ASSEMBLY OF MOLECULAR COMPLEX IN DENDRITOGENESIS

by

RYO TAMURA

(Under the Direction of Daichi Kamiyama)

ABSTRACT

Aberrant neural circuit assembly is implicated in neurodevelopmental defects. Therefore, it is essential to understand the molecular mechanisms of functional neural circuit development. Dendrites play an important role in integrating incoming signals in single neurons. Reflecting their individual roles in the brain, dendrites display a wide range of morphologies. To ensure the proper integration of dendrites into functional neural circuit, their morphogenesis is regulated in a spatially restricted manner. To identify key molecular players which underlie the spatial control of dendritic morphogenesis, we utilize the aCC (anterior corner cell) motoneuron in the *Drosophila* embryonic central nervous system (CNS), because of its highly reproducible placement of dendrite formation. With this model in hand, we have previously demonstrated that (1) the coordinate of aCC dendritic outgrowth is determined by an inter-neuronal interaction with its partner neuron, the MP1; and (2) the aCC-MP1 contact provides a scaffold for the adhesive interaction of Dscam1 receptors (Down syndrome cell adhesion molecule) expressed on the aCC and the MP1. The first part of this dissertation identifies Slit, an axon guidance molecule, as a facilitator of the Dscam1 complex assembly at the aCC-MP1 contact site, explores how this molecular complex might be assembled during development, and discusses the implication of the role of Slit/Dscam1 signaling

in a broader context of neural circuit assembly. The second part of this dissertation reports our efforts to build a color palette of split fluorescent proteins. Using rational design and directed evolution, we generate a variety of split FP variants across the entire visible spectrum and demonstrate their use in multiplexed imaging of cellular proteins. In the long term, this work will provide insights into the molecular mechanisms underlying neural circuit assembly and propose a general method to study the dynamic network of proteins in living animals.

INDEX WORDS: Dscam1, Slit, Dendritogenesis, Motoneuron, Neural Circuit Assembly, Split fluorescent protein, Multiplexed imaging

ASSEMBLY OF MOLECULAR COMPLEX IN DENDRITOGENESIS

by

RYO TAMURA

B.S., The University of Tokyo, Japan, 2015

M.S., The University of Tokyo, Japan, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

ASSEMBLY OF MOLECULAR COMPLEX IN DENDRITOGENESIS

by

RYO TAMURA

Major Professor:	Daichi Kamiyama
Committee:	Robert Haltiwanger
	Edward Kipreos
	Ping Shen

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

ACKNOWLEDGEMENTS

I would like to acknowledge Daichi for his great mentorship. His enthusiasm and positive attitude towards science has been the major driving force of my scientific journey at UGA. I would not be who I am as a scientist right now without him. He taught me and showed me firsthand how to logically think, clearly present, and stay confident. I will miss our spontaneous discussions during the day, which have helped me look at biological phenomena from different angles. I would also like to thank my committee members: Bob, Edward, and Ping. Their constructive feedback has always been a great opportunity to reflect on my work and has broadened my horizon.

I would like to thank Rie and Miyuki for supporting my Ph.D. life from all aspects possible. They always prepare samples, order reagents, keep the lab uplifting, and sometimes feed us good food (and so many other things that do not fit this page!). Thanks to their presence, we can solely focus on running our own experiments. I can't tell them how much they made my life as a Ph.D. student easier, and I will absolutely miss being spoiled by them.

I would like to acknowledge my lab mates: Melissa, Kathy, and Kota. Our daily interaction in the lab – ranging from papers we have recently read to what we should cook for dinner – have always made my lab life more colorful. Having someone that I can talk to about almost anything at any given moment is not easy to attain, and it has been an honor for me to have such friends.

Lastly, I would like to thank my parents for supporting me remotely from across the Pacific. Without their encouragement, I wouldn't have flown across the ocean and started doctoral degree abroad in the first place! I am grateful for their continuous support along the way and their never-ending supply of food (=love) to keep my stomach full.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 THE IMPORTANCE OF NEURODEVELOPMENTAL STUDY

"How is the brain wired?" has been a central subject in the neuroscience community. The human brain consists of ~10^{11} neurons [1]. The number of connections they form is significantly larger, approximately reaching ~10^{14} [2]. Even other organisms with smaller brain sizes possess enormous numbers of neurons and neuronal connections; as examples, the mouse brain contains ~10^{9} neurons with 10^{12} connections, and the fruit fly brain contains ~10^{5} neurons with ~10^{8} connections [3, 4]. During development, the gigantic number of neuronal links become established and constitute a cohesive network to execute our cognitive functions. This initial wiring process of the brain is such an indispensable step that its failure, unsurprisingly, leads to impaired brain development. This notion is evidenced by a class of pathological conditions termed neurodevelopmental disorders, including autism spectrum disorder (ASD), X-linked intellectual disability, and epilepsy, which alter affected individuals' cognition, communication, locomotion, and adaptive behaviors [5]. The recent advances in sequencing technology have accelerated the identification of causative genes in neurodevelopmental disorders [6-8]. So far, numerous genes related to neuronal connectivity, such as cell adhesion molecules and synaptic signaling-associated proteins, have been implicated in neurodevelopmental disorders, proving the importance of understanding molecular components involved in the neural circuit assembly [9]. To this end, tremendous efforts will be necessary in the community to identify such key molecular elements involved in the brain development.

1.2 DROSOPHILA AS A MODEL TO INVESTIGATE NEURAL CIRCUIT ASSEMBLY

An equally vital need is to define the biological roles of identified genes in neurodevelopment. *Drosophila melanogaster*, initially established as a model organism in the early 1900s, has been an indispensable tool for researchers to study functions of a given gene. Importantly, ~50% of the *Drosophila* genes are evolutionally conserved in humans [10]. Furthermore, the *Drosophila* genome covers ~75% of the disease-causing genes found in the human genome, making it an ideal system to simulate physiological and pathological phenomena attributed to these conserved genes [11]. Recently, an array of *de novo* variants identified in individuals with ASD have been expressed in flies and recapitulated the functional alterations in neurodevelopment, demonstrating the use of *Drosophila* to model neurodevelopmental diseases [12].

Another advantage of the *Drosophila* system is its abundance in available genetic tools, which enable the functional characterization of a gene of interest. One example is the GAL4/UAS system, a binary expression method, in which the yeast transcription factor GAL4, under the control of an endogenous tissue-specific enhancer, induces the expression of a transgene driven under the upstream activating sequences (UAS), which is activated by GAL4 binding, allowing for tissue-specific manipulation of gene dosage [13]. Additionally, *Drosophila* provides multiple options for genome engineering technology; for instance, an existing library of transposon Minos-mediated integration cassette (MiMIC) insertions replaceable via recombination allows us to label endogenous proteins with genetically encodable tags or to disrupt the targeted gene's function [14, 15]. Multiple studies have demonstrated that *Drosophila* is amenable to CRISPR/Cas9-mediated precise genome editing, expanding the range of accessible genes even further [16-21]. Altogether, these powerful genetic tools can be integrated to decipher the biological role of a target gene and the consequence of its dysregulation at the levels of single cells, tissues, or the whole system.

1.3 THE EMBRYONIC CENTRAL NERVOUS SYSTEM

A single neuron extends two distinct subcellular processes; axons typically travel long distances to innervate other cells and send signals while dendrites receive and integrate inputs from other cells. In the *Drosophila* central nervous system (CNS), formation of axons and dendrites initiates as early as ~9 hours into embryogenesis, ultimately constructing a network of 15,000 neurons, upon which the more complex adult CNS will be subsequently assembled [22]. The *Drosophila* embryonic CNS consists of a central brain and a ventral nerve cord (VNC), the latter of which has been particularly well-characterized both at the anatomical level and at the circuit level [23]. The VNC contains ~19 reiterating segments called neuromeres, each of which comprises two bilaterally symmetric hemi-neuromeres lateral to the midline cells [23]. Each hemi-neuromere originates from ~30 neuroblasts, which give rise to multiple classes of neurons upon differentiation, including interneurons and motoneurons with distinct axonal trajectories. Since the CNS neurons in the VNC can be exposed upon fillet dissection of embryos, lipophilic fluorescent dyes can be directly applied to cell bodies or neuronal projections, allowing for the investigation of axonal and dendritic organizations at the single cell resolution [24]. Using this approach, most of the CNS neurons have been identified and morphologically classified. Typically, the axons are initially organized into two commissures that run across vertically to the midline: the anterior and posterior commissures. Later on, the axons will be organized into three parallel tracts along the anterior-posterior (A-P) axis of the CNS in longitudinal connectives (**Figure 1.1**) [25]. A recent study identified ~270 distinct groups of embryonic interneurons and revealed individual axonal pathways [26]. Contrary to interneurons, motoneurons project their axons into the peripheral field. A sequence of studies over the past few decades offered a comprehensive catalog of all embryonic motoneurons, in which their axonal and dendritic projections were detailed [27, 28].

1.4 MOLECULAR MECHANISMS OF NEURONAL WIRING IN MOTONEURONS

Importantly, Landgraf and colleagues identified the unique target muscle of each motoneuron, thereby establishing the *Drosophila* neuromuscular system as an outstanding model to investigate wiring mechanisms during neurogenesis; in each neuromere, the axons of 36 motoneurons project into the peripheral muscle field and innervate 30 distinct body wall muscles to form neuromuscular junctions (NMJ) (**Figure 1.1**) [29]. Each pair of synaptic partners has been shown to link in a reproducible manner, allowing us to examine the molecular mechanisms underlying axon guidance and synaptic targeting [30]. For example, Siebert and colleagues elegantly showed that the anterior Corner Cell (aCC) motor axons navigate laterally from the CNS following an attractive cue provided by adjacent cells; the leading axon terminals express the immunoglobulin superfamily (IgSF) protein Beaten path Ia (Beat), whose heterophilic interaction with another IgSF protein Sidestep (Side) expressed by the adjacent tissue instructs the axons toward the target muscle [31]. Additionally, other cell adhesion molecules like Connectin, Capricious, and Fasciclin3 have been found to define synaptic specificities of distinct motoneuron-muscle partners through homophilic binding [32-34]. Thus, these studies along with others not mentioned here keep providing insights into molecular principles underlying axonal targeting by utilizing the neuromuscular system.

Distinct from axons that extend into the peripheral muscle field, dendritic projections of embryonic motoneurons are integrated into neuromeres in the CNS [29]. The dendritic outgrowth (dendritogenesis) begins roughly 13 hours after egg laying (around when motor axons reach their peripheral targets), producing primary dendrites. This initial growth of dendrites is followed by the elaboration of dendritic branches formed during later embryonic stages and larval stages [35]. The dendritic processes of embryonic motoneurons grow at highly invariable coordinates; as an example, aCC motoneurons form their dendrites on the medial tract, while MN24 motoneuron on

the lateral tract (**Figure 1.1**) [36]. Previous studies have shown that such segregation in targeting of dendritic branches (dendritic guidance) is achieved by different combinations of attractive and repulsive receptor/ligand pathways, which were initially discovered in axon guidance, including Netrin/Frazzled and Slit/Robo1 complex [37]; Netrin and Slit are secreted proteins produced by midline cells, while Frazzled and Robo1 are expressed by CNS neurons and detect either attractive or repulsive cues towards the midline. Previously, Furrer and colleagues showed that the attractive Netrin/Frazzled pathway and the repulsive Slit/Robo1 pathway differentially impact the dendritic targeting of three unique motoneurons: the aCC, RP2, and RP3 [38]. One of their discoveries was that the growth direction of RP3 dendrites, which usually retain on the lateral tract, shifted towards the midline in the absence of Robo1. Afterwards, Mauss and colleagues further showed that the misexpression of Robo1 in an otherwise Robo1-negative motoneuron is sufficient to re-direct its dendritic processes laterally from the midline, reinforcing the role of such instructive cues in dendritic targeting [39]. Surprisingly, however, loss-of-function mutants of Robo1 and Frazzled did not display clear alteration in the dendritic outgrowth nor in the position of primary dendrites in motoneurons, despite their defects in dendritic guidance [38]; this observation thus suggested the presence of other molecular factors that trigger dendritogenesis in a spatially regulated manner. However, contrary to what we know about dendritic guidance, little is known about what initiates dendritic outgrowth at the molecular level.

1.5 DSCAM AS A REGULATOR OF DENDRITIC MORPHOGENESIS

One of the key molecular players that plays a role in dendritic morphogenesis is Down syndrome cell adhesion molecule (DSCAM), an IgSF protein evolutionarily conserved across species [40]; it was originally identified in the human chromosome band 21q22, a genomic locus associated

with trisomy found in Down syndrome (DS), and thus deemed responsible for numerous neurological phenotypes observed in DS patients [41]. At the molecular level, DSCAM serves as a cell adhesion molecule through homophilic interaction [42-44]. Previous studies conducted using developing mouse retina have shown that the DSCAM-mediated adhesion is necessary for repulsive behaviors between dendrites in close apposition, facilitating efficient dendritic coverage in receptive fields [45-48]. Notably, both down-regulation and up-regulation of DSCAM have been shown to cause defects in dendritic patterning and induce autism-like behaviors in *in vivo* models, underscoring the importance of examining the role of DSCAM in dendritic morphogenesis and deciphering how its dysregulation leads to the etiology of DS [49-53].

Analogous to human DSCAM, *Drosophila* Dscam1 serves as a homophilic cell adhesion molecule and plays a major role in repulsive signaling [54]. In fact, most of the key discoveries regarding the biological function of DSCAM originate in *Drosophila* Dscam1 because its unusually enormous alternative splicing initially attracted the researchers' attention; although the mammalian DSCAMs do not have any isoforms, alternative splicing of the *Drosophila dscam1* results in ~38,016 isoforms [55]. Moreover, Wojtowicz and colleagues found that the homophilic interaction of Dscam1 is isoform-specific [56, 57]. Subsequently, a series of studies revealed that the isoform selection across individual neurons in the fly brain is stochastic, and therefore confines the repulsive behaviors of axons and dendrites mediated by isoform-specific interaction to single neurons [58-62]. These findings led to a paradigm, in which Dscam1 mediates repulsive signaling between neurites within a single neuron, while permitting the adjacent neurons to overlap within the same field, thereby regulating the proper connectivity of developing neurons. Nonetheless, the core function is largely conserved between DSCAM and Dscam1, particularly in the spacing of dendritic processes via its homophilic interaction.

1.6 EMERGING ROLES OF DSCAM AS A RECEPTOR FOR SECRETED LIGANDS

Despite the wealth of knowledge regarding the role of Dscam1 in dendritic patterning, we are only beginning to unravel its functions in other contexts. Recently, Andrews and colleagues found that Dscam1 functions as a receptor for Netrin and mediates attractive axon guidance of CNS neurons towards the midline when crossing the commissures [63]. Concurrently, Ly and colleagues found that human DSCAM binds Netrin-1, a human orthologue of *Drosophila* Netrin, and carries out an attractive signaling identical to the *Drosophila* counterpart [64]. These studies revealed an entirely new, conserved role of Dscam1 in axon guidance as a receptor for Netrin. Additionally, another line of studies has recently shown that Dscam1 serves as a receptor for Slit, another well-known guidance molecule; Dascenco and colleagues found that Slit interacts with Dscam1 and modulates its phosphorylation by recruiting the receptor phosphatase 69D, thus facilitating the local axonal branching proximal to the midline in larval mechanosensory neurons [65]. In addition, Alavi and colleagues revealed that Dscam1 forms a complex with Slit/Robo1 and promotes the longitudinal growth of CNS axons in the embryonic VNC [66]. In summary, these discoveries unlocked new roles of Dscam1 in promoting axon guidance.

Whether such heterophilic interactions of Dscam1 are involved in any aspects of dendritic morphogenesis, however, remains elusive. Hutchinson and colleagues caught the first glimpse of this evidence when they showed that Dscam1 is not required for the dendritic patterning, but for the dendritic outgrowth of adult motoneurons [67]. More recently, Kamiyama and colleagues performed loss-of-function screen of cell adhesion molecules and discovered that Dscam1 is required for the initial outgrowth of aCC motoneurons, further suggesting the role of Dscam1 in dendritogenesis [68]. In Chapter 2 of this dissertation, we will report our continuing effort to characterize the role of Dscam1 in aCC dendritogenesis. We hope that this study will provide

additional insights into the ever-growing list of Dscam1 functions and offer implications to the physiological and pathological function of human DSCAM.

1.7 FLUORESCENT PROTEINS AS A TOOL TO ADDRESS BIOLOGICAL QUESTIONS

(From here on, the text is modified from Tamura et al. 2023. *Methods in Molecular Biology.*)

To start asking the biological roles of target proteins identified in a functional screen in the context of neurodevelopment, or even generally in any biological context, the preliminary step one might take is to identify when and where the proteins exist in a cell or in a tissue. While the advances in single cell sequencing technology have equipped researchers with the ability to identify in which cell the target proteins are expressed, this does not inform their precise locations in the cell [69]. Examining protein localization is fundamental, since the spatial distribution of proteins predicts their functions. Additionally, the integration of the multiple spatiotemporal data adds another layer of information about functional protein networks in a certain biological context, such as protein complex assembly. Therefore, enormous efforts have been made in the scientific community to develop technologies which help us visualize how proteins cooperate to achieve distinct cellular functions. One such technology is fluorescent proteins (FPs), which have transformed cell biology with the capability to visualize proteins of interest using various fluorescence imaging techniques [70]. Previously, an array of FPs has been engineered across and beyond the visible spectrum [71]. These FPs have enabled simultaneous labeling of multiple proteins in individual live cells, which captures spatial and temporal dynamics of the proteins and provides mechanistic insights into gene expression regulation, protein modification, and protein-protein interaction, ultimately assisting in the elaboration of cellular events at the molecular level [72-76].

1.8 SPLIT FLUORESCENT PROTEINS FOR ENDOGENOUS PROTEIN TAGGING

Protein distribution studies have often been carried out through transient transfection in which proteins of interest are tagged with FPs and overexpressed in cells. However, overexpression of FP fusions tends to generate unwanted artifacts (*e.g.,* protein aggregation and mislocalization), making it difficult to interpret the obtained data [77]. Thus, endogenous gene tagging approaches are preferable [78]. One of the methods we can use to achieve endogenous FP labeling is CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9-mediated genome editing strategy. CRISPR/Cas9 system comprises a custom single guide RNA (sgRNA) and Cas9 protein, which induces a double-stranded DNA break to a specific genomic locus in cells. This consequently provides an opportunity to deliberately engineer the locus through homology-directed repair (HDR) [79]. In CRISPR/Cas9-mediated HDR, any altered DNA sequence (including insertions, deletions, and point-mutations) can be flanked by homologous DNA sequences of the corresponding genomic region, resulting in a targeted DNA repair [79]. By leveraging CRISPR/Cas9-mediated HDR, we can integrate an FP sequence into a specific protein-coding region, thus visualizing the endogenous localization of target proteins [80].

Because of the relatively large size of FPs (*e.g.,* EGFP is 717 nucleotides [nt] in length), we have to use a long HDR repair template (> 1000 nt with homologous DNA sequences), which necessitates multiple steps of molecular cloning to insert the sequence into a backbone plasmid [81]. To simplify CRISPR/Cas9-mediated HDR and provide an efficient approach for protein labeling, we have recently adapted the self-associating split GFP system [81, 82]. In the split GFP system, the beta-barrel scaffold of super-folder GFP is split into two fragments comprising one large fragment and one small fragment, termed GFP_{1-10} and GFP_{11}, respectively. Neither GFP_{1-10} nor GFP_{11} is fluorescent alone, yet upon co-expression, the two fragments undergo spontaneous

assembly and reconstitute fluorescently. The GFP_{11} fragment has been utilized as a protein tag. Because the size of the tag is only 48 nt in length [10], this small tag minimizes the length of an HDR repair template (~200 nt in length) and enhances the knock-in efficiency when used for the generation of knock-in cell lines [81]. Moreover, such a short DNA template can be commercially synthesized, making this approach cloning-free. In addition to the visualization of endogenous proteins, these FP_{11} tags have been employed in a wide range of protein analyses including, but not limited to, quantification of protein folding [83, 84], determination of protein topology [85, 86], and visualization of protein localization in living cells [87-89]. To maximize the utility of the split FP technology, others and we have developed numerous color variants of split GFP within the visible spectrum, expanding the choice of colors in split FP systems [90-95]. In Chapter 3 of this dissertation, we will share our effort to engineer a color palette of split FP variants, with the hope that the extended split FP toolkit will allow us to characterize individual protein functions and reveal functional protein networks more robustly in native contexts.

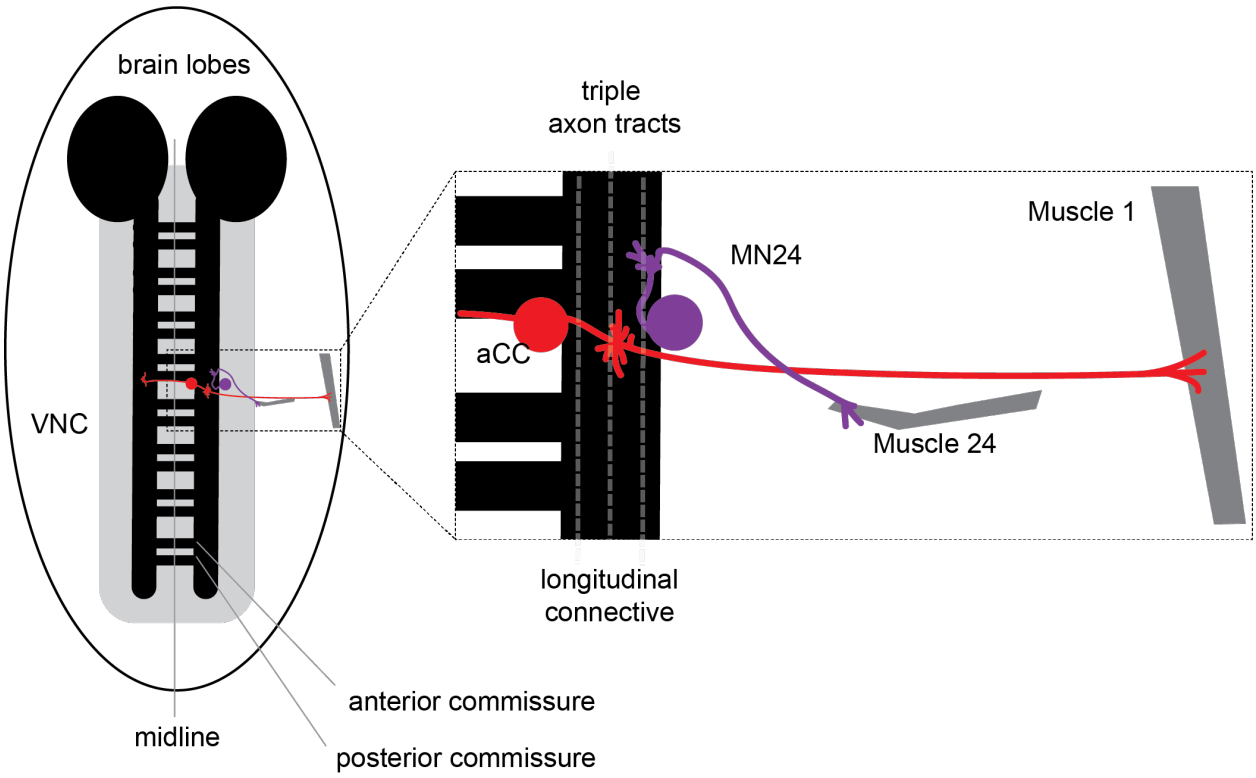


Figure 1.1 | Organization of the embryonic CNS and neurite formation of motoneurons. The embryonic CNS comprises of a brain and a VNC. There are ~19 segments (reduced to 8 in this schematic) of neuromeres in a VNC, each made up of an anterior commissure and a posterior commissure. Bilaterally symmetrical longitudinal connectives run along the midline with each containing three parallel axon tracts. There are ~36 motoneurons in each hemi-neuromere. Two motoneurons are shown as examples in red (aCC) and purple (MN24), innervating their synaptic partners, Muscle 1 and Muscle 24, respectively. Dendritic formation in the aCC will be further discussed in Chapter 2.

1.10 REFERENCES

1. Azevedo, F.A., et al., *Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain.* J Comp Neurol, 2009. **513**(5): p. 532-41.
2. Xie, Q., et al., *Transcription factor Acj6 controls dendrite targeting via a combinatorial cell-surface code.* Neuron, 2022. **110**(14): p. 2299-2314 e8.
3. Herculano-Houzel, S., et al., *Cellular scaling rules for rodent brains.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 12138-43.
4. Scheffer, L.K., et al., *A connectome and analysis of the adult Drosophila central brain.* Elife, 2020. **9**.
5. Parenti, I., et al., *Neurodevelopmental Disorders: From Genetics to Functional Pathways.* Trends Neurosci, 2020. **43**(8): p. 608-621.
6. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability.* Nature, 2014. **511**(7509): p. 344-7.
7. Wang, T., et al., *Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders.* Nat Commun, 2020. **11**(1): p. 4932.
8. Hiatt, S.M., et al., *Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders.* HGG Adv, 2021. **2**(2).
9. Bourgeron, T., *From the genetic architecture to synaptic plasticity in autism spectrum disorder.* Nat Rev Neurosci, 2015. **16**(9): p. 551-63.
10. Yamamoto, S., et al., *A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases.* Cell, 2014. **159**(1): p. 200-214.
11. Mirzoyan, Z., et al., *Drosophila melanogaster: A Model Organism to Study Cancer.* Front Genet, 2019. **10**: p. 51.

12. Marcogliese, P.C., et al., *Drosophila functional screening of de novo variants in autism uncovers damaging variants and facilitates discovery of rare neurodevelopmental diseases.* Cell Rep, 2022. **38**(11): p. 110517.

13. Brand, A.H. and Perrimon N., *Targeted gene expression as a means of altering cell fates and generating dominant phenotypes.* Development, 1993.

14. Koen, J. T., et al., *MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes.* Nat Methods, 2011.

15. Nagarkar-Jaiswal, S., et al., *A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila.* Elife, 2015. **4**.

16. Port, F., et al., *Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila.* Proc Natl Acad Sci U S A, 2014. **111**(29): p. E2967-76.

17. Kanca, O., et al., *An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms.* Elife, 2019. **8**.

18. Xia, B., et al., *CRISPR-based engineering of gene knockout cells by homology-directed insertion in polyploid Drosophila S2R+ cells.* Nat Protoc, 2020. **15**(10): p. 3478-3498.

19. Bosch, J.A., et al., *Precise genome engineering in Drosophila using prime editing.* Proc Natl Acad Sci U S A, 2021. **118**(1).

20. Kanca, O., et al., *An expanded toolkit for Drosophila gene tagging using synthesized homology donor constructs for CRISPR-mediated homologous recombination.* Elife, 2022. **11**.

21. Zirin, J., et al., *State-of-the-art CRISPR for in vivo and cell-based studies in Drosophila.* Trends Genet, 2022. **38**(5): p. 437-453.

22. Jacobs, E.R. and Goodman, C.S., *Embryonic development of axon pathways in the Drosophila CNS. II. Behavior of pioneer growth cones.* J Neurosci, 1989.

23. Crews, S.T., *Drosophila Embryonic CNS Development: Neurogenesis, Gliogenesis, Cell Fate, and Differentiation.* Genetics, 2019. **213**(4): p. 1111-1144.

24. Bossing, T. and Technau, G.M., *The fate of the CNS midline progenitors in Drosophila as revealed by a new method for single cell labelling.* Development, 1994.

25. Marta Zlatic, M.L. and Bate, M., *Genetic Specification of Axonal Arbors: atonal Regulates robo3 to Position Terminal Branches in the Drosophila Nervous System.* Neuron, 2003.

26. Christof Rickert, T.K., et al., *Morphological Characterization of the Entire Interneuron Population Reveals Principles of Neuromere Organization in the Ventral Nerve Cord of Drosophila.* J Neurosci, 2011.

27. Landgraf, M., et al., *The Origin, Location, and Projections of the Embryonic Abdominal Motorneurons of Drosophila.* J Neurosci, 1992.

28. Landgraf, M., et al., *Embryonic origins of a motor system: motor dendrites form a myotopic map in Drosophila.* PLoS Biol, 2003. **1**(2): p. E41.

29. Landgraf, M. and Thor, S., *Development of Drosophila motoneurons: specification and morphology.* Semin Cell Dev Biol, 2006. **17**(1): p. 3-11.

30. Nose, A., *Generation of neuromuscular specificity in Drosophila: novel mechanisms revealed by new technologies.* Front Mol Neurosci, 2012. **5**: p. 62.

31. Siebert, M., et al., *Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields.* Genes Dev, 2009. **23**(9): p. 1052-62.

32. Nose, A., et al., *Connectin: A homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in Drosophila.* Cell, 1992.

33. Shishido, E., et al., *Drosophila synapse formation: regulation by transmembrane protein with Leu-rich repeats, CAPRICIOUS.* Science, 1992.

34. Kose, H., et al., *Homophilic synaptic target recognition mediated by immunoglobulin-like cell adhesion molecule Fasciclin III.* Development, 1997.

35. Kim, M.D., et al., *Patterning and organization of motor neuron dendrites in the Drosophila larva.* Dev Biol, 2009. **336**(2): p. 213-21.

36. Landgraf, M., et al., *The Origin, Location, and Projections of the Embryonic Abdominal Motorneurons of Drosophila.* The Journal of Neuroscience, 1997.

37. Kim, S. and Chiba, A., *Dendritic guidance.* Trends Neurosci, 2004. **27**(4): p. 194-202.

38. Furrer, M.P., et al., *Robo and Frazzled/DCC mediate dendritic guidance at the CNS midline.* Nat Neurosci, 2003. **6**(3): p. 223-30.

39. Mauss, A., et al., *Midline signalling systems direct the formation of a neural map by dendritic targeting in the Drosophila motor system.* PLoS Biol, 2009. **7**(9): p. e1000200.

40. Schmucker, D. and Chen, B., *Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes.* Genes Dev, 2009. **23**(2): p. 147-56.

41. Yamakawa, K., et al., *DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system.* Hum Mol Genet, 1998.

42. Agarwala, K.L., et al., *Down syndrome cell adhesion molecule DSCAM mediates homophilic intercellular adhesion.* Molecular Brain Research, 2000.

43. Agarwala, K.L., et al., *Cloning and functional characterization of DSCAML1, a novel DSCAM-like cell adhesion molecule that mediates homophilic intercellular adhesion.* Biochem Biophys Res Commun, 2001. **285**(3): p. 760-72.

44. Yamagata, M. and Sanes, J.R., *Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina.* Nature, 2008. **451**(7177): p. 465-9.

45. Fuerst, P.G., et al., *Neurite arborization and mosaic spacing in the mouse retina require DSCAM.* Nature, 2008. **451**(7177): p. 470-4.

46. Fuerst, P.G., et al., *DSCAM and DSCAML1 function in self-avoidance in multiple cell types in the developing mouse retina.* Neuron, 2009. **64**(4): p. 484-97.

47. Simmons, A.B., et al., *DSCAM-mediated control of dendritic and axonal arbor outgrowth enforces tiling and inhibits synaptic plasticity.* Proc Natl Acad Sci U S A, 2017. **114**(47): p. E10224-E10233.

48. Santos, R.A., et al., *DSCAM differentially modulates pre- and postsynaptic structural and functional central connectivity during visual system wiring.* Neural Dev, 2018. **13**(1): p. 22.

49. Sago, H., et al., *Ts1Cje, a partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioral abnormalities.* Proc Natl Acad Sci U S A, 1998.

50. Belichenko, N.P., et al., *The "Down syndrome critical region" is sufficient in the mouse model to confer behavioral, neurophysiological, and synaptic phenotypes characteristic of Down syndrome.* J Neurosci, 2009. **29**(18): p. 5938-48.

51. Alves-Sampaio, A., et al., *NMDA-mediated regulation of DSCAM dendritic local translation is lost in a mouse model of Down's syndrome.* J Neurosci, 2010. **30**(40): p. 13537-48.

52. Ramón Pérez-Núñez, N.B., et al., *Overexpressed Down Syndrome Cell Adhesion Molecule (DSCAM) Deregulates P21-Activated Kinase (PAK) Activity in an In Vitro Neuronal Model of Down Syndrome: Consequences on Cell Process Formation and Extension.* HAL open science, 2016.

53. Chen, P., et al., *DSCAM Deficiency Leads to Premature Spine Maturation and Autism-like Behaviors.* J Neuosci, 2022.

54. Zipursky, S.L. and Grueber, W.B., *The molecular basis of self-avoidance.* Annu Rev Neurosci, 2013. **36**: p. 547-68.

55. Schmucker, D., et al., *Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity.* Cell, 2000.

56. Wojtowicz, W.M., et al., *Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding.* Cell, 2004. **118**(5): p. 619-33.

57. Wojtowicz, W.M., et al., *A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains.* Cell, 2007. **130**(6): p. 1134-45.

58. Zhan, X.L., et al., *Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies.* Neuron, 2004. **43**(5): p. 673-86.

59. Neves, G., et al., *Stochastic yet biased expression of multiple Dscam splice variants by individual cells.* Nat Genet, 2004. **36**(3): p. 240-6.

60. Soba, P., et al., *Drosophila sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization.* Neuron, 2007. **54**(3): p. 403-16.

61. Hughes, M.E., et al., *Homophilic Dscam interactions control complex dendrite morphogenesis.* Neuron, 2007. **54**(3): p. 417-27.

62. Miura, S.K., et al., *Probabilistic splicing of Dscam1 establishes identity at the level of single neurons.* Cell, 2013. **155**(5): p. 1166-77.

63. Andrews, G.L., et al., *Dscam guides embryonic axons by Netrin-dependent and -independent functions.* Development, 2008. **135**(23): p. 3839-3848.

64. Ly, A., et al., *DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1.* Cell, 2008. **133**(7): p. 1241-54.

65. Dascenco, D., et al., *Slit and Receptor Tyrosine Phosphatase 69D Confer Spatial Specificity to Axon Branching via Dscam1.* Cell, 2015. **162**(5): p. 1140-54.

66. Alavi, M., et al., *Dscam1 Forms a Complex with Robo1 and the N-Terminal Fragment of Slit to Promote the Growth of Longitudinal Axons.* PLoS Biol, 2016. **14**(9): p. e1002560.

67. Hutchinson, K.M., et al., *Dscam1 is required for normal dendrite growth and branching but not for dendritic spacing in Drosophila motoneurons.* J Neurosci, 2014. **34**(5): p. 1924-31.

68. Kamiyama, D., et al., *Specification of Dendritogenesis Site in Drosophila aCC Motoneuron by Membrane Enrichment of Pak1 through Dscam1.* Dev Cell, 2015. **35**(1): p. 93-106.

69. Li, J., et al., *Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators.* Cell, 2020. **180**(2): p. 373-386 e15.

70. Shaner, N.C., et al., *Advances in fluorescent protein technology.* J Cell Sci, 2007. **120**(Pt 24): p. 4247-60.

71. Rizzo, M.A., et al., *Fluorescent protein tracking and detection: applications using fluorescent proteins in living cells.* Cold Spring Harb Protoc, 2009. **2009**(12): p. pdb top64.

72. Lu, W., M, et al., *Gatekeeper function for Short stop at the ring canals of the Drosophila ovary.* Curr Biol, 2021. **31**(15): p. 3207-3220.

73. Sengupta, P., et al., *A lipid-based partitioning mechanism for selective incorporation of proteins into membranes of HIV particles.* Nat Cell Biol, 2019. **21**(4): p. 452-461.

74. Depry, C., et al., *Multiplexed visualization of dynamic signaling networks using genetically encoded fluorescent protein-based biosensors.* Pflugers Arch, 2013. **465**(3): p. 373-81.

75. Boutros, M., et al., *Microscopy-Based High-Content Screening.* Cell, 2015. **163**(6): p. 1314-25.

76. Moore, A.S., et al., *Actin cables and comet tails organize mitochondrial networks in mitosis.* Nature, 2021. **591**(7851): p. 659-664.

77. Gibson, T.J., et al., *The transience of transient overexpression.* Nat Methods, 2013. **10**(8): p. 715-21.

78. Doyon, J.B., et al., *Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells.* Nat Cell Biol, 2011. **13**(3): p. 331-7.

79. Hsu, P.D., et al., *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-1278.

80. Roberts, B., et al., *Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization.* Mol Biol Cell, 2017. **28**(21): p. 2854-2874.

81. Leonetti, M.D., et al., *A scalable strategy for high-throughput GFP tagging of endogenous human proteins.* Proc Natl Acad Sci U S A, 2016. **113**(25): p. 3501-3508.

82. Kamiyama, D., et al., *Versatile protein tagging in cells with split fluorescent protein.* Nat Commun, 2016. **7**: p. 11046.

83. Cabantous, S., et al., *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nat Biotechnol, 2005. **23**(1): p. 102-7.

84. Sitron, C.S. and Brandman. O., *CAT tails drive degradation of stalled polypeptides on and off the ribosome.* Nat Struct Mol Biol, 2019. **26**(6): p. 450-459.

85. Kaser, S., et al., *Biogenesis of the mitochondrial DNA inheritance machinery in the mitochondrial outer membrane of Trypanosoma brucei.* PLoS Pathog, 2017. **13**(12): p. e1006808.

86. Inglis, A.J., et al., *Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex.* Cell Rep, 2020. **30**(11): p. 3691-3698.

87. Park, E., et al., *Spatiotemporal Monitoring of Pseudomonas syringae Effectors via Type III Secretion Using Split Fluorescent Protein Fragments.* Plant Cell, 2017. **29**(7): p. 1571-1584.

88. Batan, D., et al., *A Multicolor Split-Fluorescent Protein Approach to Visualize Listeria Protein Secretion in Infection.* Biophys J, 2018. **115**(2): p. 251-262.

89. Li, X., et al., *Agrobacterium VirE3 Uses Its Two Tandem Domains at the C-Terminus to Retain Its Companion VirE2 on the Cytoplasmic Side of the Host Plasma Membrane.* Front Plant Sci, 2020. **11**: p. 464.

90. Feng, S., et al., *Improved split fluorescent proteins for endogenous protein labeling.* Nat Commun, 2017. **8**(1): p. 370.

91. Feng, S., et al., *Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses.* Commun Biol, 2019. **2**: p. 344.

92. Zhou, S., et al., *Improved yellow-green split fluorescent proteins for protein labeling and signal amplification.* PLoS One, 2020. **15**(11): p. e0242592.

93. Tamura, R., et al., *Multiplexed labeling of cellular proteins with split fluorescent protein tags.* Commun Biol, 2021. **4**(1): p. 257.

94. Chun, W., et al., *Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3beta activation and caspase 3 cleavage.* J Neurochem, 2007. **103**(6): p. 2529-39.

95. Koker, T., et al., *Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process.* Sci Rep, 2018. **8**(1): p. 5344.

CHAPTER 2

ROLE OF SLIT IN MEDIATING DSCAM1 SIGNALING TO PROMOTE DENDRITIC OUTGROWTH IN THE DEVELOPING ACC MOTONEURONS[1]

[1] Tamura, R. and Kamiyama, D. To be submitted to *The Journal of Neuroscience.*

2.1 ABSTRACT

Dendrites form a specialized subcellular compartment in a neuron necessary for the integration of environmental stimuli. During neural development, dendrites must emerge in a precise position to ensure their proper incorporation into the neural circuit. However, our understanding with respect to the molecular basis underlying the spatial regulation of dendritic morphogenesis is still limited. Previously, we utilized the aCC (anterior corner cell) motoneuron in the *Drosophila* embryonic central nervous system (CNS) and found that the precise location of aCC dendritic outgrowth (aCC dendritogenesis) is defined by an inter-neuronal contact between the aCC and its partner neuron, the MP1. We further discovered that Down syndrome cell adhesion molecule (Dscam1) presented by both the aCC and the MP1 serves as a cue to trigger aCC dendritogenesis. The *Drosophila dscam1* locus generates 38,016 isoforms through alternative splicing, each of which exhibits isoform-specific homophilic binding. Additionally, each individual neuron in the *Drosophila* brain has been shown to express stochastically chosen Dscam1 isoforms, indicating that the aCC and the MP1 are not likely to express matching isoforms that would directly interact with one another. Therefore, we hypothesize that the Dscam1 interaction between the aCC and the MP1 is mediated by co-factor(s), bridging non-matching isoforms of Dscam1. Here, our immunostaining assays in the CNS reveal co-localization of Dscam1 and Slit, a well-known axon guidance molecule. Loss-of-function analyses using *slit* mutants show severe defects in aCC dendritogenesis, suggesting that Slit function is required. Using transheterozygous mutants of *slit* and *dscam1*, we verify the genetic interaction between *slit* and *dscam1*. Biochemical studies show that Slit binds Dscam1 in an isoform-dependent manner and further reveal that Slit mediates isoform-independent interaction of Dscam1. Altogether, we propose a model, in which Slit functions as a co-factor for Dscam1 interaction at the aCC-MP1 contact, serving as a positional cue for aCC dendritogenesis.

2.2 INTRODUCTION

Dendrites, in contrast to axons, are a subcellular compartment of neurons required for the reception of electrochemical signals from partner neurons or environmental stimuli. The enormous diversity of dendritic morphologies underlies specialized functions of individual neurons, which are built early in development [1, 2]. Importantly, dysregulation of dendritic morphogenesis leads to severe biological consequences, evidenced by several pathological conditions including schizophrenia, Down's syndrome, and autism [3]. Therefore, understanding key molecular factors underlying dendritic morphogenesis is crucial to identify the cause of such neurodevelopmental disorders.

Morphology of dendrites arises with their initial outgrowth and the subsequent branch formation, followed by their incorporation into a functional neural circuit [4]. To ensure the neuronal connectivity, dendritic morphogenesis is instructed by extrinsic cues from the local environment, conferring spatial resolution [5, 6]. The dendritic outgrowth of the *C.elegans* motor neuron DA9 is mediated by a gradient of guidance cue, UNC-6 (Netrin), secreted by ventral neurons [7]. In addition, growing evidence suggests that dendritic morphogenesis is mediated by interactions between dendrites and their growth substrate [8-11]. For instance, the multi-dendritic PVD neurons in *C.elegans* develop a dendritic patterning, spatially constricted by an interaction between cell adhesion molecules: DMA1 present in the PVD neuron, and SAX7/MNR-1/LECT-2 present in the adjacent epidermis [11]. Recently, an axon-dendrite contact with the ALA neuron has been shown to instruct the arborization patterning of the PVD neuron [12]. However, whether such inter-cellular contacts can trigger dendritic outgrowth remains elusive.

To investigate the molecular mechanisms underlying spatially-regulated dendrite outgrowth, we utilize the anterior Corner Cell (aCC) in the *Drosophila* embryonic central nervous system (CNS). The aCC is one of the 36 motoneurons present in each abdominal hemisegment

and innervates Muscle 1, one of the two distalmost wall muscles [13]. In addition, the aCC forms ipsilateral and contralateral dendrites in the adjacent longitudinal connective and commissure, respectively. The ipsilateral dendrites consistently emerge from the axon 13 μm distal to the midline, providing an ideal system for examining spatial regulation of dendritic outgrowth [14]. Our previous study found that the precise coordinate of aCC dendritogenesis is dictated by an inter-neuronal contact between the aCC and Midline Precursor 1 (MP1) [15]. The aCC-MP1 contact further provides a scaffold for an adhesion mediated by Down syndrome cell adhesion molecule 1 (Dscam1), followed by recruitment of intracellular signaling factors, Dock and Pak1, eventually triggering downstream cytoskeletal remodeling [15].

Dscam1 is a member of the Immunoglobulin (Ig) superfamily conserved across species; it was originally identified in a human genomic region associated with Down Syndrome [16, 17]. *Drosophila* Dscam1 is remarkable in that the *dscam1* locus can generate 19,008 isoforms of ectodomain by alternative splicing [17]. The *dscam1* locus encompasses three clusters of alternative exons, exon 4, 6, 9 (**Figure 2.1a**), each encoding three variable Ig domains, Ig2, Ig3, Ig7, respectively (**Figure 2.1b**). An additional alternative cluster is encoded by exon17, generating two variable transmembrane domains (**Figure 2.1a-b**) [18]. Notably, Dscam1 exhibits homophilic binding specificity, in which a single isoform with a given combination of three variable Ig domains exclusively binds itself (**Figure 2.1c**) [19, 20]. Additionally, previous *in vivo* studies targeted for mushroom body (MB), class IV dendritic arborization neurons, and the visual system, demonstrated that *dscam1* alternative splicing is probabilistic at the level of single neurons, leading to the model that each individual neuron expresses a distinct subset of Dscam1 isoforms [21, 22]. This stochastic expression of Dscam1 isoforms across individual neurons then causes repulsive behaviors among the neurites exclusively within single neurons; neurites from the same neuron

present matching isoforms, resulting in isoform-specific interaction followed by repulsion, while neighboring neurons present non-matching isoforms, thereby being capable of occupying the overlapping field due to the lack of repulsion (self and nonself recognition) [23, 24]. This model thus predicts that a pair of contacting aCC and MP1 neurons present distinctive subsets of Dscam1 isoforms, which would preclude isoform-specific homophilic interaction between the aCC and the MP1. Therefore, we reasoned that Dscam1 signaling at the aCC-MP1 contact sites may involve other proteins that could mediate the interaction between distinctive Dscam1 isoforms. Previous studies discovered that Dscam1 functions as a receptor for secreted ligands like Slit and Netrin and promotes axonal growth, guidance, or branching [25-27]. Yet, whether such heterophilic interaction partners are involved in Dscam1 interaction remains elusive.

In this study, we explore the hypothesis that Dscam1 interaction at the aCC-MP1 contact site engages additional extracellular signaling factors. Imaging analyses in insect cell culture and the *Drosophila* embryonic CNS show that Slit is involved in Dscam1-mediated adhesion. In addition, genetic studies suggest that Slit and Dscam1 function in the identical signaling axis to promote aCC dendritogenesis. Biochemical experiments indicate that Slit interacts with Dscam1 in an isoform-independent manner. Lastly, we show that Slit mediates the interaction between different Dscam1 isoforms, which is potentially based on multiple molecules of Slit per protein complex. Altogether, we propose that Dscam1 interaction at the aCC-MP1 contact site is mediated by Slit, providing a spatial restriction to aCC dendritogenesis. Such an isoform-independent role of Dscam1 diversifies our understanding of Dscam1 functions and has additional implications for the function of human DSCAM*,* which does not possess any isoform in contrast to *Drosophila* Dscam1 [28].

2.3 RESULTS

Isoform-specific interaction of Dscam1

To investigate the possibility of inter-cellular adhesion mediated by non-identical isoforms of Dscam1, Dscam1$^{1.30.30}$ and Dscam1$^{2.4.30}$ (Entire extracellular domains of Dscam1 containing three variable Ig domains: One isoform is represented by three numbers, each denoting a single splicing variant of exon 4, 6, and 9, respectively coding for a single variable Ig domain) were selected as non-identical isoforms. The two isoforms differ in two variable domains, Ig2 and Ig3, which would eliminate direct interaction [19]. To validate that Dscam1$^{1.30.30}$ and Dscam1$^{2.4.30}$ do not directly bind, we performed bead binding to cell assay [18]. In this experiment, Fc-tagged ectodomains of the Dscam1 isoforms (Dscam1-Fc) were produced in Schneider 2 (S2) cells and coated to magnetic Protein A/G beads, followed by coupling to Alexa Fluor conjugated antibodies for fluorescence microscopy. Subsequently, Dscam1$^{1.30.30}$-beads and Dscam1$^{2.4.30}$-beads were tested for binding to Cos7 cells expressing Dscam1$^{1.30.30.2}$-GFP (GFP-tagged full-length Dscam1 with the 2nd transmembrane variant). As expected, Dscam1$^{1.30.30}$-beads showed significant binding to Dscam1$^{1.30.30.2}$-GFP-espressing cells, while Dscam1$^{2.4.30}$-beads displayed negligible binding comparable to beads not coated with Dscam1, indicating that Dscam1$^{1.30.30}$ and Dscam1$^{2.4.30}$ do not directly interact (**Figure 2.2a-b**). This result is in agreement with the previously proposed model of isoform-specific homophilic binding [18-20].

Distinct Dscam1 isoforms accumulate at cell-cell contact sites

To explore the possibility of interaction between non-identical Dscam1 isoforms in the context of cell-cell interaction, two populations of S2 cells were individually transfected with Dscam1$^{1.30.30.2}$ and Dscam1$^{2.4.30.2}$ and the Dscam1 accumulation was visualized at the cell-cell contact sites using immunofluorescence. A negative control comprised of one population of S2 cells expressing

Dscam1$^{1.30.30.2}$ and one non-transfected. Consequently, a matching Dscam1 isoform (Dscam1$^{1.30.30.2}$) expressed on two opposing cell membranes led to a substantial accumulation at the cell-cell contact sites, likely due to an isoform-specific homophilic interaction (**Figure 2.3a-b**). In addition, the non-matching Dscam1 isoforms expressed *in trans* (Dscam1$^{1.30.30.2}$ and Dscam1$^{2.4.30.2}$) led to a modest and yet significantly higher level of Dscam1 accumulation at cell-cell contact sites compared to control contact sites with a single membrane expressing Dscam1 (**Figure 2.3a-b**). These results suggest that different isoforms of Dscam1 could interact at cell adhesion sites. Considering that Dscam1$^{1.30.30.2}$ and Dscam1$^{2.4.30.2}$ are not capable of binding directly (**Figure 2.2b**), an alternative model would propose that different Dscam1 isoforms may interact through additional factors that exist in the extracellular environment.

Slit function is required in aCC dendritogenesis

Secreted factors present as ideal candidates for mediating such an isoform-independent Dscam1 interaction, due to their capacity to facilitate inter-cellular complex assembly [11]. To date, two extracellular proteins, Netrin and Slit, have been identified as ligands of Dscam1 in *Drosophila*, making them potential factors involved in Dscam1-mediated aCC dendritogenesis [27, 29, 30]. Our previous study has shown that double-knockout embryos of *netrin-A* and *netrin-B*, two *Drosophila* homologs of mammalian *Netrins*, exhibit no marked phenotype in the dendritic morphology of the aCC, suggesting that Netrins do not play a role in aCC dendritogenesis [15]. In contrast, we have previously revealed that *slit* mutant embryos showed a significant loss of aCC dendrites both at hour 13 and hour 15 (after egg laying), indicating the involvement of Slit in aCC dendritogenesis (**Figure 2.5a-b**) [31].

Previous studies have shown that Slit serves as a ligand for Dscam1 and promotes longitudinal axon guidance in the embryonic CNS [30], or instructs the formation of axon

collaterals in an adult mechanosensory neuron in a spatially-restricted manner [27]. These studies have established the role of Slit as a diffusible ligand that acts through Dscam1. However, whether or not Slit plays a role in Dscam1-mediated cell adhesion remains elusive. To test if Slit is involved in cell adhesion in general, we conducted immunofluorescence in S2 cells to visualize Slit localization. A formerly characterized antibody against the full-length Slit, C555.6D [31], revealed that Slit accumulates at cell-cell contact sites, suggesting the association of Slit in the context of cell adhesion (**Figure 2.4a**). To further examine the interplay between Slit and Dscam1, we expressed Dscam1$^{1.30.30.2}$-GFP in Slit-expressing S2 cells. A dual-color imaging of Dscam1-GFP and antibody-stained Slit displayed an overlap of Slit and Dscam1 at the cell-cell adhesion sites (**Figure 2.4b**). Taken together, these results raise the possibility that Slit is involved in Dscam1-mediated adhesion.

To investigate if Slit plays a role in aCC dendritogenesis, we performed lipophilic fluorescent dye tracing of aCC neurons, which enables high density labeling of aCC dendritic morphology at single cell resolution [32]. The loss-of-function *slit* mutant embryos showed a severe reduction in the number of dendritic tips at the aCC dendritogenesis site at hour 15, indicating that *slit* function is required for the initiation of aCC dendritogenesis (**Figure 2.5**). However, the null alleles of *slit* have been previously shown to cause a characteristic collapse of the longitudinal axon tracts towards the midline [33]. This infers that the dislocation of the MP1, which is normally harbored in the medial tract, might have affected the formation of the aCC-MP1 contact sites and the subsequent initiation of aCC dendrite formation. In fact, staining the CNS at hour 15 with anti-Fas2 antibody, which labels the innermost, medial, and outermost longitudinal axon tracts, revealed that the triplet axon tracts were no longer present (**Figure 2.6**). To eliminate the secondary effects of the loss-of-function of *slit*, we utilized hypomorphic alleles of *slit (slit^{dui})*,

which result in reduced *slit* expression. The homozygous mutants of $slit^{dui}$ maintained the global architecture of the longitudinal axon tracts intact (**Figure 2.6**); consequently, the hypomorphic *slit* mutant embryos still displayed a significant defect in the number of aCC dendritic tips (**Figure 2.5a-b**), suggesting that the *slit* function is indeed necessary for aCC dendritogenesis.

Genetic and physical interaction of Slit and Dscam1

To examine the genetic interaction between *slit* and *dscam1*, we inspected the aCC dendritic morphology in embryos transheterozygous for *slit* and *dscam1*. The control embryos individually heterozygous for either *slit* or *dscam1* showed no significant phenotype in aCC dendritogenesis compared to wild-type dendrites (**Figure 2.7a-b**). Conversely, the embryos transheterozygous for *slit* and *dscam1* showed ~50% reduction of the aCC dendritic tips, suggesting the genetic interaction between *slit* and *dscam1* (**Figure 2.7a-b**). Altogether, these results indicate that *slit* and *dscam1* function in an identical signaling pathway leading to aCC dendritogenesis.

To further investigate the molecular interaction between Slit and Dscam1 in the global context of embryonic neurogenesis, we performed a dual-color imaging of Dscam1 and Slit in the fillet-dissected embryos at hour 13. To visualize Dscam1 in endogenous contexts, we utilized a previously established strain with EGFP-coding gene inserted into the *dscam1* locus using MiMIC method (hereafter termed $Dscam1^{\mathrm{MiMICGFP}}$) [15]. Subsequently, the $Dscam1^{\mathrm{MiMICGFP}}$ embryos were stained with anti-Slit antibody to visualize secreted Slit in the CNS. The fluorescent signals of endogenous Slit and Dscam1 exhibited a substantial overlap in the bilaterally symmetric longitudinal connectives, suggesting the molecular interaction between Slit and Dscam1 in the CNS (**Figure 2.8**).

Our genetic and imaging results point to a model that Slit interacts with Dscam1 in the context of aCC dendritogenesis. To test the direct interaction between Slit and Dscam1, we co-

expressed the 3xFLAG and Fc-tagged extracellular portion of Dscam1$^{2.4.30.2}$ (termed Dscam1$^{2.4.30}$-Fc) and V5-tagged Slit in 293T cells and affinity-purified Dscam1$^{2.4.30}$-Fc using magnetic Protein A beads. In agreement with the previous studies [27, 30], both full-length Slit (hereafter termed Slit for short) and the N-terminal fragment Slit (Slit N) were pulled down with Dscam1$^{2.4.30}$-Fc, suggesting their direct interaction (**Figure 2.9a-b**). To rule out the binding of Slit to the Fc region, a control Fc protein was tested for binding against Slit, in which no binding was detected (**Figure 2.9a and 2.9c**). Altogether, we validated that Slit binds Dscam1.

Slit mediates isoform-independent Dscam1 interaction

Interestingly, we noticed that the Dscam1 isoform used in this study to test Slit binding (2.4.30) differed from the isoform used in the previous studies (1.30.30) [27, 30], hinting at the isoform-independent interaction of Dscam1 with Slit. To see whether switching variable Ig domains of Dscam1 affects Slit binding, we tested it against three different isoforms of Dscam1: 1.30.30, 2.4.30, and 3.36.25. All of the isoforms were fused to the Fc tag for comparison and co-transfected with Slit in HEK293T cells. As expected, Dscam1-Fc of all three isoforms pulled down Slit, suggesting that Slit-Dscam1 interaction occurs independently of Dscam1 isoforms (**Figure 2.10a**).

Previously, we demonstrated that *dscam1* function is required in both sides of the aCC-MP1 contact site [14]. Given the stochastic expression of Dscam1 isoforms across distinct neurons [21, 22], it is likely that the Dscam1 interaction at the aCC-MP1 contact site is mediated by distinctive, thereby non-matching isoforms. Considering the isoform-independent nature of Slit-Dscam1 interaction demonstrated in this study, we tested whether the presence of Slit causes different Dscam1 isoforms to form a complex. To assess this idea *in vitro*, we co-expressed non-matching Dscam1 isoforms (2.4.30 and 3.36.25), one in Fc-fusion and the other in transmembrane form, in the presence or absence of Slit in HEK293T cells. Subsequently, the protein complex was

isolated via Dscam1-Fc using magnetic Protein A/G beads. The following detection of isolated proteins across conditions showed that the isoform 2.4.30 was pulled down by the identical isoform, while it was pulled down substantially less by the isoform 3.36.35, in agreement with the model of isoform-specific binding (**Figure 2.10b**). Strikingly, complementation of Slit caused 2.4.30 and 3.36.25 to co-precipitate more than they did in the absence of Slit, suggesting the role of Slit in bridging the interaction between different Dscam1 isoforms (**Figure 2.10b**).

How does Slit mediate the interaction between multiple Dscam1 isoforms simultaneously? One possible explanation would be that Slit forms dimers, oligomers, or even more heterogeneous complexes containing multiple Slit proteins mediated by other molecules to maximize the number of Dscam1 binding interfaces. Indeed, previous crystallographic studies revealed dimerization of human Slit2 through its fourth leucine-rich repeat domain [34]. Additionally, Slit N has been reported to undergo oligomerization in the developing *Drosophila* musculoskeletal system [35]. However, such dimerization or oligomerization of Slit has never been fully characterized. To fill this gap in knowledge, we utilized two differentially tagged Slit proteins: Slit-FLAG-Fc and Slit-V5, which would allow us to test whether multiple molecules of Slit exist as a complex. By co-expressing the pair of Slit proteins followed by isolation of proteins using magnetic Protein A/G beads, we evaluated the possibility of multiple Slit proteins forming a cluster. Consequently, Slit-V5 was precipitated in the presence of Slit-FLAG-Fc, but not in the presence of FLAG-Fc control, suggesting that multiple molecules of Slit form a complex. Considering the isoform-independent nature of Slit-Dscam1 interaction we have just shown (**Figure 2.10a**), this data supports our hypothesis that clustered Slit may provide the scaffold for coupling multiple Dscam1 molecules regardless of the isoforms.

2.4 DISCUSSION

Herein, we showed that aCC dendritogenesis is mediated by Slit-Dscam1 signaling. Our cell biological and biochemical studies revealed that Dscam1 binds Slit in an isoform-independent manner. We further demonstrated that non-identical isoforms of Dscam1 are capable of forming a complex in the presence of Slit, potentially through oligomerized Slit molecules. We have previously shown that aCC dendritogenesis is triggered by inter-cellular Dscam1 interaction between the aCC and the MP1 [15]. Considering the stochastic nature of Dscam1 isoform selection across individual neurons [21, 22], a contacting pair of the aCC and the MP1, being two distinct neurons, are likely to express random subsets of Dscam1 isoforms, which do not directly bind one another. If so, how does the Dscam1 interaction take place at the aCC-MP1 contact site? Our novel finding described here provides a potential model, in which Slit intermediates the interaction between non-matching isoforms of Dscam1 expressed by the aCC and the MP1, thereby serving as a positional cue for aCC dendritogenesis (**Figure 2.10d**).

Slit mediates isoform-independent Dscam1 interaction

Dscam1 function has been extensively investigated from the perspective of isoform-specific homophilic interaction. Due to alternative splicing, stochastically selected variable exons in the *dscam1* locus generate up to 38,016 isoforms, only a handful of which are expressed in each individual neuron. Consequently, only neurites derived from the same neuron would possess matching Dscam1 isoforms, leading to repulsive behaviors termed self-avoidance. This ensures that sister neurites do not overlap in an identical field, thus resulting in their efficient coverage of the innervating or receptive field. Conversely, neurites from neighboring neurons would possess non-matching Dscam1 isoforms, allowing the neurites derived from the distinctive neurons to occupy the shared field. Accordingly, this isoform specificity of Dscam1 interaction serves as a

restrictive cue for shaping proper axonal and dendritic patterning in the brain [28]. In contrast, Dscam1 interaction between different isoforms is yet to be investigated. To date, the homophilic interaction of different Dscam1 isoforms, if there is any at all, has been shown to be extremely weak [18]. However, there are evidence that such an isoform-independent Dscam1 interaction plays a role in neural circuit assembly. For instance, MB axons in a corresponding dorsal lobe fasciculate normally only when multiple Dscam1 isoforms, not a single isoform, are expressed [36]. Similarly, Bolwig's nerve axons show fasciculation defects in the absence of Dscam1 [17]. Considering that adjacent and yet distinctive axons represent random, thereby non-matching Dscam1 isoforms [22], these evidence indicate that the normal fasciculation seen in the MB axons and the Bolwig's nerve axons is likely to be mediated by isoform-independent Dscam1 interactions [18]. However, there has been little evidence showing that such isoform-independent Dscam1 interaction exist at the molecular level. Our biochemical analyses herein demonstrated that Slit mediates interaction between different Dscam1 isoforms *in vitro*, providing one possible explanation as to how isoform-independent Dscam1 interaction might arise at the molecular level.

Potential role of clustered Slit molecules in mediating Dscam1 interaction

The importance regarding higher order structures of Slit has been shown in several biological contexts. One example is axon guidance during neuronal development, in which Slit dimerization facilitates Slit-Robo1 signaling in growth cones [34]. Another example has been demonstrated during the formation of the musculoskeletal system in embryogenesis, in which muscle elongation is instructed by adjacent tendon cells through an inter-cellular adhesion complex mediated by Robo receptors and secreted Slit oligomers [37]. In agreement with this notion, we showed that multiple Slit molecules form a complex *in vitro*. Specifically, in aCC dendritogenesis, such clustering of Slit potentially offers a scaffold to accommodate multiple Dscam1 molecules/isoforms and form

an adhesion complex at the aCC-MP1 contact site. This model provides further insights into the role of clustered Slit in inter-cellular communications, although solving its mechanism requires further investigation.

Slit-dependent and Slit-independent Dscam1 functions

Considering that Slit-mediated Dscam1 interaction demonstrated here underlies dendritic outgrowth, rather than dendritic patterning governed by repulsion, the presence of Slit potentially functions as a switch to differentiate distinct Dscam1 functions in a context dependent manner. This notion is in accord with recent studies showing that Slit-Dscam1 signaling facilitates axon growth restrictedly in proximity to the midline, in which Slit concentration is highest [27, 30].

How Slit-Dscam1 signaling differentiates itself from the conventional, isoform-specific Dscam1 signaling largely remains an open question. In the context of aCC dendritogenesis, Dscam1 accumulated at the aCC-MP1 contact site recruits Pak1 kinase through an adaptor protein Dock, triggering downstream cytoskeletal remodeling [15] By contrast, Dscam1-mediated repulsion in class I da neurons does not require Pak1 activity [24]. Additionally, a recent study has shown that Slit binding to Dscam1 can locally promote interactions with the receptor tyrosine phosphatase RPTP69D, leading to the dephosphorylation of Dscam1 [27]. Although whether Slit recruits such co-receptors and effectors directly or indirectly needs to be further investigated, these evidence at least suggest that differential subsets of co-receptors and/or downstream factors could differentiate Slit-Dscam1 signaling and the isoform-specific Dscam1 signaling, resulting in distinct morphological behaviors (**Figure 2.11**).

Limitation of this study

To further validate our model in which Slit mediates Dscam1 interaction at the aCC-MP1 contact site, several *in vivo* studies will be essential. Previously, we have shown that aCC dendritic defects

in *dscam1* mutants are sufficiently diminished by complementing one single isoform of Dscam1 in the aCC and the MP1. However, this did not address whether or not the aCC-MP1 contact is mediated by isoform-independent Dscam1 interaction in a Slit-dependent manner. To tackle this point, we could express different isoforms of Dscam1 in the aCC and MP1, respectively, in *dscam1* mutant background with the presence or absence of Slit. Although a series of previous studies on Dscam1 isoform selection in the developing brain suggest that the aCC and the MP1 are likely to express stochastic, thereby non-matching Dscam1 isoforms, this does not rule out the possibility that they express some degree of matching Dscam1 isoforms. To address this, we could utilize recently established transgenic reporter fly strains which enables the identification of spliced exons in individual neurons [22].

Nevertheless, our model on Slit-Dscam1 signaling could be generalizable in the developing *Drosophila* CNS. Our staining data revealed a striking co-localization of Slit and Dscam1 across the CNS neuropil, which is not limited to the proximity of the aCC dendritogenesis site. Considering that the neuropil encompasses a high concentration of Slit and is abundant in neurites representing a diverse repertoire of Dscam1 isoforms, it is conceivable that Slit-mediated Dscam1 interaction is implemented by a number of other inter-neuronal interactions besides the aCC-MP1 contact. Altogether, we propose that Dscam1 functions as a positive growth-promoting cue in the presence of Slit, distinct from its well-studied, traditional role in neurite repulsion through Dscam1 isoform interaction.

2.5 FIGURES

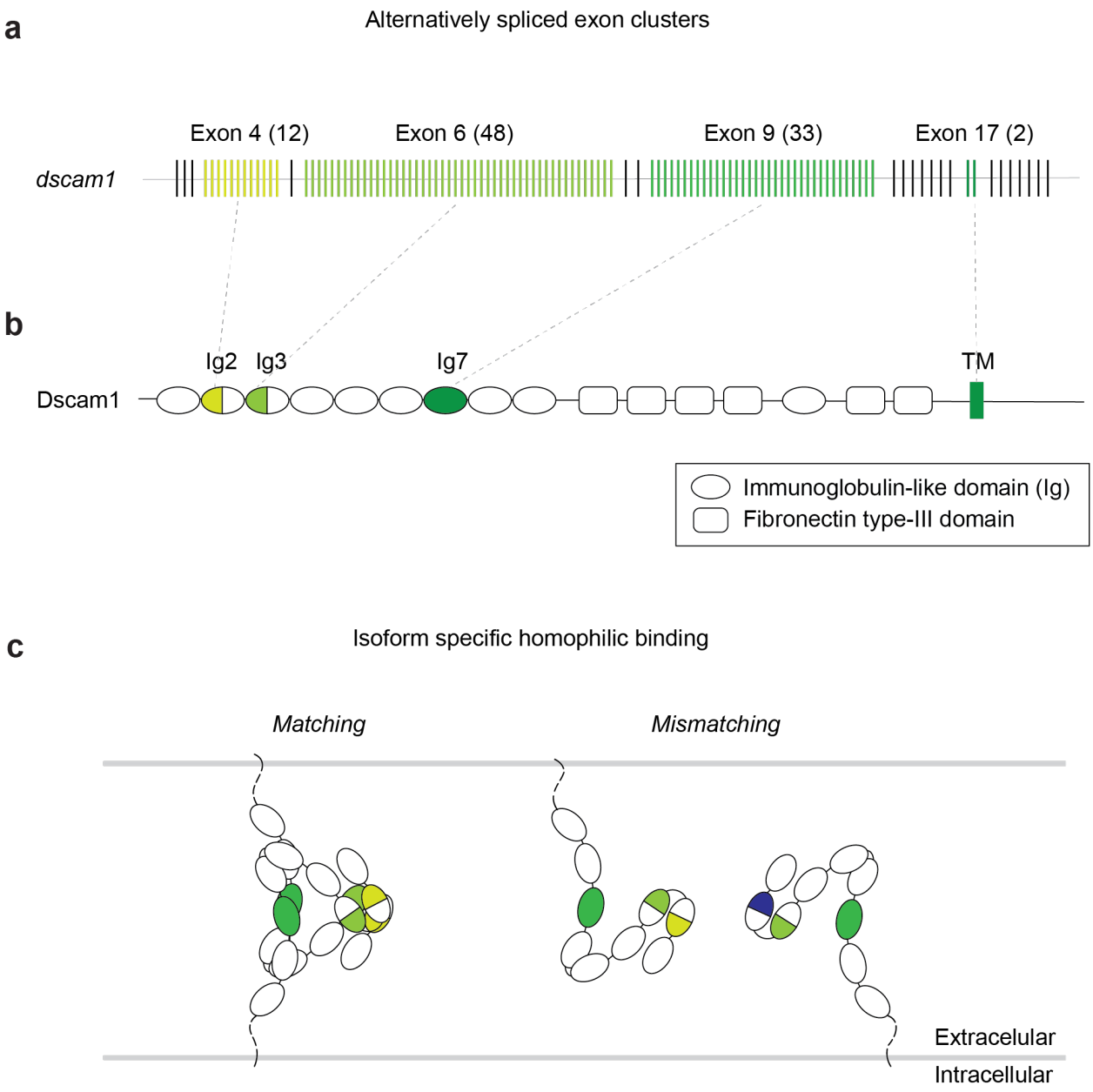


Figure 2.1 | Dscam1 alternative splicing and isoform binding specificity. (**a**) The *dscam1* locus contains 24 exons. Exons 4, 6, and 9 are composed of 12, 48, and 33 alternatively spliced exon variants, respectively, coding for 19,008 possible combinations of extracellular domains. Exon 17 has two exon variants. The numbers of splicing variants in each exon cluster are shown in

parentheses. (**b**) Domain model of Dscam1 protein, which consists of immunoglobulin-like domains (Ig), fibronectin type-III domains, a transmembrane domain, and a cytoplasmic tail. Variable protein domains encoded by the corresponding alternatively spliced exon clusters are color-coded. Ig2 and Ig3 are partially encoded by exons 4 and 6, respectively. Exon 7 encodes the entire length of Ig7. Exon 17 encodes two variable transmembrane domains (TM). (**C**) Homophilic binding of Dscam1 is isoform-specific, in which all three variable Ig domains should match between two Dscam1 molecules for the binding to occur. Even one single mismatching pair causes the opposing isoforms to not bind one another.

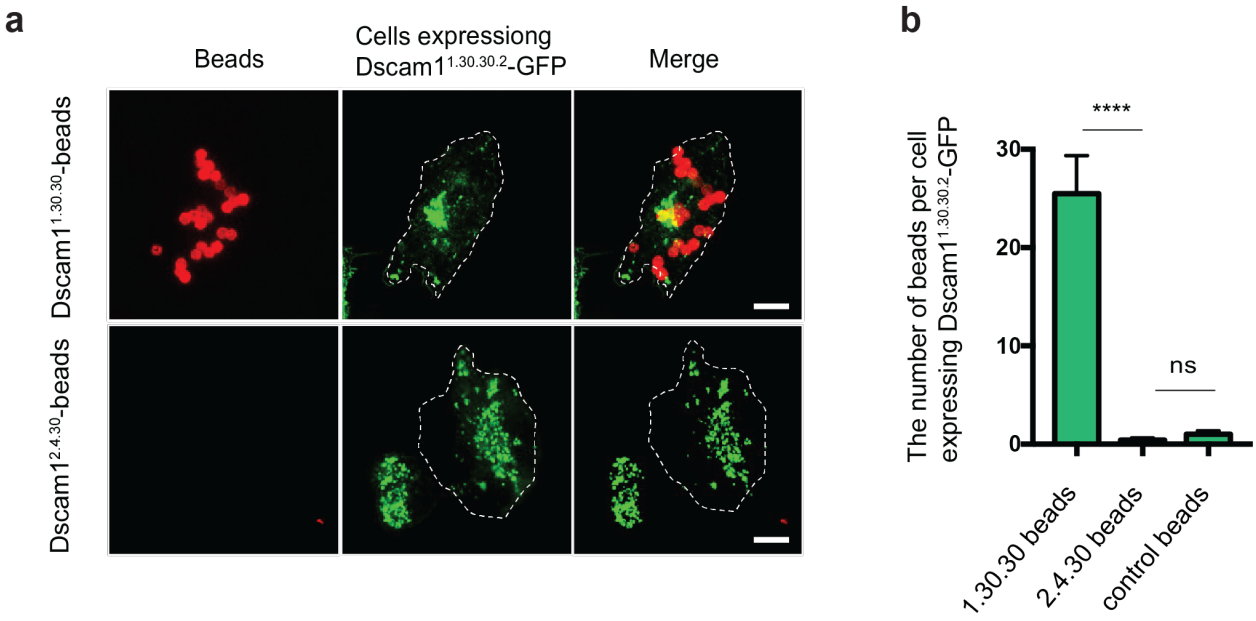


Figure 2.2 | Isoform-specific interaction of Dscam1. Binding specificities of the isoform 1.30.30 and 2.4.30 were tested using bead binding to cell assay. (**a**) Cos7 cells expressing Dscam1$^{1.30.30.2}$-GFP were incubated with red fluorescent beads coated with either Dscam1$^{1.30.30}$-Fc or Dscam1$^{2.4.30}$-Fc. Representative cells were shown in white dashed outlines. Scale bars, 10 μm. (**b**) Quantification of beads binding to Cos7 cells. Data are represented as mean ± SEM. 19 to 34 cells were counted for bead binding in each condition. ns, not significant, **** $p < 0.0001$.

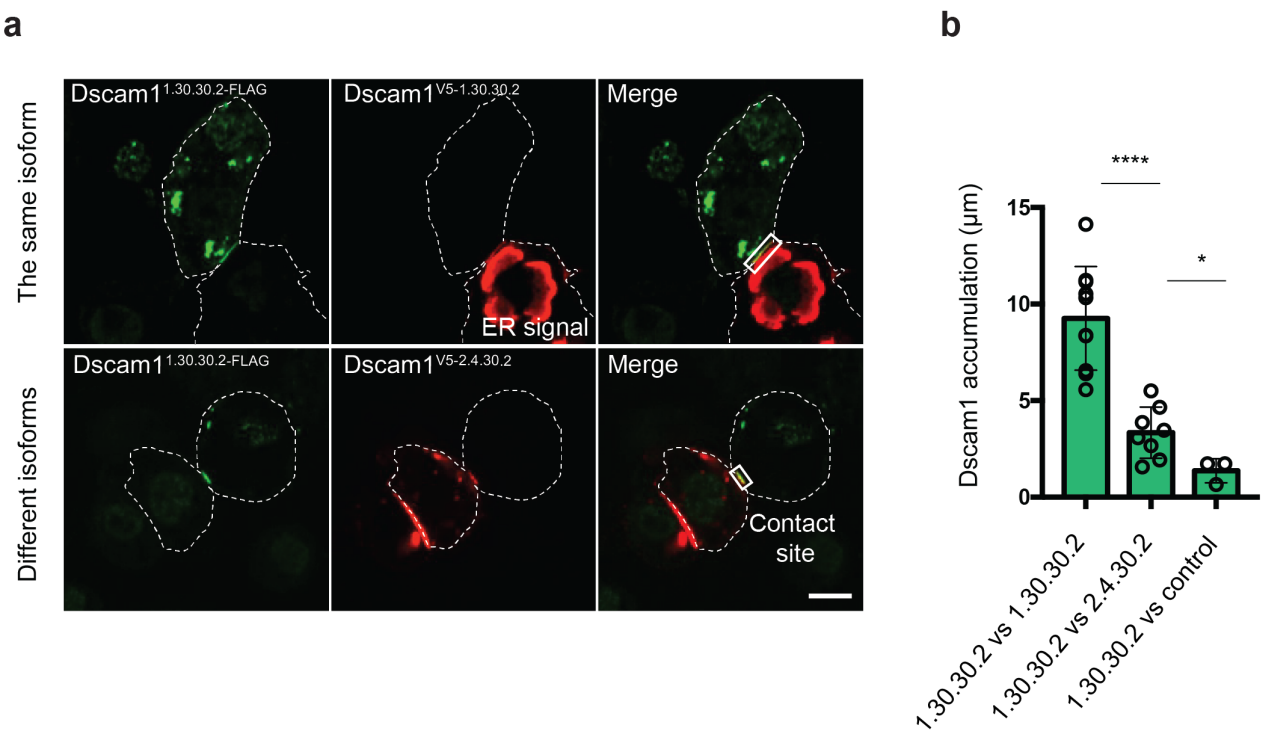


Figure 2.3 | Distinct Dscam1 isoforms accumulate at cell-cell contact sites. Binding capacities of distinct Dscam1 isoforms were assessed by localization study. (**a**) Dscam1 isoforms were expressed in two adjacent S2 cells and visualized by antibodies against FLAG or V5 tags. White boxes indicate Dscam1 accumulation at cell-cell contact sites. Strong ER accumulation of transiently expressed Dscam1 was observed at times. Scale bar, 10 μm. (**b**) Quantification of Dscam1 accumulation at cell-cell contact sites in length. Control indicates non-Dscam1 expressing cells. Data are represented as mean ± SEM. ****$p < 0.0001$, *$p < 0.05$.

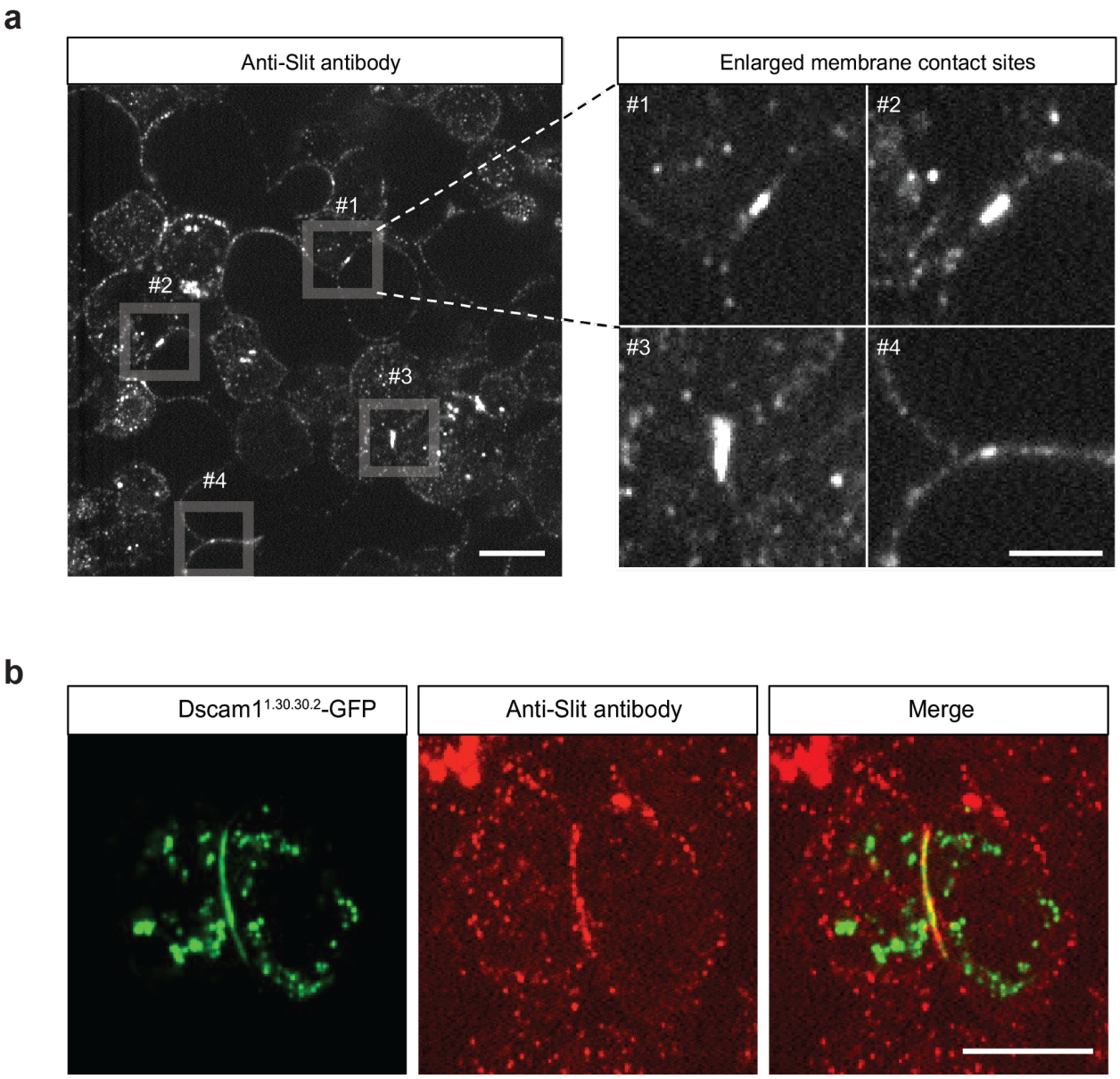


Figure 2.4 | Slit is involved in Dscam1-mediated cell adhesion. Interplay between Slit and Dscam1 was tested by localization study. (**a**) Slit-expressing S2 cells were stained with anti-Slit antibody. Slit accumulation at cell-cell contact sites were indicated in white boxes. Scale bars, 5 μm and 10 μm, from the left. (**b**) Co-localization of Dscam1 and Slit was assessed by expressing Dscam1$^{1.30.30.2}$-GFP in S2 cells, followed by anti-Slit antibody staining in the same cells. Detergent was not employed to specifically visualize extracellular Slit. Scale bar, 10 μm.

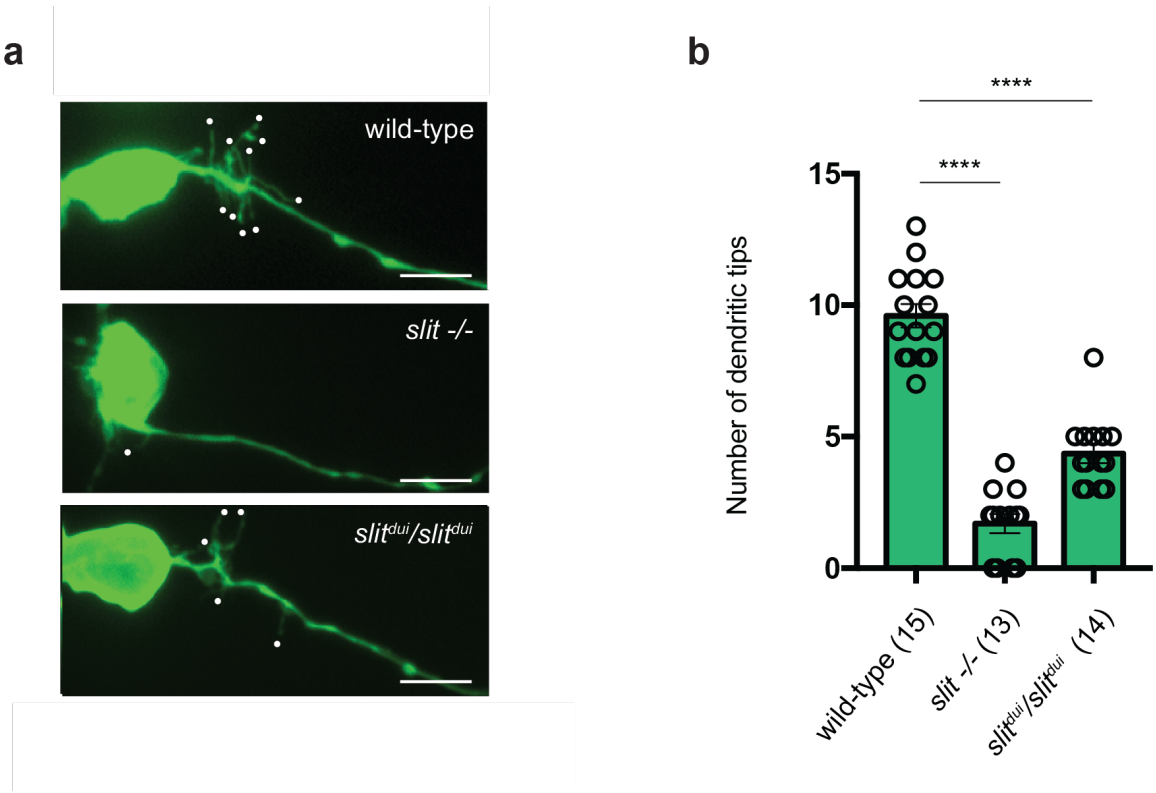


Figure 2.5 | Slit function is required in aCC dendritogenesis. To test whether Slit function is required in aCC denritogenesis, single aCC neurons were labeled by lipophilic dye tracing in wild-type, *slit -/-*, and $slit^{dui}/slit^{dui}$ embryos at hour 15 (**a**), and the number of dendritic tips were counted in each neuron (**b**). Scale bars, 5 μm. Sample sizes are indicated in parentheses. Data are represented as mean ± SEM. ****$p < 0.0001$.

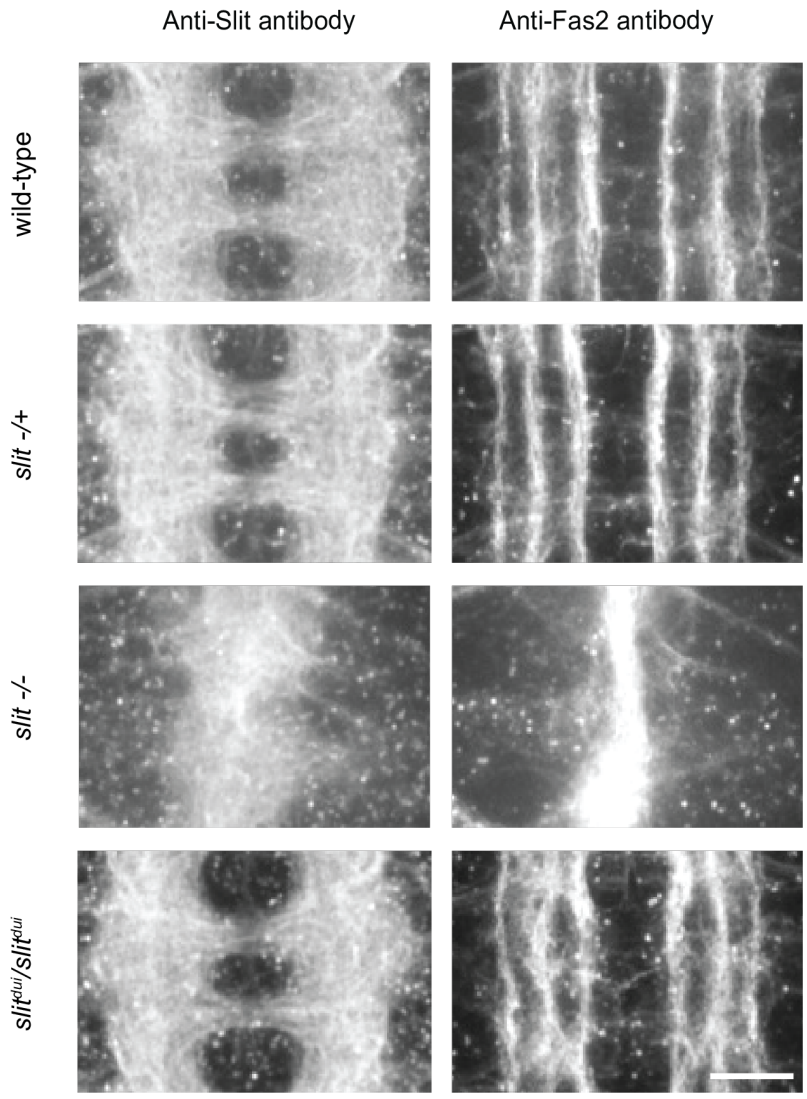


Figure 2.6 | Global CNS structure remains intact in hypomorphic *slit* mutants. *Drosophila* embryos at hour 15 were fillet-dissected and stained with anti-Slit and anti-fas2 antibodies. Bilaterally symmetrical pair of three longitudinal axon tracts (innermost, medial, and outermost) are revealed by Fas2 localization. Genotypes used are as follows: wild-type, *slit* -/+, *slit* -/-, and $slit^{dui}/slit^{dui}$. A single segment of the ventral nerve cord (VNC) is shown in each genotype. Scale bar, 10 μm.

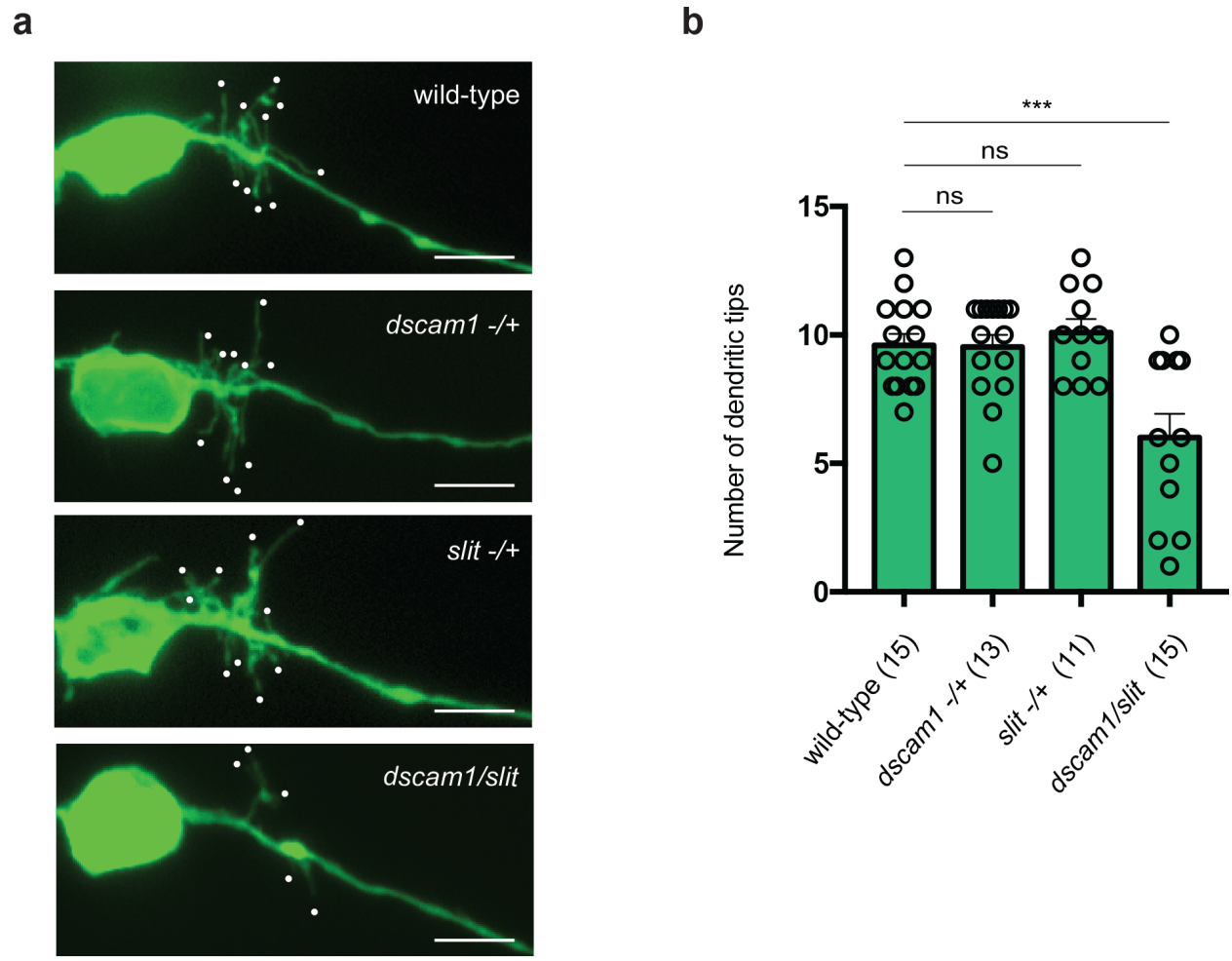


Figure 2.7 | *slit* and *dscam1* genetically interact in aCC dendritogenesis. To test the genetic interaction between *slit* and *dscam1* in aCC denritogenesis, single aCC neurons were labeled by lipophilic dye tracing in the following genotypes at hour 15: wild-type, *dscam1* -/+, *slit* -/+, and $slit^{dui}/slit^{dui}$ (**a**). The number of dendritic tips were counted in each neuron (**b**). Scale bars, 5 μm. Sample sizes are indicated in parentheses. Data are represented as mean ± SEM. ***p < 0.001.

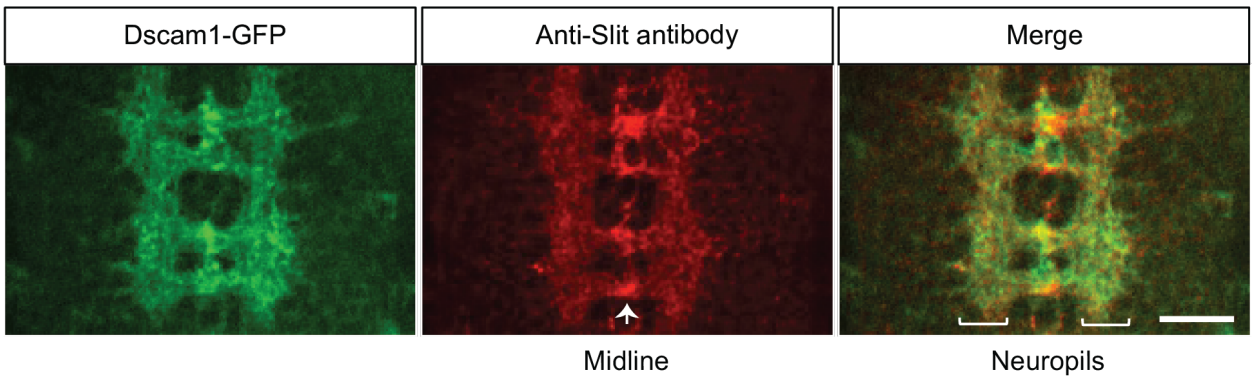


Figure 2.8 | Dscam1 and Slit co-localize in the CNS. Interplay between Slit and Dscam1 was evaluated by localization study in the CNS. Endogenous Dscam1 was visualized using the *Dscam1*^{MiMICGFP} fly line and counterstained with anti-Slit antibody. The main source of Slit secretion is the midline glia, evidenced by a strong accumulation of Slit along the midline. The midline is indicated by the arrow. Another accumulation of Slit is observed in neuropils, in which the aCC-MP1 contact transpires. Neuropils are indicated by brackets. Scale bar, 10 μm.

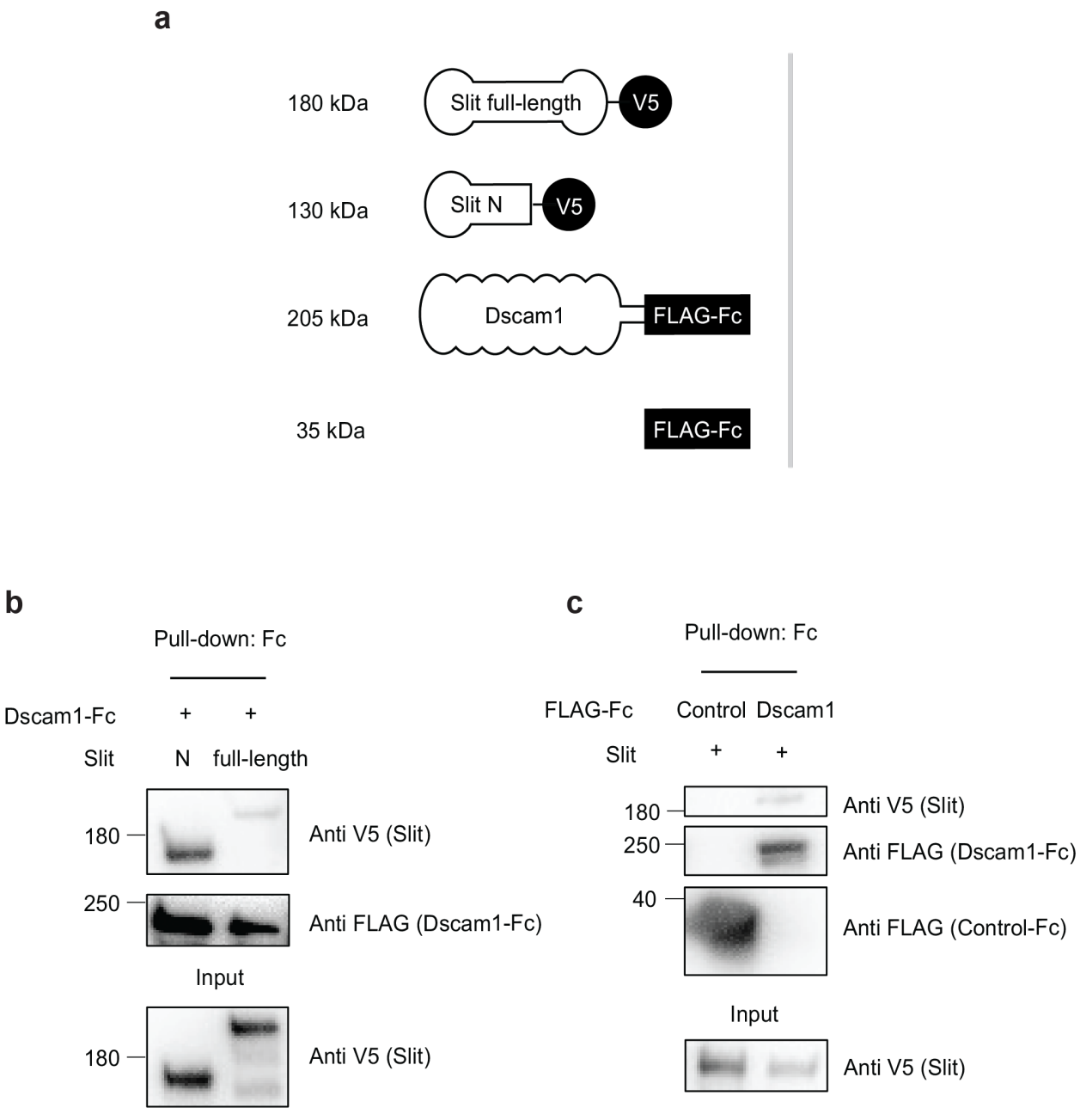


Figure 2.9 | Both full-length and N-terminal Slit form a complex with Dscam1. (**a**) Schematic representation of fusion proteins designed to test Slit-Dscam1 binding. Two versions of Slit, full-length Slit and N-terminal Slit (Slit N), were V5-tagged C-terminally. Extracellular portion of Dscam1 was concatenated with Flag tag and Fc tag. FLAG-Fc was designed for control. Molecular

weights of the corresponding designed proteins are indicated. (**b**) Full-length Slit and Slit N were co-expressed with Dscam1-Fc in HEK293T cell. Fc-fusion proteins were pulled down from the cell lysates using magnetic Protein A/G beads. Western blots were performed using anti-V5 and anti-FLAG antibodies. (**c**) To exclude the possibility of Slit binding to Fc tag, FLAG-Fc was tested for binding to Slit. No binding was detected between FLAG-Fc and Slit, suggesting that Slit specifically forms a complex with Dscam1.

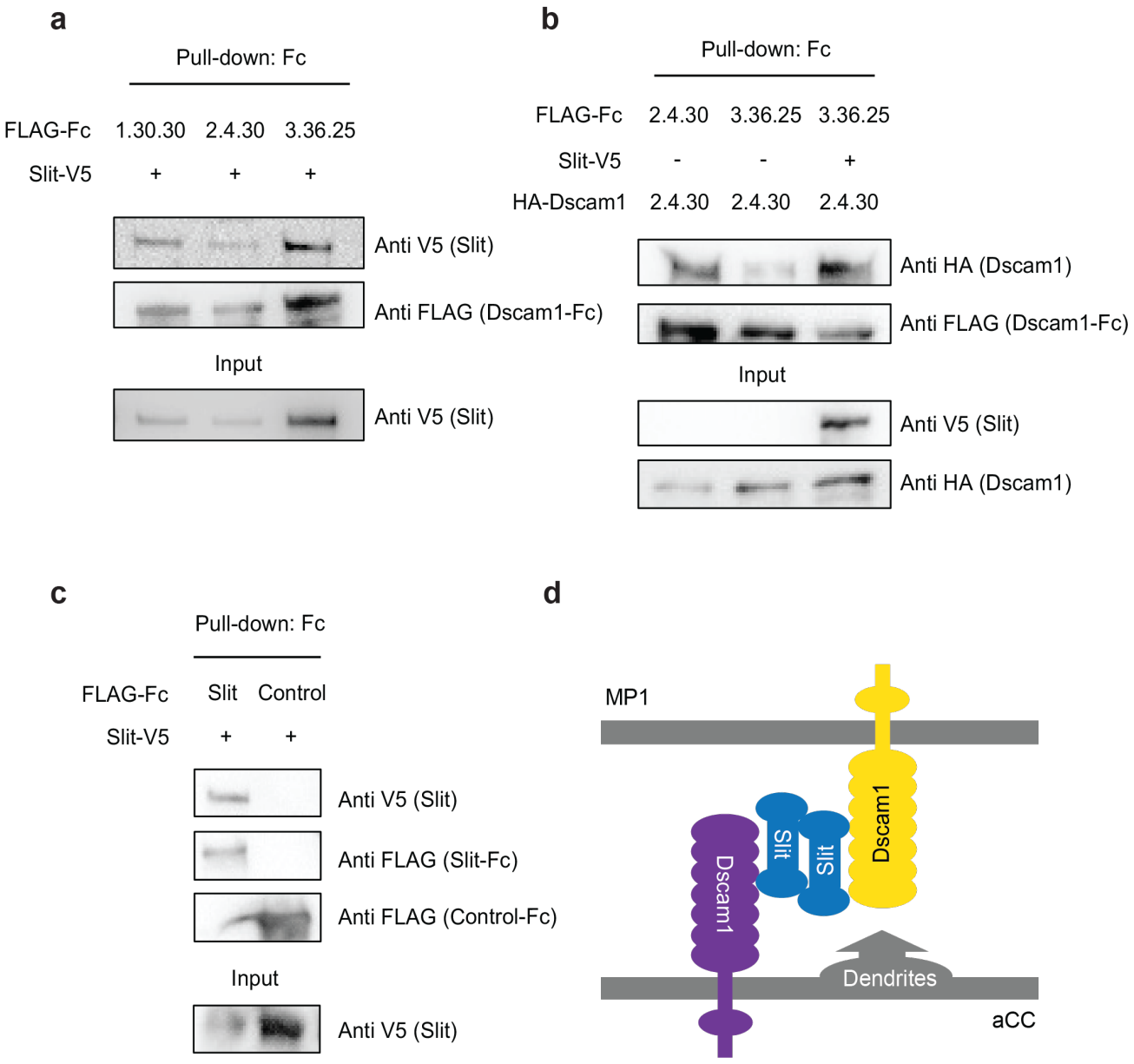


Figure 2.10 | Slit mediates isoform-independent Dscam1 interaction. (**a**) To assess the isoform specificity of Slit-Dscam1 binding, multiple isoforms of Dscam1 (1.30.30, 2.4.30, and 3.36.25) were individually co-expressed with Slit. The corresponding western blots were shown. (**b**) The isoform 2.4.30 and 3.36.35 were examined for binding in the presence or absence of Slit. To differentiate the bait isoform from the prey isoform, one was FLAG-Fc-tagged and the other was

HA-tagged as a transmembrane form. (**c**) To test if Slit undergoes oligomerization, two differentially labeled Slit proteins, Slit-FLAG-Fc and Slit-V5 were co-expressed in 293T cells and pulled down using magnetic Protein A/G beads. FLAG-Fc was used for control. (**d**) Proposed model of Dscam1 interaction at the aCC-MP1 contact sites. The aCC and the MP1 likely express non-matching Dscam1 isoforms. Slit potentially mediates Dscam1 interaction between different isoforms by forming Slit clusters (*e.g.,* dimers) to accommodate Dscam1 molecules/isoforms.

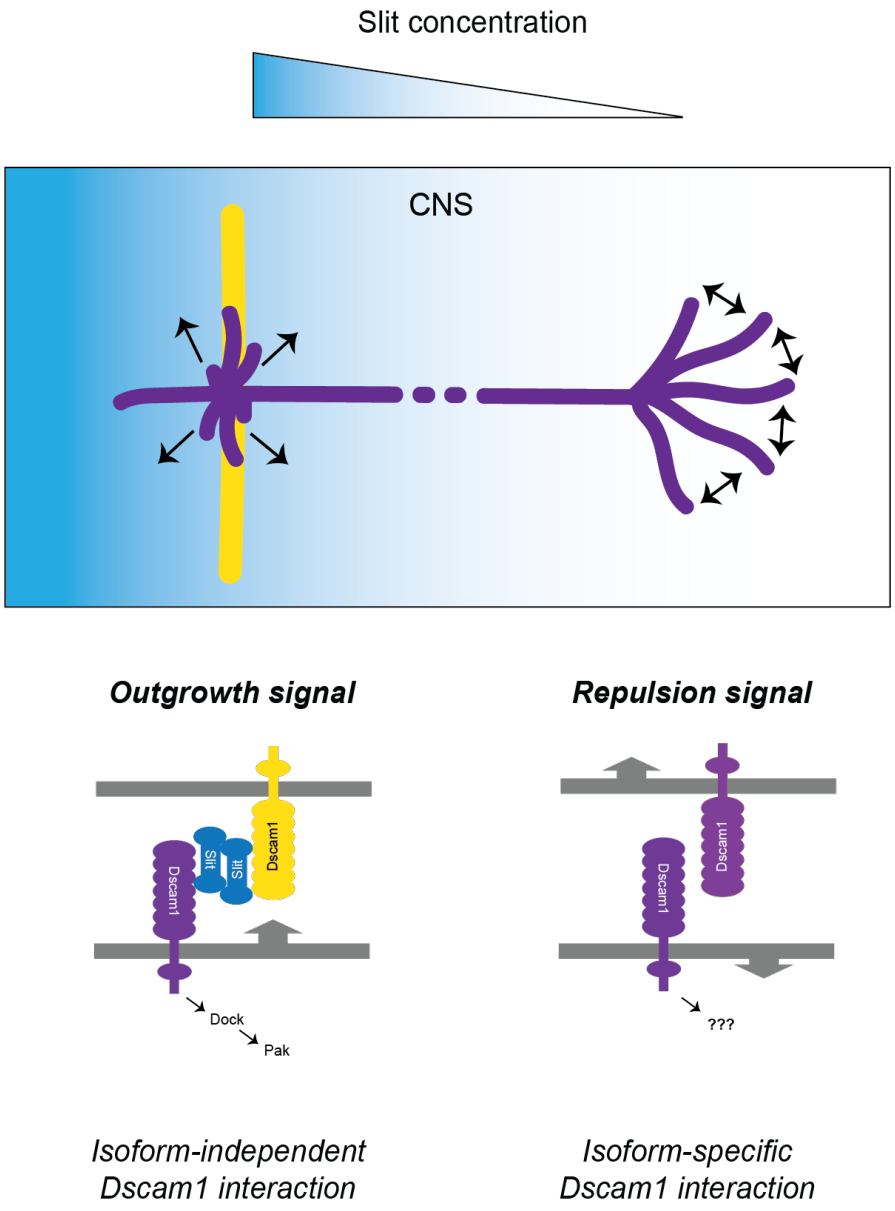


Figure 2.11 | Model of distinctive Dscam1 functions in the CNS. Schematic of distinct roles of Dscam1 signaling in neuronal wiring in the brain. Others and we have collectively demonstrated that Dscam1 signaling promotes neurite growth or branching in the presence of Slit, as opposed to its well-characterized function in repulsion. Such distinction in Dscam1 functions may arise depending on the level of ligands in the microenvironment and the level of intracellular signaling factors.

2.6 METHOD

Molecular cloning

For insect expression of Dscam1, the protein coding region of Dscam1$^{1.30.30.2}$ was amplified from *pcDNA3-V5-His* [30] using the following primers: 5'-AATTGGGAATTCGTTAACAGATCT GCGGCCGCATGAATATGCCCAACGAACGCCTCAA-3' and 5'-TCCTTCACAAAGATCCT CTAGATTACACTGCCATAGTATCGTAGGCTGTGAATCCT-3'. The amplicon was assembled into *pACUH* plasmid (Addgene) between *Not*I and *Xba*I sites to construct *pACUH-Dscam1$^{1.30.30.2}$*. *pACUH-Dscam1$^{1.30.30.2\text{-}FLAG}$* was generated by inserting the 3xFLAG tag coding sequence into *pACUH-Dscam1$^{1.30.30.2}$* using the following primers: 5'-CATGACATCGATTACA AGGATGACGATGACAAGAATGAGATCTCAGAGGCAGAGT-3' and 5'-ATCTTTATAAT CACCGTCATGGTCTTTGTAGTCACTCTCCTTGGAGTCGTTGAA-3' [18]. *pACUH-Dscam1$^{V5\text{-}1.30.30.2}$* was generated by inserting the V5 tag coding sequence into *pACUH-Dscam1$^{1.30.30.2}$* using the following primers: 5'-CCTCTCCTCGGTCTCGATTCTACGGCCAAT CCCCCAGATGCC-3' and 5'-GTTAGGGATAGGCTTACCAGCTAGGGTCTGACTACCAC AAG-3'. *pACUH-Dscam1$^{HA\text{-}1.30.30.2}$* was created by inserting the 3xHA tag coding sequence into *pACUH-Dscam1$^{1.30.30.2}$* using the following primers: 5'-CCCGGACTATGCCTATCCCTACGAC GTGCCAGACTACGCTGCCAATCCCCCAGATGCCGA-3' and 5'- ACGTCATACGGGTAG GCGTAATCGGGCACATCGTATGGGTAAGCTAGGGTCTGACTACCA-3'. The corresponding mammalian expression plasmids were constructed by subcloning the above protein coding regions into *pcDNA3.1* plasmid (V79020, Thermo Fisher Scientific) between *Not*I and *Xba*I sites.

To create *pcDNA3.1-Dscam1$^{1.30.30.2}$-GFP*, the EGFP coding sequence was amplified from *pcDNA3.1-EGFP* (Addgene) using the following primers: 5'-TTCCGGTGGAAGCGGAGGT

AGCGGC-3’ and 5’- AACCTCCGCTTCCACCGCTACCTCC-3’. The amplicon was cloned into an engineered *Xba*I site that was introduced between exon 9 and 10 using the following primers: 5’-GCCGCTACCTCCGCTTCCACCGGAACCTCCGCCACCGTAGACGTTCAGCGGCGTG GT-3’ and 5’-ACCGGTGGAGGTAGCGGTGGAAGCGGAGGTTCCGGAGTACCTCCCAGA TGGATCCTCGA-3’. To construct *pcDNA3.1-Dscam1*$^{1.30.30}$*-FLAG-Fc*, *pcDNA3.1-Dscam1*$^{1.30.30}$*-FLAG* was first created by amplifying an extracellular region of Dscam1$^{1.30.30.2}$ using two primers (5’-TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGAATATGCCCAACGAACGCCT CAAAT-3’ and 5’-TTTAAACGGGCCCTCTAGATTACTTGTCATCGTCATCCTTGTAATC GATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCAACGGTTAAGGTGGC AAAGTCGTATT-3’) and cloning the amplicon into *pcDNA3.1* between *Kpn*I and *Xba*I sites. Subsequently, the Fc region of human IgG was amplified by two primers (5’-AAGGATGACGAT GACAAGTCTAGAGACAAAACTCACACATGCCCACCGT-3’ and 5’-AGCGGGTTTAAAC GGGCCCTCTAGATTATTTACCCGGAGACAGGGAGAGGCT-3’) and the amplicon was assembled into the *Xba*I site in *pcDNA3.1-Dscam1*$^{1.30.30}$*-FLAG*. Expression plasmids coding for the other Dscam1 isoforms were constructed similarly.

For mammalian expression of full-length Slit, an uncleavable species of Slit (Slit-UC) was used to circumvent proteolytic cleavage observed in HEK293T cells [30]. *pSecTag-Slit-UC* was made by removing a nucleotide sequence encoding the 9-aa cleavage recognition site (HNMISMMYP) in *pSecTag-Slit* [30], using the following primers: 5’-CAGACGTCGCCTTG TCAAAACCACGA-3’ and 5’-GCCTTCACAGTACTTGCCCGTATAGTCGTCT-3’. Detailed information of each construct used in this study is available upon request.

Cell culture and transfection

HEK293T and Cos7 cells (gifted by Dr. Kipreos, the University of Georgia) were cultured in Dullbecco's Eagle's medium (DMEM, HyClone) with 10% fetal bovine serum (Atlanta Biologicals) in a humidified incubator at 37 °C and 5% CO_2. *Drosophila* S2 cells (*Drosophila* Genomic Resource Center) were maintained in SFX-Insect medium (HyClone) at 25 °C. For transient gene expression, HEK293T and Cos7 cells were transfected with polyethylenimine (PEI, Polysciences) and S2 cells were transfected with Effectene (Qiagen) according to the manufactures' instructions.

Bead binding to cell assay

Bead binding to cell assay was adopted from a previously described method [18]. HEK293T cells in 6-well culture plate were transfected with 1500 ng of *pcDNA3.1-Dscam1*$^{1.30.30}$*-myc-Fc* or *pcDNA3.1-Dscam1*$^{2.4.30}$*-myc-Fc* for 4 days. Culture supernatants containing Dscam1$^{1.30.30}$-myc-Fc or Dscam1$^{2.4.30}$-myc-Fc were collected and incubated with 5 μL of Protein A/G Magnetic Beads (Thermo Fisher Scientific) rotating at 25 °C overnight. Beads were washed three times with PBS-T (0.01% Triton-X100) and incubated with 0.3 μL of goat anti-Mouse antibody, Alexa Fluor™ 647 (A31571, Invitrogen). Subsequently, beads were washed three times with PBS-BSA (1% BSA) and ultimately reconstituted in 1 mL of PBS-BSA (1% BSA). Cos7 cells expressing Dscam1$^{1.30.30.2}$-GFP were seeded onto 8-well glass-bottom chamber (Nunc Lab-Tek II Chambered Cover Glass, Nalge Nunc International) coated with 1 μg/mL fibronectin (Sigma-Aldrich) and cultured overnight at 37 °C and 5% CO_2. Dscam1-Fc-coated beads were incubated with cells for 2 hr and washed three times with DMEM. Confocal microscopy was used to visualize beads binding to cells.

ProteinA/G affinity purification

HEK293T cells in T25 flask were transfected with 2000 ng of *pcDNA3.1-Dscam1-FLAG-Fc* and 4000 ng of *pSecTag-SlitFL-UC* for 5 days. Cells were collected, washed three times with PBS, and lysed with 1 mL of NP40 buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.5% NP40). Cell lysates were centrifuged and enriched with 5 μL of Protein A/G beads. Five washes were performed with NP40 buffer. Reduced eluates were run on SDS-PAGE and subjected to western blotting. Proteins were detected with anti-FLAG (F1804, Sigma-Aldrich), anti-V5 (46-0705, Invitrogen), or anti-HA (H3663, Sigma-Aldrich) mouse antibodies at 1:1000, followed by goat anti-mouse IgG antibody conjugate to horseradish peroxide (31430, Thermo Fisher Scientific) at 1:2000.

Fly strains

Lipophilic dye labeling for dendritic morphology analyses was performed using the following mutant alleles: sli^{2} (Dr. Thomas Kidd, the University of Nevada), sli^{dui} (9284, Bloomington *Drosophila* Stock Center), $dscam1^{21}$, and $dscam1^{23}$ (Dr. Jiang Wang). Endogenous localization of Dscam1 was visualized using the previously established $Dscam1^{\mathrm{MiMICGFP}}$ strain [15]. All strains were maintained at 25 °C. All crosses were performed using standard genetic procedures.

Immunofluorescence

Dscam1 accumulation in S2 cells was visualized by transiently expressing Dscam1 isoforms. S2 cells in 24-well culture plate were transfected with either 100 ng of *pACUH-*$Dscam1^{V5\text{-}1.30.30.2}$, *pACUH-*$Dscam1^{1.30.30.2\text{-}flag}$, or *pACUH-*$Dscam1^{2.4.30.2\text{-}flag}$ with 100 ng of *pAC-GAL4* (Addgene) for gene expression induction. As isoform markers, either 100 ng of *pACUH-mNeonGreen2* or *pACUH-SH2-mCherry* was co-transfected. After 48 hr, each transfectant was washed three times with PBS and diluted in SFX-Insect medium. Two transfected populations expressing matching or

non-matching isoforms marked by distinct fluorophores were co-incubated in 8-well glass-bottom chamber coated with 0.5 mg/mL Concanavalin A (Sigma-Aldrich) at 25 °C for 48 hr. Each sample was fixed with 4% paraformaldehyde in PBS, blocked in PBS-BSA (0.1% BSA) for 10 min, and stained with mouse anti-FLAG or mouse anti-V5 antibody at 1:200 overnight. Following three washes with PBS-BSA (0.1% BSA), sample was stained with goat anti-mouse IgG antibody conjugate to Alexa Fluor™ 647 at 1:1000 for 1 hr. Sample was washed five times with PBS and mounted for confocal microscopy. Embryos were fillet-dissected as previously illustrated [32]. Immunostaining of endogenous Slit in S2 cells and fillet-dissected embryos was performed similarly using mouse antibodies obtained from Developmental Studies Hybridoma Bank: anti-Slit (C555.6D), or anti-Fas2 (1D4) at 1:100.

Confocal microscopy

Confocal images of S2 cells, Cos7 cells and dissected embryos were acquired using an inverted microscope (Eclipse Ti-E, Nikon) with a 100 × 1.45 NA oil immersion objective (Plan Apo, Nikon). The microscope was linked to the Dragonfly spinning disk confocal system (CR-DFLY-501, Andor). Two excitation lasers (40 mW 488 nm and 110 mW 637 nm lasers) were coupled to a multimode fiber passing through the Andor Borealis unit. A dichroic mirror (Dragonfly laser dichroic for 405-488-561-640) and two band-pass filters (525/50 nm and 725/40 nm bandpass emission wheel filters) were utilized. Images were collected with an EM-CCD camera (iXon, Andor) and processed using Fiji software (The National Institute of Health).

Lipophilic dye labeling

For surveying dendritic morphology of aCC neurons, wild-type and mutant embryos were fillet-dissected at 15:00 AEL following the identical procedure described in the preceding section. To

visualize primary dendrites of single aCC neurons, lipophilic dye labeling was performed as previously illustrated [32].

Statistical analysis

Unpaired t-tests were implemented for quantification of bead binding to cell assay and Dscam1 accumulation at cell-cell contacts. All relevant experiments were repeated multiple times independently for validation. Statistical analyses were executed by GraphPad Prism 7. Error bars in all figures denote the standard error of the mean.

2.7 REFERENCES

1. Sugimura, K., et al., *Development of morphological diversity of dendrites in Drosophila by the BTB-zinc finger protein abrupt.* Neuron, 2004. **43**(6): p. 809-22.
2. Lefebvre, J.L., *Molecular mechanisms that mediate dendrite morphogenesis.* Curr Top Dev Biol, 2021. **142**: p. 233-282.
3. Jan, Y.N. and Jan, L.Y., *Branching out: mechanisms of dendritic arborization.* Nat Rev Neurosci, 2010. **11**(5): p. 316-28.
4. Dong, X., et al., *Intrinsic and extrinsic mechanisms of dendritic morphogenesis.* Annu Rev Physiol, 2015. **77**: p. 271-300.
5. Yogev, S. and Shen, K., *Cellular and molecular mechanisms of synaptic specificity.* Annu Rev Cell Dev Biol, 2014. **30**: p. 417-37.
6. Sanes, J.R. and Zipursky, S.L., *Design principles of insect and vertebrate visual systems.* Neuron, 2010. **66**(1): p. 15-36.
7. Teichmann, H.M. and Shen, K., *UNC-6 and UNC-40 promote dendritic growth through PAR-4 in Caenorhabditis elegans neurons.* Nat Neurosci, 2011. **14**(2): p. 165-72.

8. Dong, X., et al., *An extracellular adhesion molecule complex patterns dendritic branching and morphogenesis.* Cell, 2013. **155**(2): p. 296-307.

9. Liu, X., et al., *Receptor tyrosine phosphatase CLR-1 acts in skin cells to promote sensory dendrite outgrowth.* Dev Biol, 2016. **413**(1): p. 60-9.

10. Zou, W., et al., *A Dendritic Guidance Receptor Complex Brings Together Distinct Actin Regulators to Drive Efficient F-Actin Assembly and Branching.* Dev Cell, 2018. **45**(3): p. 362-375 e3.

11. Zou, W., et al., *A multi-protein receptor-ligand complex underlies combinatorial dendrite guidance choices in C. elegans.* Elife, 2016. **5**.

12. Ramirez-Suarez, N.J., et al., *Axon-Dependent Patterning and Maintenance of Somatosensory Dendritic Arbors.* Dev Cell, 2019. **48**(2): p. 229-244 e4.

13. Landgraf, M., et al., *The Origin, Location, and Projections of the Embryonic Abdominal Motorneurons of Drosophila.* J Neurosci, 1992.

14. Kamiyama, D. and Chiba, A., *Endogenous Activation Patterns of Cdc42 GTPase Within Drosophila Embryos.* Science, 2009.

15. Kamiyama, D., et al., *Specification of Dendritogenesis Site in Drosophila aCC Motoneuron by Membrane Enrichment of Pak1 through Dscam1.* Dev Cell, 2015. **35**(1): p. 93-106.

16. Yamakawa, K., et al., *DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system.* Hum Mol Genet, 1998.

17. Schmucker, D., et al., *Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity.* Cell, 2000.

18. Wojtowicz, W.M., et al., *Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding.* Cell, 2004. **118**(5): p. 619-33.

19. Wojtowicz, W.M., et al., *A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains.* Cell, 2007. **130**(6): p. 1134-45.

20. Wu, W., et al., *Complementary chimeric isoforms reveal Dscam1 binding specificity in vivo.* Neuron, 2012. **74**(2): p. 261-8.

21. Neves, G., et al., *Stochastic yet biased expression of multiple Dscam splice variants by individual cells.* Nature Genetics, 2004. **36**(3): p. 240-246.

22. Miura, S.K., et al., *Probabilistic splicing of Dscam1 establishes identity at the level of single neurons.* Cell, 2013. **155**(5): p. 1166-77.

23. Grueber, W.B., et al., *Tiling of the Drosophila epidermis by multidendritic sensory neurons.* Development, 2002.

24. Hughes, M.E., et al., *Homophilic Dscam interactions control complex dendrite morphogenesis.* Neuron, 2007. **54**(3): p. 417-27.

25. Andrews, G.L., et al., *Dscam guides embryonic axons by Netrin-dependent and -independent functions.* Development, 2008. **135**(23): p. 3839-3848.

26. Alavi, M., et al., *Dscam1 Forms a Complex with Robo1 and the N-Terminal Fragment of Slit to Promote the Growth of Longitudinal Axons.* PLoS Biol, 2016. **14**(9): p. e1002560.

27. Dascenco, D., et al., *Slit and Receptor Tyrosine Phosphatase 69D Confer Spatial Specificity to Axon Branching via Dscam1.* Cell, 2015. **162**(5): p. 1140-54.

28. Zipursky, S.L. and Grueber, W.B., *The molecular basis of self-avoidance.* Annu Rev Neurosci, 2013. **36**: p. 547-68.

29. Andrews, G.L., et al., *Dscam guides embryonic axons by Netrin-dependent and -independent functions.* Development, 2008. **135**(23): p. 3839-48.

30. Alavi, M., et al., *Dscam1 Forms a Complex with Robo1 and the N-Terminal Fragment of Slit to Promote the Growth of Longitudinal Axons.* PLOS Biology, 2016. **14**(9).

31. Furrer, M.P., et al., *Slit and Robo control the development of dendrites in Drosophila CNS.* Development, 2007. **134**(21): p. 3795-804.

32. Inal, M.A., et al., *Retrograde Tracing of Drosophila Embryonic Motor Neurons Using Lipophilic Fluorescent Dyes.* J Vis Exp, 2020(155).

33. Kidd, T., et al., *Slit Is the Midline Repellent for the Robo Receptor in Drosophila.* Cell, 1999.

34. Seiradake, E., et al., *Structure and functional relevance of the Slit2 homodimerization domain.* EMBO Rep, 2009. **10**(7): p. 736-41.

35. Ordan, E. and Volk, T., *A non-signaling role of Robo2 in tendons is essential for Slit processing and muscle patterning.* Development, 2015. **142**(20): p. 3512-8.

36. Zhan, X.L., et al., *Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies.* Neuron, 2004. **43**(5): p. 673-86.

37. Ordan, E. and Volk, T., *Cleaved Slit directs embryonic muscles.* Fly (Austin), 2015. **9**(2): p. 82-5.

CHAPTER 3

MULTIPLEXED LABELING OF CELLULAR PROTEINS WITH SPLIT FLUORESCENT PROTEIN TAGS[2]

[2] Tamura, R., Jiang, F., Xie, J., and Kamiyama, D. Reprinted here with permission of publisher.

3.1 ABSTRACT

Self-complementing split fluorescent proteins (split $FP_{1\text{-}10/11}$) have become an important labeling tool in live-cell protein imaging. However, current split FP systems to label multiple proteins in single cells have a fundamental limitation in the number of proteins that can be simultaneously labeled. Here, we describe an approach to expand the number of orthogonal split FP systems with spectrally distinct colors. By combining rational design and cycles of directed evolution, we expand the spectral color palette of $FP_{1\text{-}10/11}$. We also circularly permutate GFP and synthesize the β-strand 7, 8, or 10 system. These split GFP pairs are not only capable of labeling proteins but are also orthogonal to the current $FP_{1\text{-}10/11}$ pairs, offering multiplexed labeling of cellular proteins. Our multiplexing approach, using the new orthogonal split FP systems, demonstrates simultaneous imaging of four distinct proteins in single cells; the resulting images reveal nuclear localization of focal adhesion protein zyxin.

3.2 INTRODUCTION

In the self-complementing split GFP system, super-folder GFP is split between β-strands 10 and 11, rendering 214-amino acid and 16-amino acid fragments [1]. The short fragment, $GFP_{11M3\ OPT}$, is inserted into a gene of interest, acting as an epitope tag [2]. When expressed in the same cell, the $GFP_{1\text{-}10\ D7}$ and $GFP_{11M3\ OPT}$ fragments (hereafter referred to as $GFP_{1\text{-}10/11}$) spontaneously interact with each other to form a functional GFP (**Supplementary Figure 1.1**). The GFP11 fragment has been used in numerous biological studies [3, 4]: targeting nanomaterials in cells [5, 6], forming protein oligomeric structures [2, 7], verifying aggregation processes [8], and imaging protein localization in living cells [9]. With additional substitutions in the $GFP_{1\text{-}10}$ fragment, cyan and yellow spectral variants have previously been created and used to visualize localization

patterns of cellular proteins [2, 10]. The majority of substitutions which lead to the spectral shifts in these variants are located within the large fragments (i.e., CFP_{1-10} and YFP_{1-10}). These fragments also retain the capability to bind to the identical GFP_{11} fragment, so that reconstitution with GFP_{11} produces a functional cyan or yellow FP.

These self-complementing split GFP variants have already become a powerful and versatile tool for various microscopy applications. In particular, the short fragment of GFP_{11} can be efficiently introduced at the genomic locus of the gene of interest without perturbing local genomic structure, allowing us to reliably produce endogenously tagged cell lines [2, 11]. Additionally, we have successfully generated a library of human cells with GFP_{11}-tagged endogenous proteins via CRISPR/Cas9-mediated homology-directed repair (HDR) and demonstrated that GFP_{11}-tag is compatible with a wide range of cellular proteins including enzymes, receptors, transport proteins, and structural proteins [12]. However, one of the challenges is labeling multiple proteins simultaneously in single cells. Multiplexed visualization is tremendously beneficial for simultaneous comparisons of protein dynamics. Recently, great advances have been made in split super-folder Cherry ($sfCherry_{1-10/11}$) as a second, orthogonal split FP system [2, 13, 14]. The GFP_{11} and $sfCherry_{11}$ fragments allow simultaneous labeling of two different proteins. Although this multicolor approach has expanded the potential of split-FP labeling, it has a bottleneck in multiplexing because of the limited number of available orthogonal split FP systems with different colors.

In this report, we expand the color palette of self-associating split FPs. We have introduced rational mutations into the amino acid sequence of EBFP2 through site-directed mutagenesis and generated two blue-colored split FPs, $EBFP2_{1-10/11}$ and $Capri_{1-10/11}$. We have also engineered self-associating fragments of mRuby3 (mRuby3 is a red-colored FP with a shorter-wavelength than

sfCherry) [15]. We have evolved mRuby3$_{1-10}$ by a directed evolution strategy to increase its complementation with mRuby3$_{11}$. Our final optimized construct, split mRuby4, becomes an excellent fusion partner when expressed in human cells. In addition, we propose a new approach to generate more orthogonal split FPs using circularly permutated FP fragments. This approach can potentially overcome multiplexing limitations of split-FP labeling. Finally, as a proof-of-concept experiment, we applied our technique to visualize differential distribution of four proteins in single human cells and found that focal adhesion protein zyxin was often accumulated in the nucleus.

3.3 RESULTS AND DISCUSSION

Rationally designed variants of split BFP and CFP

To expand our color palette of split FPs, we began with splitting EBFP2 at the same site as GFP$_{1-10/11}$ (note that EBFP2 is 4-fold brighter and > 500-fold more photostable than EBFP [16]). While the short fragment is identical to the amino acid sequence of GFP$_{11}$, six substitutions have been introduced into the large fragment through site-directed mutagenesis (N40I/T106K/E112V/K166T/I167V/S206T; the numbering of amino acids follows that of EBFP2). These substitutions have been previously shown to enhance complementation efficiency of GFP$_{1-10}$ variants [1]. To verify *in vivo* complementation between the two fragments, we used GFP$_{11}$-tagged β-actin and histone 2B. Co-expressing each one with EBFP2$_{1-10}$ in HeLa cells, we observed blue fluorescence in images of actin stress fibers or the nucleoplasm (**Figure 3.1a-b**). Because its overall fluorescent signal is extremely weak, autofluorescence limits the usefulness of this split construct in some cases (e.g., the actin image). In particular, high autofluorescence background with UV light is often observed in the perinuclear region (**Supplementary Figure**

3.2). To improve its overall brightness, we decided to add six more substitutions to $EBFP2_{1-10}$ (S65T/Q80R/F99S/V128T/M153T/V163A; some of these have previously been characterized to promote the stability and folding rate of GFP [1, 17]). This new split FP, termed split Capri for its cyan-blue color, has the same absorption spectrum as split EBFP2 (**Supplementary Figure 3.3**). The emission spectrum, however, is red-shifted from split EBFP2 by 20 nm (λ_{abs} / λ_{em} = 379 / 469 nm). Furthermore, it has a peak extinction coefficient of 37,300 $M^{-1}cm^{-1}$ and quantum yield of 0.13, which surpass the extinction coefficient and quantum yield of split EBFP2 (**Supplementary Table 3.1**). Associated with GFP_{11}-tagged β-actin or H2B, $Capri_{1-10}$ exhibits very bright fluorescence in HeLa cells (**Figure 3.1c-d**). To assess the improvement in the resulting brightness, we co-expressed GFP_{11}-H2B in HEK 293T cells with either $EBFP2_{1-10}$ or $Capri_{1-10}$. Quantifying the fluorescence intensity of cells by flow cytometry, we found that $Capri_{1-10/11}$ had a four-fold brighter fluorescence than $EBFP2_{1-10/11}$ (**Supplementary Figure 3.4**).

In addition to BFP variants, cyan-colored FPs have been widely accepted. When we introduced substitutions into GFP_{1-10} (Y66W to make CFP_{1-10}), complementation fluorescence was observed for GFP_{11}-actin or GFP_{11}-H2B fusions co-expressing with CFP_{1-10} in HeLa cells (**Supplementary Figure 3.5**). It is, however, noticeable that the overall brightness of CFP_{1-10} is relatively weak, making it difficult to visualize thin actin filaments (A recent *in vitro* assessment also reported that split CFP has a low brightness [10]). Therefore, we sought to produce a cyan-colored FP that has enhanced brightness. A recent improvement of full-length CFP, named Cerulean, increases the brightness by ~1.6 times [18]. Because the known substitutions are located only on GFP_{1-10} (Y66W/S72A/Y145A/H148D for Cerulean), $Cerulean_{1-10}$ can associate with GFP_{11}. To evaluate the enhancement in its overall brightness for cellular microscopy, we prepared plasmids encoding $Cerulean_{1-10}$ or CFP_{1-10}. We co-transfected each one of these plasmids with a

GFP_{11}-H2B plasmid in HEK 293T cells. Imaging by confocal microscopy, we quantified the signal level of these split FPs. We found that $Cerulean_{1-10}$ signal was ~1.7 times brighter than that of CFP_{1-10} ($p < 0.001$, Student's t-test; **Supplementary Figure 3.6**). We next assessed the performance of $Cerulean_{1-10}$ when used as a fusion tag. We used GFP_{11}-fused β-actin or H2B and co-expressed each one with $Cerulean_{1-10}$ in HeLa cells. Although we observed complementation of $Cerulean_{1-10}$ in the appropriate locations, some cells exhibited thicker actin bundles, which we have never seen in cells expressing a full-length Cerulean fusion (**Figure 3.1** and **Supplementary Figure 3.7**). Because this artifact is common for dimeric or tetrameric FPs when these FPs are targeted to two-dimensional structures [19, 20], we suspect that $Cerulean_{1-10/11}$ is oligomeric. Nonetheless, an in-depth investigation is required to validate such a property in split Cerulean and under more various experimental conditions.

Engineering of a red-colored split FP variant based on mRuby3

Although developmental efforts are ongoing to improve the brightness of split sfCherry [2, 13, 14], having spectrally distinct split red FPs would foster the gross usefulness of FP_{11}-tags. Since split sfCherry2 has a far-red shifted emission peak at 610 nm, we sought to explore the evolution of orange-red FPs such as mKO2, mRuby3, mApple, and mScarlet-I [15, 21-23] in *E. coli*. Following the previously established approach [13], we inserted a 30 amino-acid spacer between the 10^{th} and 11^{th} β-strand of the four FPs. The long spacer insertion greatly diminished colony fluorescence of mKO2, mApple and mScarlet-I, while colonies expressing spacer-inserted mRuby3 were fluorescent (**Supplementary Figure 3.8**). To improve the brightness of the spacer-inserted mRuby3, we mutagenized it using error-prone PCR and then transformed into *E. coli*; the three brightest candidates were pooled and subjected to another round. After three rounds, brightness of the best candidate revamped six-fold relative to that of spacer-inserted mRuby3

(**Supplementary Figure 3.9**). We found seven substitutions in mRuby3$_{1-10}$ (M15T/Q27H/T31I/V106I/S113C/R126S/A154V), termed split mRuby4 (λ_{abs} / λ_{em} = 557 / 592 nm; see also **Supplementary Figure 3.3** for its absorbance spectrum). These efforts have resulted in the creation of a particularly bright variant, which has a higher extinction coefficient and increased quantum yield when compared to split mRuby3 (**Supplementary Table 3.1**).

To assess whether split mRuby4 could fluoresce in human cells, we over-expressed mRuby4$_{11}$-actin with mRuby4$_{1-10}$ in HeLa cells. Complemented split mRuby4 has a bright signal in fluorescent images of actin and various fusion proteins (**Figure 3.2a-f**). To determine the signal level of split mRuby4, we performed a cellular fluorescence measurement by flow cytometry and compared the signal to full-length mRuby3. With expression of spacer-inserted mRuby4 in HEK 239T cells, we found that its signal level became around 69 % of full-length mRuby3 (**Supplementary Figure 3.10**). We have also demonstrated that mRuby4$_{1-10/11}$ has sufficient efficiency to detect proteins expressed at endogenous levels. We employed CRISPR/Cas9-mediated HDR and introduced a 200-nucleotide ssDNA donor into the H2B locus of HEK 293FT cells expressing mRuby4$_{1-10}$. Subsequently, split mRuby4 complementation had a prominent signal in images of the H2B knock-in (**Supplementary Figure 3.11**).

As shown in Figure 3.2g, the emission peaks for split mRuby4 and split sfCherry2 are only 20 nm apart yet still visually distinguishable (**Supplementary Figure 3.12**). To further evaluate how many split FPs could be simultaneously visualized in different cells, we performed spectral imaging of HEK 293 cells expressing H2B fused proteins. We co-cultured six types of HEK 293 cells, each of which expressed H2B labeled with either EBFP2$_{1-10/11}$, Capri$_{1-10/11}$, Cerulean$_{1-10/11}$, GFP$_{1-10/11}$, mRuby4$_{1-10/11}$, or sfCherry2$_{1-10/11}$. Synchronized at the G2/M phase by release from a cyclin-dependent kinase inhibitor, the cells were imaged (**Supplementary Figure 3.13**). Within a

couple of hours, 20% of the cell population was in cytokinesis, consistent with previous literature [24]. Cells with each split FP fusion were then captured at different stages of mitosis (**Figure 3.2h**). Altogether, these experiments illustrate six-color spectral imaging of cellular proteins.

Evaluation of split $FP_{1-10/11}$ systems for multiplexed imaging in single cells

An orthogonal interaction between $GFP_{1-10/11}$ and $sfCherry2_{1-10/11}$ – meaning GFP_{11} can interact with GFP_{1-10}, but cannot interact with $sfCherry2_{1-10}$ – is the basis of simultaneous labeling of two different fusion proteins in single cells. With an array of $FP_{1-10/11}$ pairs developed, we sought to systematically test their binding specificities by flow cytometry. $GFP_{1-10/11}$, $sfCherry2_{1-10/11}$, $mNeonGreen2_{1-10/11}$ [13], and $mRuby4_{1-10/11}$ were examined for complementation in HEK 293T cells. Each FP_{1-10} fragment was co-expressed with any one of four FP_{11}-fused β-actin. The FP_{1-10} fragments would primarily bind to their corresponding partner FP_{11}, and the interactions were tested along the grid diagonal (**Fig. 3.3a**). As shown in Figure 3.3a, all four FP_{1-10} fragments reconstituted with their corresponding partners. Interestingly, $mRuby4_{1-10}$ and $sfCherry2_{11}$ formed complementation signal in the same way as $mRuby4_{11}$ and $sfCherry2_{1-10}$ did reciprocally. Crosstalk was somewhat expected to occur between $FP_{1-10/11}$ fragments encoded by closely related FPs (**Fig 3.3b**). To verify dual-color labeling with $mRuby4_{11}$ in single cells, we used HeLa cells co-expressing $GFP_{1-10/11}$- actin with zyxin-$mRuby4_{1-10/11}$, or $mNeonGreen2_{1-10/11}$-actin with $mRuby4_{1-10/11}$-clathrin. The two distinct fluorescence channels did not overlap in the cells (**Figure 3.3c-d**).

A strategy to create new orthogonal split FPs using circularly permutated FP fragments

In order to provide more variants of split FPs orthogonal to existing $FP_{1-10/11}$, we took advantage of circularly permutated GFP variants [25]. By linking the N- and C-termini and cutting out a single β-strand, any one of the eleven β-strands could be a split GFP-tag. We chose to measure complemented signal of the β-strands 7, 8, 9, and 10; the β-strands 1 through 6 were excluded

because the complementary fragments of these strands are unlikely to be water-soluble [26]. To this end, we prepared DNA constructs encoding each of the β-strands fused to β-actin and measured the overall complemented signal of each construct in HEK 293T cells by flow cytometry. We observed fluorescence signal reconstituted from the β-strands 7, 8, and 10 (hereafter named $GFP_{8\text{-}6/7}$, $GFP_{9\text{-}7/8}$, and $GFP_{11\text{-}9/10}$) with their corresponding partners (**Figure 3.4a**). These split GFPs retained 7 - 57% brightness of $GFP_{1\text{-}10/11}$, albeit leaving room for improvement. To validate protein labeling using the β-strands, we generated constructs encoding various cellular proteins fused with GFP_8 and co-expressed each one of them with $GFP_{9\text{-}7}$ in HeLa cells. For three proteins tested, we observed their expected localizations (**Figure 3.4b-d**).

Next, we assessed the binding specificities of $GFP_{8\text{-}6/7}$, $GFP_{9\text{-}7/8}$, $GFP_{11\text{-}9/10}$, and $GFP_{1\text{-}10/11}$. We performed the flow cytometry assay conducted in a grid format as described earlier. Either $GFP_{8\text{-}6}$, $GFP_{9\text{-}7}$, $GFP_{11\text{-}9}$, or $GFP_{1\text{-}10}$ was co-expressed with β-actin fused with the β-strands 7, 8, 9, or 11 in HEK 293T cells. In this experiment, each of the β-strands only binds to its corresponding partner (**Figure 3.4**). For instance, GFP_8 interacts with $GFP_{9\text{-}7}$, but not $GFP_{1\text{-}10}$. This orthogonal interaction was validated by dual-color imaging of U2OS cells, in which GFP_{11}-H2B and GFP_8-Lamin A/C were co-expressed with $Capri_{1\text{-}10}$ and $GFP_{9\text{-}7}$. We observed the exclusion of GFP_8-Lamin A/C from the nucleoplasm where GFP_{11}-H2B predominately localized (**Figure 3.4f**). Taken together, circularly permutated FP fragments can be used to generate additional orthogonal pairs for multiplexed split FP-labeling.

Multicolor images reveal nuclear localization of zyxin

Finally, we assessed the potential of split FP systems for multiplexed labeling of proteins in single cells. As a proof-of-principle, we used four orthogonal split FPs that we thoroughly investigated in this report ($Capri_{1\text{-}10/11}$, $GFP_{9\text{-}7/8}$, $mNeonGreen2_{1\text{-}10/11}$, and $mRuby4_{1\text{-}10/11}$). As shown in Figure

3.4g, the split FPs are distinguishable from each other by using spectral imaging methodology. U2OS cells were transfected to express these split FPs targeted to four distinct proteins (H2B, Lamin A/C, β-actin, and Zyxin); we observed their correct localization. Strikingly, a few cells displayed that some portion of zyxin proteins localized to the nucleus, although the proteins predominantly localized at focal adhesions in these cells. By visual inspection of a total of 145 cells, we found that 37 of these cells exhibited nuclear localization of zyxin during interphase (**Supplementary Figure 3.18**). Because zyxin is a relatively large molecule (> 80 kDa) but does not have a nuclear localization signal, zyxin must enter the nucleus in contact with other components. We observed a similar nuclear localization pattern of zyxin tagged with a full-length FP tag in U2OS cells (50 out of 174 cells), and this observation has also been confirmed in other cell lines [27-29].

For the initial demonstration of multiplexed labeling, split FPs were over-expressed as fusions to target proteins in single cells (**Figure 3.4g**). However, these fusion proteins might be subject to limitations because of the potential for over-expression artifacts (e.g., aberrant organelle and/or cellular morphology). To further verify our observation in the future, this approach will be extended to label endogenous proteins by methods such as CRISPR/Cas9-mediated gene knock-in [12]. Because a split FP tag is 42~63 nucleotides long (which is ~10 times smaller than the size of an intact FP), a short donor oligo can be directly synthesized, making this a cloning free approach (see also **Supplementary Figure 3.11**) [2, 12]. In addition, a small tag such as GFP_{11}-tag can be introduced into a host cell genome at high homologous recombination efficiencies [11]. Such a simple and efficient approach would facilitate the generation of multiple insertions in single cells. Sequences for multiple split FP tags could be either sequentially or simultaneously inserted

into targeted loci in individual cells stably expressing the complementary fragments, enabling multiplexed visualization of endogenous proteins.

3.4 FIGURES

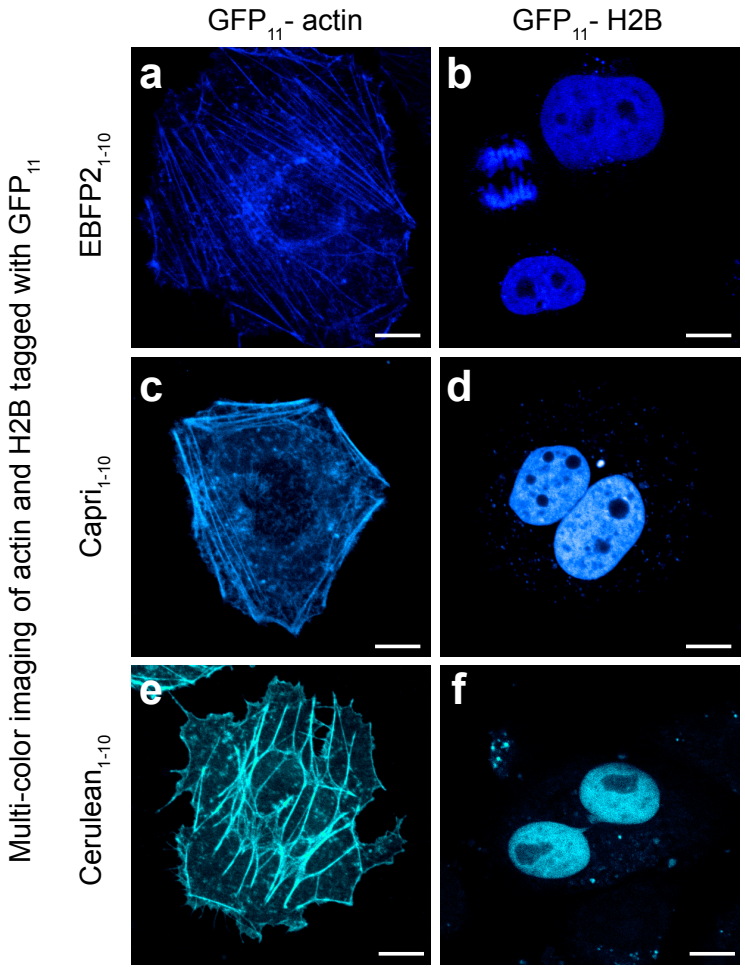
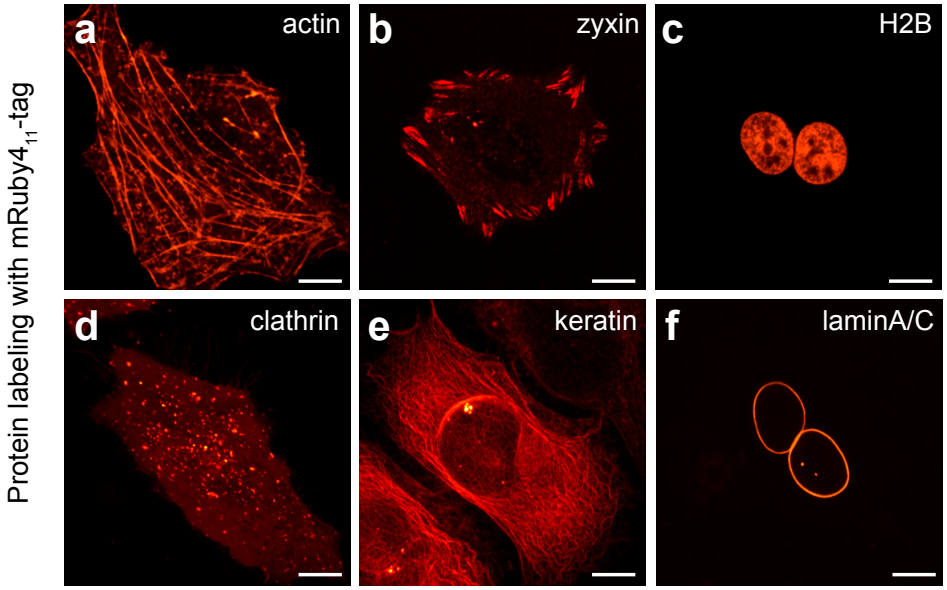
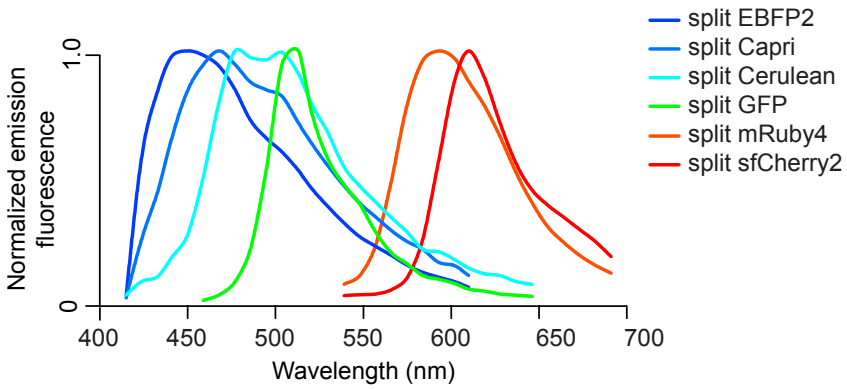


Figure 3.1 | Performance of BFP and CFP$_{1-10/11}$ variants in fusion constructs. (**a-f**) Confocal images of HeLa cells. Cells co-expressing EBFP2$_{1-10}$ (**a, b**), Capri$_{1-10}$ (**c, d**), and Cerulean$_{1-10}$ (**e, f**) with GFP$_{11}$-tagged β-actin or histone 2B.


Protein labeling with mRuby4$_{11}$-tag
a
actin
b
zyxin
c
H2B
d
clathrin
e
keratin
f
laminA/C

g


Normalized emission fluorescence
1.0
0
400
450
500
550
600
650
700
Wavelength (nm)
split EBFP2
split Capri
split Cerulean
split GFP
split mRuby4
split sfCherry2

h

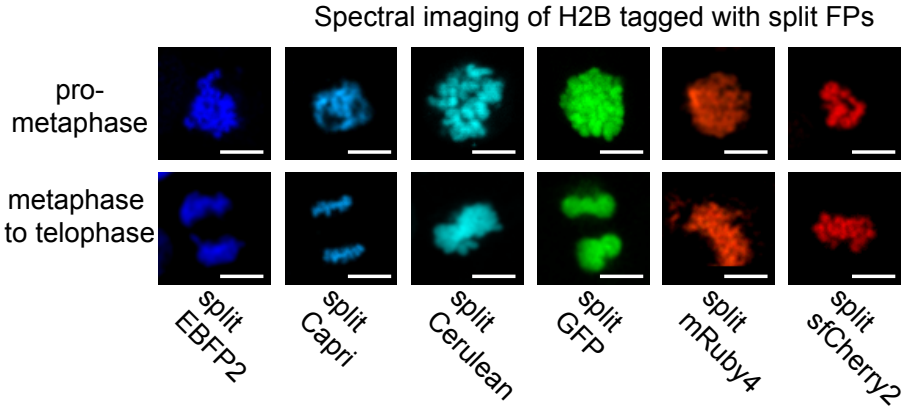

Spectral imaging of H2B tagged with split FPs
pro-metaphase
metaphase to telophase
split EBFP2
split Capri
split Cerulean
split GFP
split mRuby4
split sfCherry2

Figure 3.2 | Development of mRuby4, a new red-colored split FP (**a**-**f**) Cells co-expressing $mRuby4_{1-10}$ with $mRuby4_{11}$-fused cellular proteins. For each, the name of the fusion partner and its normal subcellular location are indicated, respectively; β-actin, actin stress fibers (**a**); zyxin, focal adhesion (**b**); histone 2B, nuclei (**c**), clathrin light chain, clathrin-coated pits (**d**); keratin, intermediate filaments (**e**); lamin A/C, nuclear envelops (**f**). (**g**) Normalized fluorescence emission spectra of $FP_{1-10/11}$ variants in HeLa cells. (**h**) HEK 293 cells expressing H2B labeled with $EBFP2_{1-10/11}$, $Capri_{1-10/11}$, $Cerulean_{1-10/11}$, $GFP_{1-10/11}$, $mRuby4_{1-10/11}$, or $sfCherry2_{1-10/11}$ were co-cultured in the same plate. Spectrally unmixed images at the different stages of mitosis are represented (see also **Supplementary Figure 3.14**). Scale bars,10 μm.

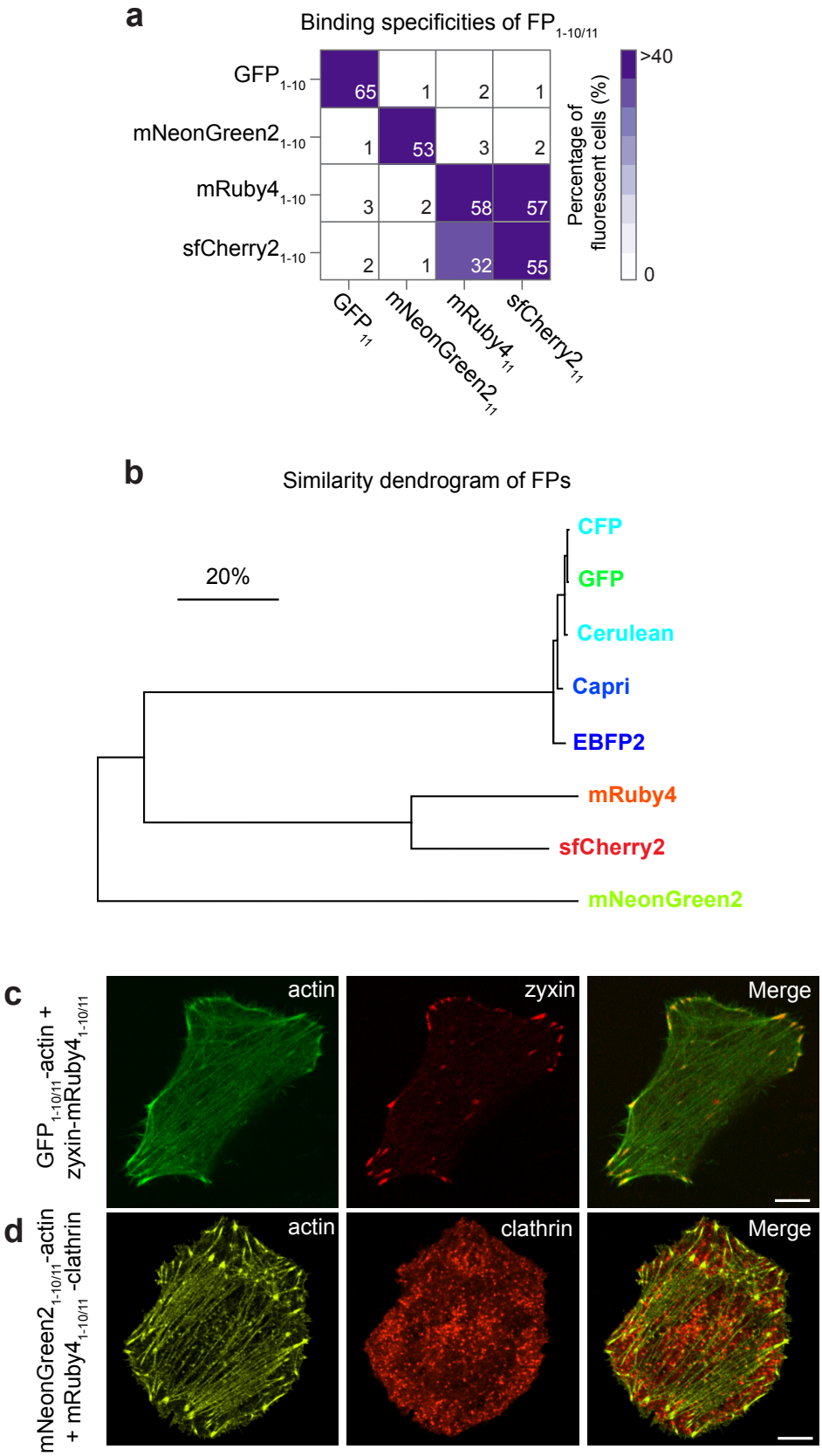

a
Binding specificities of FP1-10/11
GFP1-10
mNeonGreen21-10
mRuby41-10
sfCherry21-10
65
1
2
1
1
53
3
2
3
2
58
57
2
1
32
55
GFP11
mNeonGreen211
mRuby411
sfCherry211
Percentage of fluorescent cells (%)
>40
0
b
Similarity dendrogram of FPs
20%
CFP
GFP
Cerulean
Capri
EBFP2
mRuby4
sfCherry2
mNeonGreen2
c
GFP1-10/11-actin + zyxin-mRuby41-10/11
actin
zyxin
Merge
d
mNeonGreen21-10/11-actin + mRuby41-10/11-clathrin
actin
clathrin
Merge

Figure 3.3 | Characterizing the binding specificities of available $FP_{1-10/11}$ pairs (**a**) Characterizing the binding specificities of $GFP_{1-10/11}$, $sfCherry2_{1-10/11}$, $mNeonGreen2_{1-10/11}$, and $mRuby4_{1-10/11}$ by flow cytometry (see also **Supplementary Figure 3.15**). Each of the FP_{11} fragments was tested for complementation to all of the FP_{1-10} fragments. Complementation is indicated as the percentage of fluorescent cells by a color scale and the number in each block. (**b-c**) Dual-color fluorescence images of HeLa cells expressing $GFP_{1-10/11}$- actin and zyxin-$mRuby4_{1-10/11}$ (**b**), and $mNeonGreen2_{1-10/11}$-actin and $mRuby4_{1-10/11}$-clathrin (**c**). (**d**) This dendrogram is based on the similarities of the following fluorescence protein sequences: EBFP2, Capri, Cerulean, CFP, GFP, mNeonGreen2, mRuby4, and sfCherry2. Proteins that share sequences are separated by smaller branch lengths. Scale bar, 20% dissimilarity. The dendrogram was constructed using MEGA 7.0 software. Scale bars,10 μm.

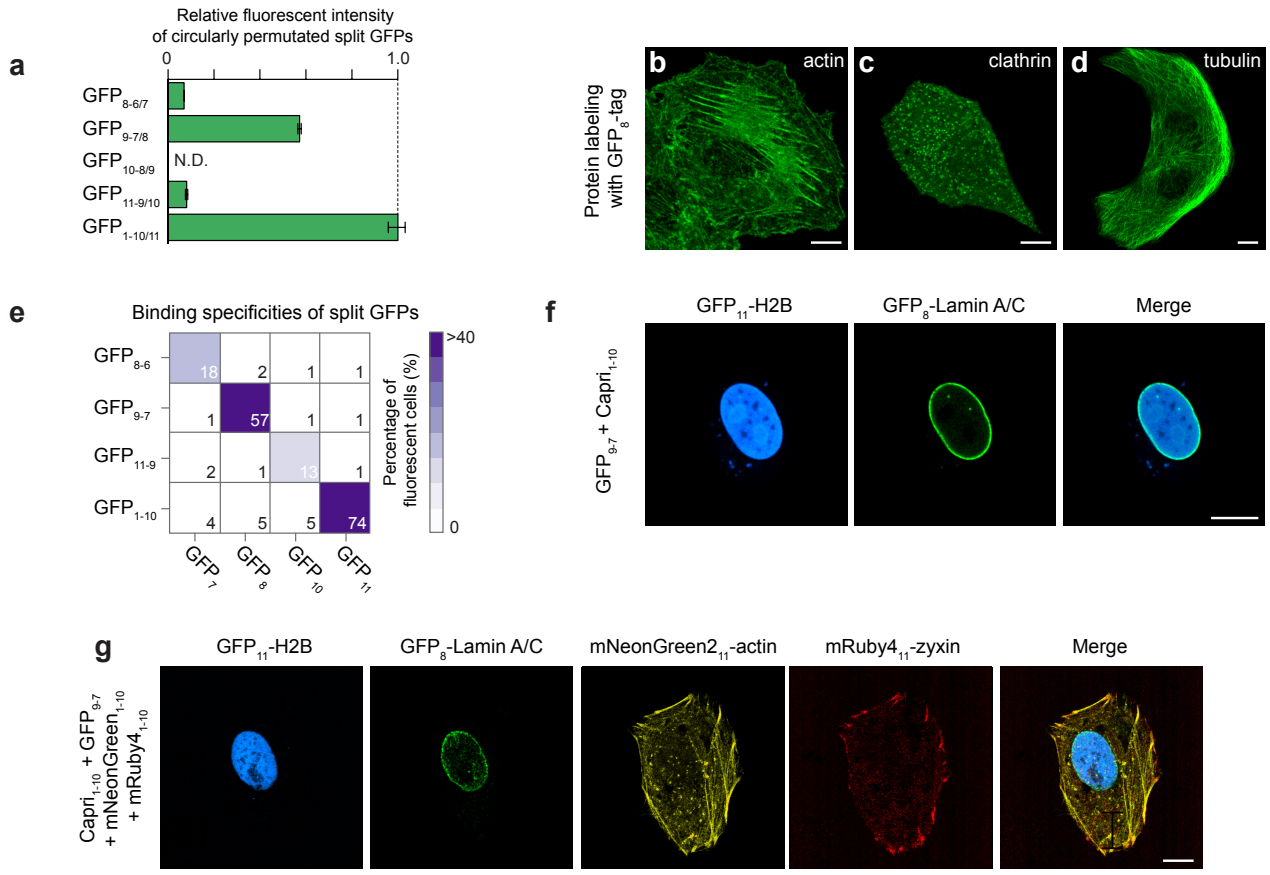
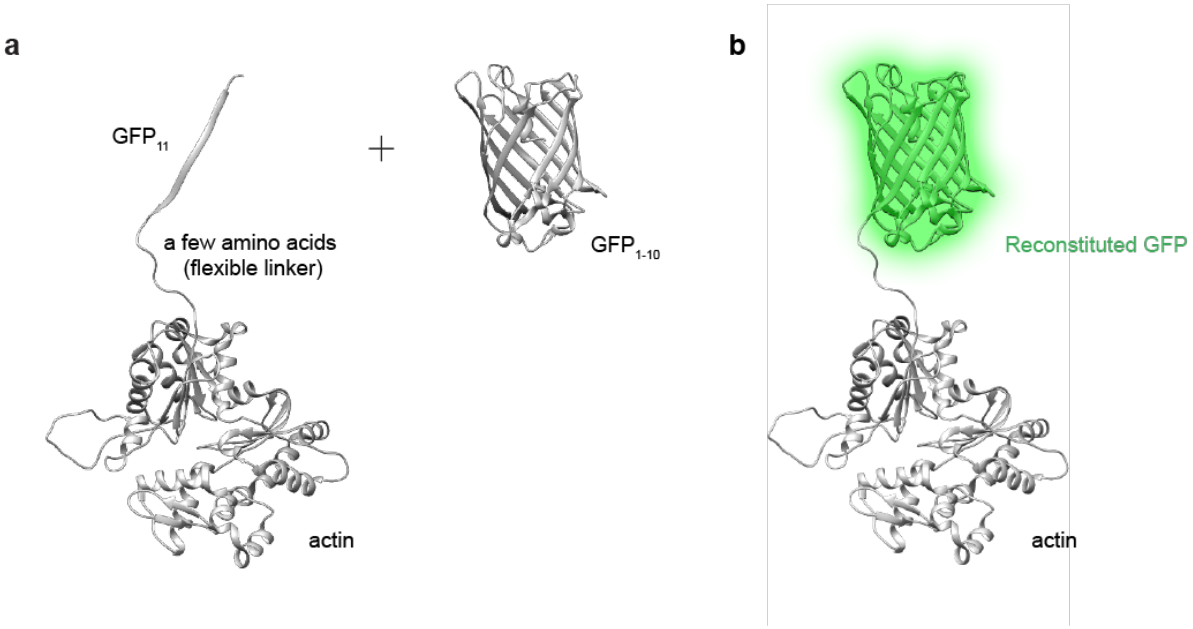
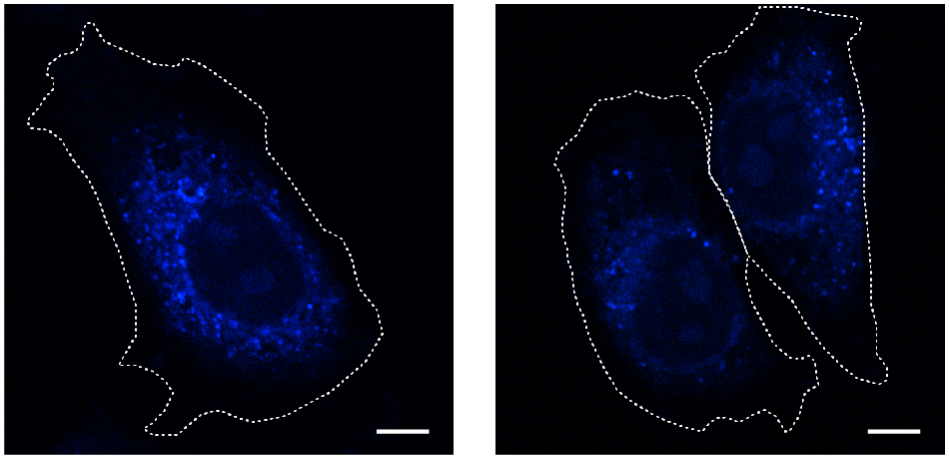


Figure 3.4 | Multiplexed labeling of cellular proteins in human cells. (**a**) Fluorescence intensity of HEK 293T cells expressing actin labeled with circularly permutated split GFP variants, measured by flow cytometry (see also **Supplementary Figure 3.16**). Error bars are SEM. (**b-d**) Confocal images of HeLa cells co-expressing GFP_{9-7} with GFP_8 fusions; β-actin (**b**), clathrin light chain (**c**), and β-tubulin (**d**). (**e**) The binding specificities of $GFP_{8-6/7}$, $GFP_{9-7/8}$, $GFP_{11-9/10}$, and $GFP_{1-10/11}$ were characterized by flow cytometry (see also **Supplementary Figure 3.17**). (**f**) Dual-color fluorescence images of a U2OS cell expressing two different fusions, GFP_{11}-H2B and GFP_8-Lamin A/C. (**g**) Four-color images of a U2OS cell co-expressing GFP_{11}-H2B, GFP_8-LaminA/C, $mNeonGreen2_{11}$-actin, and $mRuby4_{11}$-zyxin with GFP_{1-10}, GFP_{9-7}, $mNeonGreen2_{1-10}$, and $mRuby4_{1-10}$. Scale bars, 10 μm.

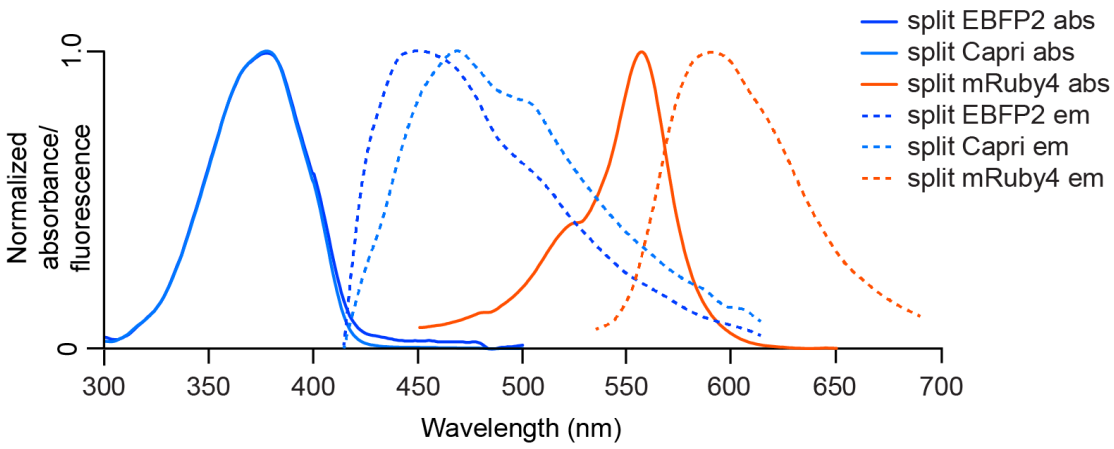
3.5 SUPPLEMENTARY FIGURES



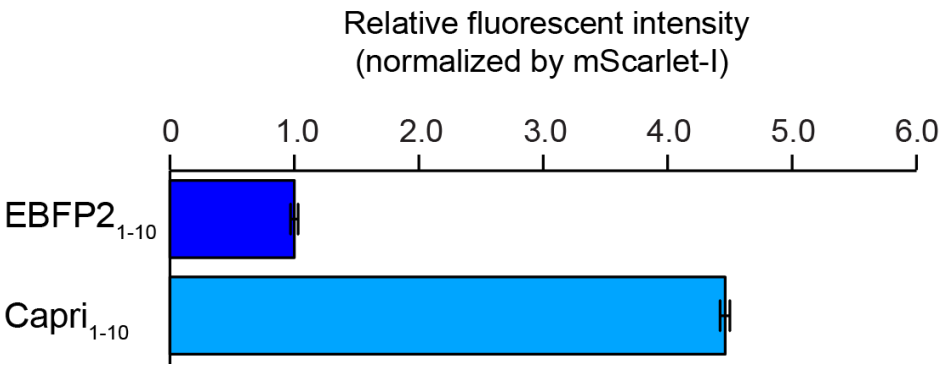
Supplementary Figure 3.1 | Schematic diagram of GFP11-tag. Schematic of labeling proteins with split GFP is illustrated. A short region of split GFP (GFP_{11}) can be inserted into a gene of interest. The remainder (GFP_{1-10}), which is needed for reconstitution of GFP fluorescence, is expressed in cells. Importantly, GFP_{1-10} does not produce background fluorescence. In this diagram, the N-terminus of β-actin is labeled with GFP_{11}-tag (**a**). The GFP_{1-10} and GFP_{11} fragments can associate by themselves and form a functional GFP (**b**).



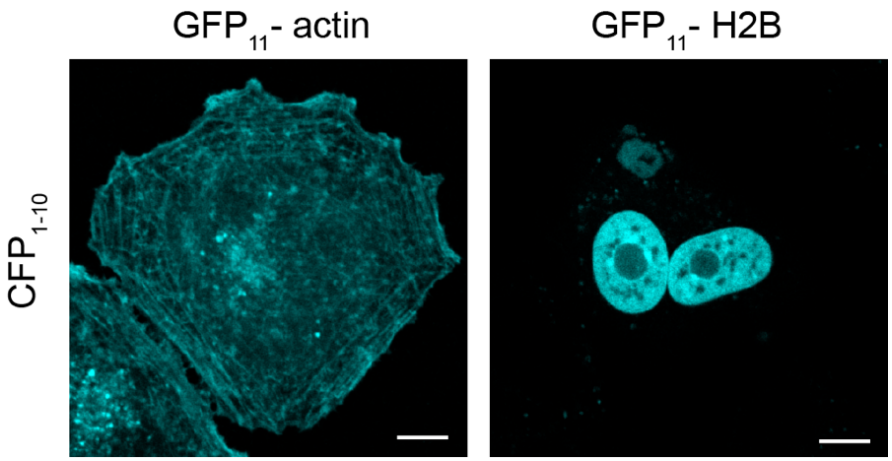
Supplementary Figure 3.2 | Auto-fluorescence in cellular imaging with illumination at 405nm. Under 405 nm illumination, blue fluorescence can frequently be collected from auto-fluorescence in the perinuclear region. Representative images of un-transfected HeLa cells in this figure and **Figure 1** were taken with the same acquisition settings. Cells are outlined by white dash lines. Scale bar, 10 μm.



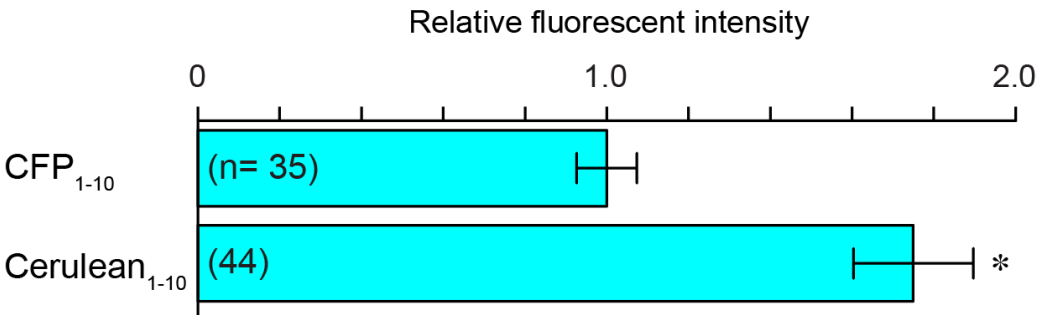
Supplementary Figure 3.3 | Absorbance measurements and fluorescence emission of split EBFP2, split Capri, and split mRuby4. Normalized absorbance (solid lines) and emission (dotted lines) spectra of split BFP2, split Capri, and split mRuby4.



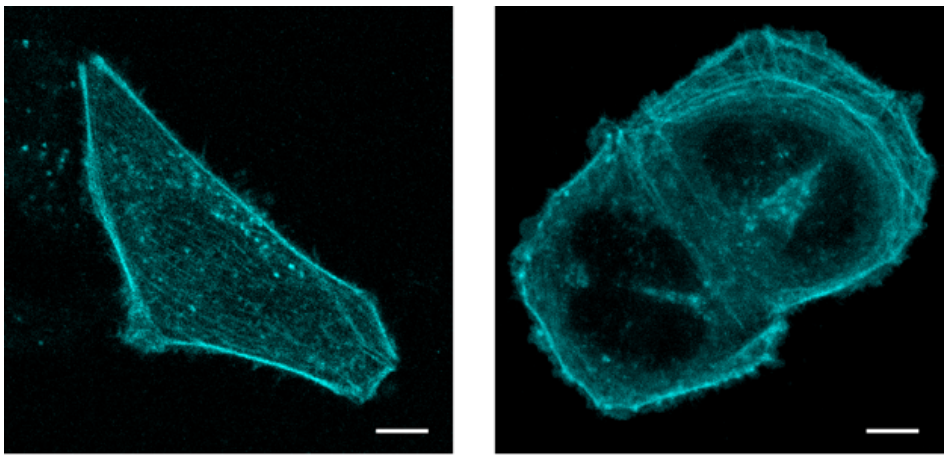
Supplementary Figure 3.4 | Cellular fluorescence measurement of split BFP variants. Relative fluorescence intensity of HEK 293T cells expressing H2B labeled with EBFP2$_{1\text{-}10/11}$ or Capri$_{1\text{-}10/11}$, measured by flow cytometry and normalized to mScarlet-I signal. EBFP2$_{1\text{-}10}$ or Capri$_{1\text{-}10}$ was directly fused to mScarlet-I, of which the signal was applied to normalize the differences of gene expression levels. n > 5000 cells for each. Error bars are SEM (Standard error of the mean).



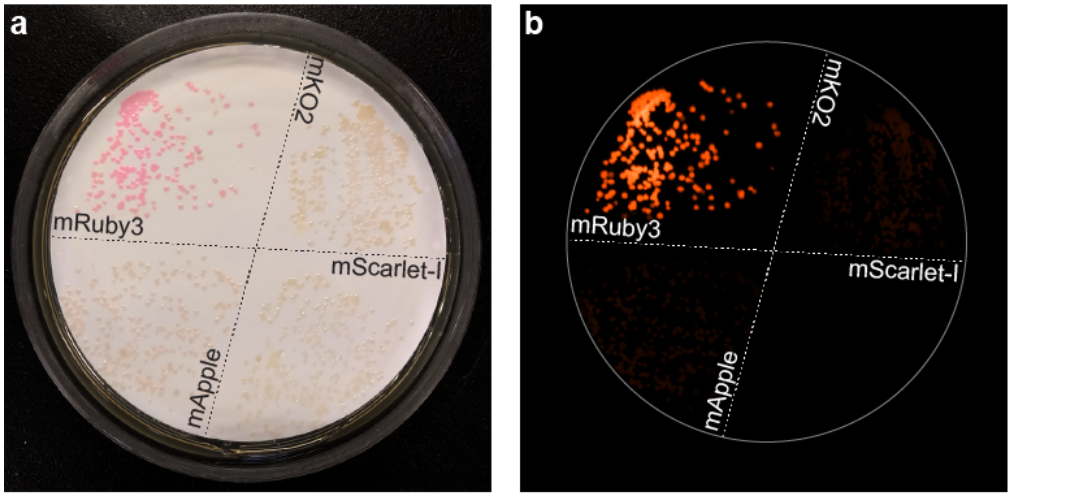
Supplementary Figure 3.5 | Confocal microscopy images of cellular proteins labeled with split CFP. Fluorescence images of HeLa cells expressing either β-actin or histone 2B fusion of $CFP_{1-10/11}$. Scale bars, 10 μm.



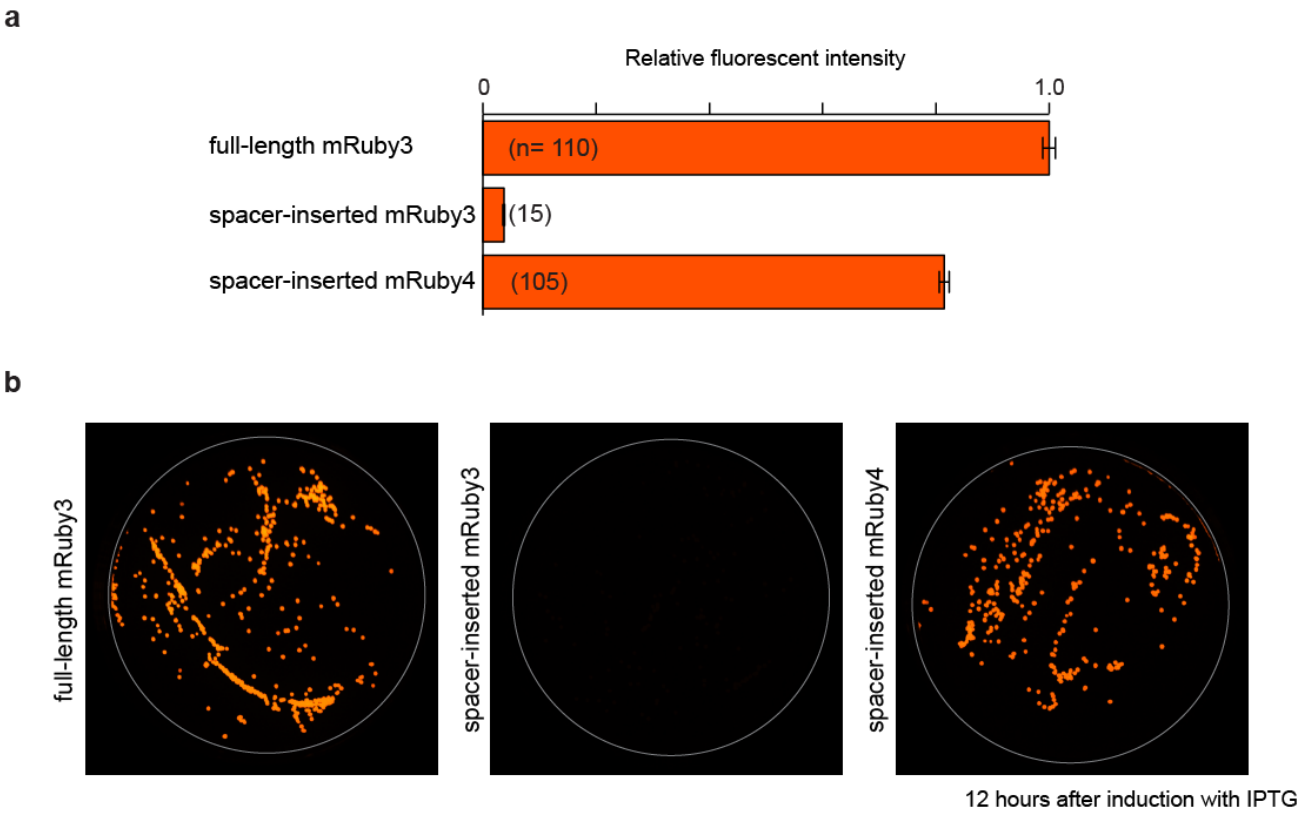
Supplementary Figure 3.6 | Cellular fluorescence measurement of split CFP and split Cerulean. Relative fluorescence intensity of HEK 293T cells co-expressing GFP_{11}-H2B with $Cerulean_{1-10}$ or CFP_{1-10}, measured by confocal microscopy. After a background correction was applied, a fluorescent intensity value was measured as the sum intensity value of the nucleus for each cell. n = 35 - 44 cells. Error bars are SEM, $^{*}P < 0.001$ (*Student's t-tests*).



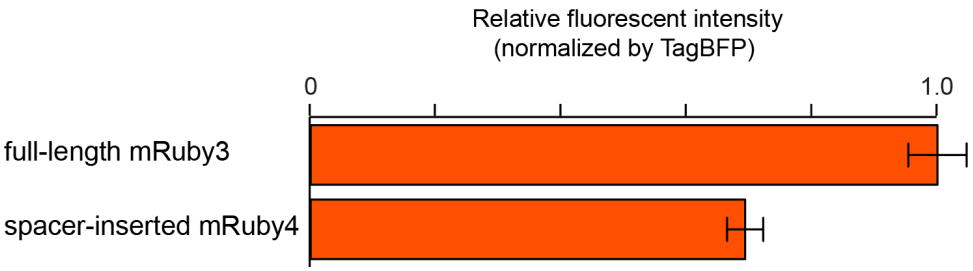
Supplementary Figure 3.7 | Full-length Cerulean-β-actin expressed in HeLa cells. Representative confocal images of full-length Cerulean-targeted actin stress fibers in HeLa cells. Scale bars, 10 μm.



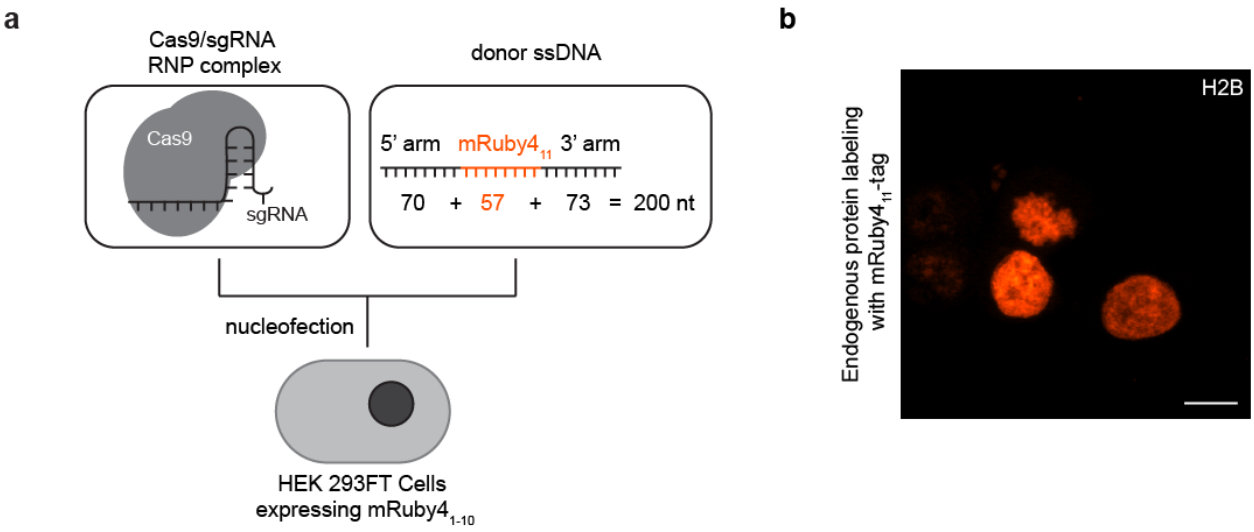
Supplementary Figure 3.8 | Colony fluorescence measurement of spacer-inserted orange-red FPs. Transformed colonies of *E. coli* with four different constructs of spacer-inserted FPs (i.e., mKO2, mRuby3, mApple, and mScarlet-I) are shown in visible light (**a**) and 520 nm light (**b**). After induction with IPTG, the plate was incubated at 37 °C for 2 days.



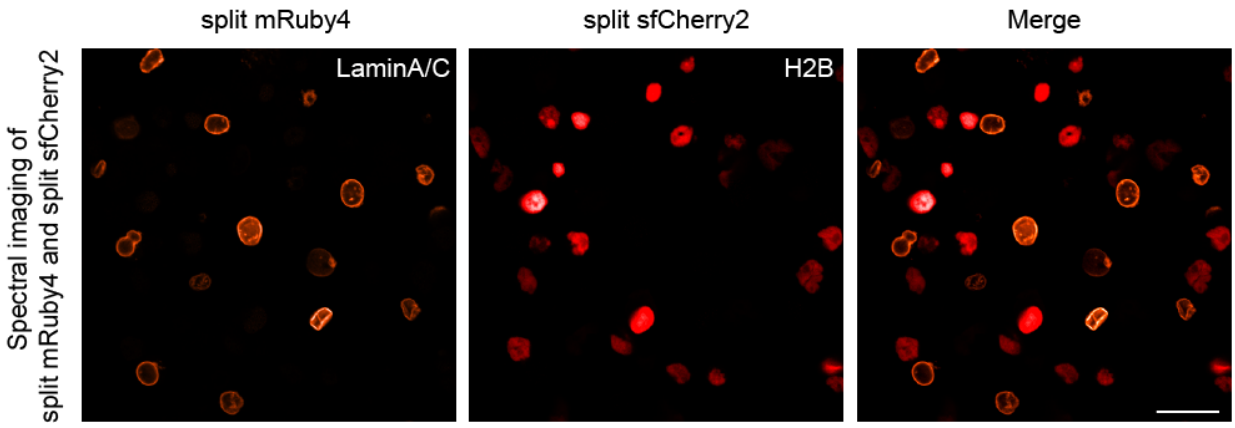
Supplementary Figure 3.9 | Engineering the self-complementing split mRuby system in *E. coli*. (**a**) Relative fluorescence intensity of *E. coli* colonies expressing full-length mRuby3, spacer-inserted mRuby3, or spacer-inserted mRuby4. n = 15 - 110 colonies. Error bars are SEM. (**b**) Representative fluorescent images of *E. coli* colonies expressing full-length mRuby3, spacer-inserted mRuby3, or spacer-inserted mRuby4. These plates were kept at 37 °C for 12 hours after induction with IPTG.



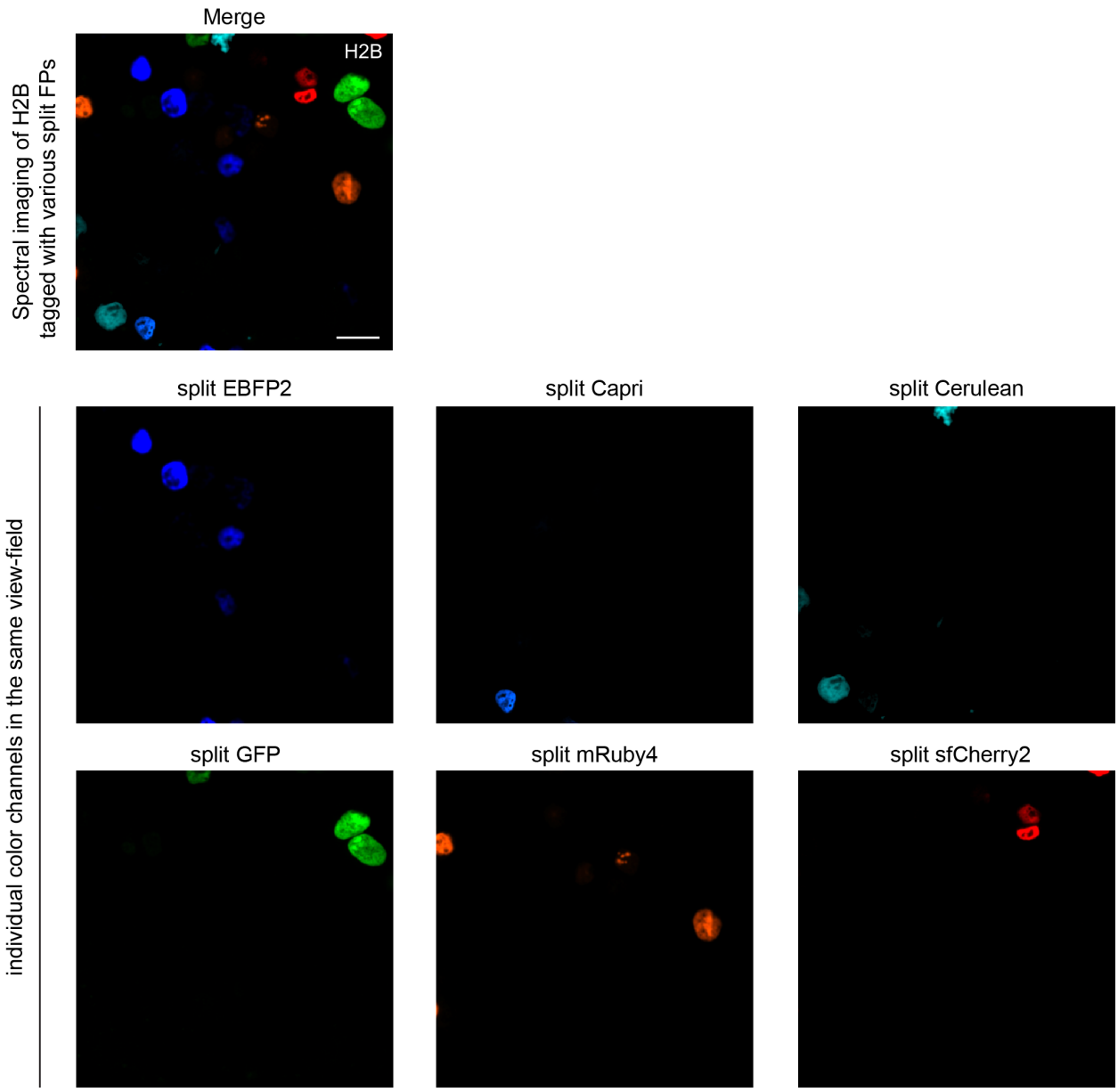
Supplementary Figure 3.10 | Cellular fluorescence measurement of full-length mRuby3 and spacer-inserted mRuby4 in HEK cells. Relative fluorescence intensity of HEK 293T cells expressing actin labeled with two different FPs (full-length mRuby3, or spacer-inserted mRuby4), measured by flow cytometry. Full-length mRuby3, and spacer inserted-mRuby4 were directly fused to TagBFP. TagBFP signal was used to normalize gene expression levels (n > 5000 cells for each). Error bars are SEM.



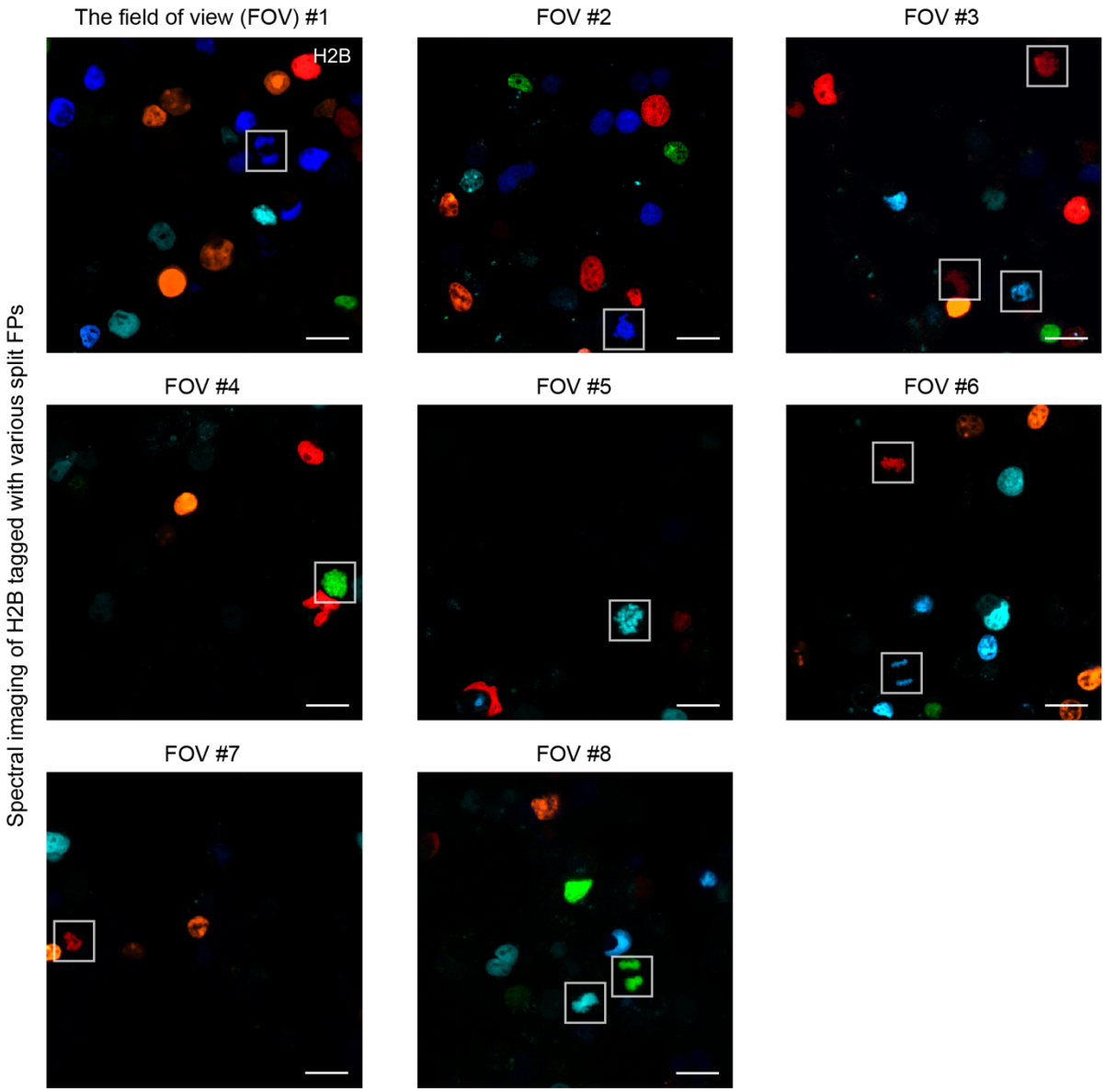
Supplementary Figure 3.11 | mRuby4$_{11}$ labeling of endogenous proteins. (**a**) Schematic diagram of endogenous protein labeling with mRuby4$_{11}$. mRuby4$_{11}$ is only 57 nt including a short amino-acid linker, and thus, we are able to adopt relatively short homology arms (70-73 nucleotides on both 5' and 3' arms). Donor DNA consequently becomes a 200-nucleotide single-strand DNA. (**b**) Fluorescent image of mRuby4$_{11}$ knock-in in HEK 293FT cells. We knocked mRuby4$_{11}$ into the H2B locus (*HIST2H2BE*). Cas9/sgRNA RNP nucleofection enabled us to obtain 0.5 % of mRuby4$_{1-10/11}$-positive cells. Scale bars, 10 μm.



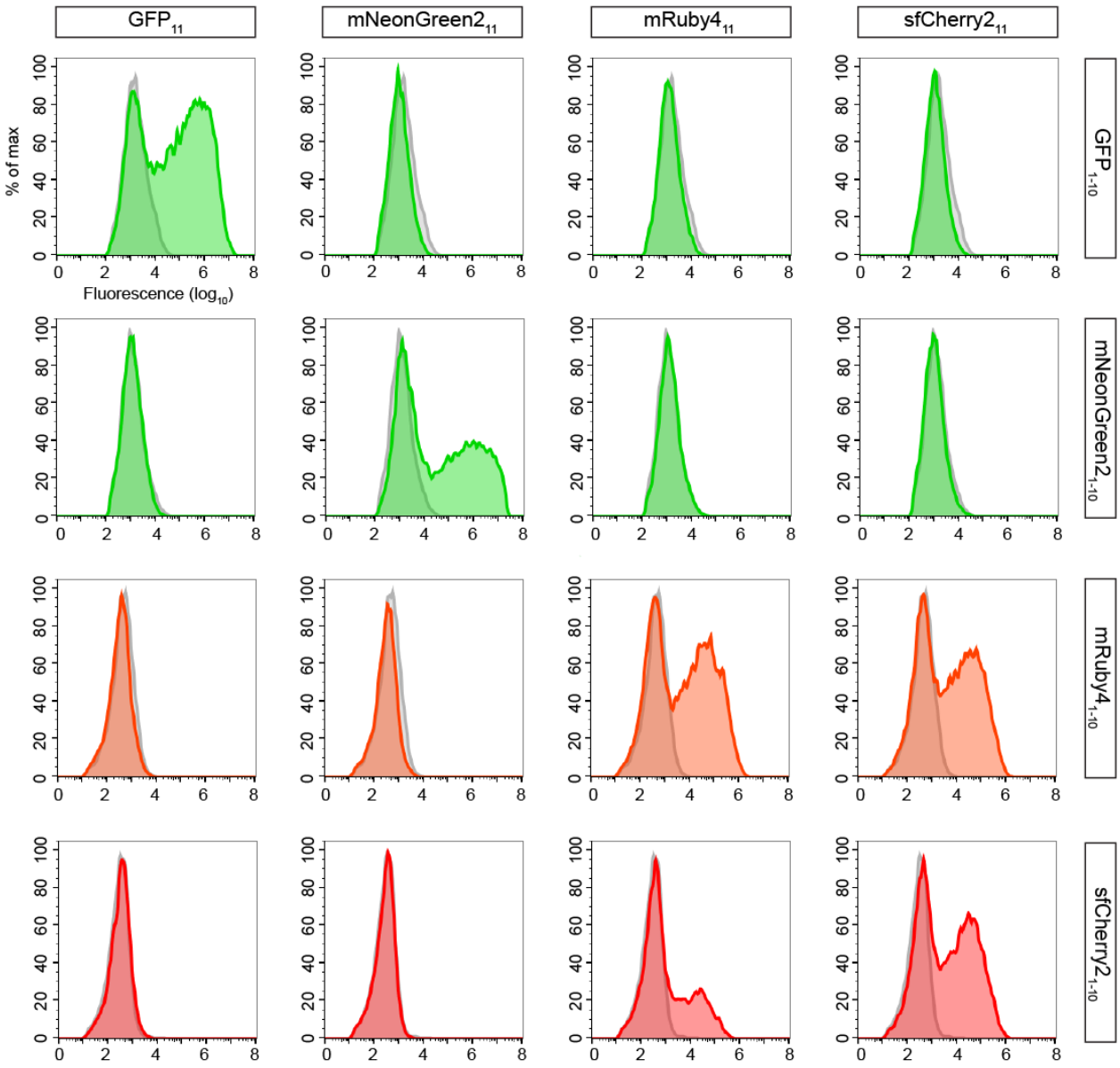
Supplementary Figure 3.12 | Distinguishing split sfCherry2 and split mRuby4 by their emission spectra. HEK 293T cells expressing either sfCherry2$_{1\text{-}10/11}$-H2B or mRuby4$_{1\text{-}10/11}$-LaminA/C were co-cultured in the same plate. Two fluorescent images were acquired using confocal microscopy with spectral detection. Emission was detected from 463 to 695 nm using excitation of 514-nm and 543-nm laser. Unmixed channels are overlaid and shown in pseudo-colors. Scale bars, 25 μm.



Supplementary Figure 3.13 | Spectral imaging of H2B fusions with multicolor split FPs. Representative images of HEK 293 cells expressing H2B labeled with either EBFP2$_{1\text{-}10/11}$, Capri$_{1\text{-}10/11}$, Cerulean$_{1\text{-}10/11}$, GFP$_{1\text{-}10/11}$, YPet$_{1\text{-}10/11}$, mRuby4$_{1\text{-}10/11}$, or sfCherry2$_{1\text{-}10/11}$. We collected images of the same view-field captured at emission wavelengths from 410 to 695 nm and subsequently obtained spectrally unmixed images of the six split FPs (see also the Methods). Scale bars, 25 μm.

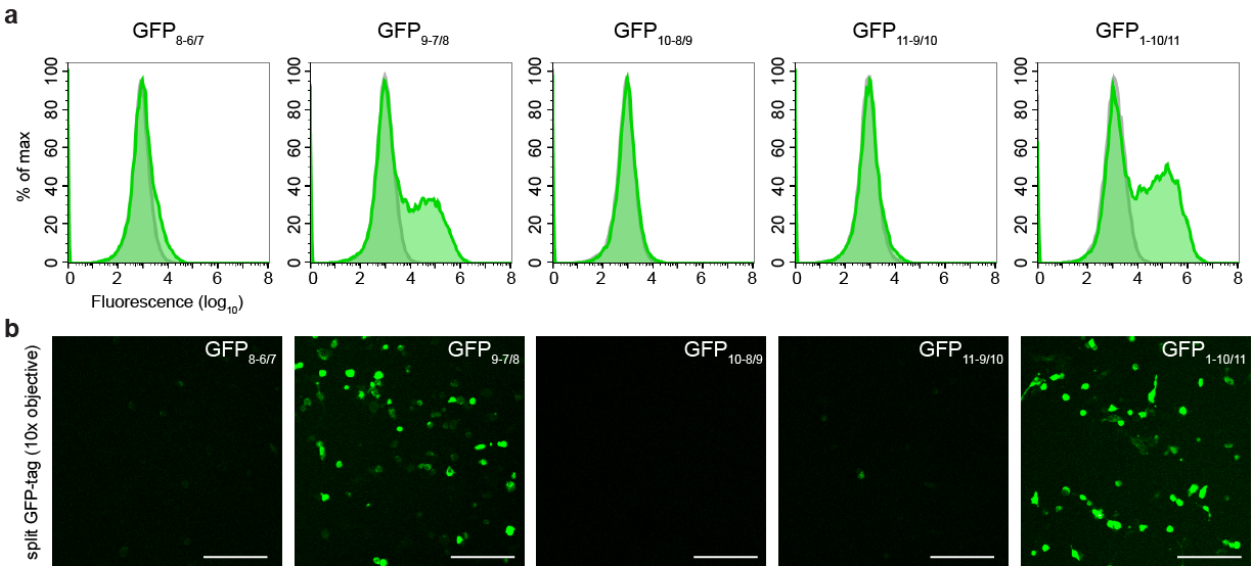


Supplementary Figure 3.14 | Unprocessed images shown in Figure 3.2h. Original images used in Figure 2h. Unmixed channels are overlaid and shown in pseudo-colors. White boxes mark the regions in the main figure. This experiment was repeated three times with similar results. Scale bars, 25 μm.

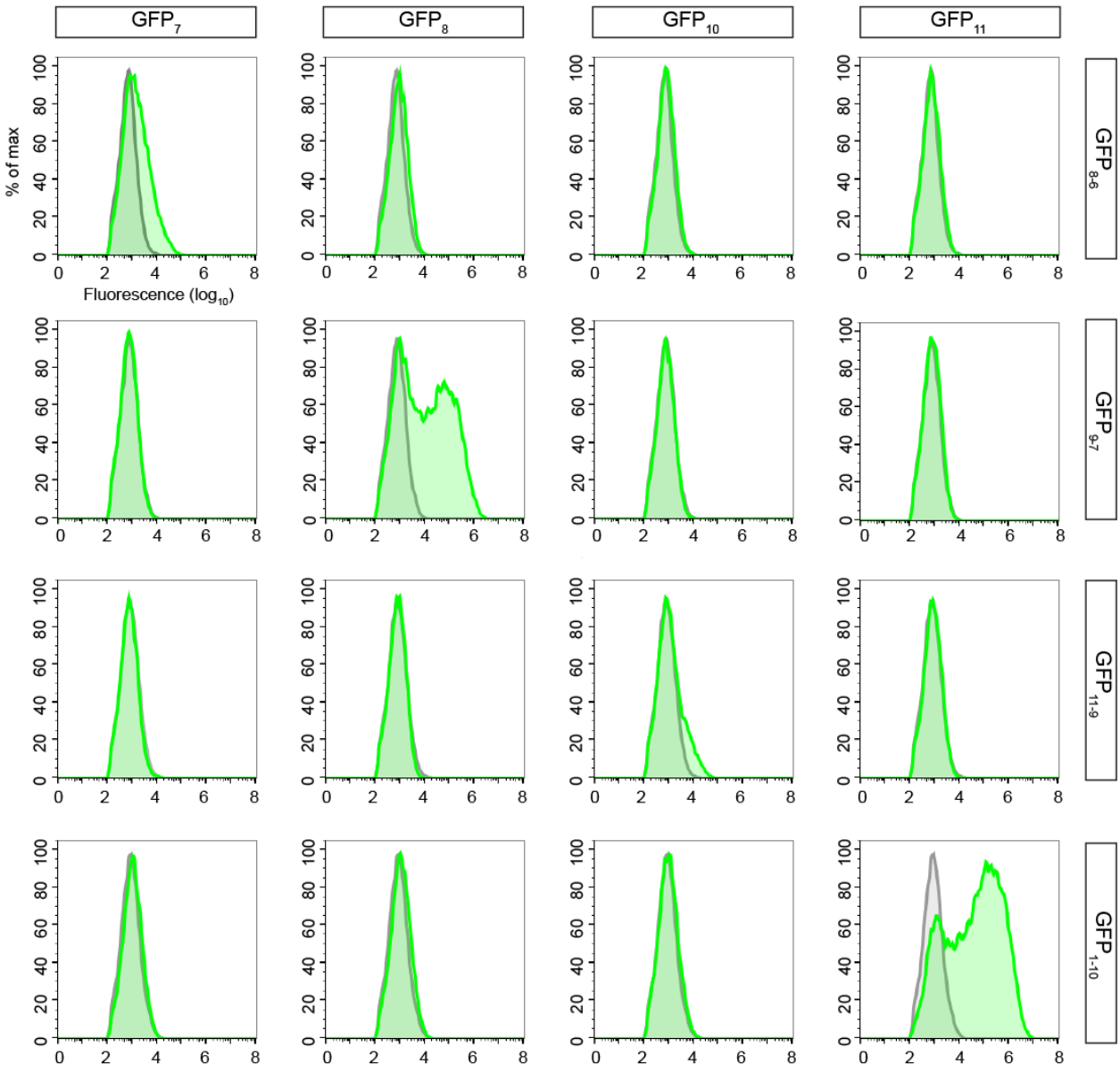


Supplementary Figure 3.15 | Testing the binding specificities of GFP$_{1-10/11}$, sfCherry2$_{1-10/11}$, mNeonGreen2$_{1-10/11}$, and mRuby4$_{1-10/11}$. Each of the FP$_{11}$ fragments was tested for complementation to all of the FP$_{1-10}$ fragments. The complemented signal was measured by flow cytometry. Samples were acquired using 488 nm or 561 nm excitation source. Measurements are represented as flow cytometry histograms. Overlay of a control population (HEK 293T cells

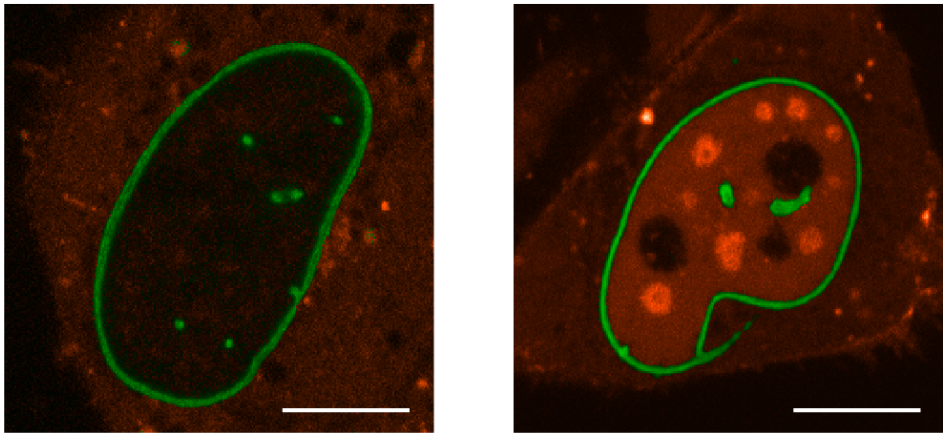
transfected with FP_{1-10} alone; gray) onto the split FP-expressing population (green, orange, and red) allows identification of the positive cells. A summary chart is shown in Figure 2c.



Supplementary Figure 3.16 | Cellular fluorescence measurement of circularly permutated split GFP variants. (**a**) Fluorescence intensity of HEK 293T cells expressing actin labeled with $GFP_{8\text{-}6/7}$, $GFP_{9\text{-}7/8}$, $GFP_{10\text{-}8/9}$, $GFP_{11\text{-}9/10}$, and $GFP_{1\text{-}10/11}$, measured by flow cytometry (green). Control cells (HEK 293T cells transfected with either $GFP_{8\text{-}6}$, $GFP_{9\text{-}7}$, $GFP_{10\text{-}8}$, $GFP_{11\text{-}9}$ or $GFP_{1\text{-}10}$ alone) are represented in each histogram overlay (gray). A summary graph is shown in Figure 2f. (**b**) Their corresponding fluorescence images of HEK 293T cells. Fluorescence images were acquired with a 10x objective (Plan Fluor, Nikon). Scale bars are 200 μm.



Supplementary Figure 3.17 | Testing the binding specificities of $GFP_{8-6/7}$, $GFP_{9-7/8}$, $GFP_{11-9/10}$, and $GFP_{1-10/11}$. The β-strands 7, 8, 9, or 11 were tested for complementation to the GFP_{8-6}, GFP_{9-7}, GFP_{11-9}, or GFP_{1-10} fragments. The complemented signal (green) was measured by flow cytometry using 488 nm excitation source. Control cells (HEK 293T cells transfected with either GFP_{8-6}, GFP_{9-7}, GFP_{11-9} or GFP_{1-10} alone) are represented in each histogram overlay (gray). A summary chart is shown in Figure 2j.



Supplementary Figure 3.18 | Nuclear localization of zyxin. Representative single confocal sections of U2OS cells expressing $GFP_{9\text{-}7/8}$-LaminA/C (green) and $mRuby4_{1\text{-}10/11}$-zyxin (red). While zyxin mostly localized outside of the nucleus (left panel), some cells display the localization of zyxin in the nucleoplasm (right panel). Scale bars, 10 μm.

3.6 SUPPLEMENTARY TABLES

Supplementary Table 3.1 | Properties of split FPs engineered in this study

	EBFP2 [b]	split EBFP2	split Capri	mRuby3	split mRuby3	split mRuby4
Absorbance peak (nm)	378	379	379	556	557	557
Emission peak (nm)	450	450	469	593	593	592
EC at peak ($mM^{-1}cm^{-1}$)	34	23	37	119	32	93
QY	0.09	0.07	0.13	0.48	0.23	0.32
Brightness [a]	3.1	1.6	4.8	57.1	7.4	29.8

EC, extinction coefficient; QY, quantum yield. [a] Calculated as the product of EC at peak and QY. [b] six substitutions (N40I/T106K/E112V/K166T/I167V/S206T) were introduced into the original EBFP2.

Supplementary Table 3.2 | Nucleotide sequences of EBFP2$_{1-10}$, Capri$_{1-10}$, Cerulean$_{1-10}$, mRuby4$_{1-10}$, GFP$_{8-6}$, GFP$_{9-7}$, and GFP$_{11-9}$

EBFP2$_{1-10}$	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGAGGGG CGAGGGCGAGGGCGATGCCACCATCGGCAAGCTGACCCTGAAGTTCAT CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC CCTGAGCCACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAA GCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA GCGCACCATCTTCTTCAAGGACGACGGCAAATACAAGACCCGCGCCGT GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG GCGTCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG TACAACTTCAACAGCCACAACATCTATATCATGGCCGTCAAGCAGAAG AACGGCATCAAGGTGAACTTCACCGTGCGCCACAACGTGGAGGACGG CAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGA CGGCCCCGTGCTGCTGCCCGACAGCCACTACCTGAGCACCCAGACCGT GCTGAGCAAAGACCCCAACGAGAAG
Capri$_{1-10}$	ATGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGAGGGGCGA GGGCGAGGGCGATGCCACCATTGGCAAGCTGACCCTGAAGTTCATCTG CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCT GACGcacGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGAGA CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC

	ACCATCTCGTTCAAGGACGACGGCAAATACAAGACCCGCGCCGTTGTG AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCAC AGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA ACTTCAACAGCCACAACATCTATATCACTGCCGTCAAGCAGAAGAACG GCATCAAGGCGAACTTCACAGTACGCCACAACGTGGAGGACGGCAGC GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC CCCGTGCTGCTGCCCGACAGCCACTACCTGAGCACCCAGACAGTGCTG AGCAAAGACCCCAACGAGAAG
Cerulean$_{1\text{-}10}$	ATGTCCAAAGGAGAAGAACTGTTTACCGGTGTTGTGCCAATTTTGGTT GAACTCGATGGTGATGTCAACGGACATAAGTTCTCAGTGAGAGGCGAA GGAGAAGGTGACGCCACCATTGGAAAATTGACTCTTAAATTCATCTGT ACTACTGGTAAACTTCCTGTACCATGGCCGACTCTCGTAACAACGCTTA CGTggGGAGTTCAGTGCTTTGCGAGATACCCAGACCATATGAAAAGAC ATGACTTTTTTAAGTCGGCTATGCCTGAAGGTTACGTGCAAGAAAGAA CAATTTCGTTCAAAGATGATGGAAAATATAAAACTAGAGCAGTTGTTA AATTTGAAGGAGATACTTTGGTTAACCGCATTGAACTGAAAGGAACAG ATTTTAAAGAAGATGGTAATATTCTTGGACACAAACTCGAATACAATG CTAATAGTGATAACGTATACATCACTGCTGATAAGCAAAAGAACGGAA TTAAAGCGAATTTCACAGTACGCCATAATGTAGAAGATGGCAGTGTTC AACTTGCCGACCATTACCAACAAAACACCCCTATTGGAGACGGTCCGG TACTTCTTCCTGATAATCACTACCTCTCAACACAAACAGTCCTGAGCAA AGATCCAAATGAAAAA

mRuby4$_{1\text{-}10}$	ATGGTGTCTAAGGGCGAAGAGCTGATCAAGGAAAATATGCGTACGAA GGTGGTCATGGAAGGTTCGGTCAACGGCCACCATTTCAAATGCATAGG TGAAGGAGAAGGCAGACCGTACGAGGGAGTGCAAACCATGAGGATCA AAGTCATCGAGGGAGGACCCCTGCCATTTGCCTTTGACATTCTTGCCAC GTCGTTCATGTATGGCAGCCGTACCTTTATCAAGTACCCGGCCGACATC CCTGATTTCTTTAAACAGTCCTTTCCTGAGGGTTTTACTTGGGAAAGAG TTACGAGATACGAAGATGGTGGAGTCATCACCGTCACGCAGGACACCT GCCTTGAGGATGGCGAGCTCGTCTACAACGTCAAGGTCAGCGGGGTAA ACTTTCCCTCCAATGGTCCCGTGATGCAGAAGAAGACCAAGGGTTGGG AGCCTAATACAGAGATGATGTATCCAGTAGATGGTGGTCTGAGAGGAT ACACTGACATCGCACTGAAAGTTGATGGTGGTGGCCATCTGCACTGCA ACTTCGTGACAACTTACAGGTCAAAAAAGACCGTCGGGAACATCAAG ATGCCCGGTGTCCATGCCGTTGATCACCGCCTGGAAAGGATCGAGGAG AGTGACAAT
GFP$_{8\text{-}6}$	ATGAAGAACGGAATTAAAGCGAATTTCACAGTACGCCATAATGTAGA AGATGGCAGTGTTCAACTTGCCGACCATTACCAACAAAACACCCCTAT TGGAGACGGTCCGGTACTTCTTCCTGATAATCACTACCTCTCAACACAA ACAGTCCTGAGCAAAGATCCAAATGAAAAACGTGACCACATGGTCCTT CATGAGTATGTAAATGCTGCTGGGATTACAGGTGGCACCGGTGGCTCT ATGTCCAAAGGAGAAGAACTGTTTACCGGTGTTGTGCCAATTTTGGTT GAACTCGATGGTGATGTCAACGGACATAAGTTCTCAGTGAGAGGCGAA GGAGAAGGTGACGCCACCATTGGAAAATTGACTCTTAAATTCATCTGT ACTACTGGTAAACTTCCTGTACCATGGCCGACTCTCGTAACAACGCTTA

	CGTACGGAGTTCAGTGCTTTTCGAGATACCCAGACCATATGAAAAGAC ATGACTTTTTTAAGTCGGCTATGCCTGAAGGTTACGTGCAAGAAAGAA CAATTTCGTTCAAAGATGATGGAAAATATAAAACTAGAGCAGTTGTTA AATTTGAAGGAGATACTTTGGTTAACCGCATTGAACTGAAAGGAACAG ATTTTAAAGAAGATGGTAATATTCTTGGACAC
GFP_{9-7}	ATGGAAGATGGCAGTGTTCAACTTGCCGACCATTACCAACAAAACACC CCTATTGGAGACGGTCCGGTACTTCTTCCTGATAATCACTACCTCTCAA CACAAACAGTCCTGAGCAAAGATCCAAATGAAAAACGTGACCACATG GTCCTTCATGAGTATGTAAATGCTGCTGGGATTACAGGTGGCACCGGT GGCTCTATGTCCAAAGGAGAAGAACTGTTTACCGGTGTTGTGCCAATT TTGGTTGAACTCGATGGTGATGTCAACGGACATAAGTTCTCAGTGAGA GGCGAAGGAGAAGGTGACGCCACCATTGGAAAATTGACTCTTAAATTC ATCTGTACTACTGGTAAACTTCCTGTACCATGGCCGACTCTCGTAACAA CGCTTACGTACGGAGTTCAGTGCTTTTCGAGATACCCAGACCATATGA AAAGACATGACTTTTTTAAGTCGGCTATGCCTGAAGGTTACGTGCAAG AAAGAACAATTTCGTTCAAAGATGATGGAAAATATAAAACTAGAGCA GTTGTTAAATTTGAAGGAGATACTTTGGTTAACCGCATTGAACTGAAA GGAACAGATTTTAAAGAAGATGGTAATATTCTTGGACACAAACTCGAA TACAATTTTAATAGTCATAACGTATACATCACTGCTGATAAGCAA

GFP_{11-9}	ATGCGTGACCACATGGTCCTTCATGAGTATGTAAATGCTGCTGGGATT ACAGGTGGCACCGGTGGCTCTATGTCCAAAGGAGAAGAACTGTTTACC GGTGTTGTGCCAATTTTGGTTGAACTCGATGGTGATGTCAACGGACAT AAGTTCTCAGTGAGAGGCGAAGGAGAAGGTGACGCCACCATTGGAAA ATTGACTCTTAAATTCATCTGTACTACTGGTAAACTTCCTGTACCATGG CCGACTCTCGTAACAACGCTTACGTACGGAGTTCAGTGCTTTTCGAGAT ACCCAGACCATATGAAAAGACATGACTTTTTTAAGTCGGCTATGCCTG AAGGTTACGTGCAAGAAAGAACAATTTCGTTCAAAGATGATGGAAAA TATAAAACTAGAGCAGTTGTTAAATTTGAAGGAGATACTTTGGTTAAC CGCATTGAACTGAAAGGAACAGATTTTAAAGAAGATGGTAATATTCTT GGACACAAACTCGAATACAATTTTAATAGTCATAACGTATACATCACT GCTGATAAGCAAAAGAACGGAATTAAAGCGAATTTCACAGTACGCCA TAATGTAGAAGATGGCAGTGTTCAACTTGCCGACCATTACCAACAAAA CACCCCTATTGGAGACGGTCCGGTA

Supplementary Table 3.3 | Amino acid sequences of full-length EBFP2, spacer-inserted EBFP2, spacer-inserted Capri, full-length mRuby3, spacer-inserted mRuby3, and spacer-inserted mRuby4

full-length EBFP2	MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTT GKLPVPWPTLVTTLSHGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFF KDDGKYKTRAVVKFEGDTLVNRIELKGVDFKEDGNILGHKLEYNFNSHNI YIMAVKQKNGIKVNFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDSH YLSTQTVLSKDPNEKRDHMVLHEYVNAAGIT
spacer-inserted EBFP2	MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTT GKLPVPWPTLVTTLSHGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFF KDDGKYKTRAVVKFEGDTLVNRIELKGVDFKEDGNILGHKLEYNFNSHNI YIMAVKQKNGIKVNFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDSH YLSTQTVLSKDPNEK**GGGGSEGGGSGGPGSGGEGSAGGGSAGGGS**RD HMVLHEYVNAAGIT
spacer-inserted Capri	MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTG KLPVPWPTLVTTLTHGVQCFARYPDHMKRHDFFKSAMPEGYVQERTISFK DDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNIY ITAVKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDSHYL STQTVLSKDPNEK**GGGGSEGGGSGGPGSGGEGSAGGGSAGGGS**RDHM VLHEYVNAAGIT
full-length Ruby3	MVSKGEELIKENMRMKVVMEGSVNGHQFKCTGEGEGRPYEGVQTMRIK VIEGGPLPFAFDILATSFMYGSRTFIKYPADIPDFFKQSFPEGFTWERVTRY

	EDGGVVTVTQDTSLEDGELVYNVKVRGVNFPSNGPVMQKKTKGWEPNT EMMYPADGGLRGYTDIALKVDGGGHLHCNFVTTYRSKKTVGNIKMPGV HAVDHRLERIEESDNETYVVQREVAVAKYSNLGGGMDELYK
spacer- inserted Ruby3	MVSKGEELIKENMRMKVVMEGSVNGHQFKCTGEGEGRPYEGVQTMRIK VIEGGPLPFAFDILATSFMYGSRTFIKYPADIPDFFKQSFPEGFTWERVTRY EDGGVVTVTQDTSLEDGELVYNVKVRGVNFPSNGPVMQKKTKGWEPNT EMMYPADGGLRGYTDIALKVDGGGHLHCNFVTTYRSKKTVGNIKMPGV HAVDHRLERIEESDN**GGGGSEGGGSGGPGSGGEGSAGGGSAGGGS**ET YVVQREVAVAKYSNLGGGMDELYK
spacer- inserted Ruby4	MVSKGEELIKENMRTKVVMEGSVNGHHFKCIGEGEGRPYEGVQTMRIKV IEGGPLPFAFDILATSFMYGSRTFIKYPADIPDFFKQSFPEGFTWERVTRYE DGGVITVTQDTCLEDGELVYNVKVSGVNFPSNGPVMQKKTKGWEPNTE MMYPVDGGLRGYTDIALKVDGGGHLHCNFVTTYRSKKTVGNIKMPGVH AVDHRLERIEESDN**GGGGSEGGGSGGPGSGGEGSAGGGSAGGGS**ETY VVQREVAVAKYSNLGGGMDELYK

30-aa spacer

Supplementary Table 3.4 | List of primers used in this study

Split FP-tag

Start/stop codon

pcDNA3.1_GFP$_7$_15 amino acid (aa)-linker_Actin

- Forward primer:

  GTTTAAACTTAAGCTTGGTACCGGATCCATGAAACTCGAATACAATTTTAATA

  GTCATAACGTATACATCACTGCTGATAAGCAAACCGGTGGCAGCGGTGGA

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

pcDNA3.1_GFP$_8$_15aa-linker_Actin

- Forward primer:

  GTTTAAACTTAAGCTTGGTACCGGATCCATGAAGAACGGAATTAAAGCGAAT

  TTCACAGTACGCCATAATGTAACCGGTGGCAGCGGTGGA

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

pcDNA3.1_GFP$_9$_15aa-linker_Actin

- Forward primer:

  GTTTAAACTTAAGCTTGGTACCGGATCCATGGAAGATGGCAGTGTTCAACTTG

  CCGACCATTACCAACAAAACACCCCTATTGGAGACGGTCCGGTAACCGGTGG

  CAGCGGTGGA

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

pcDNA3.1_GFP$_{10}$_15aa-linker_Actin

- Forward primer:

  GTTTAAACTTAAGCTTGGTACCGGATCCATGCTTCTTCCTGATAATCACTACC

  TCTCAACACAAACAGTCCTGAGCAAAGATCCAAATGAAAAAACCGGTGGCA

  GCGGTGGA

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

pcDNA3.1_GFP$_{11}$_15aa-linker_Actin

- Forward primer:

  GTTTAAACTTAAGCTTGGTACCGGATCCATGCGTGACCACATGGTCCTTCATG

  AGTATGTAAATGCTGCTGGGATTACAACCGGTGGCAGCGGTGGA

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

pcDNA3.1_GFP$_{8}$_15aa-linker_Tubulin

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGAAGAACGGAATTAAAGC

  GAATTTCACAGTACGCCATAATGTAACCGGTGGCAGCGGTGGAGGCAGCGCA

  TCCGGCGGAAGCGGAAGCGTGCGTGAGTGCATCTCCAT

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCTTAGTATTCCTCTCCTTCTTCCTCACCCTCT

pcDNA3.1_GFP$_{8}$_15aa-linker_Clathrin

- Forward primer:

GTTTAAACTTAAGCTTGGTACCGGATCCATGAAGAACGGAATTAAAGCGAAT

TTCACAGTACGCCATAATGTAACCGGTGGCAGCGGTGGA

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCCTAGCGGGACAGTGGCGTCT

pcDNA3.1_GFP$_8$_15aa-linker_Lamin A

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGAAGAACGGAATTAAAGC

  GAATTTCACAGTACGCCATAATGTAACCGGTGGCAGCGGTGGAGGCAGCGCA

  TCCGGCGGAAGCGGAAGCGAGACCCCGTCCCAGCGGCGCGCCA

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCTTACATGATGCTGCAGTTCTGGGGGCTCTG

  GGTT

pcDNA3.1_mRuby4$_{11}$_15aa-linker_Clathrin

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGGAAACCTACGTAGTGCA

  AAGAGAAGTGGCAGTTGCCAAATACAGCAACCTTGGTGGTGGCATGGACGA

  GCTGTACAAGACCGGTGGCAGCGGTGGAGGCAGCGCATCCGGCGGAAGCGG

  AAGCATGGCTGATGACTTTGGCTTCTT

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCCTAGCGGGACAGTGGCGTCT

pcDNA3.1_mRuby4$_{11}$_15aa-linker_H2B

- Forward primer:

TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGGAAACCTACGTAGTGCA

AAGAGAAGTGGCAGTTGCCAAATACAGCAACCTTGGTGGTGGCATGGACGA

GCTGTACAAGACCGGTGGCAGCGGTGGAGGCAGCGCATCCGGCGGAAGCGG

AAGCCCAGAGCCAGCGAAGTCTGCT

- Reverse primer:

TGTGCTGGATATCTGCAGAATTCTTACTTAGCGCTGGTGTACTTGGTGAT

pcDNA3.1_mRuby4$_{11}$_15aa-linker_ Lamin A

- Forward primer:

TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGGAAACCTACGTAGTGCA

AAGAGAAGTGGCAGTTGCCAAATACAGCAACCTTGGTGGTGGCATGGACGA

GCTGTACAAGACCGGTGGCAGCGGTGGAGGCAGCGCATCCGGCGGAAGCGG

AAGCGAGACCCCGTCCCAGCGGCGCGCCA

- Reverse primer:

TGTGCTGGATATCTGCAGAATTCTTACATGATGCTGCAGTTCTGGGGGCTCTG

GGTT

pcDNA3.1_Keratin_8aa-linker_mRuby4$_{11}$

- Forward primer:

TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGACTACCTGCAGCCGCCAG

TTCA

- Reverse primer:

TGTGCTGGATATCTGCAGAATTCTTACTTGTACAGCTCGTCCATGCCACCACC

AAGGTTGCTGTATTTGGCAACTGCCACTTCTCTTTGCACTACGTAGGTTTCAG

ACCCCCCGCCAGCGCTGGATCCCGTGTTCTTGGTGCGAAGGACCTGCT

pcDNA3.1_Zyxin_8aa-linker_mRuby4$_{11}$

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCTCTCGAGCTCAAGCTTCGAAT

  TCCATG

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCTTACTTGTACAGCTCGTCCATGCCACCACC

  AAGGTTGCTGTATTTGGCAACTGCCACTTCTCTTTGCACTACGTAGGTTTCAG

  ACCCCCCGCCAGCGCTGGATCCCGTCTGGGCTCTAGCAGTGTGGCACTT

pcDNA3.1_sfCherry2$_{11}$_15aa-linker_H2B

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGTACACCATCGTGGAGCA

  GTACGAGAGAGCCGAGGCCAGACACAGCACCACCGGTGGCAGCGGTGGAGG

  CAGCGCATCCGGCGGAAGCGGAAGCCCAGAGCCAGCGAAGTCTGCT

- Reverse primer:

  TGTGCTGGATATCTGCAGAATTCTTACTTAGCGCTGGTGTACTTGGTGAT

pcDNA3.1_mNeonGreen2$_{11}$_15aa-linker_Actin

- Forward primer:

  TAGCGTTTAAACTTAAGCTTGGTACCGGATCCATGACCGAGCTCAACTTCAAG

  GAGTGGCAAAAGGCCTTTACCGATATGATGAGATCTGGCAGCGGTGGAGGCA

  GCGCAT

- Reverse primer:

  ACTGTGCTGGATATCTGCAGAATTCCTAGAAGCATTTGCGGTGGA

Supplementary Table 3.5 | Amino acid sequences of split FP-tags.

GFP_7	KLEYNFNSHNVYITADKQ
GFP_8	KNGIKANFTVRHNV
GFP_{10}	LLPDNHYLSTQTVLSKDPNEK
$EBFP2_{11}$ $Capri_{11}$ CFP_{11} $Cerulean_{11}$ GFP_{11} YFP_{11}	RDHMVLHEYVNAAGIT
$mNeonGreen2_{11}$	TELNFKEWQKAFTDMM
$mRuby4_{11}$ *	ETYVVQREVAVAKYSN
$sfCherry2_{11}$	YTIVEQYERAEARHST

* The original 11th β-strand of $mRuby4_{11}$ consists of 26 amino acids (ETYVVQREVAVAKYSN**LGGGMDELYK**), and we removed the last 10 amino acids from it. These amino acids are unlikely to be essential, because they form an unstructured polypeptide chain and do not provide the structural rigidity to the β-strand.

3.7 METHODS

Molecular cloning

The amino acid sequence of EBFP2 was obtained from a published report [16] and slightly altered (see **Supplementary Table 3.2**). EBFP2 was split at the same site at $GFP_{1-10/11}$. We introduced six substitutions to $EBFP2_{1-10}$ (S65T/Q80R/F99S/V128T/M153T/V163A), which altered the $EBFP2_{1-10/11}$ spectral property to $Capri_{1-10/11}$. $EBFP2_{1-10}$ and $Capri_{1-10}$ were synthesized (Integrated DNA Technologies) and cloned into the mammalian expression vector pcDNA3.1 (Invitrogen) between KpnI and EcoRI (NEB) by using an In-Fusion HD cloning kit (Takara Bio). In order to clone specific DNA fragments into a plasmid vector, we used the In-Fusion HD cloning system in this entire study.

For the expression of CFP_{1-10} in mammalian cells, we introduced the corresponding point mutations into pcDNA3.1-GFP_{1-10} (Addgene #70219) using Q5 High-Fidelity DNA Polymerase (NEB). Primers used were as follows: CFP_{1-10}_forward (5' - ACGCTTACGTggGGA GTTCAGTGC - 3'), and CFP_{1-10}_reverse (5' - TGTTACGAGAGTCGGCCA - 3'). To express $Cerulean_{1-10}$, we synthesized and cloned their DNA fragment into the KpnI/EcoRI sites of pcDNA3.1. For the nucleotide sequence of $Cerulean_{1-10}$, see Supplementary Table 2.

GFP permits circular permutation of the amino acid sequence [25]. By linking the N- and C-termini and cutting out a single β-strand, any one of the eleven β-strands would become a new split GFP-tag. Huang et al. [26] previously investigated all possible β-strands, and measured the solubility and relative reconstituted fluorescence intensity of each split GFP construct in *E. coli*. We tested Huang's design of the β-strands 7, 8, 9, or 10 system in human cells. We constructed plasmids encoding each of the β-strands fused to β-actin. We used mEmerald-Actin-C-18 (Addgene #53978) as the template for PCR amplification of β-actin. The actin gene was amplified using

primers, in which DNA sequences encoding the β-strands were included in part (For the sequence information of the primers, see **Supplementary Table 3.4**). The resultant PCR products were cloned into the KpnI/EcoRI sites of pcDNA3.1.

GFP_{8-6}, GFP_{9-7}, GFP_{10-8}, and GFP_{11-9} were amplified from super-folder GFP OPT [1] by PCR and inserted into the KpnI/EcoRI sites of pcDNA3.1. For the nucleotide sequences of GFP_{8-6}, GFP_{9-7}, and GFP_{11-9}, see Supplementary Table 2.

To demonstrate the usefulness of $mRuby4_{11}$, $sfCherry2_{11}$, $mNeonGreen2_{11}$ and GFP_8, we labeled several cellular proteins with these split FP-tags. To construct plasmids of split FP-tag fusions, DNA fragments encoding cellular proteins were amplified by PCR with sets of primers (see also **Supplementary Table 3.4**) and cloned into the KpnI/EcoRI sites of pcDNA3.1. For the PCR amplification, we used the following DNA templates: mEmerald-Actin-C-18; mEmerald-Clathrin-15 (Addgene #54040); sfGFP-H2B-C-10 (Addgene #56367); pBABE-puro-GFP-wt-lamin A (Addgene #17662); mEmerald-Keratin-17 (Addgene #54134); sfGFP-Zyxin-6 (Addgene #56491).

The amino acid sequences of split FP-tags (i.e., GFP_7, GFP_8, GFP_{10}, GFP_{11}, $mNeonGreen2_{11}$, $sfCherry2_{11}$, and $mRuby4_{11}$) were listed in Supplementary Table 5.

$sfCherry2_{1-10}$, $mNeonGreen2_{1-10}$, and sfGFP-Zyxin plasmids [13] that can be transfected in mammalian cells are available through Addgene (#82603, #82610, and #56491, respectively).

Mutagenesis and screening of libraries

When engineering split orange-red FP variants, we adopted a complementation assay previously described to optimize split mCherry2 in *E. coli* [13]. We inserted a 30-aa spacer (GGGGSEGGGSGGPGSGGEGSAGGGSAGGGS) between the tenth and eleventh β strands of each of the following fluorescent proteins; mKO2, mRuby3, mApple, and mScarlet-I [15, 21-23].

The corresponding DNA sequences were directly synthesized and then cloned into the BamHI/XhoI sites of the *E. coli* expression vector pET28a (Novagen).

The longer spacer insertion eliminated colony fluorescence of mKO2, mApple and mScarlet-I, whereas colonies expressing spacer-inserted mRuby3 gave low signal (**Supplementary Figure 3.8**). To improve the brightness of spacer-inserted mRuby3, we mutagenized it by using a GeneMorph II Random Mutagenesis Kit (Agilent). Mutants were expressed and screened in pET28a. Plasmids were transformed into *E. cloni* EXPRESS Electrocompetent Cells (Lucigen). Transformation was performed by the Gene Pulser Electroporation Systems (BioRad). Colonies were grown on LB agar media (30 μg/mL Kanamycin) at 37°C for 24 hours and for additional 12-48 hours at 37°C after induction with 1 mM IPTG. For each round of mutagenesis, the number of colonies screened was at least 1 x 10^4. Colonies expressing spacer-inserted mRuby3 variants were screened for fluorescence with the ChemiDoc Imaging System (BioRad). The imaging system was equipped with an Epi-green 520-545 nm excitation source, a Green Epi 605/50 filter, and a cooled CCD camera.

Through library screening, we obtained an extremely bright variant of spacer-inserted mRuby3, which we named spacer-inserted mRuby4 (**Supplementary Figure 3.9**). The $mRuby4_{1-10}$ sequence of spacer-inserted mRuby4 was amplified by PCR and cloned into the KpnI/EcoRI sites of pcDNA3.1. For the nucleotide sequence of $mRuby4_{1-10}$, see Supplementary Table 2.

Protein production and characterization of FPs

For spectral characterization of FPs, we produced and purified recombinant proteins: full-length EBFP2, spacer-inserted EBFP2, spacer-inserted Capri, full-length mRuby3, spacer-inserted mRuby3, and spacer-inserted mRuby4 (the amino acid sequences of those were listed in Supplementary Table 3). We designed pET plasmids such that recombinant proteins were labeled

at the C termini with poly-histidine tags. The plasmids were introduced into BL21(DE3) Competent *E. coli* cells (NEB) via transformation. Cells were grown in 250 mL LB medium at 37 °C for 6 hours (OD_{600} = 0.5), induced with IPTG (1mM) for 4 hours, and harvested by centrifugation. Cell pellets were lysed by French press. His-tagged proteins were purified with HisPur Cobalt Resin (Pierce). Proteins were further desalted into PBS pH7.4 using a GE Healthcare illustra NAP column (GE Healthcare). Extinction coefficients were calculated using Beer-Lambert law [10]. Quantum yields were determined using EBFP2 [16], and Rhodamine B (Wako) as reference fluorophores. The absorbance signals of samples and reference were measured using a microreader (Biotek Synergy 2). Diluted samples and reference were added into a quartz fluorescence cuvette (Thorlabs), and their integrated fluorescence intensities were measured by a fluorescence spectrophotometer (Hitachi F-7100). With the quantum yield of reference to be known, the final quantum yields of samples were attained using:

$Qs = Qr \times (Ar/As) \times (Es/Er) \times (ns/nr)^2$ ["*r*" and "*s*" refer to the reference and samples]

where Q is the quantum yield, n is the refractive index, A is the absorbance of solution, and E is the integrated fluorescence intensity of emitted light.

Fluorescence Imaging

Confocal microscopy images of $mRuby4_{1\text{-}10/11}$, $GFP_{1\text{-}10/11}$ and $mNeonGreen2_{1\text{-}10/11}$ were acquired on an inverted fluorescence microscope (Ti-E, Nikon) with a 100x 1.45 NA oil immersion objective (Plan Apo, Nikon). The microscope was attached to the Dragonfly Spinning disk confocal unit (CR-DFLY-501, Andor). Two excitation lasers (40 mW 488 nm and 50 mW 561 nm lasers) were coupled to a multimode fiber passing through the Andor Borealis unit. A dichroic mirror (Dragonfly laser dichroic for 405-488-561-640) and band-pass filters (525/50 nm and

600/50 nm bandpass emission wheel filters) were selected for two-color imaging. The images were recorded with an EM-CCD camera (iXon, Andor).

Confocal microscopy images of EBFP2$_{1\text{-}10/11}$, Capri$_{1\text{-}10/11}$, CFP$_{1\text{-}10/11}$, and Cerulean$_{1\text{-}10/11}$, were collected by using an upright microscope (Axio imager Z2, Zeiss) with a 63x 1.4 NA oil immersion objective (Plan Apo, Zeiss). The upright microscope had the LSM 880 Scan-head (Zeiss) with 32 channel GaAsP spectral PMT detector. It was equipped with six laser lines (Diode 405 nm; Argon 458, 488, 514 nm; HeNe 543, 633 nm). We used the 405-nm diode laser for EBFP2$_{1\text{-}10/11}$, and Capri$_{1\text{-}10/11}$ (main beam filter MBS-405 and 488/543, 409-491 nm barrier filter), the 458-nm Argon line for CFP$_{1\text{-}10/11}$, and Cerulean$_{1\text{-}10/11}$ (main beam filter MBS-458/514, 454-518 nm barrier filter). To characterize a spectrum from each individual FP, spectral images were acquired. We used 8.9 nm channel widths, meaning that each of the 32 channels on the spectral detector captured light over 8.9 nm bandwidth of the visible and near infrared spectrum. Sequential spectral image acquisitions were achieved in order of ascending wavelength of the excitation laser: 405, 488, 543 nm. A typical data set consisted of 32 images (1024 x 1024 pixels), corresponding to different wavelengths from 410 to 695 nm. We then performed linear unmixing (**Figure 3.2h, Figure 3.4g, and Supplementary Figure 3.12, 13, and 14**) using Carl Zeiss' Zen software. Confocal microcopy images showed average intensity z projections, unless otherwise noted in the figure legends. Analysis of the confocal images was performed on Fiji software (NIH).

Flow cytometry

Fluorescence of cells was measured using CytoFLEX (Beckman Coulter) in the CTEGD Cytometry Shared Resource at the University of Georgia (UGA). The instrument had four excitation lasers (405 nm, 488 nm, 561 nm, 610 nm) and three band-pass filters (450/45 nm, 525/40 nm, 610/20 nm). Post-acquisition analysis was carried out using FlowJo software (Treestar, Inc.)

Cell culture, transfection and drug treatment

HEK 293, HEK 293T, HeLa and U2OS cells (gifted from Drs. Eggenschwiler and Kipreos, UGA) were grown in Dullbecco's Eagle's medium (HyClone), supplemented with fetal bovine serum (10%, *v/v*; Atlanta Biologicals) and penicillin/streptomycin (100 U/mL penicillin and 100 μg/mL streptomycin; HyClone). Cells were cultured at 37°C and 5% CO_2 in a humidified incubator. Plasmids were transfected at 400-800 ng DNA per well with Lipofectamine 2000 (3 μL, Invitrogen) or polyethylenimine (3 μL of 1 mg/mL PEI, Polysciences, Inc.) into Nunc Lab-Tek II Chambered Cover Glass (size: 8 wells, Nalge Nunc International) or Corning Costar Cell Culture Plates (size: 12 or 24 wells, Corning). In particular, for the quantitative comparisons in cellular fluorescence intensity (Fig. 4a, and Supplementary Figs 4 and 6), we transfected the plasmids of FP-tags and their partners at 400ng and 800ng, respectively. Cells were fixed with buffered 4 % paraformaldehyde (Electron Microscope Sciences), mounted with PBS, and imaged by confocal microscopy.

For spectral imaging of multicolor H2B fusions (**Figure 3.2h, and Supplementary Figure 3.13 and 14**), we used a Cdk1 inhibitor (10 μM of RO-3306, Sigma-Aldrich) to synchronize HEK 293 cells. HEK 293 cells were treated with the inhibitor for 18 hours, blocked in the G2/M phase. For release from the inhibitor, we washed the culture five times with prewarmed culture media. Released cells returned to normal cell cycle progression, and were eventually fixed with 100% ice-cold methanol and mounted with PBS for microscopy.

Knock-in cell creation

For knock-in of mRuby4$_{11}$ into the H2B locus, we ordered 200-nt HDR templates in single-stranded DNA (5' - gcccggcgagctggccaagcacgccgtgtccgagggcaccaaggcggtcaccaagtacaccagctc

caagGGTGGCGGCGAAACCTACGTAGTGCAAAGAGAAGTGGCAGTTGCCAAATACAG CAACtgagtccctgccgggacctggcgctcgctcgctcgagtcgccggctgcttgactccaaaggctcttttcagag - 3', Integrated DNA Technologies). Cas9 protein was expressed in *E. coli* and purified by the Kipreos laboratory at UGA as described previously [11]. sgRNA and Cas9/sgRNA ribonucleoprotein complexes were prepared as describe before [12]. After the treatment of HEK293 FT cells with nocodazole (200 ng/mL, Sigma-Aldrich) for 16 hours, we performed electroporation on an Amaxa Nucleofector 2b device with Nucleofector Solution V reagents (Lonza).

Nocodazole-treated cells were resuspended at a concentration of 1 x 10^4 cells/μL in 100 μL of Nucleofector Solution V. We added cells to the RNP/donor template mixture (50 μL), electroporated using the Q-001 program, and quickly transferred to 12-well plates with pre-warmed media. Electroporated cells were cultured for 2-5 days and transfected with mRuby4$_{1-10}$ plasmid.

Data availability

Relevant plasmids and sequences have been deposited in Addgene (www.addgene.org). The raw data referring to the plots shown in the main figures are provided in Supplementary figures. All relevant data are available for authors upon requests.

Statistics and reproducibility

All experiments for the measurement of signal levels were replicated multiple times independently. Statistical analyses were performed using GraphPad Prism 7. Error bars in all figures refer to the standard error of the mean (SEM).

3.8 ACKNOWLEDGEMENTS

We thank Zahra Abdul Nawaz and Dr. Edward Kipreos (University of Georgia) for assistance in *in vitro* Cas9 expression and purification, Dr. Rick Tarleton (University of Georgia) for use of his Nucleofector equipment, and members of the Kamiyama Lab for comments on the manuscript. We also acknowledge the assistance of the Biomedical Microscopy Core and the CTEGD Cytometry Shared Resource Laboratory. This work was supported by the University of Georgia Faculty Seed Grant (to D.K.) and an NIH R01 NS107558 (to R.T., and D.K.). R.T. was supported by a predoctoral fellowship from the Nakajima Foundation.

3.9 AUTHOR CONTRIBUTIONS

R.T. and D.K. conceived and designed the experiments. R.T. performed random mutagenesis, protein labeling, CRISPR-mediated knock-in, flow cytometry, and imaging experiment. R.T. analyzed the data. F.J., R.T., and J.X. performed *in vitro* characterizations of split FPs. D.K. wrote the manuscript.

3.10 COMPETING FINANTIAL INTERESTS

The authors declare no competing financial interests.

3.11 REFERENCES

1. Cabantous, S., et al., *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nat Biotechnol, 2005. **23**(1): p. 102-7.

2. Kamiyama, D., et al., *Versatile protein tagging in cells with split fluorescent protein.* Nat Commun, 2016. **7**: p. 11046.

3. Romei, M.G. and Boxer, S.G., *Split green fluorescent proteins: scope, limitations, and outlook.* Annu. Rev. Biophys., 2019. **48**.

4. Pedelacq, J.D. and Cabantous, S., *Development and applications of superfolder and split fluorescent protein detection systems in biology.* Int J. Mol. Sci., 2019. **20**.

5. Pinaud, F. and Dahan, M., *Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins.* Proc Natl Acad Sci U S A, 2011. **108**(24): p. E201-10.

6. Koker, T., et al., *Cellular imaging by targeted assembly of hot-spot SERS and photoacoustic nanoprobes using split-fluorescent protein scaffolds.* Nat Commun, 2018. **9**(1): p. 607.

7. Kim, Y.E., et al., *Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency.* Nat Commun, 2015. **6**: p. 7134.

8. Chun, W., et al., *Split GFP complementation assay for quantitative measurement of tau aggregation in situ.* Methods Mol Biol, 2011. **670**: p. 109-23.

9. Van Engelenburg, S.B. and Palmer, A.E., *Imaging type-III secretion reveals dynamics and spatial segregation of Salmonella effectors.* Nat Methods, 2010. **7**(4): p. 325-30.

10. Koker, T., et al., *Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process.* Sci Rep, 2018. **8**(1): p. 5344.

11. Paix, A., et al., *Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks.* Proc Natl Acad Sci U S A, 2017. **114**(50): p. E10745-E10754.

12. Leonetti, M.D., et al., *A scalable strategy for high-throughput GFP tagging of endogenous human proteins.* Proc Natl Acad Sci U S A, 2016. **113**(25): p. E3501-8.

13. Feng, S., et al., *Improved split fluorescent proteins for endogenous protein labeling.* Nat Commun, 2017. **8**(1): p. 370.

14. Feng, S., et al., *Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses.* Commun Biol, 2019. **2**: p. 344.

15. Bajar, B.T., et al., *Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting.* Sci Rep, 2016. **6**: p. 20889.

16. Ai, H.W., et al., *Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins.* Biochemistry, 2007. **46**(20): p. 5904-10.

17. Pedelacq, J.D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nat Biotechnol, 2006. **24**(1): p. 79-88.

18. Rizzo, M.A., et al., *An improved cyan fluorescent protein variant useful for FRET.* Nat Biotechnol, 2004. **22**(4): p. 445-9.

19. Day, R.N. and Davidson, M.W., *The fluorescent protein palette: tools for cellular imaging.* Chem Soc Rev, 2009. **38**(10): p. 2887-921.

20. Cranfill, P.J., et al., *Quantitative assessment of fluorescent proteins.* Nat Methods, 2016. **13**(7): p. 557-62.

21. Sakaue-Sawano, A., et al., *Visualizing spatiotemporal dynamics of multicellular cell-cycle progression.* Cell, 2008. **132**(3): p. 487-98.

22. Shaner, N.C., et al., *Improving the photostability of bright monomeric orange and red fluorescent proteins.* Nat Methods, 2008. **5**(6): p. 545-51.

23. Bindels, D.S., et al., *mScarlet: a bright monomeric red fluorescent protein for cellular imaging.* Nat Methods, 2017. **14**(1): p. 53-56.

24. Vassilev, L.T., et al., *Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1.* Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10660-5.

25. Baird, G.S., et al., *Circular permutation and receptor insertion within green fluorescent proteins.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11241-6.

26. Huang, Y.M., et al., *Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein.* Protein Sci, 2011. **20**(11): p. 1775-80.

27. Fujita, Y., et al., *Zyxin is a novel interacting partner for SIRT1.* BMC Cell Biol, 2009. **10**: p. 6.

28. Cattaruzza, M., et al., *Focal adhesion protein zyxin is a mechanosensitive modulator of gene expression in vascular smooth muscle cells.* Hypertension, 2004. **43**(4): p. 726-30.

29. Hervy, M., et al., *The LIM Protein Zyxin Binds CARP-1 and Promotes Apoptosis.* Genes Cancer, 2010. **1**(5): p. 506-515.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 SUMMARY OF DISSERTATION

Slit-Dscam1 interaction underlies aCC dendritogenesis

In Chapter 2, we revealed a new role of Slit/Dscam1 signaling in aCC dendritogenesis. Previous studies have shown that the position of aCC dendritogenesis is dictated by an interneuronal interaction between the aCC and the MP1 through Dscam1. Given the diversity of Dscam1 isoforms and the randomness of Dscam1 isoform selection across individual neurons, it is likely that the aCC-MP1 contact involves distinct Dscam1 isoforms, which would not bind directly. Our main question was whether such isoform-independent Dscam1 interaction might exist, and if it does, how? Firstly, we tested if Dscam1 isoforms bind one another in an isoform-independent manner. We visualized two distinct isoforms of Dscam1, which do not directly bind, in S2 cells and found that they accumulate at cell-cell contact sites (**Figure 2.3**). This raised the possibility of the existence of co-factors that mediate the isoform-independent interaction. Secondly, we showed that Slit, a secreted molecule with a known axon guidance function, co-localizes with Dscam1 both in cell culture and the embryonic CNS, indicating the interplay between Slit and Dscam1 during neurogenesis (**Figure 2.4 and 2.8**). Thirdly, we employed *slit* mutant embryos to test the involvement of Slit in aCC dendritogenesis. Both loss-of-function or reduced function of *slit* led to defects in the number of aCC dendritic tips, indicating the role of Slit in aCC dendritogenesis (**Figure 2.5**). In addition, transheterozygous mutants of *slit* and *dscam1* resulted in obvious defects

in aCC dendritogenesis, demonstrating the genetic interaction of Slit and Dscam1 (**Figure 2.7**). Fourthly, we characterized Slit-Dscam1 interaction using biochemical assays. We tested if Slit interacts with multiple Dscam1 isoforms using co-immunoprecipitation and discovered that Slit-Dscam1 binding is isoform-independent (**Figure 2.10a**). Subsequently, we showed that different Dscam1 isoforms form a complex in the presence of Slit, suggesting the role of Slit as a co-factor mediating isoform-independent interaction of Dscam1 (**Figure 2.10b**). To gain more insights into how Slit could accommodate multiple Dscam1 isoforms at once, we tested whether Slit molecules form a complex by themselves. Indeed, we observed that Slit pulled down other Slit molecules, suggesting that multiple Slit proteins can exist in a complex through either oligomerization or other unidentified co-factors (**Figure 2.10c-d**). This result further supports the possibility that Slit mediates the interaction between multiple different Dscam1 isoforms. Altogether, we proposed a paradigm, in which Slit mediates Dscam1 interaction at the aCC-MP1 contact site and how this molecular complex would play a role in aCC dendritogenesis (**Figure 2.11**).

Development of split FP color palette

In Chapter 3, we expanded the toolkit of available split FP variants to facilitate multiplexed protein labeling. Firstly, we engineered blue and cyan variants of $GFP_{1-10/11}$, $EBFP_{1-10/11}$, $Capri_{1-10/11}$, and $Cerulean_{1-10/11}$, by introducing existing color-shifting mutations (**Figure 3.1**). Secondly, we developed a new red split FP variant, $mRuby4_{1-10/11}$, by directed evolution and demonstrated its use in endogenous protein tagging via CRISPR/Cas9-mediated homology directed repair (**Figure 3.2 and Supplementary Figure 3.11**). Thirdly, we showed that the newly created split FP variants ($EBFP_{1-10/11}$, $Capri_{1-10/11}$, $Cerulean_{1-10/11}$, and $mRuby4_{1-10/11}$) and existing variants ($GFP_{1-10/11}$ and $sfCherry2_{1-10/11}$) possess fluorescent properties distinguishable by spectral imaging (**Figure 3.2**). Fourthly, we characterized the binding specificities of different combinations of FP_{1-10} and FP_{11}

fragments and found that $FP_{1-10/11}$ pairs with a high amino acid sequence similarity (*e.g.,* $mRuby4_{1-10/11}$ and $sfCherry2_{1-10/11}$) exhibit cross-reactivity (**Figure 3.3**). Fifthly, to tackle the issue of cross-reactivity, we generated split GFP variants with distinct binding specificities ($GFP_{8-6/7}$, $GFP_{9-7/8}$, and $GFP_{11-9/10}$) from circular-permutated GFPs (**Figure 3.4a-f**). Lastly, as a proof of concept, we conducted 4-color imaging of cellular proteins labeled with 4 split FP variants of distinct binding and spectral properties (**Figure 3.4g**). Altogether, we not only expanded the spectral landscape of the split FP system, but also proposed a generalizable pipeline to fashion orthogonal split FP pairs with distinct binding properties, starting from circular-permutated FPs. In theory, this strategy demonstrated here can be applied to any available fluorescent proteins of a given user's choice.

4.2 FUTURE DIRECTIONS

Slit/Dscam1 signaling in neural circuit assembly

In this dissertation, we proposed that the heterophilic complex composed of Slit and Dscam1 promotes dendritic outgrowth in the embryonic aCC motoneuron. This is in stark contrast to the well-studied function of Dscam1 in dendritic repulsion through isoform-specific homophilic interaction. This finding suggests a model, in which the presence of Slit serves as a switch from the repulsion to the outgrowth of dendritic processes. This leads to the next question of what the molecular basis of this switch is. One possible explanation is context-dependent protein complex assembly; that is, Dscam1 protein complex may form in different compositions of ligands, receptors or intracellular molecules across different tissues or different developmental timepoints. In fact, recent studies found that Dscam1 interacts with other membrane proteins such as the receptor tyrosine phosphatase 69D and Robo1 in a Slit-dependent manner, indicating that Slit binding is accompanied by recruitment of other molecules [1, 2]. To further examine the context-

dependent compositional difference in Dscam1 complex assembly, cell-type-specific proteomics methods like split TurboID could be implemented. Briefly, TurboID is a promiscuous biotin ligase capable of biotinylating proteins in proximity [3]. When genetically fused to a protein of interest, TurboID labels interactors with biotin, allowing for downstream affinity-purification and identification of interactors using mass spectrometry. Split TurboID can further confer spatial and temporal specificity to proteomics depending on when the two fragments are co-expressed at once in a system [4]. This method can be employed to identify components of Dscam1 complex in different contexts; for example, Dscam1 can be endogenously tagged with one fragment of split TurboID, and the other fragment can be expressed in a specific tissue using the GAL4/UAS system, resulting in a tissue-specific reconstitution of TurboID activity. Eventually, such cell-type-specific proteomics will allow us to ask which difference in molecular composition causes divergent consequences of Dscam1 signaling, and ultimately help us understand in part the molecular basis of neural circuit assembly.

Our endeavors to characterize the role of Slit/Dscam1 signaling will provide insights into that of human DSCAM as well. Given the significant homology within the extracellular domains of *Drosophila* Dscam1 and human DSCAM (~40%) [5], it is probable that Slit-Dscam1 interaction is conserved across the two species. In fact, the interaction between Dscam1 and Netrin, a midline-secreted axon guidance protein, is conserved between *Drosophila* and mammals [6, 7], suggesting the conserved role of Dscam1 as a receptor for secreted ligands in neural development, apart from its conserved role in repulsive signaling through homophilic interaction. Additionally, our biochemical studies showed that Dscam1 binds Slit in an isoform-independent manner, suggesting that the Slit-Dscam1 binding interface may lie within invariable regions of Dscam1. Considering that human DSCAM does not display any isoform diversity, such an isoform-independent nature

of Slit-Dscam1 binding found in *Drosophila* is more than likely to be preserved in humans as well. To evaluate this possibility, we can conduct biochemical assays to examine whether the human counterparts, DSCAM and SLITs, (three paralogues exist) interact or not. Future investigation to in rodent models will further reveal the physiological and pathological role of DSCAM signaling, potentially conserved across species.

Visualizing molecular complex assembly *in vivo* using the diversified split FP system

While our demonstration of newly developed split FP variants in Chapter 3 were restricted to mammalian cells, our expanded split FP toolkit can be readily adopted to other systems, including the *Drosophila* system. The split FP system becomes even more powerful when combined with binary expression systems like the GAL4/UAS method; briefly, the first step is the insertion of the smaller split FP fragment (*e.g.,* GFP_{11}) into a genomic locus. This gene insertion procedure could be achieved by precise genome editing methods like CRISPR/Cas9-mediated homology directed repair. The second step is to express the larger split FP fragment (*e.g.,* GFP_{1-10}) in a specific tissue using the corresponding GAL4/UAS system. Consequently, the target protein becomes fluorescent upon tissue-specific FP reconstitution. Building on this principle, Kamiyama and colleagues have recently adopted multicolor split FP variants in *Drosophila* and simultaneously visualized multiple endogenous synaptic proteins in a subset of neuroendocrine cells [8]. This case study further demonstrated the potential of the split FP system in visualizing protein complex assembly *in vivo*.

Success of endogenous protein tagging using split FP largely depends on the expression of a protein of interest [9]. In case that the target protein expression is extremely low, the split FP system provides multiple solutions. One solution is to link the smaller FP fragment in tandem and recruit multiple copies of the complementary fragment to the target protein. Another solution is to engineer brighter split FP variants, which can be readily accomplished by implementing the

directed evolution method shown in Chapter 3. Additionally, future engineering work will further make the split FP system available across model systems to help us understand the dynamics of functional protein networks.

4.3 REFERENCES

1. Dascenco, D., et al., *Slit and Receptor Tyrosine Phosphatase 69D Confer Spatial Specificity to Axon Branching via Dscam1.* Cell, 2015. **162**(5): p. 1140-54.
2. Alavi, M., et al., *Dscam1 Forms a Complex with Robo1 and the N-Terminal Fragment of Slit to Promote the Growth of Longitudinal Axons.* PLoS Biol, 2016. **14**(9): p. e1002560.
3. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.
4. Mair, A., et al., *Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID.* Elife, 2019. **8**.
5. Schmucker, D., et al., *Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity.* Cell, 2000.
6. Andrews, G.L., et al., *Dscam guides embryonic axons by Netrin-dependent and -independent functions.* Development, 2008. **135**(23): p. 3839-3848.
7. Ly, A., et al., *DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1.* Cell, 2008. **133**(7): p. 1241-54.
8. Kamiyama, R., et al., *Cell-type-specific, multicolor labeling of endogenous proteins with split fluorescent protein tags in Drosophila.* Proc Natl Acad Sci U S A, 2021. **118**(23).
9. Leonetti, M.D., et al., *A scalable strategy for high-throughput GFP tagging of endogenous human proteins.* Proc Natl Acad Sci U S A, 2016. **113**(25): p. 3501-3508.