mNG2–10/11 and sfCherry2–10/11), the first term dominates, resulting in log–log scatter plots.
SpyTag/SpyCatcher-assisted complementation of split sfCherry2. Our model indicates that the complementation efficiency of split mNG2\textsubscript{10/11} and sfCherry2\textsubscript{10/11} improves with raised local concentration of fragments, leading to enhanced overall fluorescence signal. Therefore, we sought to utilize a pair of high-affinity binding partners to bring the two fragments into spatial proximity. Because a major advantage of the split FP\textsubscript{10/11} is to label endogenous proteins through knocking-in the short FP\textsubscript{11} peptide, it is preferable to have a small binding partner for the FP\textsubscript{11} fragment. For this purpose, we chose the SpyTag/SpyCatcher\textsuperscript{15} system, a peptide–protein pair that undergoes irreversible binding through formation of an isopeptide bond. The 13-amino-acid (aa) SpyTag is sufficiently short that even when concatenated with GFP\textsubscript{11}, the resulting sequence remains small enough for knock-in using synthetic oligo donor DNAs\textsuperscript{16}.

We examined SpyTag/SpyCatcher-assisted complementation on sfCherry2\textsubscript{10/11} using the recently improved Spy002 pair\textsuperscript{17} (Fig. 2a). We fused SpyCatcher to the N-terminus of sfCherry2\textsubscript{10/11} (the C-terminus is the split site) through a flexible linker in either 6 aa or 15 aa length (Fig. 2a). We generated concatenated tags with SpyTag on either the N- or C-terminus of sfCherry2\textsubscript{11} with a double-glycine linker. We performed similar flow cytometry experiments as described earlier except that mIFP was replaced by TagBFP. Among the four possible combinations of binders/tags (Supplementary Fig. 2), SpyCatcher-6aa-sfCherry2\textsubscript{10/11} with SpyTag-sfCherry2\textsubscript{11}-TagBFP demonstrated the most pronounced shift towards a trend line with a slope of 1 in the scatter plot (Fig. 2b). We have further verified that a shorter SpyTag-sfCherry2\textsubscript{11} fusion without the double-glycine spacer behaved as efficiently (data not shown).

We validated the improvement in overall brightness for cellular microscopy by labeling the N-terminus of histone 2B (H2B) with SpyTag-sfCherry2\textsubscript{11}-TagBFP and co-expressing it with either sfCherry2\textsubscript{10/11} or SpyCatcher-6aa-sfCherry2\textsubscript{10} fusion in HEK 293T cells. We observed the Spy-assisted system (Fig. 2c) could mark the nuclei with much stronger fluorescence signal with the same expression vectors. We further demonstrated that the sfCherry2\textsubscript{11}-SpyTag can also be fused the C-terminus of the protein of interest (Supplementary Fig. 4).

Fig. 1 Characterization of split fluorescent proteins. a) Single-molecule brightness measurement of split FPs and their full-length counterparts using fluorescence fluctuation spectroscopy. N = 10 (mNG2 and sfCherry5) or 15 (GFP) measurements. Error bars are standard deviations. See Supplementary Data for the list of data values. b-d) Flow cytometry analysis of whole-cell fluorescence in HEK 293T expressing either b) GFP\textsubscript{10/11}, c) mNG2\textsubscript{10/11}, and d) sfCherry2\textsubscript{10/11} or their full-length counterparts. The x-axis is the log-scale infrared fluorescence intensity indicating the expression level, and the y-axis is the log-scale green (or red) fluorescence intensity. The gray dashed trend lines have a slope of 1 and intercepts are set to best follow the points in the right (full-length) panels. The pink dashed trend lines have a slope of 2 and intercepts set to best follow the points in the left (split) panels with the exception of GFP\textsubscript{10/11}. e) Expression levels of two fragments are proportional within a wide range of expression levels in a co-transfect experiment. The gray dashed trend line has a slope of 1.

Following more closely to slope-2 trend lines, matching our observations (Fig. 1c, d).
Engineered split sfCherry3 for better complementation. Previously, we engineered sfCherry2-10/11 using a spacer-insertion strategy. This strategy was based on inserting a 32-aa spacer between the sfCherry1-10 and sfCherry11 coding regions. Beyond allowing mutagenesis of both fragments in a single PCR amplicon, the spatial constraints imposed by the linker were hypothesized to assist in the detection of the original mutations by raising the local concentrations of complementary fragments. To increase the screening stringency for complementation-enhancing mutations, we chose to express the fragments of sfCherry2 separately from two promoters using a pETDuet vector (Fig. 3a). Considering short peptides are prone to degradation in Escherichia coli, we fused the sfCherry2-11 sequence to the N-terminus of a well-folded carrier protein (SpyCatcher in our case).

We subjected the sfCherry2-10 fragment to four rounds of error-prone PCR mutagenesis and screening. In every round, a mixture of ~20 brightest variants were selected for the next round. The final isolated mutants were then subjected to one round of DNA shuffling. We have not mutated sfCherry2-11 so that all variants still bind the identical sfCherry2-11 peptide. In the end, sfCherry3C with five substitutions, K45R, G52D, T106A, K182R, N194D (numbering starts from the first Glu after the starting codon Met) was identified as the best variant after the directed evolution (Fig. 3b). All mutations were mapped to either surface oriented residues (T106A, K182R) or locations potentially interacting with the sfCherry2-11 peptide (K45R, G52D, and N194D).

To further improve the complementation efficiency of sfCherry3C, we introduced rational mutations inspired by the mCherry mutant named cp193g7 that is tolerant of circular permutations near our split site, because this mutant contains multiple similar mutations as in sfCherryA and sfCherry2 (ref. 9). We combinatorially introduced the remaining mutations of cp193g7 (I7F, F65L, and L83W) into sfCherry3C through site-directed mutagenesis. Only the variant containing a single L83W mutation gave brighter signal than sfCherry3C, which we designated as sfCherry3V (Fig. 3c). Complemented sfCherry3C-10/11 has an identical emission spectrum as that of sfCherry2 (both split and full-length), whereas the emission spectrum of complemented sfCherry3V-10/11 is blue-shifted by 5 nm (Fig. 3d). Fluorescence fluctuation spectroscopy indicates that sfCherry3C-10/11 has the same single-molecule brightness as sfCherry2 (Supplementary Fig. 5). On the other hand, sfCherry3V-10/11 is dimmer at the single-molecule level (Supplementary Fig. 5), which might be attributed to a difference in two-photon excitation of the blue-shifted chromophore at 1000 nm in the fluorescence fluctuation spectroscopy measurement.

Next, we performed the flow cytometry analysis (Fig. 3e) on HEK 293T cells co-transfected with mIFP-P2A sfCherry2-11 Carrier and sfCherryX1-10 (X being 2, 3C, or 3V). By fitting the data points (excluding those below a threshold above the scattering background) to a line with a fixed slope of 2 (Supplementary Fig. 6; see Supplementary Notes), we found a substantial up-shift of the fitted lines from sfCherry2 to sfCherry3C and sfCherry3V. The shifts corresponded to an increase of complemented signal by 2.5-fold and 8.2-fold, respectively. Assuming the same single-molecule brightness under one-photon excitation for flow cytometry, this signal enhancement also means the same fold of increase in the effective affinity between FP fragments.

We have further characterized other properties of the new split sfCherry3 variants that are important in practical applications. Split sfCherry3V demonstrated similar photobleaching rate as mCherry, whereas split sfCherry3C has a slightly slower photobleaching rate (Supplementary Fig. 7). pH stability is another factor that can affect FP performance in specific situations. Split sfCherry3C showed a relative low pKa (5.0), close to that of mCherry (pKa ~ 4.8) which is known to be acid tolerant, whereas split sfCherry3V presented a higher pKa (5.9), making it still fully fluorescent at neutral pH but more sensitive to the acidic
cellular environment such as in the lumen lysosomes (Supplementary Fig. 8). This pH sensitivity is actually beneficial in tagging proteins in secretory pathways, which can suffer from fluorescence accumulation in lysosomes due to the inability for lysosomal protease to digest mCherry-derived fluorescent proteins. Finally, we compared the temperature dependence of fluorescence in full length and split versions of sfCherry2, as well as split sfCherry3C, in E. coli cultures under either 37 or 25 °C (Supplementary Fig. 9). Although full-length sfCherry2 demonstrated similar fluorescence under different temperatures, the split sfCherry2 had lower intensity under the 37 °C. However, split sfCherry3C was able to recover the fluorescence in the physiological temperature 37 °C, which makes it advantageous for live-cell imaging applications.

Endogenous protein labeling in human cells using sfCherry3.

One unique application of FP11 tag is to generate library-scale fluorescently labeled endogenous proteins through genetic knock-in by homology-directed DNA repair. The small 16-aa size of FP11 allows us to fit its DNA sequence and short homology arms (~70 nt on either side) into commercially available 200 nt single-strand oligo-DNA (ssDNA). By electroporating Cas9/ssRNA ribonucleoprotein (RNP) and donor ssDNA into cells constitutively expressing the corresponding FP11 fragment, robust generation of FP-labeled endogenous proteins is fast and cost-effective. Multicolor knock-in has also been demonstrated by using orthogonal split FP systems to visualize differential distribution and interaction of multiple endoplasmic reticulum proteins.

The overall increased brightness of complemented sfCherry3 variants make them superior to sfCherry2 in the application of labeling endogenous proteins through knock-in. Because sfCherry3 and sfCherry2 share the same FP11 fragment, we adopted a reversed strategy as our previously reported one: knock-in of sfCherry2 into HEK 293T cell lines were created, with knock-ins at: lamin A/C (LMNA, inner nuclear membrane), clathrin light chain A (CLTA), RAB11A, heterochromatin protein 1 β (HP1b), endoplasmic reticulum proteins SEC61b (translocon complex), and ARL6IP1 (tubular ER). We compared fluorescence-activated cell sorting (FACS) enrichment efficiency for each cell line after infection with lentivirus (Fig. 4b–g). In all examined targets, the sfCherry3C1–10 and sfCherry3V1–10 groups displayed remarkable population enhancement in the red-fluorescence-positive gate compared to sfCherry21–10, with the sorting process substantially faster. Practically, for targets like CLTA, SEC61b, or ARL6IP1, we were able to gate the fluorescent population around the clear peak and have ~10-fold higher yield of isolating cells with successful knock-in.

We further confirmed our knock-ins were on-target through confocal microscopy imaging (Fig. 4b–g). Consistent with its higher pKa, sfCherry3V1–10 groups have a reduced tendency to show fluorescent puncta from lysosomes (Fig. 4g) which is reported in our previous work. This makes the sfCherry3V1–10 the preferred protein fragment when labeling endogenous proteins involved in the endomembrane system.
NLG-1 CLASP visualizes specific synapses in live animals. To visualize synapses between specific sets of pre- and postsynaptic neurons in live animals, the trans-synaptic marker NLG-1 GRASP was designed using split GFP fragments. Complementary split GFP1-10/11 fragments were linked via a flexible linker to the transmembrane synaptic protein Neuroligin, which localizes to both pre- and postsynaptic sites in Caenorhabditis elegans. When the two neurons in which the complementary pre- and postsynaptic markers are expressed form synapses, the split GFP fragments come into contact, reconstitute, and fluoresce (Fig. 5a). Using NLG-1 GRASP, we discovered that the recognition between two synaptic partners, the PHB sensory neurons and the AVA interneurons, is mediated the secreted ligand UNC-6/Netrin, its canonical receptor UNC-40/Deleted in Colorectal Cancer, and the receptor protein tyrosine phosphatase (RPTP) CLR-113. NLG-1 GRASP has also been adapted to many other systems, indicating that this technology is transferable. The addition of a red fluorescent trans-synaptic marker would greatly expand this system, allowing us to differentially label different subsets of synapses within the same animals, and potentially even within the same neuron.

Leveraging the split-sfCherry3 tagging system, we have developed the split sfCherry3-based NLG-1 CLASP. The left and right PHB sensory neurons, located in the posterior of C. elegans, form the majority of their synapses with the left and right AVA and PVC interneurons11,12. In this study, we focused on synapses formed between the two PHB neurons with the two AVA interneurons that within the posterior half of the worm is specific for these presynaptic neurons (Supplementary Fig. 10). Animals carrying PHB-AVA NLG-1 CLASP and AVA-VA/DA NLG-1 GRASP had synaptic enrichment.

Analysis of FACS sorting efficiency in six targets, b lam A/C, c clathrin light chain A, d RAB11A, e heterochromatin protein 1 β, f ER translocon complex SEC61b, and g ER tubule protein ARL6IP1, and visualization of sorted knock-in cells through confocal fluorescence microscopy. Scale bar: 10 μm
the \textit{clr-1/RPTP} loss-of-function mutation into the PHB-AVA NLG-1 CLASP and AVA-VA/DA NLG-1 GRASP transgene-carrying animals. We found that fluorescence of both markers was severely reduced, consistent with these markers being synaptic.

**Discussion**

In this study, we have developed bright sfCherry\textsubscript{3}1–10/11 (and its variant sfCherry\textsubscript{3}V1–10/11) with enhanced complementation efficiency, enabling the high-efficient generation of human cell lines with endogenously sfCherry-labeled proteins. Moreover, we have transformed the split sfCherry3 into a trans-synaptic marker called NLG-1 CLASP and have demonstrated the visualization of specific synapses in living nervous systems.

The proposed two-step model, consisting of a dynamic association/dissociation equilibrium followed by an irreversible folding/maturation process, can be generalized to other split fluorescent proteins, including the non-self-associating ones used to monitor protein–protein interaction in bimolecular fluorescence complementation (BiFC) assays\textsuperscript{25,26}. In fact, there is no definitive boundary between the non-self-associating and self-associating split FPs. Instead, their only difference is in the

---

**Fig. 5** NLG-1 CLASP visualizes specific subsets of synapses in live \textit{C. elegans}. \textit{a, b} Schematic diagrams of GFP\textsubscript{1-10/11}-based NLG-1 GRASP and sfCherry\textsubscript{3-10/11}-based NLG-1 CLASP expressed in PHB and AVA neurites. \textit{c, d} Schematics and micrographs of NLG-1 GRASP and NLG-1 CLASP specifically labeling synaptic contacts between the two PHB and the two AVA neurons in \textit{C. elegans} (one neuron from each pair is shown for simplicity). PHB-AVA NLG-1 GRASP and NLG-1 CLASP fluorescence intensity are dramatically reduced in the synaptic partner recognition mutant \textit{clr-1/RPTP}. \textit{e, f} Quantification of the reduction in relative fluorescence intensities of NLG-1 GRASP and NLG-1 CLASP in \textit{clr-1/RPTP} mutants. \(N\) = animals in each case, ***\(P < 0.001\), **\(P < 0.01\), Mann-Whitney U-test. See Supplementary Data for the list of data values.
spontaneous binding affinity, which can be characterized by $K_D'$ in our model. For BiFC analysis, this affinity is one of the major determinants of the sensitivity. If it is too low, the split FP will fail to produce sufficient complementation signal even when the probed molecular interaction does occur. On the other hand, the opposite extreme leads to high background complementation as observed in certain split constructs. While this $K_D'$ is not straightforward to measure biochemically in vitro due to the overall irreversible nature of complementation, our steady-state flow cytometry analysis provides a reliable way to characterize it in cells. Our model also shows that the overall complemented signal is determined not only thermodynamically by the initial binding affinity and local fragment concentrations, but also kinetically by the rates of folding, chromophore maturation and protein degradation.

The assisted complementation demonstrated by SpyTag/Spy-Catcher presents a simple way to improve complementation efficiency between FP fragments. This strategy can be expanded to multicolor imaging using FP tagging using orthogonal binder/tag pairs such as SnoopTag/SnoopCatcher (engineered by splitting an adhesin from *Streptococcus pneumoniae*⁷) or SsrA/SspB (a degradation tag and its adaptor protein from bacterial ClpX ATPase)⁸,⁹.

The substantial improvement of complementation efficiency in sfCherry_{10/11} is beneficial for a wide variety of applications ranging from protein labeling to scaffolding protein complexes, as
as monitoring cell–cell connections. A highly efficient complementation process not only guarantees an enhanced overall brightness, but also enables us to tune down the expression level of FP1–10 fragments which might otherwise exhaust the cellular machinery that maintaining the protein homeostasis (biogenesis, folding, trafficking, and degradation of proteins). This benefit is essential in scenarios that are sensitive to the expression of exogenous proteins, such as tagging endogenous proteins in embryos. Moreover, our engineering platform based on pETDuet vector can be utilized to optimize other self-associating split FPs with insufficient complementation efficiency, such as mNG2,10–11.

Utilizing the split sfCherry31–1011 construct, we have established a completely orthogonal, red-colored trans-synaptic marker called NLG-1–CLASP. We demonstrate that this marker can be used in combination with the original green fluorescent trans-synaptic marker NLG-1 GRASP to label subsets of synapses made between a neuron and different synaptic partners. Similar cyan and yellow trans-synaptic markers (called dual-sGRASP) have recently been generated in vertebrates10. The use of NLG-1 CLASP in this trans-synaptic and yellow-trans-synaptic markers has the potential to allow simultaneous and differential labeling of synapses between a single neuron and three or more synaptic partners in live animals. Since many synaptic connections within the nervous system have overlapping localizations within the nerve ring and other nerve bundles,11,12 this tool will allow accurate visualization of multiple subsets of a single neuron’s connections. These tools may similarly be of use in denser innervated regions of the vertebrate nervous system, such as the hippocampus and cortex. Thus, we propose that NLG-1 CLASP will be a powerful tool with which to probe the development and plasticity of neural circuits within live animals.

Taken together, our work greatly expands the SAsFP toolbox, not only by providing a new, red SAsFP with improved complementation signal, but also by mechanistic elucidation of the complementation process to laid down rational engineering routes. These advances have led to a drastic efficiency improvement in red-tagging of endogenous proteins via gene editing to systematically visualize protein–protein colocalization and interactions. Moreover, they also create a completely orthogonal color channel for the detection of neuronal synapses, allowing simultaneous and differential labeling of synapses between multiple synaptic partners in live animals, paving the way for mapping complex neuronal connectivity networks, and understanding the genes and environmental factors that shape and influence neuronal connectivity.

Methods

Molecular cloning. The DNA sequence of SpyCatcher 002 and SpyTag 002 (based on the reported sequence from ref.17) were directly synthesized (Integrated DNA Technologies, IDT). The DNAs of histone H2B, TOMM20, TagBFP, and mIFP were subcloned from mEmerald, TagBFP, or mIFP fusion plasmids (cDNA source: the Kintzer laboratory at UCSF). The DNAs of histone H2B, TOMM20, TagBFP, and mIFP were co-transfected into HEK 293T cells using FuGENE HD (Promega) following the manufacturer’s recommended protocol. The virus containing supernatant is harvested 48 h after transfection and was centrifuged to pellet all packaging cells. Virus containing medium is used immediately or stored in −80 °C freezer for future use.

For single-molecule brightness measurement sample preparation, U2OS cells were grown in 24-well plates with #1.5 glass coverslip bottoms (In Vitro Scientific) and transfected ~24 h before measurement using transfection reagent (SignaGen Laboratories) according to the manufacturer’s instructions. Immediately before measurement, the growth media was exchanged with PBS buffer with calcium and magnesium (Gibco).

Sample preparation and data analysis in flow cytometry. To characterize the relationship between complementation efficiency and expression level in split GFP, split mNG2, and split sfCherry2, we made pSFFV-mIFP-P2A_fP10,10; pSFFV-sfCherry3C10,10 using the following primers: MVP861 (GGGAGGCTAGCTTCTCAGTGACGATCCTCTCAGG) and MVP862 (GCATCCGGGCCGATGACGCTGTCTGTCG). The fragment was subcloned into pSFFV–mIFP10–10 (MVC228), replacing GFP10–10, using the Nhel and SacI sites. To generate pSFFV–sfCherry3C10–10 (MVC233) construct, site-directed mutagenesis (SDM) was performed using the QuickChange II Site-Directed Mutagenesis Kit on pSFFV–sfCherry3V10–10 using the following primers: MVP861 (GGGAGGCTAGCTTCTCAGTGACGATCCTCTCAGG) and MVP862 (GCATCCGGGCCGATGACGCTGTCTGTCG). The fragment was subcloned into pSFFV–sfCherry3C10–10 (MVC233), replacing GFP10–10, using the Nhel and SacI sites. To generate pSFFV–sfCherry3V10–10 (MVC233) construct, site-directed mutagenesis (SDM) was performed using the QuickChange II Site-Directed Mutagenesis Kit on pSFFV–sfCherry3V10–10 using the following primers: MVP861 (GGGAGGCTAGCTTCTCAGTGACGATCCTCTCAGG) and MVP862 (GCATCCGGGCCGATGACGCTGTCTGTCG).
SpyTag-sfCherry21, TagBFP plus 200 ng pCDAO-sfCherry21, or 100 ng pSSFFV-SpyTag-sfCherry21. TagBFP plus 200 ng SpyCatcher-6aa-sfCherry21, or 200 ng pSSFFV-SpyTag-sfCherry21 to get the brightest clone. The emission spectrum of sfCherry variants and split sfCherry2, we made pSSFFV-sfCherry3C10 and pSSFFV-sfCherry3V10 constructs. Corresponding to each scatter plots in Fig. 3c, same cells were co-transfected with (left) 100 ng pSSFFV-miFP_P2A sfCherry21 and SpyCatcher with 200 ng pSSFFV-miFP_P2A sfCherry3V10 (middle) 1000 ng pSSFFV-miFP_P2A sfCherry3C10 and SpyCatcher with 200 ng pSSFFV-miFP_P2A sfCherry3V10 (right) 100 ng pSSFFV-miFP_P2A sfCherry21 and SpyCatcher with 200 ng pSSFFV-miFP_P2A sfCherry3V10. For flow cytometry analysis, 48 h after transfection, transfected HEK 293T cells were diagnosed with EDDA (0.35%) by a confocal single-color system reconstituted in 0.5 ml PBS solution. 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 by the scattering channel and plotting) was conducted using the FlowJo software (FlowJo LLC).

Single-molecule brightness measurement and data analysis. Fluorescence brightness measurements were performed on a homebuilt two-photon microscope, which has been previously described35. Pulsed laser light (100 fs pulses with a repetition frequency of 80 MHz) from a mode-locked Ti-Sapphire laser (Mai-Tai, Spectra Physics) was focused by a 636 Cs-Apochromat water immersion objective (NA = 1.2, Zeiss) to create two-photon excitation. The emitted fluorescence was collected using a water immersion and separated by dichroic mirror (675DCXSR, Chroma Technology). The fluorescence emission was separated into two detection channels by a 580 nm dichroic mirror (585DCXSR, Chroma Technology) and the green channel was further filtered by an 84nm-wide bandpass filter centered at 510 nm (FF01-500/84-25, Semrock). Avalanche photodiodes (SPCM-AQR-15, PerkinElmer) detected the fluorescence signal, and the fluorescence were recorded by a data acquisition card (FLEX02X, Correlator.com) for ~60 s with a sampling frequency of 200 kHz. All measurements were carried out at an excitation wavelength of 1000 nm and a measured power after the objective of ~0.46 mW. The photon count record was supposed to be analyzed and found on a probability distribution written on Excel (Microsoft).

Mutagenesis and screening. The amino-acid sequence of sfCherry2 and the split site were from our previous published literature28. The sfCherry2_10_10 sequence was subjected to random mutagenesis. PCR (forward primer: AGAGATAGACGTTGTTGTTCAAACTTGGCATGTCGACGAGGAGGAGGAGA, reverse primer: TCTCGGGCCTAGCTTCCTGTTG) using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies). A high mutation rate protocol suggested in the instruction manual was adapted, with an initial target DNA amount of 0.2 µg and 300 cycle amplification. The cDNA library pool was then digested and ligated into a PCR-linearized pETDuet vector (forward primer: TGCAGATGGGCTAGCATGTTGTT, reverse primer: CATGGGATCCATCTGTCC) filtered by an 84nm-wide bandpass filter centered at 510 nm (FF01-500/84-25, Semrock). Avalanche photodiodes (SPCM-AQR-15, PerkinElmer) detected the fluorescence signal, and the fluorescence were recorded by a data acquisition card (FLEX02X, Correlator.com) for ~60 s with a sampling frequency of 200 kHz. All measurements were carried out at an excitation wavelength of 1000 nm and a measured power after the objective of ~0.46 mW. The photon count record was supposed to be analyzed and found on a probability distribution written on Excel (Microsoft).

Fluorescence microscopy. In preparation of cell samples for the imaging purpose, to achieve better cell attachment, the eight-well glass bottom chamber (Thermo Fisher Scientific) was coated with Fibronectin (Sigma Aldrich) for 45 min, washed three times with PBS, and subjected to air-dry before seeding cells. To validate the performance of Spy-assisted complementation in enhancing the overall brightness of protein labeling (Fig. 2c, Supplementary Fig. 4), we transfected either HEK 293T cells or HeLa cells (0.8–1.5 × 10⁶ per well) grown on an eight-well chamber with FuGene HD according to the manufacturer’s protocol (Promega). Total plasmid amount of 180 ng per well with the FPLC1 to FPLC10 ratio in 1:2 was used to achieve optimal expression and labeling. For SpyTag-sfCherry21 linked to the N-terminal of H2B, HEK 293T cells were fixed and stained with 4% paraformaldehyde and then imaged on a Nikon Ti-E inverted wide-field fluorescence microscope equipped with an LED light source (Excitell X-Cite XLED1), a ×40 0.55 NA oil objective (Nikon), a motorized stage (ASI), and an sCMOS camera (iXon Extreme 897, Andor). The negative control was treated with fresh medium supplemented with the same concentration of polybrene. Twenty-four hours after infection, the viral supernatant was swopped with fresh medium. After another 48–72 hours, the infected cells were harvested for flow cytometry analysis and cell sorting.

Crispr/Cas9-mediated gene editing. We generated synthetic single-strand DNA oligos from Integrated DNA Technologies (IDT). We prepared sgRNAs and Cas9/sgRNA RNP complexes following our published methods. Specifically, sgRNAs were obtained by in vitro transcription of DNA oligos (Tucsen). For sfCherry211-SpyTag linked to the C-terminal of TOMM20, 48 h after transfection, HeLa cells were imaged live on a TCF inverted microscope (UCSF Nikon Imaging Center), a Yokogawa CSU-W1 confocal scanner unit, a Plan Apo VC ×100 1.4 NA oil immersion objective, a stage incubator, an Andor Zyla.
acetate sodium (Sigma Aldrich) and acetic acid (Acros Organics). pH value was calculated by hydrochloric acid (Fisher Scientific). The physiological buffer. Acidic PBS (pH 7.4) was used to fix the nuclear morphology, and 2× physiological buffer (pH 7.4) was used for the same conditions. After acquiring the image of pH 7.4 was taken, during the whole experiment. We measured the photobleaching properties of mCherry, sfCherry2C, and split sfCherry3V. The half-life of each fluorescent protein was calculated by fitting every trajectory to a single exponential decay model using GraphPad Prism.

**pH stability measurement.** To characterize the pH stability of split sfCherry3C and split sfCherry3V and compare them with full-length mCherry, we measured their fluorescent intensity in labeling the H2B structure of mammalian cells after fixation under physiological (pH = 7.4) and acidic (pH = 6 and 5) conditions. We used commercial phosphate-buffered saline (PBS, pH = 7.4, 1x, Gibco) as the physiological buffer. Acidic PBS (pH = 6) was prepared from PBS (pH = 7.4) and hydrochloric acid (Fisher Scientific); acetic buffer (pH = 5) was prepared from acetic sodium (Sigma Aldrich) and acetic acid (Acros Organics). pH value was calculated by measuring the pH of each sample.

**SDS behavior assay.** The SDS behavioral assay was used to test the function of the PHB circuit in Supplementary Figure 10, and has been previously described. Briefly, animals are placed on partially dried plates, and induced to move backwards by touching their nose with a hair-pick. A drop of M13 control buffer or 0.1% sodium dodecyl sulfate (SDS) diluted in M13 buffer is placed behind the animal, and quickly absorbs into the dry plate. The time the animal backs into the dry drop before stopping is then measured. It is important that the drop is dry so that it does not wick along the animal and activate sensory neurons outside the PHB circuit. Wild-type animals usually back up into M13 buffer for a second, but stop backing into 0.1% SDS in less than a second. Eighty wild-type and 80 marker-carrying animals were tested in this study. Forty animals from each genotype were tested with control M13 buffer, and 40 animals from each genotype were tested with 0.1% SDS in M13 buffer. The relative response index is calculated by dividing the response time to 0.1% SDS by the average response time to M13 buffer for the same genotype. The response index for marker-carrying animals was normalized to the response index for wild-type animals tested on the same day.

**Statistics and reproducibility.** For all characterizations of FP and split FP constructs by flow cytometry, fluorescence fluctuation spectroscopy and microscopy, we have repeated the experiments with at least two biological replicates which have yielded consistent results.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** All data are available from the authors upon request (M.V. for data in Figs. 5 and 6 and Supplementary Fig. 10, J.D.M. for data in Fig. 1a, and B.H. for all other data). Fluorescent protein constructs are available from B.H. or AddGene. C. elegans lines are available upon request from M.V.

Received: 29 January 2019 | Revised: 6 August 2019 | Published: online 17 September 2019

**Generation of NLG-1 CLASP in C. elegans.** All C. elegans strains were maintained using standard protocols and were raised on 60 mm Nematode Growth Media plates seeded with OP50 E. coli. Wild-type strains were C. elegans variety Bristol, strain N2, and the mutant strain used for this study was br-1(e1745) II. Transgenic strains included wyIs1892 (ref. 21), yIs163, yIs368, yIs180, cbr-1, yIs163, yIs368, cbr-1, wyIs163 carries the extra-chromosomal PHB-AVA NLG-1 GRASP marker (gpa-6::nlg-1::GFP1_10 (60 ng μL⁻¹), gfp-18::nlg-1::GFP5_10 (30 ng μL⁻¹), gfp-18::mCherry (10 ng μL⁻¹), mCherry (5 ng μL⁻¹) and mCherry (20 ng μL⁻¹)); yIs368 contains the extra-chromosomal PHB-AVA NLG-1 GRASP marker (gpa-6::nlg-1::GFP1_10 (60 ng μL⁻¹), gfp-18::nlg-1::GFP5_10 (30 ng μL⁻¹), gfp-18::mCherry (10 ng μL⁻¹), mCherry (5 ng μL⁻¹) and mCherry (20 ng μL⁻¹)); and the AVA-IADA NLG-1 GRASP (gunc-4::nlg-1::GFP1_10 (20 ng μL⁻¹), gfp-18::nlg-1::GFP5_10 (30 ng μL⁻¹) and mCherry (20 ng μL⁻¹)).

A Zeiss Axio Image.A1 compound fluorescence microscope was used to capture images of live C. elegans under x63 magnification. Worms were paralyzed on 2% agarose pads using a 2:1 ratio of 0.3 M 2,3-butanedione monoxime (BDM) and 10 mM levamisole in M9 buffer. All micrographs taken of PHB-AVA NLG-1 GRASP and NLG-1 CLASP markers were of larval stage 4 animals. All data from micrographs were quantified using NIH Image software. Intensity of PHB-AVA NLG-1 GRASP and PHB-AVA NLG-1 CLASP was measured as previously described. Briefly, the intensity at each pixel within each syncytium puncta was measured using NIH Image. To account for differences in background fluorescence, background intensity was estimated by calculating the minimum intensity value in a region immediately adjacent to the puncta. This minimum intensity value was then subtracted from the intensity for each pixel, and the sum of the adjusted values was calculated. For control, pictures of wild-type animals were also taken on the same day using the same settings.

**SDS behavior assay.** The SDS behavioral assay was used to test the function of the PHB circuit in Supplementary Figure 10, and has been previously described. Briefly, animals are placed on partially dried plates, and induced to move backwards by touching their nose with a hair-pick. A drop of M13 control buffer or 0.1% sodium dodecyl sulfate (SDS) diluted in M13 buffer is placed behind the animal, and quickly absorbs into the dry plate. The time the animal backs into the dry drop before stopping is then measured. It is important that the drop is dry so that it does not wick along the animal and activate sensory neurons outside the PHB circuit. Wild-type animals usually back up into M13 buffer for a second, but stop backing into 0.1% SDS in less than a second. Eighty wild-type and 80 marker-carrying animals were tested in this study. Forty animals from each genotype were tested with control M13 buffer, and 40 animals from each genotype were tested with 0.1% SDS in M13 buffer. The relative response index is calculated by dividing the response time to 0.1% SDS by the average response time to M13 buffer for the same genotype. The response index for marker-carrying animals was normalized to the response index for wild-type animals tested on the same day.

**Statistics and reproducibility.** For all characterizations of FP and split FP constructs by flow cytometry, fluorescence fluctuation spectroscopy and microscopy, we have repeated the experiments with at least two biological replicates which have yielded consistent results.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** All data are available from the authors upon request (M.V. for data in Figs. 5 and 6 and Supplementary Fig. 10, J.D.M. for data in Fig. 1a, and B.H. for all other data). Fluorescent protein constructs are available from B.H. or AddGene. C. elegans lines are available upon request from M.V.

**References**

1. Cabantous, S., Terveriller, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* 23, 102–107 (2005).

2. 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).

3. Kamiyama, D. et al. Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* 7, 11046 (2016).
4. 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**, E3001–3008 (2016).

5. Cabanou, S. et al. Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis. *J. Struct. Funct. Genomics* **6**, 113–119 (2005).

6. Chuan, W., Waldo, G. S. & Johnson, G. V. Split GFP COMplementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3β activation and caspase 3 cleavage. *J. Neurochem.* **103**, 2529–2539 (2007).

7. Ishikawa, H., Meng, F., Kondo, N., Iwamoto, A. & Matsuda, Z. Generation of a dual-functional split-reporter protein for monitoring membrane fusion using self-associating split GFP. *Protein Eng. Des. Sel.* **25**, 813–820 (2012).

8. Feinberg, E. H. et al. GFP Reconstitution Across Sypaptic Partn ers (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **57**, 353–363 (2008).

9. Feng, S. et al. Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017).

10. Batan, D. et al. A multicolor split-fluorescent protein approach to visualize listeria protein secretion in infection. *Biophys. J.* **115**, 251–262 (2018).

11. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).

12. Li, F., D. H. & Russell, R. L. The posterior nervous system of the nematode Caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions. *J. Neurosci.* **11**, 1–22 (1991).

13. Varshney, A. et al. The receptor protein tyrosine phosphatase CLR-1 is required for synaptic partner recognition. *PLoS Genet.* **14**, e1007312 (2018).

14. Slaughter, B. D. & Li, R. Toward quantitative fluorescence fluctuation spectroscopy for the study of protein interactions within living cells. *Biophys. J.* **99**, 979–988 (2010).

15. Yu, D. et al. An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat. Commun.* **5**, 3626 (2014).

16. Kerppola, T. K. Bimolecular fluorescence complementation: visualisation of molecular interactions in living cells. *Methods Cell Biol.* **85**, 431–470 (2008).

17. Lee, H., Oh, W. C., Seong, J. & Kim, J. Advanced fluorescence microscopy. *Sci. Rep.* **6**, 11781 (2016).

18. Carlson, H. J., Cotton, D. W. & Campbell, R. E. Circularly permuted fluorescent proteins for endogenous protein labeling. *Biophys. J.* **109**, 6960–6997 (2012).

19. Pessino, V., Citron, Y. R., Feng, S. & Huang, B. Covalent protein labeling by SpyTag-SpyCatcher in fixed cells for super-resolution microscopy. *Chem Sci.* **8**, 1460–1465 (2017).

20. Klee, A. H. et al. Evolving accelerated amimation by SpyTag/SpyCatcher to analyse membrane dynamics. *Angew. Chem. Int. Ed. Engl.* **56**, 16521–16525 (2017).

21. Carlson, H. J., Cotton, D. W. & Campbell, R. E. Circularly permuted fluorescent proteins with new termini in the beta-sheet. *Protein Sci.* **19**, 1490–1499 (2010).

22. Nguyen, H. B., Hung, L. W., Yeates, T. O., Tervilliger, 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).

23. Costantini, L. M. et al. A palette of fluorescent proteins optimized for diverse cellular environments. *Nat. Commun.* **6**, 7670 (2015).

24. Park, J. et al. A conserved juxtransin signal regulates synaptic partner recognition in Caenorhabditis elegans. *Neural Dev.* **6**, 2 (2011).

25. Lee, H., Oh, W. C., Seong, J. & Kim, J. Advanced fluorescence protein-based synapse-detectors. *Front Syst. Neurosci.* **8**, 16 (2014).

26. Jansen, G. et al. The complete family of genes encoding G proteins of Caenorhabditis elegans. *Nat. Genet.* **21**, 414–419 (1999).

27. Rogers, C. et al. Inhibition of Caenorhabditis elegans social feeding by FMRFamide-related peptide activation of NPR-1. *Nat. Neurosci.* **6**, 1178–1185 (2003).

28. Kerppola, T. K. Bimolecular fluorescence complementation: visualization of molecular interactions in living cells. *Methods Cell Biol.* **85**, 431–470 (2008).

29. 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).

30. Veggiani, G. et al. Programmable polyproteins built using twin peptide superglues. *Proc. Natl. Acad. Sci. USA* **113**, 1202–1207 (2016).

31. Levenshuk, I., Grant, R. A., Wah, D. A., Sauer, R. T. & Baker, T. A. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol. Cell* **12**, 365–372 (2003).

32. Gontis, G. et al. Engineering an improved light-induced dimer (ILID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci. USA* **112**, 112–117 (2015).

33. Wu, B., Chen, Y. & Muller, J. D. Fluorescence fluctuation spectroscopy of mCherry in living cells. *Biophys. J.* **96**, 2391–2404 (2009).

34. Sanchez-Andres, A., Chen, Y. & Muller, J. D. Molecular brightness determined from a generalized form of Mandel’s Q-parameter. *Biophys. J.* **89**, 3531–3547 (2005).

35. Yu, D. et al. An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat. Commun.* **5**, 3626 (2014).

36. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).