



(86) **Date de dépôt PCT/PCT Filing Date:** 2015/07/14
(87) **Date publication PCT/PCT Publication Date:** 2016/01/21
(85) **Entrée phase nationale/National Entry:** 2017/01/11
(86) **N° demande PCT/PCT Application No.:** US 2015/040439
(87) **N° publication PCT/PCT Publication No.:** 2016/011070
(30) **Priorité/Priority:** 2014/07/14 (US62/024,241)

(51) **Cl.Int./Int.Cl. A61K 49/00** (2006.01)
(71) **Demandeur/Applicant:**
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA,
US
(72) **Inventeurs/Inventors:**
TANENBAUM, MARVIN E., US;
GILBERT, LUKE A., US;
QI, LEI S., US;
WEISSMAN, JONATHAN S., US;
VALE, RONALD D., US
(74) **Agent:** SMART & BIGGAR

(54) **Titre : SYSTEME DE MARQUAGE DE PROTEINE POUR L'IMAGERIE MONOMOLECULAIRE IN VIVO ET LA REGULATION DE
LA TRANSCRIPTION GENIQUE**
(54) **Title: A PROTEIN TAGGING SYSTEM FOR IN VIVO SINGLE MOLECULE IMAGING AND CONTROL OF GENE
TRANSCRIPTION**

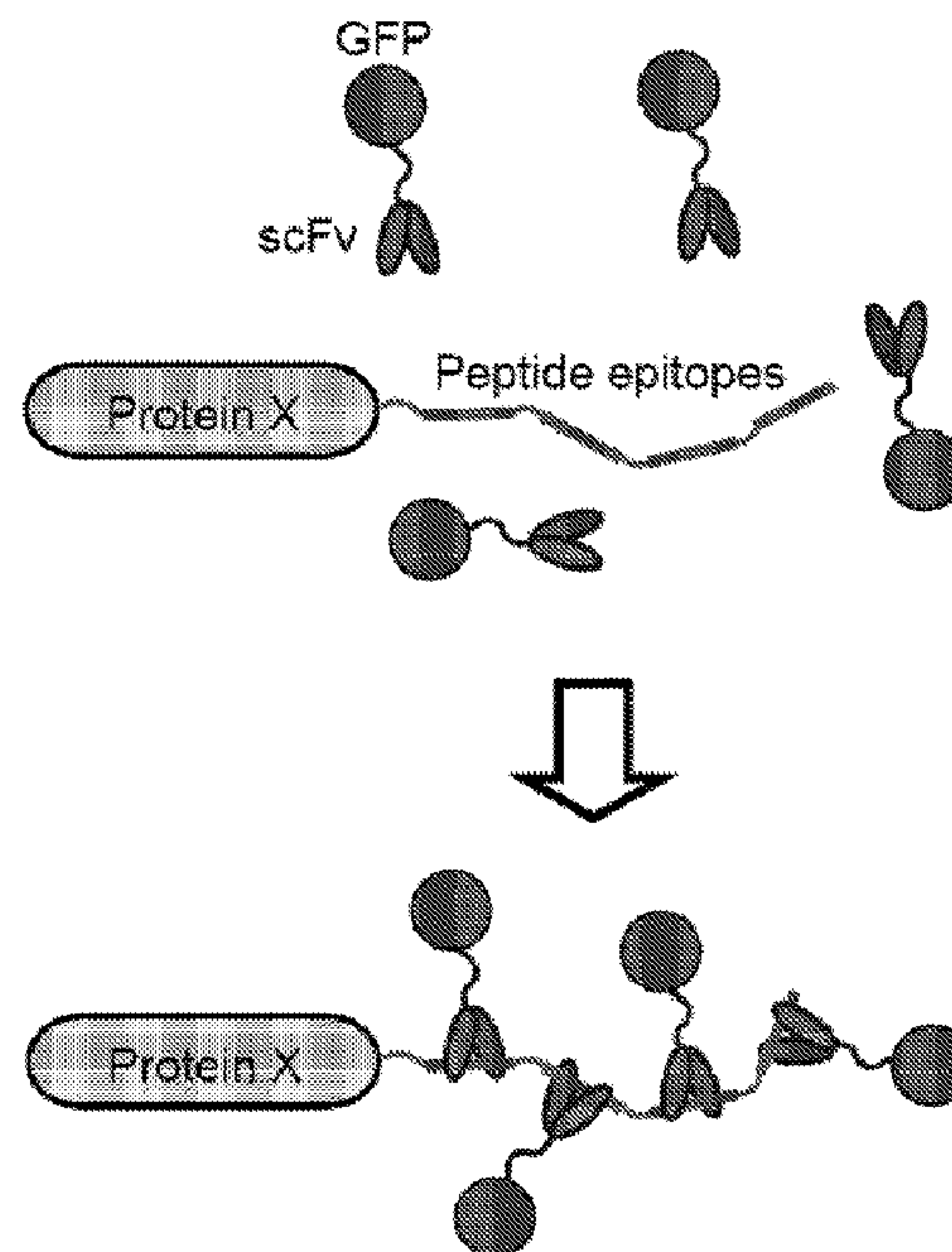


FIG. 1A

(57) **Abrégé/Abstract:**

Methods, compositions, and kits are provided for imaging a polypeptide of interest. Methods, compositions, and kits are also provided for site-specific transcriptional regulation of one or more genetic elements.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau(10) International Publication Number
WO 2016/011070 A3(43) International Publication Date
21 January 2016 (21.01.2016)(51) International Patent Classification:
C07K 16/18 (2006.01)(21) International Application Number:
PCT/US2015/040439(22) International Filing Date:
14 July 2015 (14.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/024,241 14 July 2014 (14.07.2014) US(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607 (US).(72) Inventors: **TANENBAUM, Marvin E.**; 201 Laguna Street, Apt. 12, San Francisco, California 94102 (US). **GILBERT, Luke A.**; 222 Haight Street, #3, San Francisco, California 94102 (US). **QI, Lei S.**; 1111 FranklinStreet, 12th Floor, Oakland, California 94607 (US). **WEISSMAN, Jonathan S.**; 912 Castro Street, San Francisco, California 94114 (US). **VALE, Ronald D.**; 22 Hillcrest Road, Tiburon, California 94920 (US).(74) Agents: **GOETZ, David H.** et al.; Kilpatrick Townsend and Stockton LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111 (US).(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: A PROTEIN TAGGING SYSTEM FOR IN VIVO SINGLE MOLECULE IMAGING AND CONTROL OF GENE TRANSCRIPTION

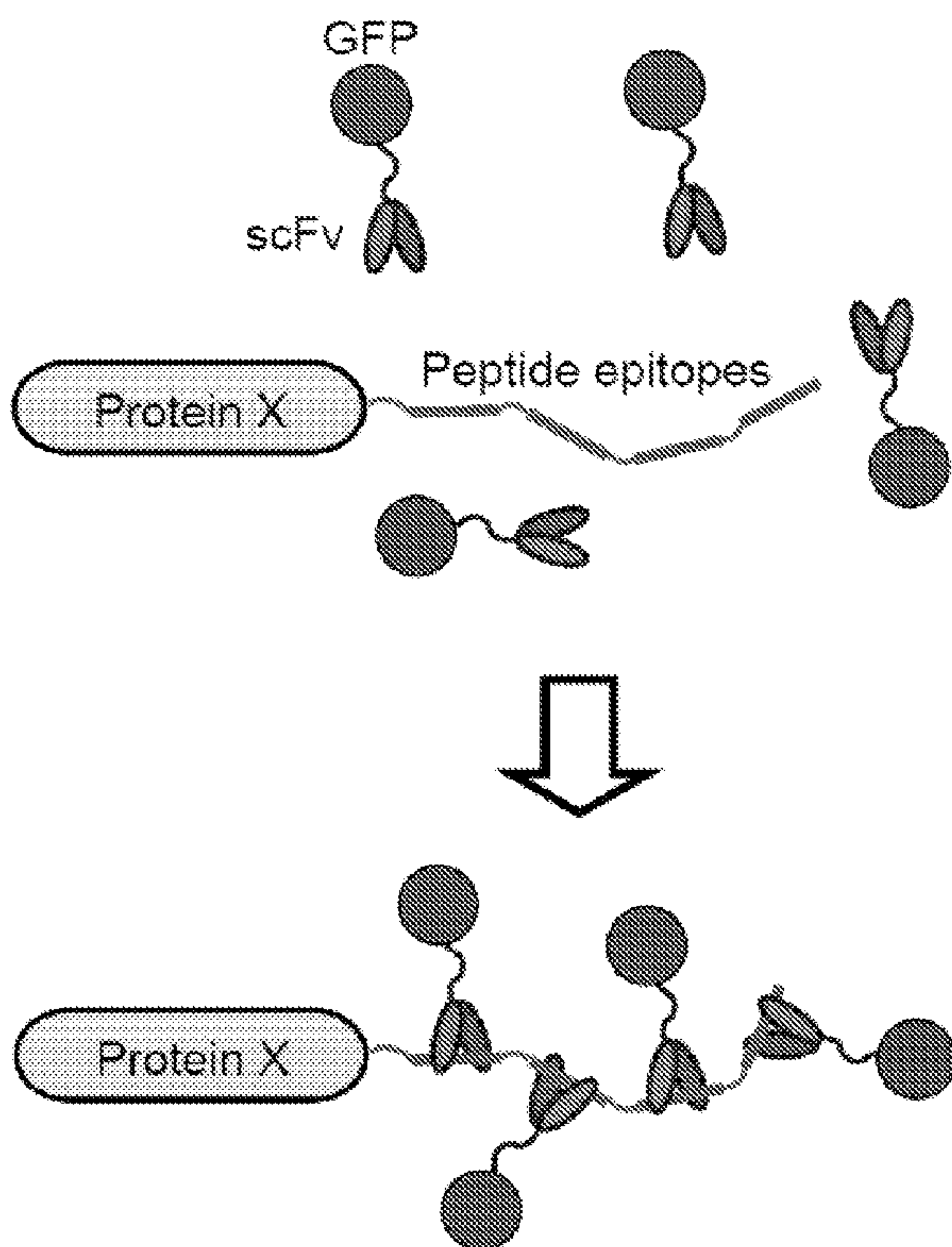


FIG. 1A

(57) **Abstract:** Methods, compositions, and kits are provided for imaging a polypeptide of interest. Methods, compositions, and kits are also provided for site-specific transcriptional regulation of one or more genetic elements.

WO 2016/011070 A3

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:

3 March 2016

**A PROTEIN TAGGING SYSTEM FOR IN VIVO SINGLE MOLECULE
IMAGING AND CONTROL OF GENE TRANSCRIPTION**

**STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER
FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT**

[0001] This invention was made with government support under grant nos. P50 GM102706, R01 DA036858, OD017887 and R37 GM038499 awarded by the National Institutes of Health. The government has certain rights in the invention.

CROSS-REFERENCES TO RELATED APPLICATIONS

[0002] This application claims priority to U.S. Provisional Application No. 62/024,241, filed on July 14, 2014, the contents of which are hereby incorporated by reference in the entirety for all purposes.

REFERENCE TO SUBMISSION OF A SEQUENCE LISTING

[0003] This application includes a Sequence Listing as a text file named "SEQ_81906-950428_ST25" created July 14, 2015 and containing 429,403 bytes. The material contained in this text file is incorporated by reference in its entirety for all purposes.

BACKGROUND OF THE INVENTION

[0004] Methods and compositions for imaging and detection of proteins in cells or cellular extract are useful in a wide array of research and diagnostic techniques. Similarly, methods and compositions for transcriptional regulation (*e.g.*, activation or inhibition) of genetic elements in a cell or cellular extract are useful in a wide array of research, diagnostic, and clinical techniques. Generally, however, such methods can fail to provide sufficient sensitivity and/or specificity.

BRIEF SUMMARY OF THE INVENTION

[0005] In some embodiments, the present invention provides a composition for recruiting one or more effector domains to a polypeptide of interest in a cell or cell extract, the composition comprising: the polypeptide of interest fused to a multimerized epitope; and an affinity agent fusion protein, wherein the affinity agent fusion protein comprises: an affinity domain that specifically binds the epitope; and an effector domain. In some cases, the polypeptide of interest comprises dCas9 (SEQ ID NO:9). In some cases, the multimerized epitope comprises SEQ ID NO: 10, 11, or 12.

[0006] In some cases, the effector domain is an enzyme (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional activator, or a transcriptional repressor. In some cases, the multimerized epitope contains multiple copies of an epitope of at least 5 amino acids in length. In some cases, the multimerized epitope contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the epitope. Each epitope of the multimerized epitope can be separated by a linker. In some cases, the linker is at least 5 amino acids in length. In some cases, the multimerized epitope comprises SEQ ID NO:1 or 2 and SEQ ID NO:2 or 3. In some cases, the multimerized epitope comprises: at least one copy of SEQ ID NO:3 or 4; and: at least two copies of SEQ ID NO:1; at least two copies of SEQ ID NO:2; or at least one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

[0007] In some cases, wherein the affinity domain is an antibody or a single-chain antibody that specifically binds the epitope. In some cases, the antibody or single-chain antibody is stable under the reducing conditions of a cell or cellular extract. In some cases, the affinity domain comprises a single chain antibody of SEQ ID NO:5. In some cases the effector domain comprises a fluorophore. For example, the effector domain can be a fluorescent protein. In some cases, the affinity domain is a single-chain antibody fused to a solubility enhancing domain. For example, the solubility enhancing domain can be a GB1 polypeptide (SEQ ID NO:6). In some cases, the solubility enhancing domain is a solubility enhanced effector domain. For example, the solubility enhanced effector domain can be superfolder-GFP (SEQ ID NO:7). In some cases, the affinity domain is fused to an N-terminal solubility enhancing domain and a C-terminal solubility enhancing domain. In some cases, the N-terminal solubility enhancing domain is a GB1 polypeptide (SEQ ID NO:6) and the C-

terminal solubility enhancing domain is superfolder-GFP (SEQ ID NO:7). In some cases, the N-terminal solubility enhancing domain is superfolder-GFP (SEQ ID NO:7) and the C-terminal solubility enhancing domain is a GB1 polypeptide (SEQ ID NO:6). In some cases, the affinity agent fusion protein comprises the amino acid sequence of SEQ ID NO:8.

5 [0008] In some embodiments, the present invention provides a cell or cell extract comprising any one of the foregoing compositions. In some embodiments, the present invention provides an isolated polynucleotide encoding SEQ ID NO:5 or SEQ ID NO:8.

[0009] In some embodiments, the present invention provides an isolated polynucleotide encoding a polypeptide of interest fused to a multimerized epitope, wherein the multimerized
10 epitope contains multiple copies of an epitope of at least 5 amino acids in length. In some cases, the multimerized epitope contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the epitope. In some cases, each epitope of the multimerized epitope is separated by a linker. In some cases, the multimerized epitope comprises SEQ ID NO:1 or 2 and SEQ ID NO:3 or 4. In some cases, the multimerized
15 epitope comprises: at least one copy of SEQ ID NO:3 or 4; and: at least two copies of SEQ ID NO:1; at least two copies of SEQ ID NO:2; or at least one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

[0010] In some embodiments, the present invention provides one or more expression cassettes, the expression cassettes containing one or more promoters (*e.g.*, heterologous
20 promoters) operably linked to one or more polynucleotides encoding: (i) any one of the foregoing polypeptides fused to a multimerized epitope; and/or (ii) any one of the foregoing affinity agent fusion proteins.

[0011] In some embodiments, the present invention provides a host cell transformed with one or more expression cassettes, the expression cassettes encoding: (i) any one of the
25 foregoing polypeptides fused to a multimerized epitope; and/or (ii) any one of the foregoing affinity agent fusion proteins. In some cases, one or more of the one or more of the expression cassettes of the host cell are inducible. In some cases, the host cell comprises a tet-transactivator, and the host cell further comprises a tet-inducible expression cassette.

[0012] In some embodiments, the present invention provides a kit comprising: (i) an
30 expression cassette comprising a heterologous promoter operably linked to a polynucleotide encoding an affinity agent fusion protein, wherein the affinity agent fusion protein comprises: an affinity domain that specifically binds the epitope; and a effector domain; and/or (ii) an

expression cassette encoding: (a) a heterologous promoter, a cloning site, and a multimerized epitope, wherein the cloning site is configured to allow cloning of a polypeptide of interest operably linked to the promoter and fused to the multimerized epitope; or (b) a heterologous promoter operably linked to a polypeptide of interest fused to a multimerized epitope.

5 [0013] In some cases, the effector domain is an enzyme (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional activator, or a transcriptional repressor. In some cases, the affinity domain comprises the single chain antibody of SEQ ID NO:5. In
10 some cases, the affinity agent fusion protein comprises the amino acid sequence of SEQ ID NO:8. In some cases, the multimerized epitope contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the epitope. In some cases, each epitope of the multimerized epitope is separated by a linker. In some cases, the linker is at least 5 amino acids in length. In some cases, the multimerized epitope comprises SEQ ID
15 NO:1 or 2 and SEQ ID NO:3 or 4. In some cases, the multimerized epitope comprises: at least one copy of SEQ ID NO:3 or 4; and: at least two copies of SEQ ID NO:1; at least two copies of SEQ ID NO:2; or at least one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

[0014] In some cases, the kit comprises an expression cassette encoding a small guide RNA
20 (sgRNA) or an sgRNA scaffold. In some cases, the expression cassette encoding an sgRNA scaffold comprises from 5' to 3': a 5' promoter; a cloning site; a 5' hairpin region; a 3' hairpin region; and a transcription termination region, wherein the cloning site is configured to operably link a binding region to the 5' promoter and the 3' regions, when the binding region is cloned into the cloning site.

25 [0015] In some embodiments, the present invention provides, a method for recruiting one or more effector domains to a polypeptide of interest in a cell or cell extract, the method comprising: contacting the cell or cell extract with any one of the foregoing compositions for recruiting one or more effector domains under conditions suitable to permit binding of multiple copies of the affinity agent fusion protein to the multimerized epitope fused to the
30 polypeptide of interest, thereby bringing multiple copies of the effector domain in proximity to the polypeptide of interest.

[0016] In some cases, the method comprises detecting the effector domain. In some cases, the detecting comprises directing incident light into the cell or cell extract, thereby inducing fluorescence from the effector domain and detecting the fluorescence. In some cases, the detecting comprises measuring upregulation or downregulation of transcription at or near a target binding site of the sgRNA. In some cases, the method comprises binding at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the affinity agent fusion protein to the multimerized epitope, thereby binding said number of copies of the effector domain to the polypeptide of interest. In some cases, the method comprises single molecule detection of the polypeptide of interest.

[0017] In some embodiments, the present invention provides a composition for site-specific transcriptional activation of a genetic element comprising: a dCas9 domain fused to a multimerized epitope; and an affinity agent fusion protein, wherein the affinity agent fusion protein comprises: an affinity domain that specifically binds the epitope; and a transcriptional activator domain.

[0018] In some cases, the multimerized epitope contains multiple copies of an epitope of at least 5 amino acids in length. In some cases, wherein the multimerized epitope contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the epitope. In some cases, each epitope of the multimerized epitope is separated by a linker of at least 5 amino acids in length. In some cases, the linker is at least 5 amino acids in length. In some cases, the multimerized epitope comprises SEQ ID NO:1 or 2 and SEQ ID NO:3 or 4. In some cases, the multimerized epitope comprises: at least one copy of SEQ ID NO:3 or 4; and: at least two copies of SEQ ID NO:1; at least two copies of SEQ ID NO:2; or at least one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

[0019] In some cases, the dCas9 fused to a multimerized epitope comprises the amino acid sequence of SEQ ID NO:9. In some cases, the dCas9 fused to a multimerized epitope comprises the amino acid sequence of SEQ ID NO:9 and the amino acid sequence of SEQ ID NO:10, 11, or 12. In some cases, the dCas9 fused to a multimerized epitope comprises the amino acid sequence of SEQ ID NO:13.

[0020] In some cases, the affinity domain is an antibody or a single-chain antibody that specifically binds the epitope. In some cases, the antibody or single-chain antibody is stable under the reducing conditions of a cell or a cellular extract. In some cases, the transcriptional activator domain comprises a VP16 domain. In some cases, the transcriptional activator

domain comprises at least 2, 3, 4, or more VP16 domains. In some cases, the affinity domain is a single-chain antibody fused to solubility enhancing domain. In some cases, the solubility enhancing domain is a GB1 polypeptide (SEQ ID NO:6). In some cases, the affinity agent fusion protein comprises SEQ ID NO:5. In some cases, the composition further comprises a small guide RNA (sgRNA).

[0021] In some embodiments, the present invention provides one or more expression cassettes, the expression cassettes containing one or more promoters (*e.g.*, heterologous promoters) operably linked to one or more polynucleotides encoding: (i) an sgRNA; (ii) a dCas9 fused to a multimerized epitope; and/or (iii) an affinity agent fusion protein of any one of the foregoing affinity agent fusion protein compositions.

[0022] In some embodiments, the present invention provides a host cell transformed with one or more expression cassettes, the expression cassettes encoding: (i) an sgRNA; (ii) a dCas9 fused to a multimerized epitope; and/or (iii) an affinity agent fusion protein of any one of the foregoing affinity agent fusion protein compositions. In some cases, one or more of the expression cassettes are inducible. In some cases, the host cell comprises a tet-transactivator, and the host cell further comprises a tet-inducible expression cassette encoding dCas9 fused to a multimerized epitope.

[0023] In some embodiments, the present invention provides a kit for activating transcription of a genetic element, the kit comprising one or more expression cassettes encoding: (i) a small guide RNA (sgRNA) or an sgRNA scaffold; (ii) a dCas9 fused to a multimerized epitope; and/or (iii) an affinity agent fusion protein of any one of the foregoing affinity agent fusion protein compositions. In some cases, the kit comprises an expression cassette encoding a small guide RNA (sgRNA) or an sgRNA scaffold. In some cases, the expression cassette encoding an sgRNA scaffold comprises from 5' to 3': a 5' promoter; a cloning site; a 5' hairpin region; a 3' hairpin region; and a transcription termination region, wherein the cloning site is configured to operably link a binding region to the 5' promoter and the 3' regions, when the binding region is cloned into the cloning site.

[0024] In some embodiments, the present invention provides a method of site-specific transcriptional activation of a genetic element in a cell or cell extract comprising: contacting the cell or cell extract with any one of the foregoing compositions containing dCas9 fused to a multimerized epitope, wherein the composition further comprises a small guide RNA (sgRNA) that specifically binds the genetic element, or a region proximal to the genetic

element, under conditions suitable to permit the binding of the sgRNA to the genetic element or region, the binding of the sgRNA to the dCas9 domain fused to the multimerized epitope, and the binding of multiple copies of the affinity agent fusion protein to the multimerized epitope, thereby bringing multiple copies of the transcriptional activator domain in proximity to the genetic element. In some cases, the method comprises binding at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of the affinity agent fusion protein to the multimerized epitope, thereby bringing said number of copies of the transcription activator domain in proximity to the genetic element.

[0025] In some embodiments, the present invention provides a composition comprising dCas9 fused to a multimerized effector domain. In some cases, the multimerized effector domain comprises two or more (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) copies of an effector domain. In some cases, the effector domain is an enzyme (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional activator, or a transcriptional repressor.

[0026] In some embodiments, the present invention provides a kit comprising one or more expression cassettes encoding: (i) a dCas9 fused to a multimerized effector domain of any one of foregoing compositions; and optionally (ii) a small guide RNA (sgRNA) or an sgRNA scaffold.

[0027] In some embodiments, the present invention provides a method for site-specific recruitment of effector domains to a genetic element in a cell or cell extract comprising: contacting the cell or cell extract with any one of the foregoing compositions containing dCas9 fused to a multimerized effector domain, wherein the composition further comprises a small guide RNA (sgRNA) that specifically binds the genetic element, or a region proximal to the genetic element, under conditions suitable to permit the binding of the sgRNA to the genetic element or region, and the binding of the sgRNA to the dCas9 domain fused to the multimerized effector domain, thereby bringing multiple copies of the effector domain in proximity to the genetic element.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] **Figure 1. Identification of an antibody-peptide pair that binds tightly *in vivo*.**

[0029] A) Schematic of the antibody-peptide labeling strategy. A protein of interest (protein X) is tagged with 4-24 copies of a short peptide (peptide epitopes), and is co-

expressed with the single chain antibody tagged with GFP that recognizes the short peptide and can be recruited in multiple copies. B). A schematic of an experiment in which the mitochondrial targeting domain of mitoNEET (mito) is fused to mCherry and 4 tandem copies of a peptide, which binds to mitochondria and labels them with a red fluorescent protein. The matching antibodies are tagged with GFP and expressed in the same cell. If binding occurs between antibody and peptide, then GFP labeling of the mitochondria should be observed. C) Indicated GFP-tagged antibodies are co-expressed with mitochondrial-targeted, mCherry-tagged 4xpep arrays in U2OS cells, and cells were imaged using spinning disk confocal microscopy. The GCN4 and V1 antibody-GFP fusions succeed in recognizing their corresponding peptide arrays on the mitochondria but the C4 antibody-GFP fusion does not. D) As a control, scFv-GCN4-GFP is co-expressed with a mito-mCherry plasmid in which the GCN4 peptides have been swapped for the FKBP protein, which does not bind the antibody. Scale bars, 10 μ m.

[0030] Figure 2 Mitoneet N-terminal domain targets proteins to the mitochondria

[0031] U2OS cells were transfected with a construct encoding the N-terminus of mitoNEET fused to GFP and incubated with mitotracker to stain mitochondria. Scale bars, 10 μ m.

[0032] Figure 3. Characterization of the off-rate and stoichiometry of the binding interaction between the scFv-GCN4 antibody and the GCN4 peptide array *in vivo*.

[0033] A) Mito-mCherry-24xGCN4pep was co-transfected in U2OS cells along with scFv-GCN4-GFP and their co-localization on mitochondria in a single cell is shown at time -10 sec. At 0 sec, the GFP signal from half of this cell was photobleached, and fluorescence recovery was followed by time-lapse microscopy. Scale bar, 5 μ m. B) The fluorescence recovery after photobleaching was quantified (shown is an average of FRAP recovery curves from 6 cells). A small amount of recovery is observed in the first 10 sec, which may be due to recovery of unbound GFP-tagged antibody which is freely diffusing in the cytoplasm in the vicinity of the mitochondria. C-E) Indicated constructs were transfected in U2OS cells and images were acquired 24 hr after transfection with equivalent image acquisition settings. Representative images are shown in C). Note that the GFP signal intensity in the mito-mCherry-24xGCN4pep + scFv-GCN4-GFP is highly saturated when the same scaling is used as in the other panels. Bottom row shows a zoom of a region of interest: dynamic scaling was different for the GFP and mCherry signals, so that both could be observed. Scale bars, 10 μ m.

D-E) Quantifications of the GFP:mCherry fluorescence intensity ratio on mitochondria after normalization (The average GFP:mCherry ratio for the sfGFP-linker-mCherry fusion protein was set to 1, see methods section). Each dot represents a single cell and dashed lines indicates the average value. All scale bars, 10 μ m.

5 **[0034] Figure 4. Optimizing the GCN4 antibody-peptide pair**

[0035] A) HEK293 cells were transfected with the indicated constructs and 24 hr after transfection, images were acquired using spinning disk confocal microscopy. Maximum intensity Z-projections are shown. All scale bars, 10 μ m. B) U2OS cells were transfected with a sfGFP-linker-mCherry fusion protein and images were acquired on a spinning disk
10 confocal microscope. GFP and mCherry fluorescence intensities for single cells were quantified and values were plotted after background subtraction.

[0036] Figure 5. sunGFP allows long-term single molecule fluorescence imaging in the cytoplasm.

[0037] A-H) U2OS cells were transfected with indicated SunTag constructs, all containing
15 24 copies of the GCN4 peptide, and were imaged by spinning disk confocal microscopy 24 hr after transfection. To decrease cytoplasmic background fluorescence of unbound scFv-GCN4-GFP, a nuclear localization signal was added to the scFv-GCN4-GFP to shuttle unbound antibody from the cytoplasm to the nucleus. A) A representative image of SunTag_{24x}-IFP-CAAX-GFP is shown (top), as well as the fluorescence intensities
20 quantification of the foci (bottom). Dotted line marks the outline of the cell. Scale bar, 10 μ m. B) Cells expressing K560-SunTag_{24x}-GFP were followed by spinning disk confocal microscopy (image acquisition every 200 ms). Movement is revealed by a maximum intensity projection of 50 time-points (left) and a kymograph (right). Scale bar, 10 μ m. C-D) Cells expressing both EB3-tdTomato and K560-SunTag_{24x}-GFP were imaged and moving
25 particles were tracked manually. Tracks indicate movement towards the cell interior and periphery (C). Scale bar, 5 μ m. Dots in (D) represent fraction of movement towards the interior from individual cells with between 5-20 moving particles scored per cell. The mean and standard deviation is indicated. (E-F) Cells expressing Kif18b-SunTag_{24x}-GFP were imaged with a 250 ms time interval. Images in (E) show a maximum intensity projection (50
30 time-points (left)) and a kymograph (right). Speeds of moving molecules were quantified from 10 different cells (F). (G-H) Cells expressing both mCherry- α -tubulin and K560rig-SunTag_{24x}D were imaged with a 600 ms time interval. The entire cell is shown in (G), while

H shows stills of a time series from the same cell. Open circles track two foci on the same microtubule, which is indicated by the dashed line. Asterisks indicate stationary foci. Scale bars, 10 and 2 μm (G and H), respectively.

[0038] Figure 6. Single molecule imaging using the SunTag.

5 **[0039]** A) Representative images of cells expressing either scFv-GCN4-GFP alone or together with IFP-SunTag_{24x} are shown. Bottom panels are enlargements of boxed areas. B-C) Run length (B) and speed (E) of K560-SunTag_{24x} were calculated in at least 10 different cells.

[0040] Figure 7. An optimized peptide array for high expression.

10 **[0041]** A) Indicated constructs were transfected in HEK293 cells and imaged 24 hr after transfection using wide-field microscopy. All images were acquired using identical acquisition parameters. B) Sequence of the first and second generation GCN4 peptide. C-D) Indicated constructs were transfected in HEK293 (C) or U2OS (D) cells and imaged 24 hr after transfection using wide-field (C) or spinning disk confocal (D) microscopy. E) U2OS cells were transfected with scFv-GCN4-GFP together with mito-mCherry-SunTag_{10x_v4}. 24 hr
15 after transfection, GFP signal on mitochondria was photobleached and fluorescence recovery was determined over time. The graph represents an average of 6 cells. The results are overlaid with the fluorescence recovery measurements shown in fig. 3B. Cells expressing K560-SunTag_{24x_v4}-GFP were followed by time-lapse microscopy (acquisition at 100 msec intervals); a maximum intensity projection of 25 time-points (left) or a kymograph (right) is
20 shown. Scale bars in A and C, 50 μm , scale bars in D, 10 μm .

[0042] Figure 8. dCas9-SunTag allows genetic rewiring of cells through activation of endogenous genes.

[0043] A) Schematic of gene activation by dCas9-VP64 and dCas9-SunTag-VP64. dCas9 binds to a gene promoter through its sequence specific sgRNA. Direct fusion of VP64 to
25 dCas9 (top) results in a single VP64 domain at the promoter which weakly activates transcription of the downstream gene. In contrast, recruitment of many VP64 domains using the SunTag potently activates transcription of the gene (bottom). (B-D) K562 cells stably expressing dCas9-VP64 or dCas9-SunTag_{10x}-VP64 were infected with lentiviral particles encoding indicated sgRNAs, as well as BFP and a puromycin resistance gene and selected
30 with 0.7 $\mu\text{g/ml}$ puromycin for 3 days. B) Cells were stained for CXCR4 using a directly labeled α -CXCR4 antibody and fluorescence analyzed by FACS. C) Levels of CXCR4,

analyzed as indicated in panel B, were determined with several sgRNAs. (D) Trans-well migration assays were performed with the same set of sgRNAs as in panel C (see methods). (E) dCas9-VP64 or dCas9-SunTag_{10x}-VP64 induced transcription of CDKN1B with several sgRNAs. mRNA levels were quantified by qPCR. (F) Growth competition assays were

5 performed by infecting around 30% of cells with indicated sgRNA/BFP, as well as a control sgRNA. Two days after infection the percentage of BFP positive cells was determined for each population. Cells were then grown for 2 weeks and the percentage of BFP positive cells was determined again. From the decrease in BFP/sgRNA positive cells over time, combined with the cell doubling time (which was determined in parallel to be on average 27 hr) the

10 percentage growth reduction was determined. Note that the control sgRNA did not affect the doubling time of cells. Graphs in B, D, and F are averages of three independent experiments. Graph in E is average of two biological replicates, each with two or three technical replicates. Error bars indicated standard error of the mean (SEM).

[0044] Figure 9. dCas9-SunTag can recruit many copies of scFv-GCN4-GFP to a

15 **genomic locus.**

[0045] A-B) HEK293 cells were transfected with dCas9-SunTag_{24x}, scFv-GCN4-GFP and indicated sgRNAs. 24 hr after transfection, cells were imaged by spinning disk confocal microscopy. Images are maximum intensity projections of Z-stacks (A). Intensities of individual telomere foci was measured in ImageJ and telomere fluorescence was calculated

20 by subtraction of diffuse nuclear background. Vertical set of dots in (B) represents individual telomere intensities in a single cell. Scale bars, 5 μ m.

DEFINITIONS

[0046] As used in this specification and the appended claims, the singular forms "**a**," "**an**,"

25 and "**the**" include plural reference unless the context clearly dictates otherwise.

[0047] The term "**nucleic acid**" or "**polynucleotide**" refers to deoxyribonucleic acids (DNA) or ribonucleic acids (RNA) and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and

30 are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologs, SNPs, and

complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzner *et al.*, *Nucleic Acid Res.* **19**:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* **260**:2605-2608 (1985); and Rossolini *et al.*, *Mol. Cell. Probes* **8**:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[0048] The term “**gene**” means the segment of DNA involved in producing a polypeptide chain. It may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0049] A “**promoter**” is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription.

[0050] An “**expression cassette**” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular polynucleotide sequence in a host cell. An expression cassette may be part of a plasmid, viral genome, or nucleic acid fragment. Typically, an expression cassette includes a polynucleotide to be transcribed, operably linked to a promoter. The promoter can be a heterologous promoter. In the context of promoters operably linked to a polynucleotide, a “heterologous promoter” refers to a promoter that would not be so operably linked to the same polynucleotide as a product of nature (*i.e.*, in a wild-type organism).

[0051] A “**reporter gene**” encodes proteins that are readily detectable due to their biochemical characteristics, such as enzymatic activity or chemifluorescent features. One specific example of such a reporter is green fluorescent protein. Fluorescence generated from this protein can be detected with various commercially-available fluorescent detection systems. Other reporters can be detected by staining. The reporter can also be an enzyme that generates a detectable signal when contacted with an appropriate substrate. The reporter can be an enzyme that catalyzes the formation of a detectable product. Suitable enzymes include, but are not limited to, proteases, nucleases, lipases, phosphatases and hydrolases. The

reporter can encode an enzyme whose substrates are substantially impermeable to eukaryotic plasma membranes, thus making it possible to tightly control signal formation. Specific examples of suitable reporter genes that encode enzymes include, but are not limited to, CAT (chloramphenicol acetyl transferase; Alton and Vapnek (1979) Nature 282: 864-869);

5 luciferase (lux); β -galactosidase; LacZ; β -glucuronidase; and alkaline phosphatase (Toh, et al. (1980) Eur. J. Biochem. 182: 231-238; and Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), each of which are incorporated by reference herein in its entirety. Other suitable reporters include those that encode for a particular epitope that can be detected with a labeled antibody that specifically recognizes the epitope.

10 [0052] The term “**amino acid**” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have
15 the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. “Amino acid mimetics” refers to chemical
20 compounds having a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0053] There are various known methods in the art that permit the incorporation of an unnatural amino acid derivative or analog into a polypeptide chain in a site-specific manner, *see, e.g.*, WO 02/086075.

25 [0054] Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0055] “**Polypeptide**,” “**peptide**,” and “**protein**” are used interchangeably herein to refer to a
30 polymer of amino acid residues. All three terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-

naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins, wherein the amino acid residues are linked by covalent peptide bonds.

[0056] “Conservatively modified variants” applies to both amino acid and nucleic acid

5 sequences. With respect to particular nucleic acid sequences, “conservatively modified variants” refers to those nucleic acids that encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA,
10 GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are “silent variations,” which are one species of conservatively modified variations. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of
15 the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

20 **[0057]** As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables
25 providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention. In some cases, conservatively modified variants of Cas9 or sgRNA can have an increased stability, assembly, or activity as described herein.

[0058] The following eight groups each contain amino acids that are conservative
30 substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);

- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 5 8) Cysteine (C), Methionine (M)

(see, e.g., Creighton, *Proteins*, W. H. Freeman and Co., N. Y. (1984)).

[0059] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly
10 accepted single-letter codes.

[0060] In the present application, amino acid residues are numbered according to their relative positions from the left most residue, which is numbered 1, in an unmodified wild-type polypeptide sequence.

[0061] As used in herein, the terms “**identical**” or percent “**identity**,” in the context of
15 describing two or more polynucleotide or amino acid sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same. For example, a core small guide RNA (sgRNA) sequence responsible for assembly and activity of a sgRNA:nuclease complex has at least 80% identity, preferably 85%, 90%, 91%, 92%, 93, 94%, 95%, 96%, 97%, 98%, 99%, or
20 100% identity, to a reference sequence, e.g., one of SEQ ID NOs:42-45), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. As another example, a Cas9 sequence responsible for assembly and activity of a sgRNA:nuclease complex has at least 80% identity, preferably
25 85%, 90%, 91%, 92%, 93, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity, to a reference sequence, e.g., one of SEQ ID NOs:46-50), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be “substantially identical.” With regard to polynucleotide
30 sequences, this definition also refers to the complement of a test sequence. With regard to amino acid sequences, preferably, the identity exists over a region that is at least about 50

amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

[0062] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters. For sequence comparison of nucleic acids and proteins, the BLAST and BLAST 2.0 algorithms and the default parameters discussed below are used.

[0063] A “**comparison window**”, as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

[0064] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, (1990) *J. Mol. Biol.* 215: 403-410 and Altschul *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available at the National Center for Biotechnology Information website, ncbi.nlm.nih.gov. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*).

These initial neighborhood word hits acts as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching
5 residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of
10 either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word size (W) of 28, an expectation (E) of 10, M=1, N=-2, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word size (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix
15 (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0065] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between
20 two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0066] An indication that two nucleic acid sequences or polypeptides are substantially
25 identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two
30 molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence. Yet another indication

that two polypeptides are substantially identical is that the two polypeptides retain identical or substantially similar activity.

[0067] A “**translocation sequence**” or “**transduction sequence**” refers to a peptide or protein (or active fragment or domain thereof) sequence that directs the movement of a protein from one cellular compartment to another, or from the extracellular space through the cell or plasma membrane into the cell. Translocation sequences that direct the movement of a protein from the extracellular space through the cell or plasma membrane into the cell are “**cell penetration peptides**.” Translocation sequences that localize to the nucleus of a cell are termed “**nuclear localization**” sequences, signals, domains, peptides, or the like.

Examples of translocation sequences include, without limitation, the TAT transduction domain (see, e.g., S. Schwarze et al., Science 285 (Sep. 3, 1999); penetratins or penetratin peptides (D. Derossi et al., Trends in Cell Biol. 8, 84-87); Herpes simplex virus type 1 VP22 (A. Phelan et al., Nature Biotech. 16, 440-443 (1998), and polycationic (e.g., poly-arginine) peptides (Cell Mol. Life Sci. 62 (2005) 1839-1849). Further translocation sequences are known in the art. Translocation peptides can be fused (e.g. at the amino or carboxy terminus), conjugated, or coupled to a compound of the present invention, to, among other things, produce a conjugate compound that may easily pass into target cells, or through the blood brain barrier and into target cells.

[0068] The “**CRISPR/Cas**” system refers to a widespread class of bacterial systems for defense against foreign nucleic acid. CRISPR/Cas systems are found in a wide range of eubacterial and archaeal organisms. CRISPR/Cas systems include type I, II, and III subtypes. Wild-type type II CRISPR/Cas systems utilize the RNA-mediated nuclease, Cas9 in complex with guide and activating RNA to recognize and cleave foreign nucleic acid.

[0069] Cas9 homologs are found in a wide variety of eubacteria, including, but not limited to bacteria of the following taxonomic groups: *Actinobacteria*, *Aquificae*, *Bacteroidetes-Chlorobi*, *Chlamydiae-Verrucomicrobia*, *Chloroflexi*, *Cyanobacteria*, *Firmicutes*, *Proteobacteria*, *Spirochaetes*, and *Thermotogae*. An exemplary Cas9 protein is the *Streptococcus pyogenes* Cas9 protein. Additional Cas9 proteins and homologs thereof are described in, e.g., Chylinksi, et al., RNA Biol. 2013 May 1; 10(5): 726–737 ; Nat. Rev. Microbiol. 2011 June; 9(6): 467-477; Hou, et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Sampson et al., Nature. 2013 May 9;497(7448):254-7; and Jinek, et al., Science. 2012 Aug 17;337(6096):816-21.

[0070] As used herein, “**activity**” in the context of CRISPR/Cas activity, Cas9 activity, sgRNA activity, sgRNA:nuclease activity and the like refers to the ability to bind to a target genetic element and recruit effector domains to a region at or near the target genetic element. Such activity can be measured in a variety of ways as known in the art. For example,

5 expression, activity, or level of a reporter gene, or expression or activity of a gene encoded by the genetic element can be measured. As another example, a signal (*e.g.*, a fluorescent signal) provided by a recruited effector domain (*e.g.*, a recruited fluorescent protein) can be detected.

[0071] As used herein, the term “**effector domain**” refers to a polypeptide that provides an effector function. Exemplary effector functions include, but are not limited to, enzymatic activity (*e.g.*, nuclease, methylase, demethylase, acetylase, deacetylase, kinase, phosphatase, ubiquitinase, deubiquitinase, luciferase, or peroxidase activity), fluorescence, binding and recruitment of additional polypeptides or organic molecules, or transcriptional modulation (*e.g.*, activation, enhancement, or repression). Thus, exemplary effector domains include, but

10 are not limited to enzymes (*e.g.*, nucleases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, ubiquitinases, deubiquitinases, luciferases, or peroxidases), adaptor proteins, fluorescent proteins (*e.g.*, green fluorescent protein), transcriptional enhancers, transcriptional activators, or transcriptional repressors. Adaptor protein effector domains can function to bind, and thus recruit other polypeptides, organic

15 molecules, *etc.*

20

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0072] Recruitment of multiple copies of a protein to a target substrate (*e.g.* DNA, RNA, or protein) is used to amplify signals in biological systems. For example, recruitment of multiple

25 copies of a transcription factor to a single gene promoter can dramatically enhance transcriptional activation of the target gene (Anderson and Freytag, 1991; Chen *et al.*, 1992; Pettersson and Schaffner, 1990). Similarly, the recruitment of multiple copies of an RNA binding protein to an mRNA can result in potent regulation of translation (Pillai *et al.*, 2004;

30 Pique *et al.*, 2008). Protein localization and interactions also can be modulated by the copy number of interaction sites within a polypeptide sequence. For example, many nuclear proteins contain multiple nuclear localization signal (NLS) sequences, which control

robustness of nuclear import (Luo *et al.*, 2004). Similarly, in receptor-mediated signaling, multimerization of receptors in response to ligand binding helps to elicit a downstream response (Boniface *et al.*, 1998). Downstream of the receptors, adapter proteins with multiple SH2/SH3 domains can generate multivalent interactions of interacting signaling molecules (Li *et al.*, 2012), which is thought to facilitate the signaling response

[0073] Protein multimerization also has been widely used in synthetic biology. A commonly used method to study RNA localization, even at the single molecule level, is to insert many copies of the MS2 binding aptamer (as many as 24), which then recruit many MS2-GFP fusion proteins (Bertrand *et al.*, 1998; Fusco *et al.*, 2003). Similarly, the activity of a RNA-binding protein can be studied by artificially tethering it to an RNA in multiple copies using the MS2 system (Coller and Wickens, 2007). Similar multimerization approaches have also been used to fluorescently label a specific region of a chromosome. For example, the LacO operon can be inserted into a chromosomal locus in many tandem repeats and then visualized by the recruitment of many copies of GFP-LacI (Gordon *et al.*, 1997). More recently, several studies have shown that GFP-tagged engineered DNA-binding proteins, like TALEs or the CRISPR effector protein Cas9, can also be used to fluorescently label an endogenous DNA sequence when its binding site is present in many tandem repeats in the DNA (Chen *et al.*, 2013; Ma *et al.*, 2013; Miyanari *et al.*, 2013). Furthermore, as with native transcriptional regulation, a gene can be artificially activated when a binding site for a synthetic transcription factor is placed upstream of a gene in multiple copies; this principle is employed in the “tet-on” system for inducible transgene expression (Huang *et al.*, 1999; Sadowski *et al.*, 1988). Taken together, these studies demonstrate the power of introducing multiple copies of protein binding sites within RNA or DNA for the purpose of signal amplification.

[0074] Despite the success of multimerizing nucleic acid based motifs within RNA and DNA for protein recruitment, no comparable and generic system exists for controlling copy number of protein-protein interactions. For fluorescence imaging, the fusion of 3 copies of GFP to a protein of interest has been used to increase signal intensity, but a further increase in the copy number of fluorescent proteins is challenging due to their size (~25 kDa) and bacterial recombination when constructing DNA plasmids encoding such proteins. Here, we describe a new synthetic system for recruiting as many as 24 copies of a protein to a target polypeptide chain. We demonstrate that this approach can be used to create bright fluorescent signals for single molecule protein imaging in living cells, through the

recruitment of 24 copies of GFP to a target protein. We also demonstrate that the system can be used to modulate gene expression through the recruitment of multiple copies of gene regulatory effector domains to a modified CRISPR/Cas9 protein targeted to specific sequences in the genome. The ability to multimerize proteins in a controlled fashion on a polypeptide backbone will likely have many additional uses in biotechnology.

II. Compositions

[0075] Described herein are compositions useful as components of a system for recruiting one or more effector domains to a polypeptide of interest. The components can be used to target the effector domains to the polypeptide of interest, or a binding partner of the polypeptide of interest. Thus, for example, the components can be used to target the effector domains to a region of interest such as a genomic region, an intracellular compartment (*e.g.*, nucleus, cytoplasm, endoplasmic reticulum, *etc.*), or a membrane (*e.g.*, cytoplasmic, nuclear, or mitochondrial, *etc.*). The polypeptide of interest can be any natural, recombinant, or synthetic polypeptide. The components include epitopes, multimerized epitopes, affinity agents, Cas9 domains (including dCas9 domains), sgRNAs, and effector domains.

A. Epitopes and Multimerized Epitopes

[0076] Described herein are epitopes and multimerized epitopes for recruiting affinity agents to a polypeptide of interest. Typically, the epitopes are fused to the polypeptide of interest. The epitopes can be fused to one or more of the N-terminus of the polypeptide of interest, the C-terminus of the polypeptide of interest, or inserted into the polypeptide of interest. For example, the epitopes can be inserted into a region of the polypeptide of interest that is solvent accessible when the polypeptide is in a folded conformation. Such regions include, but are not limited to protein surface loops or linker regions between discrete protein domains. A polypeptide of interest can be fused to an epitope, multiple copies of an epitope, more than one different epitope, or multiple copies of more than one different epitope as further described herein.

[0077] The epitopes can be any polypeptide sequence that is specifically recognized by an affinity agent. Such epitopes include, but are not limited to the c-Myc affinity tag, an HA affinity tag, a His affinity tag, an S affinity tag, a methionine-His affinity tag, an RGD-His affinity tag, a 7x His tag, a FLAG octapeptide, a strep tag or strep tag II, a V5 tag, or a VSV-G epitope. An exemplary epitope includes, but is not limited to, a GCN4 epitope (*e.g.*, SEQ ID NOs:1 or 2).

[0078] Epitopes, such as the epitopes described herein can be multimerized. For example, the a polypeptide of interest can be fused to a multimerized epitope containing 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of an epitope. In some cases, the polypeptide of interest is fused to a first epitope or multimerized epitope. In
 5 some cases, the polypeptide of interest is fused to a first epitope or multimerized epitope and a second epitope or multimerized epitope. Multimerized epitopes include, but are not limited to multimerized epitopes containing 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies of a GCN4 epitope. An exemplary epitopes include, but are not limited to, a 24xGNC4 epitope (*e.g.*, SEQ ID NOs:10 or 11) or a 10xGCN4 epitope
 10 (*e.g.*, SEQ ID NO:12)

[0079] The individual epitopes of a multimerized epitope can be separated by a linker region. Suitable linker regions are known in the art. In some cases, the linker is configured to allow the binding of affinity agents to adjacent epitopes without, or without substantial, steric hindrance. In some cases, the linker sequences are configured to provide an
 15 unstructured or linear region of the polypeptide. For example, the linker sequence can comprise one or more glycines and/or serines. The linker sequences can be at least about 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acids in length. In some cases, the linker sequences are, or comprise, one or more of the linkers disclosed on the world wide web at parts.igem.org/Protein_domains/Linker. Exemplary linkers include, but are not limited to,
 20 SEQ ID NOs:3 or 4.

[0080] Also described herein are expression cassettes and vectors for producing one or more epitopes or multimerized epitopes described herein (*e.g.*, a polypeptide of interest fused to an epitope or multimerized epitope) in a host cell. The expression cassettes can contain a promoter (*e.g.*, a heterologous promoter) operably linked to a polynucleotide encoding an
 25 epitope or multimerized epitope. The promoter can be inducible or constitutive. The promoter can be tissue specific. In some cases, the promoter is a strong promoter. For example, the promoter can be a CMV promoter, an SFFV long terminal repeat promoter, or the human elongation factor 1 promoter (EF1A). In some cases, the polynucleotide encoding the epitope or multimerized epitope of the expression cassette further encodes the polypeptide
 30 of interest. In some cases, an expression cassette is provided for cloning a polynucleotide encoding a polypeptide of interest in frame with an epitope or multimerized epitope. The expression cassette can include one or more localization sequences. In some cases, the polypeptide of interest provides a localization function. The expression cassette can be in a

vector, such as a plasmid, a viral vector, a lentiviral vector, *etc.* In some cases, the expression cassette is in a host cell. The expression cassette can be episomal or integrated in the host cell.

B. Affinity Agents

5 [0081] Described herein are affinity agents for recruiting effector functions to a polypeptide fused to an epitope or multimerized epitope. A wide variety of affinity agents can be utilized. Generally, the affinity agent is stable under the reducing conditions present in the intracellular environment of the cell. Additionally, the affinity agent should specifically bind to its corresponding epitope with minimal cross-reactivity. In some cases, 10 the affinity agent is an antibody, such as an scFv. In some cases, the affinity agent is an antibody (*e.g.*, scFv) that has been optimized for stability in the intracellular environment. For example, the affinity agent (*e.g.*, scFv) can be an intrabody (see, *e.g.*, Lo *et al.*, Handb. Exp. Pharm. 2008;(181):343-73). An exemplary affinity agent comprises the anti-GCN4 scFv domain of SEQ ID NO:5. In some cases, the affinity agent comprises an affinity 15 domain (*e.g.*, an anti-GCN4 scFv domain such as SEQ ID NO:5) and a linker (*e.g.*, a linker such as SEQ ID NO:58), wherein the linker links the affinity domain to an effector domain.

[0082] The affinity agent can contain one or more solubility enhancing domains. For example, the affinity agent can be fused at the N- and/or C-terminus to a highly soluble, and/or a highly stable, polypeptide. Exemplary solubility enhancing domains include, 20 without limitation, superfolder GFP (Pédélecq *et al.*, Nat Biotechnol. 2006 Jan; 24(1):79-88), maltose binding protein, albumin, hen egg white lysozyme, glutathione S-transferase, the protein G B1 domain (SEQ ID NO:6), protein D, the Z domain of protein A, thioredoxin, bacterioferritin, DhaA, HaloTag, and GrpE.

[0083] The affinity agent can be fused (*e.g.*, at the N- or C-terminus) to one or more 25 effector domains. Such effector domains include, but are not limited to enzymes (*e.g.*, nucleases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, ubiquitinases, deubiquitinases, luciferases, or peroxidases), fluorescent proteins (*e.g.*, green fluorescent protein), transcriptional enhancers, transcriptional activators, or transcriptional repressors. An exemplary effector domain is fluorescent protein such as green fluorescent 30 protein (GFP). In some cases, the effector domain is optimized for expression (*e.g.*, codon optimized) or stability. For example, the fluorescent effector domain can be superfolder green fluorescent protein (superfolder GFP (sfGFP), SEQ ID NO:7).

[0084] In some embodiments, the affinity agent effector domain comprises a transcriptional modulator domain. For example, the affinity agent can contain an affinity domain (*e.g.*, an scFv domain) and a transcriptional modulator (*e.g.*, transcriptional activator or repressor) domain. In some cases, the affinity agent contains an affinity domain fused to one or more
5 copies of a Herpes Simplex Virus Viral Protein 16 (VP16) domain, or a portion thereof. In some cases, the affinity agent contains an anti-GCN4 affinity domain fused to one or more (*e.g.*, at least 2, 3, 4, or more) copies of a VP16 domain. A polypeptide containing 4 copies of the Herpes Simplex Virus Viral Protein 16 (VP16) domain is known as a VP64 domain. An exemplary affinity agent fused to a VP64 domain is an anti-GCN4 antibody fused to
10 sfGFP and VP64 (*e.g.*, SEQ ID NO:16).

[0085] Also described herein are expression cassettes and vectors for producing one or more affinity agents described herein in a host cell. The expression cassettes can contain a promoter (*e.g.*, a heterologous promoter) operably linked to a polynucleotide encoding an affinity agent. The promoter can be inducible or constitutive. The promoter can be tissue
15 specific. In some cases, the promoter is a strong promoter. For example, the promoter can be a CMV promoter, an SFFV long terminal repeat promoter, or the human elongation factor 1 promoter (EF1A). In some cases, the polynucleotide encoding an affinity agent of the expression cassette further encodes one or two localization sequences (*e.g.*, nuclear
20 localization sequences) to ensure that the affinity agent localizes at or near the polypeptide of interest fused to the epitope or multimerized epitope. For example, the polynucleotide can encode an affinity agent having one or more localization sequences at the N- and/or C-terminus. The expression cassette can be in a vector, such as a plasmid, a viral vector, a lentiviral vector, *etc.* In some cases, the expression cassette is in a host cell. The expression cassette can be episomal or integrated in the host cell.

25 C. Cas9

[0086] Described herein are guide RNA dependent nucleases and derivatives thereof. In some embodiments, the guide RNA dependent nucleases can serve as a polypeptide of interest fused to an epitope or multimerized epitope. In some embodiments, the guide RNA dependent nucleases can serve as a polypeptide of interest fused to a multimerized effector
30 domain. In some cases, the sgRNA-mediated nuclease is a Cas9 protein. For example, the sgRNA-mediated nuclease can be a type I, II, or III Cas9 protein. In some cases, the sgRNA-mediated nuclease can be a modified Cas9 protein. Cas9 proteins can be modified by any method known in the art. For example, the Cas9 protein can be codon optimized for

expression in host cell or an *in vitro* expression system. Additionally, or alternatively, the Cas9 protein can be engineered for stability, enhanced target binding, or reduced aggregation.

[0087] The Cas9 can be a nuclease defective Cas9 (*i.e.*, dCas9). For example, certain Cas9 mutations can provide a nuclease that does not cleave or nick, or does not substantially cleave or nick the target sequence. Exemplary mutations that reduce or eliminate nuclease activity include one or more mutations in the following locations: D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, or A987, or a mutation in a corresponding location in a Cas9 homologue or ortholog. The mutation(s) can include substitution with any natural (*e.g.*, alanine) or non-natural amino acid, or deletion. An exemplary nuclease defective dCas9 protein is Cas9D10A&H840A (Jinek, *et al.*, Science. 2012 Aug 17;337(6096):816-21; Qi, *et al.*, Cell. 2013 Feb 28;152(5):1173-83).

[0088] dCas9 proteins that do not cleave or nick the target sequence can be utilized in combination with an sgRNA, such as one or more of the sgRNAs described herein, to form a complex that is useful for targeting, detection, or transcriptional modulation of target nucleic acids as further explained below. The dCas9 can be targeted to one or more genetic elements by virtue of the binding regions encoded on one or more sgRNAs. Recruitment of dCas9 can therefore provide recruitment of additional effector domains as provided by polypeptides fused to the dCas9 domain. For example, a polypeptide comprising an effector domain can be fused to the N and/or C-terminus of a dCas9 domain. In some cases, the polypeptide encodes a transcriptional activator or repressor. In other cases, the polypeptide encodes an epitope or multimerized epitope fusion that can be used to recruit one or more copies of an affinity agent. In some cases, the affinity agent is fused to one or more copies of an effector domain, such as an enzyme (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional activator, or a transcriptional repressor.

[0089] In one embodiment, the dCas9 is a transcriptional activator and comprises a dCas9 domain and a multimerized transcriptional activator domain. In some cases, the dCas9 domain is fused to two or more copies of a p65 activation domain (p65AD). In some cases, the dCas9 domain transcriptional activator comprises a dCas9 domain fused to two or more copies of a VP16 or VP64 activation domain. In some cases, the dCas9 domain is fused to at

least one copy of a first activation domain (*e.g.*, p65AD) and at least one copy of a second activation domain (*e.g.*, VP16 or VP64).

[0090] In some embodiments, the dCas9 is a transcriptional repressor and comprises a dCas9 domain and a multimerized transcriptional repressor domain. In some cases, the dCas9 domain is fused to two or more copies of a Krüppel associated box (KRAB) repressor domain. In some cases, the dCas9 domain is fused to two or more copies of a chromoshadow domain (CSD) repressor. In some cases, the dCas9 is fused to at least one copy of a first repressor domain (*e.g.*, a KRAB domain) and at least one copy of a second repressor domain (*e.g.*, a CSD domain).

[0091] In some embodiments, the dCas9 transcriptional modulator is a dCas9 domain fused to an epitope fusion polypeptide. The epitope fusion polypeptide can contain one or more copies (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies) of an epitope. In some cases, the epitope fusion polypeptide contains multiple copies of an epitope separated by one or more linker sequences.

[0092] The amino acid sequence of the epitope can be any sequence that is specifically recognized by a corresponding affinity agent. Thus, the dCas9 domain fused to the epitope fusion polypeptide will recruit one or more copies of the corresponding fusion agent. This can result in an amplification of any signal or effector function provided by the affinity agent. For example, the affinity agent can be a fusion protein comprising an affinity domain and a transcriptional modulation domain. The dCas9 epitope fusion can form a complex with an sgRNA specific for a target genetic element and recruit multiple copies of the transcriptional modulation domain via the affinity domain to the targeted genetic element. As another example, the affinity agent can be a fusion protein comprising an affinity domain and a fluorescent protein. The dCas9 epitope fusion can form a complex with an sgRNA specific for a target genetic element and recruit multiple copies of the fluorescent protein via the affinity domain to the targeted genetic element.

[0093] In some cases, the dCas9 domain fused to an epitope fusion polypeptide contains one or more copies of a GCN4 epitope. In some cases, the epitope fusion polypeptide contains multiple copies of a GCN4 epitope separated by one or more copies of one or more linker sequences. In some cases, the linker is configured to allow the binding of affinity agents to adjacent GCN4 epitopes without, or without substantial, steric hindrance. An exemplary dCas9 fused to a GCN4 epitope fusion domain is or comprises SEQ ID NO:13. In

some cases, the dCas9 fused to a GCN4 epitope fusion domain is at least about 90%, 95%, or 99% identical, or identical, to SEQ ID NO:13.

[0094] In some embodiments, the epitope fusion polypeptide contains one or more copies of two or more different epitopes. In such cases, the dCas9 can recruit multiple different effector functions. For example, the epitope fusion polypeptide can contain a first epitope that recruits an affinity agent fused to a transcriptional activator. The epitope fusion polypeptide can further contain a second epitope that recruits an affinity agent fused to different effector function (*e.g.*, a different transcriptional activator, a chromatin modifier, or a regulator of DNA methylation). For example, the epitope fusion polypeptide can recruit a p65 activation domain (p65AD) and a VP64 activation domain, or a VP64 activation domain and a regulator of histone or DNA methylation. In some cases, the epitope fusion polypeptide containing one or more copies of two or more different epitopes can be used to enhance the specificity of a CRISPR/Cas interaction. For example, one epitope can recruit an affinity agent fused to one half of an obligate dimer effector domain, while the other epitope recruits an affinity agent fused to the other half of the obligate dimer effector domain. In some cases, the obligate dimer can be a transcription factor, a transcriptional activator, a transcriptional repressor, a fluorescent protein (*e.g.*, GFP), a recombinase (*e.g.*, CRE recombinase), a luciferase, thymidine kinase, TEV protease, or dihydrofolate reductase.

[0095] Also described herein are expression cassettes and vectors for producing a small guide RNA-mediated nuclease (*e.g.*, Cas9 or dCas9), including Cas9 or dCas9 fusion proteins, in a host cell. The expression cassettes can contain a promoter (*e.g.*, a heterologous promoter) operably linked to a polynucleotide encoding Cas9 or dCas9. The promoter can be inducible or constitutive. The promoter can be tissue specific. In some cases, the promoter is a weak mammalian promoter as compared to the human elongation factor 1 promoter (EF1A). In some cases, the weak mammalian promoter is a ubiquitin C promoter, a *vav* promoter, or a phosphoglycerate kinase 1 promoter (PGK). In some cases, the weak mammalian promoter is a TetOn promoter in the absence of an inducer. In some cases, when a TetOn promoter is utilized, the host cell is also contacted with a tetracycline transactivator.

[0096] In some embodiments, the strength of the selected small guide RNA-mediated nuclease promoter is selected to express an amount of small guide RNA-mediated nuclease (*e.g.*, Cas9 or dCas9) that is proportional to the amount of sgRNA or amount of sgRNA expression. In some embodiments, the strength of the selected promoter is selected to

express an amount of small guide RNA-mediated nuclease epitope fusion protein that expresses an amount of epitopes that is proportional to the amount of corresponding affinity agent. For example, if a dCas9 epitope fusion protein contains ten copies of an epitope, then the dCas9 promoter can be selected to express $1/10^{\text{th}}$ the amount of dCas9 as compared to
 5 corresponding affinity agent (or less). In some cases, the a weak promoter can be selected to reduce cytotoxicity induced by expression of the Cas9 or dCas9 gene.

[0097] In some cases, the polynucleotide encoding a small guide RNA-mediated nuclease of the expression cassette further encodes one or two localization sequences. For example, the polynucleotide can encode a Cas9 or dCas9 protein having a nuclear localization
 10 sequence at the N- and/or C-terminus. The expression cassette can be in a vector, such as a plasmid, a viral vector, a lentiviral vector, *etc.* In some cases, the expression cassette is in a host cell. The expression cassette can be episomal or integrated in the host cell.

D. sgRNAs

[0098] Described herein are small guide RNAs (sgRNAs). The sgRNAs can contain from
 15 5' to 3': a binding region, a 5' hairpin region, a 3' hairpin region, and a transcription termination sequence. The sgRNA can be configured to form a stable and active complex with a small guide RNA-mediated nuclease (*e.g.*, Cas9 or dCas9). In some cases, the sgRNA is optimized to enhance expression of a polynucleotide encoding the sgRNA in a host cell.

[0099] The 5' hairpin region can be between about 15 and about 50 nucleotides in length
 20 (*e.g.*, about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or about 50 nucleotides in length). In some cases, the 5' hairpin region is between about 30-45 nucleotides in length (*e.g.*, about 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45 nucleotides in length). In some cases, the 5' hairpin region is, or is at least about, 31 nucleotides in length (*e.g.*, is at least about 31, 32,
 25 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45 nucleotides in length). In some cases, the 5' hairpin region contains one or more loops or bulges, each loop or bulge of about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In some cases, the 5' hairpin region contains a stem of between about 10 and 30 complementary base pairs (*e.g.*, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 complementary base pairs).

[0100] In some embodiments, the 5' hairpin region can contain protein-binding, or small molecule-binding structures. In some cases, the 5' hairpin function (*e.g.*, interacting or assembling with a sgRNA-mediated nuclease) can be conditionally activated by drugs,

growth factors, small molecule ligands, or a protein that binds to the protein-binding structure of the 5' stem-loop. In some embodiments, the 5' hairpin region can contain non-natural nucleotides. For example, non-natural nucleotides can be incorporated to enhance protein-RNA interaction, or to increase the thermal stability or resistance to degradation of the sgRNA.

[0101] The sgRNA can contain an intervening sequence between the 5' and 3' hairpin regions. The intervening sequence between the 5' and 3' hairpin regions can be between about 0 to about 50 nucleotides in length, preferably between about 10 and about 50 nucleotides in length (*e.g.*, at a length of, or about a length of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides). In some cases, the intervening sequence is designed to be linear, unstructured, substantially linear, or substantially unstructured. In some embodiments, the intervening sequence can contain non-natural nucleotides. For example, non-natural nucleotides can be incorporated to enhance protein-RNA interaction or to increase the activity of the sgRNA:nuclease complex. As another example, natural nucleotides can be incorporated to enhance the thermal stability or resistance to degradation of the sgRNA.

[0102] The 3' hairpin region can contain an about 3, 4, 5, 6, 7, or 8 nucleotide loop and an about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotide or longer stem. In some cases, the 3' hairpin region can contain a protein-binding, small molecule-binding, hormone-binding, or metabolite-binding structure that can conditionally stabilize the secondary and/or tertiary structure of the sgRNA. In some embodiments, the 3' hairpin region can contain non-natural nucleotides. For example, non-natural nucleotides can be incorporated to enhance protein-RNA interaction or to increase the activity of the sgRNA:nuclease complex. As another example, natural nucleotides can be incorporated to enhance the thermal stability or resistance to degradation of the sgRNA.

[0103] In some embodiments, the sgRNA includes a termination structure at its 3' end. In some cases, the sgRNA includes an additional 3' hairpin region, *e.g.*, before the termination and after a first 3' hairpin region, that can interact with proteins, small-molecules, hormones, *etc.*, for stabilization or additional functionality, such as conditional stabilization or conditional regulation of sgRNA:nuclease assembly or activity.

[0104] In some embodiments, the sgRNA forms an sgRNA:Cas9 or dCas9 complex that has increased stability and/or activity as compared to previously known sgRNAs or an sgRNA substantially identical to a previously known sgRNA. In some cases, the sgRNA forms an sgRNA:Cas9 or dCas9 complex that has increased stability and/or activity as compared to as an sgRNA encoded by:

SEQ ID NO:42 [N]₅-

₁₀₀GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC

GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUU, where [N]

represents a target specific binding region of between about 5-100 nucleotides (*e.g.*, about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, or 90 nucleotides) that is complementary or substantially complementary to the target genetic element. In some embodiments, the binding region of the sgRNA is, or is about, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 or more nucleotides in length. In some cases, the binding region of the sgRNA is between about 19 and about 21 nucleotides in length.

[0105] Generally, the binding region is designed to complement or substantially complement the target genetic element or elements. In some cases, the binding region can incorporate wobble or degenerate bases to bind multiple genetic elements. In some cases, the binding region can be altered to increase stability. For example, non-natural nucleotides, can be incorporated to increase RNA resistance to degradation. In some cases, the binding region can be altered or designed to avoid or reduce secondary structure formation in the binding region. In some cases, the binding region can be designed to optimize G-C content. In some cases, G-C content is preferably between about 40% and about 60% (*e.g.*, 40%, 45%, 50%, 55%, 60%). In some cases, the binding region, can be selected to begin with a sequence that facilitates efficient transcription of the sgRNA. For example, the binding region can begin at the 5' end with a G nucleotide. In some cases, the binding region can contain modified nucleotides such as, without limitation, methylated or phosphorylated nucleotides.

[0106] In some cases, the sgRNAs described herein form an sgRNA:nuclease complex with enhanced stability or activity as compared to SEQ ID NO:42, or an sgRNA 90, 95, 96, 97, 98, or 99% or more identical to SEQ ID NO:42. In some cases, the optimized sgRNAs described herein form an sgRNA:nuclease complex with enhanced stability or activity as compared to SEQ ID NO:42, or an sgRNA with fewer than 5, 4, 3, or 2 nucleotide substitutions, additions,

or deletions of SEQ ID NO:42. As used herein, identity of an sgRNA to another sgRNA, such as an sgRNA to SEQ ID NO:42 is determined with reference to the identity to the nucleotide sequences outside of the binding region. For example, two sgRNAs with 0% identity inside the binding region and 100% identity outside the binding region are 100% identical to each other. Similarly, as used herein, the number of substitutions, additions, or deletions of an sgRNA as compared to another, such as an sgRNA compared to SEQ ID NO:42 is determined with reference to the nucleotide sequences outside of the binding region. For example, two sgRNAs with multiple additions, substitutions, and/or deletions inside the binding region and 100% identity outside the binding region are considered to contain 0 nucleotide substitutions, additions, or deletions.

[0107] In some embodiments, the sgRNA can be optimized for expression by substituting, deleting, or adding one or more nucleotides. In some cases, a nucleotide sequence that provides inefficient transcription from an encoding template nucleic acid can be deleted or substituted. For example, in some cases, the sgRNA is transcribed from a nucleic acid operably linked to an RNA polymerase III promoter. In such cases, sgRNA sequences that result in inefficient transcription by RNA polymerase III, such as those described in Nielsen *et al.*, Science. 2013 Jun 28;340(6140):1577-80, can be deleted or substituted. For example, one or more consecutive uracils can be deleted or substituted from the sgRNA sequence. In some cases, the consecutive uracils are present in the stem portion of a stem-loop structure. In such cases, one or more of the consecutive uracils can be substituted by exchanging the uracil and its complementary base. For example, if the uracil is hydrogen bonded to a corresponding adenine, the sgRNA sequence can be altered to exchange the adenine and uracil. This “A-U flip” can retain the overall structure and function of the sgRNA molecule while improving expression by reducing the number of consecutive uracil nucleotides. In some cases, the sgRNA containing an A-U flip is encoded by:

SEQ ID NO:43 [N]₅.

₁₀₀GUUUAAGAGCUAGAAAUAGCAAGUUUAAAUAAGGCUAGUCC

GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUU, where the A-U flipped nucleotides are underlined. In some cases, the optimized sgRNA is at least 90, 91,

92, 93, 94, 95, 96, 97, 98, or 99% identical or more to SEQ ID NO:43, or contains fewer than 10, 9, 8, 7, 6, 5, 4, 3, or 2 nucleotide additions, deletions, or substitutions compared to SEQ ID NO:43. Alternatively, the A-U pair can be replaced by a G-C, C-G, A-C, G-U pair. In some cases, the sgRNA is designed so that, with the exclusion of the transcription terminator

sequence, it does not contain any run of four or more consecutive nucleotides of the same type (*e.g.*, four or more consecutive U nucleotides; four or more consecutive A nucleotides; four or more consecutive G nucleotides; four or more consecutive C nucleotides; or a combination thereof).

- 5 [0108] In some embodiments, the sgRNA can be optimized for stability. Stability can be enhanced by optimizing the stability of the sgRNA:nuclease interaction, optimizing assembly of the sgRNA:nuclease complex, removing or altering RNA destabilizing sequence elements, or adding RNA stabilizing sequence elements. In some embodiments, the sgRNA contains a 5' stem-loop structure proximal to, or adjacent to, the binding region that interacts with the
10 sgRNA-mediated nuclease. Optimization of the 5' stem-loop structure can provide enhanced stability or assembly of the sgRNA:nuclease complex. In some cases, the 5' stem-loop structure is optimized by increasing the length of the stem portion of the stem-loop structure. An exemplary sgRNA containing an optimized 5' stem-loop structure is encoded by:

SEQ ID NO:44 [N]₅.

- 15 100 GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAU
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
U, where the nucleotides contributing to the elongated stem portion of the 5' stem-loop structure are underlined. In some cases, the optimized sgRNA is at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identical or more to SEQ ID NO:44, or contains fewer than 10, 9, 8, 7,
20 6, 5, 4, 3, or 2 nucleotide additions, deletions, or substitutions compared to SEQ ID NO:44.

[0109] In some embodiments, the 5' stem-loop optimization is combined with mutations for increased transcription to provide an optimized sgRNA. For example, an A-U flip and an elongated stem loop can be combined to provide an optimized sgRNA. An exemplary sgRNA containing an A-U flip and an elongated 5' stem-loop is encoded by:

- 25 SEQ ID NO: 45 [N]₅.

- 100 GUUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAAU
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
U, where the A-U flipped nucleotides and the nucleotides contributing to the elongated stem portion of the 5' stem-loop structure are underlined. In some cases, the optimized sgRNA is
30 at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identical or more to SEQ ID NO:45, or contains fewer than 10, 9, 8, 7, 6, 5, 4, 3, or 2 nucleotide additions, deletions, or substitutions compared to SEQ ID NO:45.

[0110] sgRNAs can be modified by methods known in the art. In some cases, the modifications can include, but are not limited to, the addition of one or more of the following sequence elements: a 5' cap (*e.g.*, a 7-methylguanylate cap); a 3' polyadenylated tail; a riboswitch sequence; a stability control sequence; a hairpin; a subcellular localization
 5 sequence; a detection sequence or label; or a binding site for one or more proteins. Modifications can also include the introduction of non-natural nucleotides including, but not limited to, one or more of the following: fluorescent nucleotides and methylated nucleotides.

[0111] Also described herein are expression cassettes and vectors for producing sgRNAs in a host cell. The expression cassettes can contain a promoter (*e.g.*, a heterologous promoter)
 10 operably linked to a polynucleotide encoding an sgRNA. The promoter can be inducible or constitutive. The promoter can be tissue specific. In some cases, the promoter is a U6, H1, or spleen focus-forming virus (SFFV) long terminal repeat promoter. In some cases, the promoter is a weak mammalian promoter as compared to the human elongation factor 1 promoter (EF1A). In some cases, the weak mammalian promoter is a ubiquitin C promoter
 15 or a phosphoglycerate kinase 1 promoter (PGK). In some cases, the weak mammalian promoter is a TetOn promoter in the absence of an inducer. In some cases, when a TetOn promoter is utilized, the host cell is also contacted with a tetracycline transactivator. In some embodiments, the strength of the selected sgRNA promoter is selected to express an amount of sgRNA that is proportional to an amount of Cas9 or dCas9. The expression cassette can be
 20 in a vector, such as a plasmid, a viral vector, a lentiviral vector, *etc.* In some cases, the expression cassette is in a host cell. The sgRNA expression cassette can be episomal or integrated in the host cell.

E. Effector Domains

[0112] Described herein are effector domains for recruitment to a polypeptide of interest or
 25 a genetic target of interest. One or more effector domains, or one or more copies of an effector domain, can be fused to an affinity agent and recruited to a polypeptide of interest that is fused to an epitope or multimerized epitope recognized by the affinity agent. Alternatively, one or more effector domains, or one or more copies of an effector domain can be fused to a small guide RNA-mediated nuclease (*e.g.*, dCas9 or Cas9) and recruited to an
 30 sgRNA that specifically binds to a genetic target of interest. Effector domains can be any polypeptide that provides a desired effector function. Exemplary effector domains include, but are not limited to enzymes, adaptor proteins, fluorescent proteins, transcriptional activators, and transcriptional repressors.

III. Methods

[0113] Described herein are methods for recruiting effector domains to a polypeptide of interest. The recruitment can be performed *in vivo*, *e.g.*, in a cell, or *in vitro*, *e.g.*, in a cell extract. In one embodiment, the recruitment is performed in a cultured cell. In some
 5 embodiments, the recruitment is performed by contacting a cell (*e.g.*, a cell in culture or a cell in an organism) or cell extract with a composition containing a polypeptide of interest fused to an epitope or multimerized epitope; and an affinity agent fusion protein, wherein the affinity agent fusion protein contains an affinity domain that specifically binds one or more epitopes that are fused to the polypeptide of interest, and one or more effector domains or one
 10 or more copies of an effector domain. The method can include recruiting 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more affinity agents, and their fused effector domains to the epitope or multimerized epitope, and thus the polypeptide of interest.

[0114] The contacting can be performed by contacting the cell or cell extract with one or
 15 more expression cassettes that contain a promoter operably linked to a polynucleotide that encodes one or more components of the composition. In some cases, each component of the composition is encoded in a polynucleotide in a separate expression cassette. In some cases, an expression cassette can contain one or more polynucleotides that encode multiple components of the composition. In some cases, one or more of the expression cassettes are in
 20 a vector, such as a lentiviral vector. For example, a cell or population of cells can be transiently or stably transfected with a vector (*e.g.*, lentiviral vector) containing an expression cassette having a promoter operably linked to a polynucleotide encoding a polypeptide of interest (*e.g.*, dCas9 or any other polypeptide of interest) fused to, *e.g.*, a multimerized epitope or a multimerized effector domain. The cell or population of cells can optionally be
 25 subject to a selection step to select against a cell that has not been transfected. Stably or transiently transfected cells can be transfected with a second vector (*e.g.*, lentiviral vector) containing an expression cassette with a promoter operably linked to a polynucleotide encoding an affinity agent that specifically binds to the multimerized epitope and is fused to an effector domain. Alternatively, the second vector can contain an expression cassette with
 30 a promoter operably linked to a polynucleotide encoding an sgRNA. One of skill in the art can appreciate that expression vectors described herein can be used in any order, or simultaneously to contact a cell or cell extract with a polypeptide of interest fused to an epitope or multimerized epitope. For example a cell can be first transfected with an

expression vector with a promoter operably linked to a polynucleotide encoding an sgRNA and then transfected with an expression vector with a promoter operably linked to a polynucleotide encoding a dCas9 fused to a multimerized epitope or multimerized effector domain.

- 5 [0115] Recruitment of effector domains to the polypeptide of interest can be detected by a variety of methods known in the art. In some cases, the effector domain is a fluorescent protein, and the method includes directing incident excitation light onto the cell or cell extract and detection of emission light from the cell or cell extract to detect recruitment of the fluorescent protein to the polypeptide of interest. In other cases, the effector domain is a transcriptional modulator and recruitment can be detected by a change in expression of a target genetic element or a change in cellular phenotype.
- 10

IV. Kits

- [0116] Also described herein are kits for performing methods described herein or obtaining or using a composition described herein. Such kits can include one or more polynucleotides encoding one or more compositions described herein (*e.g.*, an sgRNA, a dCas9, an epitope or multimerized epitope, an affinity agent, one or more effector domains or multimerized effector domains), or portions thereof. The polynucleotides can be provided as expression cassettes with promoters operably linked to one or more of the foregoing polynucleotides. The expression cassettes can be provided in one or more vectors for transfecting a host cell.
- 15
- 20 In some embodiments, the kits provide a host cell transfected with one or more polynucleotides encoding one or more compositions described herein.

- [0117] For example, a kit can contain a vector containing an expression cassette with a promoter operably linked to a polynucleotide encoding an sgRNA scaffold and a cloning region. A binding region of the sgRNA can be cloned into the cloning region, thereby generating a polynucleotide encoding an sgRNA that targets a desired genetic element. Alternatively, or in addition, the kit can contain an expression cassette with a promoter operably linked to a polynucleotide encoding an sgRNA. As another example, a kit can contain a vector containing an expression cassette with a promoter operably linked to a polynucleotide encoding a cloning region and an epitope or multimerized epitope or effector domain or multimerized effector domain. A polypeptide of interest or an affinity domain can be cloned into the cloning region thereby fusing the polypeptide of interest or affinity domain to the epitope, multimerized epitope, effector domain, or multimerized effector domain.
- 25
- 30

[0118] In one embodiment, the kit contains (i) an expression cassette with a heterologous promoter operably linked to a polynucleotide encoding an affinity agent fusion protein, wherein the affinity agent fusion protein comprises: an affinity domain that specifically binds the epitope; and a effector domain; and/or (ii) an expression cassette encoding: (a) a
5 heterologous promoter, a cloning site, and a multimerized epitope, wherein the cloning site is configured to allow cloning of a polypeptide of interest operably linked to the promoter and fused to the multimerized epitope; or (b) a heterologous promoter operably linked to a polypeptide of interest fused to a multimerized epitope.

[0119] All patents, patent applications, and other publications, including GenBank
10 Accession Numbers, cited in this application are incorporated by reference in the entirety for all purposes.

EXAMPLES

[0120] The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of non-critical parameters
15 that could be changed or modified to yield essentially the same or similar results.

Example 1

Introduction

[0121] Signal amplification is important for many biological processes as well as bioengineering applications. Outputs from transcriptional and signaling pathways can be
20 amplified by recruiting multiple copies of regulatory proteins to a site of action. Taking advantage of this principle, we have developed a novel protein scaffold (a repeating peptide array termed SunTag) that can recruit multiple copies of an antibody-fusion protein. We show that the SunTag can be used to recruit a variety of proteins to the protein scaffold, including GFP, which allows tagging of a single protein molecule with up to 24 copies of GFP, thereby
25 enabling long-term imaging of single protein molecules in living cells. We also used the SunTag to create a potent synthetic transcription factor by recruiting multiple copies of a transcriptional activation domain to a modified CRISPR/Cas9 protein and demonstrate strong activation of endogenous gene expression with this system. Thus, SunTag provides a versatile platform for multimerizing proteins on a target protein scaffold and is likely to have many
30 potential applications in imaging and in controlling biological outputs.

Materials and Methods

Cell culture, transfection and viral infection

[0122] HEK293 and U2OS cells were grown in DMEM supplemented with 10% FCS and Pen/Strep. K562 cells were grown in RPMI containing 25 mM HEPES supplemented with 10% FCS and Pen/Strep. HEK293 and U2OS cells were transfected with PEI (Sigma) and Eugene 6 (Roche), respectively. To generate lentivirus, HEK293 cells were plated in 6-well plates, and 24 hr after plating, cells were transfected with lentiviral packaging plasmids. 24 hr after transfection, the cell culture medium was replaced, and 72 hr after transfection the cell medium containing lentiviral particles was harvested and either used directly to infect cells or frozen at -80° C. To generate K562 cells stably expressing dCas9-SunTag_{10x_v4} and scFv-GCN4-GFP-NLS-VP64, cells were infected with freshly harvested lentivirus diluted 1:3 in RPMI cell culture medium and incubated for 24 hr in virus-containing medium. Our initial experiments with the polyclonal K562 cell line expressing dCas9-SunTag_{10x_v4} and scFv-GCN4-GFP-NLS-VP64 generated in this way revealed that only ~40% of cells showed robust transcriptional activation, possibly due to cell-to-cell variation in transgene expression level. We therefore plated the K562 cells expressing dCas9-SunTag_{10x_v4} and scFv-GCN4-GFP-VP64 at one cell per well in a 96-well plate and isolated several monoclonal cell lines that showed uniform transcriptional activation. One clone (E3) was selected for further experiments. For all experiments involving transcriptional activation, K562 cells expressing dCas9-SunTag_{10x_v4} and scFv-GCN4-GFP-VP64 were infected with lentivirus encoding for a gene-specific sgRNA together with a puromycin resistance gene and either BFP or mCherry at an multiplicity of infection (MOI) of less than one, so most cells received a single lentivirus. Cells were then treated with 1 µg/ml puromycin for 3 days to select for cells that expressed an sgRNA.

Plasmids and cloning

[0123] Sequences of all constructs used in this paper are provided in the sequence listing as SEQ ID NOs:14-41 and 56-57.

Microscopy

[0124] Cells were grown in 96-well glass bottom dishes and were imaged on an inverted Nikon TI spinning disk confocal microscope with the Nikon Perfect Focus system which was operated by Micro-Manager software (Edelstein *et al.*, 2010). Epifluorescence images were

acquired using widefield epifluorescence illumination using a 20x air objective combined with a Hamamatsu CMOS Flash 4.0 camera. All other images were obtained using spinning disk microscopy and were acquired using a 100x 1.45NA oil objective combined with an EM-CCD camera (Andor). For time-lapse microscopy cells were grown in DMEM:F12 medium without phenol red, supplemented with 20 mM HEPES to maintain correct pH in the absence of added CO₂ and were imaged in a thermally-controlled chamber heated to 37° C. For single molecule imaging of the SunTag, 2x2 pixel binning was applied, resulting in a pixel size of 166 nm. For photobleaching experiments, a single point was illuminated for 500 ms using a dedicated 488 nm photobleaching laser which was run at 5 mW. Image acquisition before and after photobleaching was performed using spinning disk confocal microscopy as described above. Fluorescence intensities of GFP before and after photobleaching were determined for each time point and corrected for cellular background fluorescence signal.

Quantitative image analysis

[0125] To determine the number of antibodies bound to a single peptide array, a sfGFP-mCherry fusion protein was created, in which sfGFP and mCherry were separated by a long linker to prevent energy transfer between the two fluorophores. Image acquisition parameters were chosen so that GFP and mCherry fluorescence intensities were approximately equal. Imaging of the mito-mCherry-peptide arrays with GFP-tagged antibody and the sfGFP-mCherry fusion protein was performed on the same day using the same acquisition parameters to allow a quantitative comparison. In all cases, background fluorescence was subtracted first. The sfGFP:mCherry fluorescence intensity ratio for the sfGFP-mCherry fusion protein of all cells was averaged and was set to 1. The GFP:mCherry ratio of individual cells was then normalized to this average.

[0126] To measure spot fluorescence intensities of either single SunTag foci associated with the cell membrane or of individual telomeres, a circular region of interest (ROI) was generated with a diameter of 0.5 µm. The ROI was centered over the individual fluorescent foci and the average fluorescence intensity of the ROI was measured. For a background measurement, the same ROI was positioned in five different areas of the cell (or the nucleus in the case of the telomere measurements) that did not contain any fluorescent foci and the average intensity of those measurements was used as a background value that was subtracted from the foci intensities.

[0127] To determine kinesin run lengths and speeds, maximal intensity projections were generated of the single color time-series to identify kinesin runs. Kymographs were then created along the motor trajectories in these maximal intensity projections and the run length and speed were then calculated from the length and angle of the bright fluorescence lines then
5 were apparent in the kymographs.

[0128] In experiments in which the fraction of inward and outward moving particles was determined, a line was drawn halfway in between the cell nucleus and the most distal part of the cell and the number of particles that crossed the line, either moving towards the nucleus, or moving towards the cell periphery was scored.

10 *Quantification of protein and mRNA levels*

[0129] To determine the levels of CXCR4 and CDKN1B transcriptional activation, K562 cells stably expressing either dCas9-VP64-BFP or dCas9-SunTag_{10x_v4} together with scFv-GCN4-GFP-VP64 were infected with lentivirus encoding individual sgRNAs targeting the upstream region of the CXCR4 and CDKN1B transcripts, as well as BFP and a puromycin
15 resistance gene. Cells were then selected with 1 µg/ml puromycin for 3 days. Measurements of CXCR4 protein levels was then performed by FACS as described previously (Gilbert *et al.*, 2013). For the measurement of CDKN1B mRNA levels, total RNA was isolated with Trizol (Ambion) and cDNA was synthesized using the Superscript cDNA synthesis kit VILO (Life Technologies). qPCR was then performed using the following CDKN1B specific
20 primers: Fw GAGTGGCAAGAGGTGGAGAA (SEQ ID NO:46) and Rev GCGTGTCCTCAGAGTTAGCC (SEQ ID NO:47) as described previously (Gilbert *et al.*, 2013). sgRNA sequences used in this study are: Control TTCTCTTGCTGAAAGCTCGA (SEQ ID NO:48), CXCR4 #1 GCCTCTGGGAGGTCCTGTCCGGCTC (SEQ ID NO:49), CXCR4 #2 GCGGGTGGTCGGTAGTGAGTC (SEQ ID NO:50), CXCR4 #3
25 GCAGACGCGAGGAAGGAGGGCGC (SEQ ID NO:51), CDKN1B #1 AAGGTCGCCGGCAGCTCGCT (SEQ ID NO:52), CDKN1B #2 GAAGCCGGGACCTGGACCAG (SEQ ID NO:53), CDKN1B #3 CTGCGTTGGCGGGTTCGCCG (SEQ ID NO:54), CDKN1B #4 GGGCCCGGCGCTGCGTTGG (SEQ ID NO:55).

30 *Transwell migration*

[0130] Recombinant human SDF-1alpha (Peprotech) was used as a chemoattractant for the migration assay. K562 cells were cultured in RPMI-1640 with 2% serum for 16 hr. 75,000

cells were counted and resuspended in RPMI-1640 with 2% serum and added to the upper chamber of 24-well Transwell inserts (8-micron pore size polyethylene terephthalate, Millipore), and 200 ng/mL SDF-1a was added to the lower chamber. The number of K562 cells that migrated to the lower chamber was quantified after 5 hr by flow cytometry on a BD Bioscience LSR-II flow cytometer. Results are displayed as the fold change in directional migrating cells over control cell migration.

K562 growth rate measurements

[0131] K562 cells stably expressing either dCas9-VP64-BFP alone or dCas9-SunTag_{10x_v4} together with scFv-GCN4-GFP-VP64 were infected with lentivirus encoding indicated sgRNAs together with BFP at an MOI of approximately 0.3. Three days after infection, the fraction of BFP positive cells was determined by FACS for each sample. Cells were then grown for two weeks, after which the fraction of BFP positive cells was re-measured. In cells infected with a control sgRNA, the fraction of BFP positive cells remained constant over time, indicating that infection with a lentivirus encoding control sgRNA and BFP did not reduce cell proliferation rate as compared to the uninfected cells within the same dish. In contrast, in dCas9-SunTag_{10x_v4}-VP64 expressing cells infected with 3/4 sgRNAs targeting CDKN1B, the fraction of the BFP positive cells was substantially reduced over time, indicating they had a reduced growth rate compared to uninfected cells in the same dish. In parallel the cell doubling time of uninfected cells was determined. Using the cell doubling time and the fraction of BFP positive cells at day 3 and day 14, the growth rate of BFP positive cells was determined compared to uninfected control cells.

Results

Development of SunTag, a system for recruiting multiple protein copies to a polypeptide scaffold

[0132] Protein multimerization on a single RNA or DNA template is made possible by identifying protein domains that bind with high affinity to a relatively short nucleic acid motif. We therefore sought a protein-based system with similar properties, specifically a protein that can bind tightly to a short peptide sequence. Antibodies are capable of binding to short, unstructured peptide sequences with high affinity and specificity, and, importantly, peptide epitopes can be designed that differ from naturally occurring sequences in the genome. Furthermore, while antibodies generally do not fold properly in the cytoplasm, single chain variable fragment (scFv) antibodies, in which the epitope binding regions of the

light and heavy chains of the antibody are fused to form a single polypeptide, have been successfully expressed in soluble form in cells (Colby *et al.*, 2004a; Lecerf *et al.*, 2001; Worn *et al.*, 2000).

[0133] We expressed three previously developed single-chain antibodies fused to GFP in U2OS cells to see if they would recognize their cognate peptide (multimerized in 4 tandem copies) fused to the cytoplasmic side of the mitochondrial protein mitoNEET (referred to here as Mito) (Colca *et al.*, 2004). We then assayed by fluorescence microscopy whether the antibody-GFP fusion proteins would be recruited to the mitochondria, which would indicate binding between antibody and peptide (Fig. 1A and Fig. 2A). The three antibody-peptide tested were: 1) A single chain variable fragment (scFv) antibody, developed using *in vitro* evolution, which binds with very high affinity to a 22 amino acid monomeric fragment of the yeast transcription factor GCN4 (scFv-GCN4) (Hanes *et al.*, 1998), 2) V1 12.3-Htt, an antibody light chain domain, that binds to a 20 amino acid fragment of the N-terminus of huntingtin (Colby *et al.*, 2004a; Colby *et al.*, 2004b) and 3) scFv-C4-Htt, a single chain variable fragment antibody that binds to the N-terminal 17 amino acids of huntingtin (Lecerf *et al.*, 2001). The GFP-tagged GCN4 antibody-peptide and the V1 12.3-Htt antibody-peptide pairs, but not the scFv-C4-Htt pair, were recruited to mitochondria, indicating that these antibodies were binding to their cognate peptides *in vivo* (Fig. 1B-C). However, expression of the Htt peptide-Mito fusion, even without the antibody-GFP being expressed, disrupted mitochondrial organization (Fig. 1B). This effect was likely due to aggregation of the Htt 4xpep, as expression of the 4xHtt peptide lacking the mitochondrial targeting domain resulted in large perinuclear aggregates (data not shown), making this antibody-peptide pair unsuitable. In contrast, the GCN4 peptide showed no detectable aggregation and the scFv-GCN4-GFP was not recruited to mitochondria in the absence of its cognate peptides, confirming the specificity of the interaction (Fig. 1C). Thus, we focused our further efforts on the GCN4 antibody-peptide pair

[0134] The GCN4 antibody was optimized to allow intracellular expression in yeast (Worn *et al.*, 2000). In human cells however, we still observed some protein aggregates of scFv-GCN4-GFP at high expression levels (Fig. 4A). To improve scFv-GCN4 stability, we added a variety of N-and C-terminal fusion proteins known to enhance protein solubility, and found that fusion of superfolder-GFP (sfGFP) along with the small solubility tag GB1 to the C-terminus of the GCN4 antibody almost completely eliminated protein aggregation, even at

very high expression levels (Fig. 4A). Thus, we performed all further experiments with scFv-GCN4-sfGFP-GB1 (hereafter referred to as scFv-GCN4-GFP).

[0135] Very tight binding of the antibody-peptide pair *in vivo* is critical for the formation of multimers on a protein scaffold backbone. To determine the dissociation rate of the GCN4 antibody-peptide interaction, we performed fluorescence recovery after photobleaching (FRAP) experiments on scFv-GCN4-GFP bound to the mitochondrial-localized mito-mCherry-4xGCN4pep. After photobleaching, very slow GFP recovery was observed (half-life of ~5-10 min (Fig. 3A-B)), indicating that the antibody bound very tightly to the peptide. We next sought to optimize the spacing of the scFv-GCN4 binding sites within the protein scaffold so that they could be saturated by scFv-GCN4, since steric hindrance of neighboring peptide binding sites was a concern. We varied the spacing between neighboring GCN4 peptides and quantified the antibody occupancy on the peptide array using the mitochondrial localization assay described above combined with quantitative fluorescence microscopy. The ratio of GFP fluorescence (from the scFv-GCN4-GFP antibody) to mCherry fluorescence (present in one copy on the mito-4xGCN4pep scaffold) on the mitochondria provided a measure of the number of antibodies recruited to the protein scaffold. This ratio was normalized to the GFP-mCherry ratio of a control protein in which GFP and mCherry were directly fused (Fig 3. A-C and 4B). We compared a short (GGSGG; SEQ ID NO:3) and long (GGSGGSGGTGGTGG; SEQ ID NO:59) linker and found an average GFP:mCherry molar ratio of 3.4 and 2.9, respectively (Fig. 3C-D). This experiment indicates that a spacer as short as five amino acids sufficiently separates peptides to allow binding of antibodies to neighboring peptides. Importantly, in a peptide array containing 24 tandem copies of the peptide, separated by 5 a.a. linkers, we found an average GFP:mCherry molar ratio of ~24 (Fig. 3C,E). These results show that full antibody occupancy can be achieved with as many as 24 copies of a 22 a.a peptide binding site, separated by a 5 residue linker, fused to the parent polypeptide chain (an 24x peptide tag is thus ~70 kDa). Taken together, these results show that this optimized GCN4 antibody-peptide pair meets all the requirements for an effective system for recruiting many copies of a protein to a polypeptide scaffold. As the GCN4 antibody-peptide pair allows ultra-bright fluorescent labeling of molecules, we named the tagging system SUpErNova (SunTag) after the very bright stellar explosion.

Single molecule imaging in living cells using SunTag

[0136] Single molecule imaging is a powerful emerging tool in biology; in our first application of the SunTag, we tested whether SunTag_{24x} (24 copies of the peptide binding site) could be used for single molecule imaging in living cells. We first fused a cytoplasmic protein, infrared fluorescent protein (IFP), to the C-terminus of the SunTag_{24x} (SunTag_{24x}-IFP) and added a plasma membrane targeting domain (CAAX) to SunTag_{24x}-IFP (SunTag_{24x}-IFP-CAAX) and co-expressed the scFv-GCN4-GFP antibody (referred to as SunTag_{24x}-IFP-CAAX-GFP) which resulted in localization to the plasma membrane. By spinning disk microscopy, individual fluorescent punctae could be visualized diffusing in the plane of the membrane (Fig. 5A); their intensities were very homogeneous (Fig. 5A-B), suggesting that they are single polypeptides and not a mixture of aggregates. Single GFP molecules at the plasma membrane are routinely imaged by total internal fluorescence (TIRF) microscopy, but these molecules bleach in several seconds. In contrast, with multiple GFP copies bound to a single SunTag_{24x}-IFP-CAAX-GFP, we could still observe single molecules at the plasma membrane after several minutes of continuous imaging.

[0137] Single molecule imaging in the interior of the cell is more difficult than at the plasma membrane due to lower signal to background and the inability to use TIRF microscopy. We therefore tested whether SunTag could be used to image single molecules deep inside the cell. We imaged U2OS cells expressing low levels of the mitoNEET mitochondrial targeting domain fused to the SunTag_{24x} together with scFv-GCN4-GFP using spinning disk confocal microscopy. Bright punctae of uniform fluorescence intensity were observed that colocalized with mitochondria and showed very rapid diffusion in the mitochondrial membrane. Similarly, when the SunTag_{24x} was fused to a cytoplasmic protein (infrared fluorescent protein IFP-SunTag_{24x}-GFP) or a nuclear protein (NLS-IFP-SunTag_{24x}-GFP), bright foci were observed that rapidly diffused in the cytoplasm or nucleus, respectively (Fig. 6A). Together, these results show that the SunTag_{24x} can be used to image single protein molecules in different regions of the cell.

[0138] We next tested whether the SunTag could be used to make single molecule measurements of cytoskeletal motors moving *in vivo*. Previous studies have imaged single motor proteins fused to three copies of GFP using TIRF microscopy (Cai *et al.*, 2009), but the signal is relatively weak and imaging by TIRF microscopy is limited to molecules that are very close to the glass surface (<200 nm). We first fused SunTag_{24x} to a truncated version of

kinesin-1 (termed K560), which is a processive motor that lacks its cargo binding domain (Friedman and Vale, 1999). Spinning disk confocal imaging (10 frames/sec) of K560-SunTag_{24x}-GFP revealed bright foci moving unidirectionally throughout the cell with an average speed of $1.29 \pm 0.24 \mu\text{m/s}$ (Fig. 5B and 6B). Due to the very low photobleaching, we were able to accurately measure run lengths of single K560-SunTag_{24x}-GFP molecules, revealing an average run length of $1.28 \pm 0.63 \mu\text{m}$ (Fig. 6C), which is consistent with previous measurements (Cai *et al.*, 2009; Courty *et al.*, 2006). These results show that the SunTag allows long-term single molecule imaging of function cytoskeletal motor proteins *in vivo*.

[0139] Interestingly, when we imaged motility of K560-SunTag_{24x}-GFP (which moves exclusively towards plus-ends of microtubules), we found that a substantial fraction of K560-SunTag_{24x}-GFP motors moved towards the cell interior, indicating that the microtubule tracks for these motors have their plus-ends directed inwards (Fig. 5C-D). This was surprising, as microtubules are generally thought to be oriented with their plus-ends outwards. Indeed, in these same cells, when microtubule polarity was assessed using a conventional method of visualizing EB3-GFP, which tracks along the growing plus ends of microtubule, then microtubules plus-ends were found to be oriented almost exclusively towards the cell periphery (Fig. 5C-D). These results reveal that cells contain a subpopulation of microtubules that have inverted polarity and are not growing and hence not interacting with EB3. Furthermore, these results show that K560-SunTag_{24x}-GFP can be used as a general tool to dissect microtubule polarity *in vivo*.

[0140] We next sought to test whether the SunTag could be used to study cytoskeletal motors whose motility have not been characterized. KIF18b is a member of the kinesin superfamily which has been shown to track with growing microtubule plus-ends and regulate their dynamics (Stout *et al.*, 2011; Tanenbaum *et al.*, 2011). However, it is currently unclear how Kif18b tracks the growing plus-ends. Robust accumulation at microtubule plus-ends of Kif18b requires both direct binding to the microtubule plus-end tracking protein EB1, as well as Kif18b's motor domain (Akhmanova and Steinmetz, 2008; Stout *et al.*, 2011; Tanenbaum *et al.*, 2011), suggesting Kif18b may be initially recruited to plus-ends by EB1 and subsequently individual molecules of Kif18b remain at the tip of the growing microtubule by transporting itself along the microtubule at a rate equal to the speed of microtubule growth. However, while KIF18b motility has not been directly measured, all the homologs of Kif18b were found to move at rates that are far too slow to keep up with microtubule growth *in vitro*

(<100 nm/s), arguing against this model. To analyze Kif18b's motility *in vivo*, we expressed full length KIF18b with a C-terminal SunTag_{24x} in U2OS cells. Surprisingly, and unlike what was reported for its homologs, single KIF18b-SunTag_{24x}-GFP molecules moved highly processively and at fast speeds (635 ± 163 nm/s; mean $\pm$ s.d.) (Fig. 5E-F), demonstrating that individual molecules of Kif18b are sufficiently fast and processive to remain at the tip of microtubules as they grow through its own plus-end directed motility, explaining the requirement of Kif18b's motor domain for its ability to track growing microtubule plus-ends. Taken together, our results for kinesin-1 and KIF18b show that the SunTag is a versatile tool for imaging single molecule motility in living cells.

[0141] We also tested whether the SunTag could be used to image single cytoskeletal filament dynamics in dense networks using fluorescence speckle microscopy (FSM). FSM visualizes and tracks identifiable fluorescent "speckles" that arise from the stochastic variations in the incorporation of fluorescently-labeled actin or tubulin monomers into complex cytoskeletal networks (Waterman-Storer *et al.*, 1998). However, due to the stochastic nature of the labeling in traditional FSM, signal-to-noise is generally suboptimal and fluorescent speckles can contain fluorescently labeled monomers that are present in different filaments. Therefore, a FSM strategy that allows very bright labeling of single filaments would be a great improvement. We examined whether we could follow the movements of microtubules in living cells by creating positional marks using single SunTagged molecules. For this purpose, we fused SunTag_{24x} to a K560 ATP hydrolysis blocked, rigor mutant (K560rig) that binds tightly to microtubules but does not translocate along them (Rice *et al.*, 1999). As K560rig-SunTag_{24x}-GFP binds statically to a microtubules, a movement of a K560rig-SunTag_{24x}-GFP foci reveals the translocation of the entire microtubule. Expression of K560rig-SunTag_{24x}-GFP at low levels resulted in sparse labeling of the microtubule network (visualized by α -tubulin-mCherry), in which individual K560rig-SunTag_{24x}-GFP molecules could be observed colocalizing with microtubules (Fig. 5G-H). While the microtubule network appeared largely static when imaging the microtubules directly with mCherry-tubulin, imaging of K560rig-SunTag_{24x}-GFP revealed many microtubules undergoing translocation events in cells (Fig. 5H). As many microtubules had two or more K560rig-SunTag_{24x}-GFP molecules bound, changes in angle of the microtubule axis also could be observed (Fig. 5H). These results reveal that the SunTag provides a powerful tool to study movements of individual microtubule filaments in dense microtubule networks in living cells.

Optimizing protein expression levels of the SunTag

[0142] The first generation construct of SunTag_{24x} described in the previous sections was expressed at extremely low levels, usually only a few hundred protein copies per cell (based on the number of foci observed when the SunTag_{24x} is co-expressed with scFv-GCN4-GFP).

5 Indeed, when SunTag_{24x} peptide array was fused directly to sfGFP and transfected in HEK293 cells, the GFP signal was extremely low compared to sfGFP expressed alone (Fig. 7A). While such low level expression is ideal for single molecule imaging, other applications for controlled protein multimerization could benefit from higher expression. The very low expression level of the SunTag_{24x} may be due to either a problem with the mRNA (poor
10 synthesis, stability or translation) or an instability of the peptide array after its translation. To distinguish between these possibilities, we inserted a viral P2A ribosome skipping sequence in between the 24xGCN4 peptide array and GFP, which allows synthesis of two distinct proteins (i.e. 24xGCN4 peptide array and GFP) from the same mRNA (Kim *et al.*, 2011). Insertion of the P2A site in between 24xGCN4 peptide and GFP dramatically increased GFP
15 expression (Fig. 7A), indicating that the mRNA is present and efficiently translated. This result strongly suggests that poor protein stability explains the low expression of the 24xGCN4 peptide array.

[0143] The GCN4 peptide contains many hydrophobic residues (Fig. 7B) and is largely unstructured in solution (Berger *et al.*, 1999); thus, the poor expression of the peptide array
20 could be due to its unstructured and hydrophobic nature. To test this idea, we designed several modified peptide sequence that were predicted to increase α -helical propensity and reduce hydrophobicity. One of these optimized peptides (v4, Fig. 7B) was expressed moderately well as a 24x peptide array although somewhat higher expression was achieved with a 10x peptide array (Fig. 7C). Importantly, the GCN4 v4 peptide array still bound the
25 antibody with similar affinity as the original peptide (Fig. 4D-E). Furthermore, robust single molecule motility could be observed when K560 was tagged with the optimized v4 24x peptide array, suggesting that the optimized v4 peptide array did not interfere with protein function. Together, these results identify a new version of the peptide array that can be used for both single molecule imaging as well as applications requiring higher expression.

Activation of gene transcription using Cas9-SunTag

[0144] Since the SunTag system can be used to amplify a fluorescence signal, we wondered whether it also could be used to amplify other outputs from biological systems.

Gene transcription is enhanced by recruiting multiple copies of transcriptional activators to endogenous or artificial gene promoters (Anderson and Freytag, 1991; Chen *et al.*, 1992; Pettersson and Schaffner, 1990). Thus, we thought that activation of gene transcription might also be achieved by recruiting multiple copies of a synthetic transcriptional activator to a

5 gene. Recently, a highly versatile, synthetic transcriptional activator was developed by fusing the herpes virus transcriptional activation domain VP16 (or 4 copies of VP16, termed VP64) to a nuclease-deficient mutant of the CRISPR effector protein Cas9 (dCas9), which can be targeted to any sequence in the genome using sequence specific small guide RNAs (sgRNAs) (Cheng *et al.*, 2013; Farzadfard *et al.*, 2013; Gilbert *et al.*, 2013; Hu *et al.*, 2014; Kearns *et*

10 *al.*, 2014; Maeder *et al.*, 2013; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013). While targeting of dCas9-VP64 was able to increase transcription of the targeted gene, the level of gene activation using dCas9-VP64 was generally very low, most often less than 50% (Cheng *et al.*, 2013; Hu *et al.*, 2014; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013), thus severely limiting the potential use of this system. Intriguingly, several studies found that recruitment of multiple

15 copies of dCas9-VP64 to a single promoter, using multiple non-overlapping sgRNAs could enhance transcriptional activation (Cheng *et al.*, 2013; Hu *et al.*, 2014; Maeder *et al.*, 2013; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013), consistent with the fact that multiple transcriptional activators are required to stimulate robust transcription. We therefore wondered whether recruitment of multiple VP64 domains to a single molecule of dCas9

20 using the SunTag would enhance the ability of dCas9 to activate endogenous transcription (See Fig. 8A).

[0145] To test whether dCas9 could be tagged with the SunTag, dCas9-SunTag_{24x_v4} was co-expressed with scFv-GCN4-GFP and targeted to telomeres using a telomere-specific sgRNA. When examined by fluorescence microscopy, very bright dots were observed in the

25 nucleus, similar to previous work with dCas9 directly labeled with GFP (dCas9-GFP) (Chen *et al.*, 2013) (Fig. 9A). Comparison of dCas9-SunTag_{24x_v4}-GFP with dCas9-GFP, showed that telomere labeling was ~20-fold brighter when dCas9 was labeled with the SunTag compared to dCas9 directly fused to GFP, consistent with the recruitment of ~24 copies of GFP to a single dCas9 molecule (Fig. 9A-B). As a control, in the absence of the sgRNA

30 targeting the telomere, nuclear GFP fluorescence was diffuse (Fig. 9A). Thus, dCas9-SunTag can efficiently recruit multiple proteins to a single genomic locus and can be used for very bright labeling of telomeres.

[0146] Next, scFv-GCN4-GFP was fused to VP64 to test whether recruitment of multiple VP64 domains to a promoter would enhance transcription of the downstream gene. K562 cell lines were generated expressing either dCas9-VP64 (Gilbert *et al.*, 2013) alone or co-expressing dCas9_{10x_v4} with GCN4-sfGFP-NLS-VP64 (hereafter referred to as dCas9-SunTag-VP64). dCas9-SunTag_{10x_v4} was used for these experiments, as we found similar maximal activation and less cell-to-cell variation in gene expression than the dCas9-SunTag_{24x_v4} (see also Fig. 7C). As a target gene, we selected CXCR4, a transmembrane receptor known to stimulate cell migration, which is normally poorly expressed in K562 cells. dCas9-VP64 and dCas9-SunTag_{10x_v4}-VP64 expressing cells were infected with a lentivirus that encoded either a control sgRNA or an sgRNA targeting CXCR4 (sgCXCR4; three different sgRNA were tested). Five days after lentivirus infection, the levels of CXCR4 protein were determined. We found little or no activation of CXCR4 expression using dCas9-VP64 with the three sgRNAs tested (Fig. 8B-C), consistent with previous studies. In contrast, strong activation (10-50-fold) was observed with all three CXCR4 sgRNAs using dCas9-SunTag_{10x_v4}-VP64 (Fig. 8B-C). These results show that robust transcriptional activation can be achieved by SunTag-dependent multimerization of transcriptional activation domains at an endogenous gene promoter.

[0147] We next wished to test whether transcriptional regulation using dCas9-SunTag_{10x_v4}-VP64 could induce a biological response. CXCR4 is a chemokine receptor which can stimulate cell migration in response to activation by SDF1a (Brenner *et al.*, 2004). We tested whether activation of CXCR4 in K562 could induce migration in response to SDF1 using a transwell migration assay. We found that activating CXCR4 expression using dCas9-SunTag_{10x_v4}-VP64 dramatically stimulated cell migration by an order of magnitude (Figure 8D). In contrast, very weak (<2-fold) enhancement of cell migration was observed using CXCR4 activation by dCas9-VP64 (data not shown). This result indicates that dCas9-SunTag_{10x_v4}-VP64-dependent gene activation is sufficiently potent to affect the behavior of these cells. Surprisingly, cells expressing the highest level of CXCR4 showed less cell migration, suggesting there may be an optimum level of CXCR4 expression for stimulation of cell migration (compare Fig. 8B with 8D).

[0148] CXCR4 is normally expressed at very low levels in K562 cells, so we tested whether the expression of a well-expressed gene, the cell cycle inhibitor CDKN1B (also known as p27kip1), could also be increased using SunTag-dependent transcriptional activation. Four different sgRNAs were designed that target CDKN1B, and their effects on

CDKN1B mRNA expression level were determined in both dCas9-VP64 and dCas9-SunTag-VP64 cells. Very little activation of CDKN1B transcription was observed using dCas9-VP64 (28% increase in mRNA at best) (Fig. 8E), while 3/4 sgRNAs robustly activated CDKN1B in dCas9-SunTag_{10x_v4}-VP64 cells (330% for the best sgRNA) (Fig. 8E). Furthermore, as
 5 expected for increased levels of the cell cycle inhibitor CDKN1B, activation of CDKN1B with dCas9-SunTag_{10x_v4}-VP64 significantly reduced cell growth (Fig. 8F). In contrast, activation of CDKN1B with dCas9-VP64 had little impact on cell growth (Fig. 8F). Taken together, these results show that the SunTag-dependent signal amplification robustly enhances transcriptional activation by dCas9-VP64 and allows functional re-engineering of
 10 cell behavior through precise control of gene expression.

Discussion

[0149] Amplification of biological signal is crucial for many biological processes as well as for bioengineering. Here, we have developed a versatile protein tagging system, the SunTag, which can be used to increase fluorescence of genetically-encoded proteins as well as amplify
 15 gene expression. The SunTag system provides a proof-of-concept of the power of controlled protein multimerization, and could form the basis for developing other protein multimerization strategies.

Imaging applications of the SunTag

[0150] SunTag represents the brightest genetically-encoded fluorescent tagging system
 20 available and has several major advantages over existing imaging methods. First, due to its extremely high signal, a low expression level of SunTag-proteins is sufficient for imaging and thus avoids potential problems associated with protein overexpression. For example, we have found that overexpression of GFP-mitoNEET is detrimental to mitochondrial function (data not shown). However, we have achieved very bright images of mitochondria with much
 25 lower expression of mitoNEET-SunTag than can be achieved by single copy GFP tagging. Second, bright labeling of both organelles and single molecules allows imaging with much lower light illumination, which reduces photobleaching and minimizes phototoxicity, allowing long-term tracking. Third, automated tracking algorithms are very sensitive to signal-to-noise ratios, and bright labeling using the SunTag will likely be beneficial for such
 30 analyses, especially for single molecule tracking *in vivo*. Fourth, the SunTag allows single molecule imaging deep inside the cytoplasm and nucleus. In contrast, single molecule imaging of GFP in TIRF microscopy is only applicable to molecules that are located very

close to the cell membrane (for examples, see (Cai *et al.*, 2009; Douglass and Vale, 2005)). Finally, our analysis of microtubule translocation in the cytoplasm provides a proof-of-concept that the SunTag, when expressed at low levels to sparsely label dense or complex structures, can be used follow the movement of individual cytoskeletal filaments. Because
 5 SunTag speckles are brighter and more homogeneous and labels only a single filament, this method might have advantages over traditional FSM, which relies on stochastic fluctuations in fluorophore distribution (Waterman-Storer *et al.*, 1998).

[0151] We also show that SunTag is a powerful single molecule reporter of intracellular processes. For example, analysis of K560-SunTag movements revealed a stable subset of
 10 microtubules with reversed polarity, which was not evident from tracking growing microtubules with EB3-GFP. The K560rig-SunTag allowed visualization of microtubule movement in dense microtubule networks. These applications could be especially powerful during mitosis, when the high microtubule density in the mitotic spindle makes analysis of single microtubules very difficult. Similarly, labeling of genomic loci by dCas9 by the
 15 SunTag allows much brighter labeling of genomic loci than dCas9 directly fused to GFP (Fig. 9) (Chen *et al.*, 2013). SunTag potentially could be used to image non-repetitive DNA loci as well using single dCas9 molecules; however, our preliminary attempts to observe single dCas9-SunTag_{24x} molecules binding to a non-repetitive DNA sequence have been unsuccessful, possibly due to the large amount of unbound dCas9 in the nucleus, which
 20 obscured detection of the bound molecule. Overall, these results show that the SunTag is a versatile tool for single molecule imaging and very bright labeling of intracellular structures and organelles.

Using SunTag to Engineer Gene Transcription and Cell Behavior

[0152] The second application of the SunTag, for which we provide a proof-of-concept, is
 25 the amplification of biological signaling pathways. Transcriptional regulation is a powerful example, as transcriptional output is strongly dependent on the number of transcriptional activators recruited to the gene promoter (Anderson and Freytag, 1991; Chen *et al.*, 1992; Pettersson and Schaffner, 1990). Indeed, previous attempts to activate transcription of endogenous genes using a single dCas9 or TALE fused to the transcriptional activation
 30 domain VP64 generally resulted in very weak or no transcriptional activation. However, several studies showed that robust gene activation was possible when multiple sgRNAs targeting the same promoter were co-expressed, in effect targeting multiple copies of dCas9-

VP64 to the promoter (Cheng *et al.*, 2013; Hu *et al.*, 2014; Maeder *et al.*, 2013; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013). In contrast, our results demonstrate that the dCas9-SunTag transcriptional system can robustly activate the expression of a gene using a single sgRNA, which not only simplifies single gene activation, but also opens the possibilities of activation of multiple genes simultaneously, potentially allowing complex genetic re-wiring of cells or organisms. For example, generation of induced pluripotent stem cells (iPS) requires expression of four proteins (Takahashi and Yamanaka, 2006), and it will be very interesting to test whether such iPS cells can be generated through activation of the endogenous genes using the SunTag, rather than through gene overexpression with transfected plasmids.

[0153] The ability to upregulate gene expression using dCas9-SunTag with a single sgRNA opens the door to large scale genetic screens to uncover phenotypes that result from increased gene expression. This application will be especially important for understanding the effects of gene upregulation in cancer. In addition, large scale activation screens could be used to identify proteins that promote induced pluripotency (Takahashi and Yamanaka, 2006) or, conversely, promote differentiation to a specific lineage.

[0154] Here, we have applied the SunTag to transcriptional activation, but a similar approach could be used to enhance dCas9-dependent transcriptional silencing. Previous work found that the fusion of dCas9 to a transcriptional silencing domain was able to inhibit gene-specific transcription (Gilbert *et al.*, 2013), but in most cases residual transcription was still observed. Possibly recruitment of many transcriptional silencing domains to a single promoter could enhance gene silencing and could be a powerful tool for loss-of-function studies. This could provide a parallel approach to gene knockout that is possible through the nuclease activity of wildtype Cas9, and could be especially useful to study essential genes and non-coding RNAs, which are both more difficult to study using Cas9-dependent DNA cleavage. In addition, multiple types of transcriptional activators or repressors could be recruited to a single scaffold, which may provide maximal or enhanced transcriptional activation or repression.

References

Akhmanova, A., and Steinmetz, M.O. (2008). Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nature reviews Molecular cell biology* 9, 309-322.

Anderson, G.M., and Freytag, S.O. (1991). Synergistic activation of a human promoter in vivo by transcription factor Sp1. *Molecular and cellular biology* 11, 1935-1943.

Berger, C., Weber-Bornhauser, S., Eggenberger, J., Hanes, J., Pluckthun, A., and Bosshard, H.R. (1999). Antigen recognition by conformational selection. *FEBS letters* 450, 149-153.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular cell* 2, 437-445.

- 5 Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature biotechnology* 22, 575-582.

- Boniface, J.J., Rabinowitz, J.D., Wulfig, C., Hampl, J., Reich, Z., Altman, J.D., Kantor, R.M., Beeson, C., McConnell, H.M., and Davis, M.M. (1998). Initiation of signal
10 transduction through the T cell receptor requires the multivalent engagement of peptide/MHC ligands [corrected]. *Immunity* 9, 459-466.

- Brenner, S., Whiting-Theobald, N., Kawai, T., Linton, G.F., Rudikoff, A.G., Choi, U., Ryser, M.F., Murphy, P.M., Sechler, J.M., and Malech, H.L. (2004). CXCR4-transgene expression significantly improves marrow engraftment of cultured hematopoietic stem cells. *Stem Cells*
15 22, 1128-1133.

Cai, D., McEwen, D.P., Martens, J.R., Meyhofer, E., and Verhey, K.J. (2009). Single molecule imaging reveals differences in microtubule track selection between Kinesin motors. *PLoS biology* 7, e1000216.

- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J.,
20 Blackburn, E.H., Weissman, J.S., Qi, L.S., *et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479-1491.

Chen, X., Azizkhan, J.C., and Lee, D.C. (1992). The binding of transcription factor Sp1 to multiple sites is required for maximal expression from the rat transforming growth factor alpha promoter. *Oncogene* 7, 1805-1815.

- 25 Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research* 23, 1163-1171.

- Colby, D.W., Chu, Y., Cassady, J.P., Duennwald, M., Zazulak, H., Webster, J.M., Messer,
30 A., Lindquist, S., Ingram, V.M., and Wittrup, K.D. (2004a). Potent inhibition of huntingtin

aggregation and cytotoxicity by a disulfide bond-free single-domain intracellular antibody. Proceedings of the National Academy of Sciences of the United States of America *101*, 17616-17621.

Colby, D.W., Garg, P., Holden, T., Chao, G., Webster, J.M., Messer, A., Ingram, V.M., and
 5 Wittrup, K.D. (2004b). Development of a human light chain variable domain (V(L)) intracellular antibody specific for the amino terminus of huntingtin via yeast surface display. Journal of molecular biology *342*, 901-912.

Colca, J.R., McDonald, W.G., Waldon, D.J., Leone, J.W., Lull, J.M., Bannow, C.A., Lund, E.T., and Mathews, W.R. (2004). Identification of a novel mitochondrial protein
 10 ("mitoNEET") cross-linked specifically by a thiazolidinedione photoprobe. American journal of physiology Endocrinology and metabolism *286*, E252-260.

Coller, J., and Wickens, M. (2007). Tethered function assays: an adaptable approach to study RNA regulatory proteins. Methods in enzymology *429*, 299-321.

Courty, S., Luccardini, C., Bellaiche, Y., Cappello, G., and Dahan, M. (2006). Tracking
 15 individual kinesin motors in living cells using single quantum-dot imaging. Nano letters *6*, 1491-1495.

Douglass, A.D., and Vale, R.D. (2005). Single-molecule microscopy reveals plasma membrane microdomains created by protein-protein networks that exclude or trap signaling molecules in T cells. Cell *121*, 937-950.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer control of microscopes using microManager. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 14*, Unit14 20.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. ACS synthetic biology *2*, 604-613.

25 Friedman, D.S., and Vale, R.D. (1999). Single-molecule analysis of kinesin motility reveals regulation by the cargo-binding tail domain. Nature cell biology *1*, 293-297.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S.M., Blanchard, J.M., Singer, R.H., and Bertrand, E. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Current biology : CB *13*, 161-167.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442-451.

5 Gordon, G.S., Sitnikov, D., Webb, C.D., Teleman, A., Straight, A., Losick, R., Murray, A.W., and Wright, A. (1997). Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell* 90, 1113-1121.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R., and Pluckthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proceedings of the National Academy of Sciences of the United States of America* 95, 14130-14135.

Hu, J., Lei, Y., Wong, W.K., Liu, S., Lee, K.C., He, X., You, W., Zhou, R., Guo, J.T., Chen, X., *et al.* (2014). Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic acids research* 42, 4375-4390.

15 Huang, C.J., Spinella, F., Nazarian, R., Lee, M.M., Dopp, J.M., and de Vellis, J. (1999). Expression of green fluorescent protein in oligodendrocytes in a time- and level-controllable fashion with a tetracycline-regulated system. *Mol Med* 5, 129-137.

Kearns, N.A., Genga, R.M., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* 141, 219-223.

20 Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., and Choi, S.Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one* 6, e18556.

25 Lecerf, J.M., Shirley, T.L., Zhu, Q., Kazantsev, A., Amersdorfer, P., Housman, D.E., Messer, A., and Huston, J.S. (2001). Human single-chain Fv intrabodies counteract in situ huntingtin aggregation in cellular models of Huntington's disease. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4764-4769.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., *et al.* (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336-340.

- Luo, M., Pang, C.W., Gerken, A.E., and Brock, T.G. (2004). Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import. *Traffic* 5, 847-854.
- Ma, H., Reyes-Gutierrez, P., and Pederson, T. (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences of the United States of America* 110, 21048-21053.
- Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature methods* 10, 977-979.
- Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* 31, 833-838.
- Miyanari, Y., Ziegler-Birling, C., and Torres-Padilla, M.E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. *Nature structural & molecular biology* 20, 1321-1324.
- Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., *et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* 10, 973-976.
- Pettersson, M., and Schaffner, W. (1990). Synergistic activation of transcription by multiple binding sites for NF-kappa B even in absence of co-operative factor binding to DNA. *Journal of molecular biology* 214, 373-380.
- Pillai, R.S., Artus, C.G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA* 10, 1518-1525.
- Pique, M., Lopez, J.M., Foissac, S., Guigo, R., and Mendez, R. (2008). A combinatorial code for CPE-mediated translational control. *Cell* 132, 434-448.
- Rice, S., Lin, A.W., Safer, D., Hart, C.L., Naber, N., Carragher, B.O., Cain, S.M., Pechatnikova, E., Wilson-Kubalek, E.M., Whittaker, M., *et al.* (1999). A structural change in the kinesin motor protein that drives motility. *Nature* 402, 778-784.
- Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335, 563-564.

Stout, J.R., Yount, A.L., Powers, J.A., Leblanc, C., Ems-McClung, S.C., and Walczak, C.E. (2011). Kif18B interacts with EB1 and controls astral microtubule length during mitosis. *Molecular biology of the cell* 22, 3070-3080.

5 Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676.

Tanenbaum, M.E., Macurek, L., van der Vaart, B., Galli, M., Akhmanova, A., and Medema, R.H. (2011). A complex of Kif18b and MCAK promotes microtubule depolymerization and is negatively regulated by Aurora kinases. *Current biology : CB* 21, 1356-1365.

10 Waterman-Storer, C.M., Desai, A., Bulinski, J.C., and Salmon, E.D. (1998). Fluorescent speckle microscopy, a method to visualize the dynamics of protein assemblies in living cells. *Current biology : CB* 8, 1227-1230.

Worn, A., Auf der Maur, A., Escher, D., Honegger, A., Barberis, A., and Pluckthun, A. (2000). Correlation between in vitro stability and in vivo performance of anti-GCN4 intrabodies as cytoplasmic inhibitors. *The Journal of biological chemistry* 275, 2795-2803.

15 Wozniak, M.J., Bola, B., Brownhill, K., Yang, Y.C., Levakova, V., and Allan, V.J. (2009). Role of kinesin-1 and cytoplasmic dynein in endoplasmic reticulum movement in VERO cells. *Journal of cell science* 122, 1979-1989.

WHAT IS CLAIMED IS:

1 1. A composition for recruiting one or more effector domains to a
 2 polypeptide of interest in a cell or cell extract, the composition comprising:
 3 the polypeptide of interest fused to a multimerized epitope; and
 4 an affinity agent fusion protein, wherein the affinity agent fusion protein
 5 comprises:

6 an affinity domain that specifically binds the epitope; and
 7 the effector domain.

1 2. The composition of claim 1, wherein the effector domain is an enzyme
 2 (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a
 3 phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent
 4 protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional
 5 activator, or a transcriptional repressor.

1 3. The composition of claim 1, wherein the multimerized epitope contains
 2 multiple copies of an epitope of at least 5 amino acids in length.

1 4. The composition of claim 1, 2, or 3, wherein the multimerized epitope
 2 contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or
 3 more copies of the epitope.

1 5. The composition of claim 1, 2, 3, or 4, wherein each epitope of the
 2 multimerized epitope is separated by a linker.

1 6. The composition of claim 5, wherein the linker is at least 5 amino acids
 2 in length.

1 7. The composition of claim 1, wherein the multimerized epitope
 2 comprises SEQ ID NO:1 or 2 and SEQ ID NO:2 or 3.

1 8. The composition of claim 7, wherein the multimerized epitope
 2 comprises:

3 at least one copy of SEQ ID NO:3 or 4; and

4 at least:

5 two copies of SEQ ID NO:1;

6 two copies of SEQ ID NO:2; or
7 one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

1 9. The composition of claim 1, wherein the affinity domain is an antibody
2 or a single-chain antibody that specifically binds the epitope.

1 10. The composition of claim 9, wherein the antibody or single-chain
2 antibody is stable under reducing conditions of an intracellular environment or a cellular
3 extract.

1 11. The composition of claim 10, wherein the affinity domain comprises a
2 single chain antibody of SEQ ID NO:5.

1 12. The composition of claim 1, wherein the effector domain comprises a
2 fluorophore.

1 13. The composition of claim 12, wherein the fluorophore is a fluorescent
2 protein.

1 14. The composition of claim 1, wherein the affinity domain is a single-
2 chain antibody fused to a solubility enhancing domain.

1 15. The composition of claim 14, wherein the solubility enhancing domain
2 is a GB1 polypeptide (SEQ ID NO:6).

1 16. The composition of claim 14, wherein the solubility enhancing domain
2 is a solubility enhanced effector domain.

1 17. The composition of claim 16, wherein the solubility enhanced effector
2 domain is superfolder-GFP (SEQ ID NO:7).

1 18. The composition of claim 14, wherein the affinity domain is fused to
2 an N-terminal solubility enhancing domain and a C-terminal solubility enhancing domain.

1 19. The composition of claim 18, wherein the N-terminal solubility
2 enhancing domain is a GB1 polypeptide (SEQ ID NO:6) and the C-terminal solubility
3 enhancing domain is superfolder-GFP (SEQ ID NO:7).

1 20. The composition of claim 18, wherein the N-terminal solubility
2 enhancing domain is superfolder-GFP (SEQ ID NO:7) and the C-terminal solubility
3 enhancing domain is a GB1 polypeptide (SEQ ID NO:6).

1 21. The composition of claim 18, wherein the affinity agent fusion protein
2 comprises the amino acid sequence of SEQ ID NO:8.

1 22. The composition of claim 1, wherein the polypeptide of interest
2 comprises dCas9 (SEQ ID NO:9).

1 23. The composition of claim 1, wherein the multimerized epitope
2 comprises SEQ ID NO: 10, 11, or 12.

1 24. A cell or cell extract comprising any one of the compositions of claims
2 1 - 23.

1 25. An isolated polynucleotide encoding SEQ ID NO:5 or SEQ ID NO:8.

1 26. An isolated polynucleotide encoding a polypeptide of interest fused to
2 a multimerized epitope, wherein the multimerized epitope contains multiple copies of an
3 epitope of at least 5 amino acids in length.

1 27. The isolated polynucleotide of claim 26, wherein the multimerized
2 epitope contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,
3 24, or more copies of the epitope.

1 28. The isolated polynucleotide of claim 26, wherein each epitope of the
2 multimerized epitope is separated by a linker.

1 29. The isolated polynucleotide of claim 26, wherein the multimerized
2 epitope comprises:

3 at least one copy of SEQ ID NO:3 or 4; and

4 at least:

5 two copies of SEQ ID NO:1;

6 two copies of SEQ ID NO:2; or

7 one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

1 30. A host cell transformed with one or more expression cassettes, the
2 expression cassettes encoding:

3 (i) any one of the polypeptides fused to a multimerized epitope of claims 1 -
4 23; and/or

5 (ii) any one of the affinity agent fusion proteins of claims 1 - 23.

1 31. The host cell of claim 30, wherein one or more of the expression
2 cassettes are inducible.

1 32. The host cell of claim 31, wherein the host cell comprises a tet-
2 transactivator, and the host cell further comprises a tet-inducible expression cassette.

1 33. A kit comprising:

2 (i) an expression cassette comprising a heterologous promoter operably linked
3 to a polynucleotide encoding an affinity agent fusion protein, wherein the affinity agent
4 fusion protein comprises:

5 an affinity domain that specifically binds the epitope; and

6 a effector domain; and/or

7 (ii) an expression cassette encoding:

8 (a) a heterologous promoter, a cloning site, and a multimerized
9 epitope, wherein the cloning site is configured to allow cloning of a polypeptide of
10 interest operably linked to the promoter and fused to the multimerized epitope; or

11 (b) a heterologous promoter operably linked to a polypeptide of
12 interest fused to a multimerized epitope.

1 34. The kit of claim 33, wherein the effector domain is an enzyme (*e.g.*, a
2 nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a phosphatase, a
3 ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent protein (*e.g.*, a
4 green fluorescent protein), a transcriptional enhancer, a transcriptional activator, or a
5 transcriptional repressor.

1 35. The kit of claim 33, wherein the affinity domain comprises the single
2 chain antibody of SEQ ID NO:5.

1 36. The kit of claim 33, wherein the affinity agent fusion protein comprises
2 the amino acid sequence of SEQ ID NO:8.

1 37. The kit of any one of claims 33 - 36, wherein the multimerized epitope
2 contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or
3 more copies of the epitope.

1 38. The kit of any one of claims 33 - 37 wherein each epitope of the
2 multimerized epitope is separated by a linker.

1 39. The kit of claim 38, wherein the linker is at least 5 amino acids in
2 length.

1 40. The kit of claim 33, wherein the multimerized epitope comprises SEQ
2 ID NO:1 or 2 and SEQ ID NO:3 or 4.

1 41. The kit of claim 40, wherein the multimerized epitope comprises:
2 at least one copy of SEQ ID NO:3 or 4; and
3 at least:
4 two copies of SEQ ID NO:1;
5 two copies of SEQ ID NO:2; or
6 one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

1 42. The kit of claim 33, wherein the kit comprises an expression cassette
2 encoding a small guide RNA (sgRNA) or an sgRNA scaffold.

1 43. The kit of claim 45, wherein the expression cassette encoding an
2 sgRNA scaffold comprises from 5' to 3':
3 a 5' promoter;
4 a cloning site;
5 a 5' hairpin region;
6 a 3' hairpin region; and
7 a transcription termination region, wherein the cloning site is configured to
8 operably link a binding region to the 5' promoter and the 3' regions, when the binding region
9 is cloned into the cloning site.

1 44. A method for recruiting one or more effector domains to a polypeptide
2 of interest in a cell or cell extract, the method comprising:

3 contacting the cell or cell extract with any one of the compositions of claims 1
4 - 23 under conditions suitable to permit binding of multiple copies of the affinity agent fusion
5 protein to the multimerized epitope fused to the polypeptide of interest, thereby bringing
6 multiple copies of the effector domain in proximity to the polypeptide of interest.

1 45. The method of claim 44, wherein the method comprises detecting the
2 effector domain.

1 46. The method of claim 45, wherein the detecting comprises directing
2 incident light into the cell or cell extract, thereby inducing fluorescence from the effector
3 domain and detecting the fluorescence.

1 47. The method of claim 45, wherein the detecting comprises measuring
2 upregulation or downregulation of transcription at or near a target binding site of the sgRNA.

1 48. The method of claim 44, wherein the method comprises binding at
2 least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies
3 of the affinity agent fusion protein to the multimerized epitope, thereby binding said number
4 of copies of the effector domain to the polypeptide of interest.

1 49. The method of claim 44, wherein the method comprises single
2 molecule detection of the polypeptide of interest.

1 50. A composition for site-specific transcriptional activation of a genetic
2 element comprising:

3 a dCas9 domain fused to a multimerized epitope; and
4 an affinity agent fusion protein, wherein the affinity agent fusion protein
5 comprises:
6 an affinity domain that specifically binds the epitope; and
7 a transcriptional activator domain.

1 51. The composition of claim 50, wherein the multimerized epitope
2 contains multiple copies of an epitope of at least 5 amino acids in length.

1 52. The composition of claim 50 or 51, wherein the multimerized epitope
2 contains at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or
3 more copies of the epitope.

1 53. The composition of claim 50, 51, or 52, wherein each epitope of the
2 multimerized epitope is separated by a linker of at least 5 amino acids in length.

1 54. The composition of claim 50, wherein the multimerized epitope
2 comprises SEQ ID NO:1 or 2 and SEQ ID NO: 3 or 4.

1 55. The composition of claim 54, wherein the multimerized epitope
2 comprises:

3 at least one copy of SEQ ID NO:3 or 4; and

4 at least:

5 two copies of SEQ ID NO:1;

6 two copies of SEQ ID NO:2; or

7 one copy of SEQ ID NO:1 and at least one copy of SEQ ID NO:2.

1 56. The composition of claim 50, wherein the dCas9 fused to a
2 multimerized epitope comprises the amino acid sequence of SEQ ID NO:9.

1 57. The composition of claim 56, wherein the dCas9 fused to a
2 multimerized epitope comprises the amino acid sequence of SEQ ID NO:9 and the amino
3 acid sequence of SEQ ID NO:10, 11, or 12.

1 58. The composition of claim 57, wherein the dCas9 fused to a
2 multimerized epitope comprises the amino acid sequence of SEQ ID NO:13.

1 59. The composition of claim 50, wherein the affinity domain is an
2 antibody or a single-chain antibody that specifically binds the epitope.

1 60. The composition of claim 59, wherein the antibody or single-chain
2 antibody is stable under the reducing conditions of a cell or a cellular extract.

1 61. The composition of claim 50, wherein the transcriptional activator
2 domain comprises a VP16 domain.

1 62. The composition of claim 61, wherein the transcriptional activator
2 domain comprises at least 2, 3, 4, or more VP16 domains.

1 63. The composition of claim 50, wherein the affinity domain is a single-
2 chain antibody fused to solubility enhancing domain.

1 64. The composition of claim 63, wherein the solubility enhancing domain
2 is a GB1 polypeptide (SEQ ID NO:6).

1 65. The composition of any one of claims 59 - 64, wherein the affinity
2 agent fusion protein comprises SEQ ID NO:5.

1 66. The composition of any one of claims 50 - 65, wherein the
2 composition further comprises a small guide RNA (sgRNA).

1 67. A host cell transformed with one or more expression cassettes, the
2 expression cassettes encoding:

- 3 (i) an sgRNA;
4 (ii) a dCas9 fused to a multimerized epitope; and/or
5 (iii) an affinity agent fusion protein of any one of the compositions of claims
6 50 - 66.

1 68. The host cell of claim 67, wherein one or more of the expression
2 cassettes are inducible.

1 69. The host cell of claim 68, wherein the host cell comprises a tet-
2 transactivator, and the host cell further comprises a tet-inducible expression cassette encoding
3 dCas9 fused to a multimerized epitope.

1 70. A kit for activating transcription of a genetic element, the kit
2 comprising one or more expression cassettes encoding:

- 3 (i) a small guide RNA (sgRNA) or an sgRNA scaffold;
4 (ii) a dCas9 fused to a multimerized epitope; and/or
5 (iii) an affinity agent fusion protein of any one of the compositions of claims
6 50 - 66.

1 71. The kit of claim 70, wherein the kit comprises an expression cassette
2 encoding a small guide RNA (sgRNA) or an sgRNA scaffold.

1 72. The kit of claim 71, wherein the expression cassette encoding an
2 sgRNA scaffold comprises from 5' to 3':

3 a 5' promoter;
4 a cloning site;
5 a 5' hairpin region;
6 a 3' hairpin region; and
7 a transcription termination region, wherein the cloning site is configured to
8 operably link a binding region to the 5' promoter and the 3' regions, when the binding region
9 is cloned into the cloning site.

1 73. A method of site-specific transcriptional activation of a genetic
2 element in a cell or cell extract comprising:

3 contacting the cell or cell extract with any one of the compositions of claims
4 50 - 66, wherein the composition further comprises a small guide RNA (sgRNA) that
5 specifically binds the genetic element, or a region proximal to the genetic element, under
6 conditions suitable to permit the binding of the sgRNA to the genetic element or region, the
7 binding of the sgRNA to the dCas9 domain fused to the multimerized epitope, and the
8 binding of multiple copies of the affinity agent fusion protein to the multimerized epitope,
9 thereby bringing multiple copies of the transcriptional activator domain in proximity to the
10 genetic element.

1 74. The method of claim 73, wherein the method comprises binding at
2 least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more copies
3 of the affinity agent fusion protein to the multimerized epitope, thereby bringing said number
4 of copies of the transcription activator domain in proximity to the genetic element.

1 75. A composition comprising dCas9 fused to a multimerized effector
2 domain.

1 76. The composition of claim 75, wherein the multimerized effector
2 domain comprises two or more (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) copies of an effector
3 domain.

1 77. The composition of claim 76, wherein the effector domain is an
2 enzyme (*e.g.*, a nuclease, a methylase, a demethylase, an acetylase, a deacetylase, a kinase, a
3 phosphatase, a ubiquitinase, a deubiquitinase, a luciferase, or a peroxidase), a fluorescent
4 protein (*e.g.*, a green fluorescent protein), a transcriptional enhancer, a transcriptional
5 activator, or a transcriptional repressor.

1 78. A kit comprising one or more expression cassettes encoding:
2 (i) a dCas9 fused to a multimerized effector domain of any one of claims 75 -
3 77; and optionally
4 (ii) a small guide RNA (sgRNA), or an sgRNA scaffold.

1 79. A method for site-specific recruitment of effector domains to a genetic
2 element in a cell or cell extract comprising:
3 contacting the cell or cell extract with any one of the compositions of claims
4 75 - 77, wherein the composition further comprises a small guide RNA (sgRNA) that
5 specifically binds the genetic element, or a region proximal to the genetic element, under
6 conditions suitable to permit the binding of the sgRNA to the genetic element or region, and
7 the binding of the sgRNA to the dCas9 domain fused to the multimerized effector domain,
8 thereby bringing multiple copies of the effector domain in proximity to the genetic element.

1 / 22

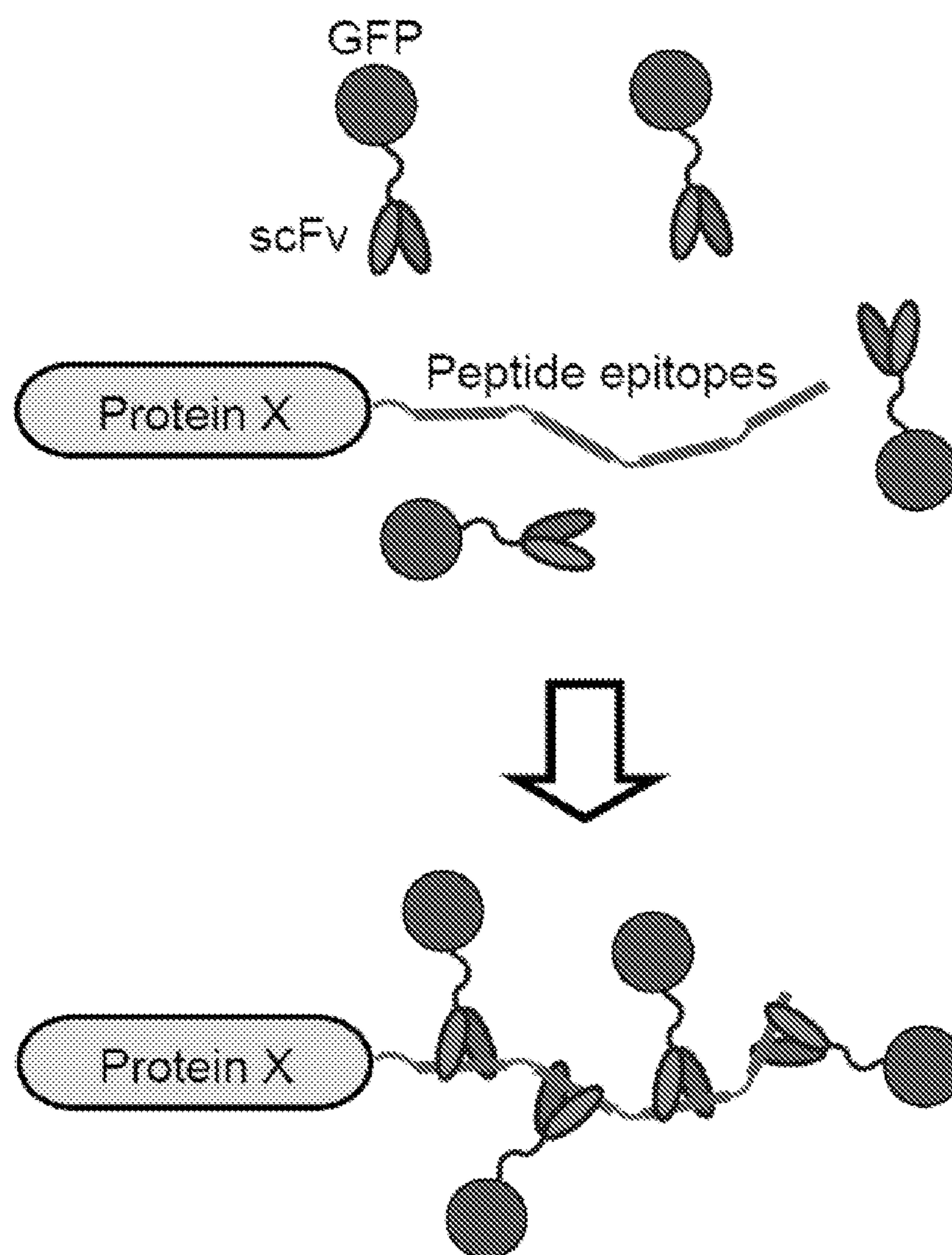


FIG. 1A

2 / 22

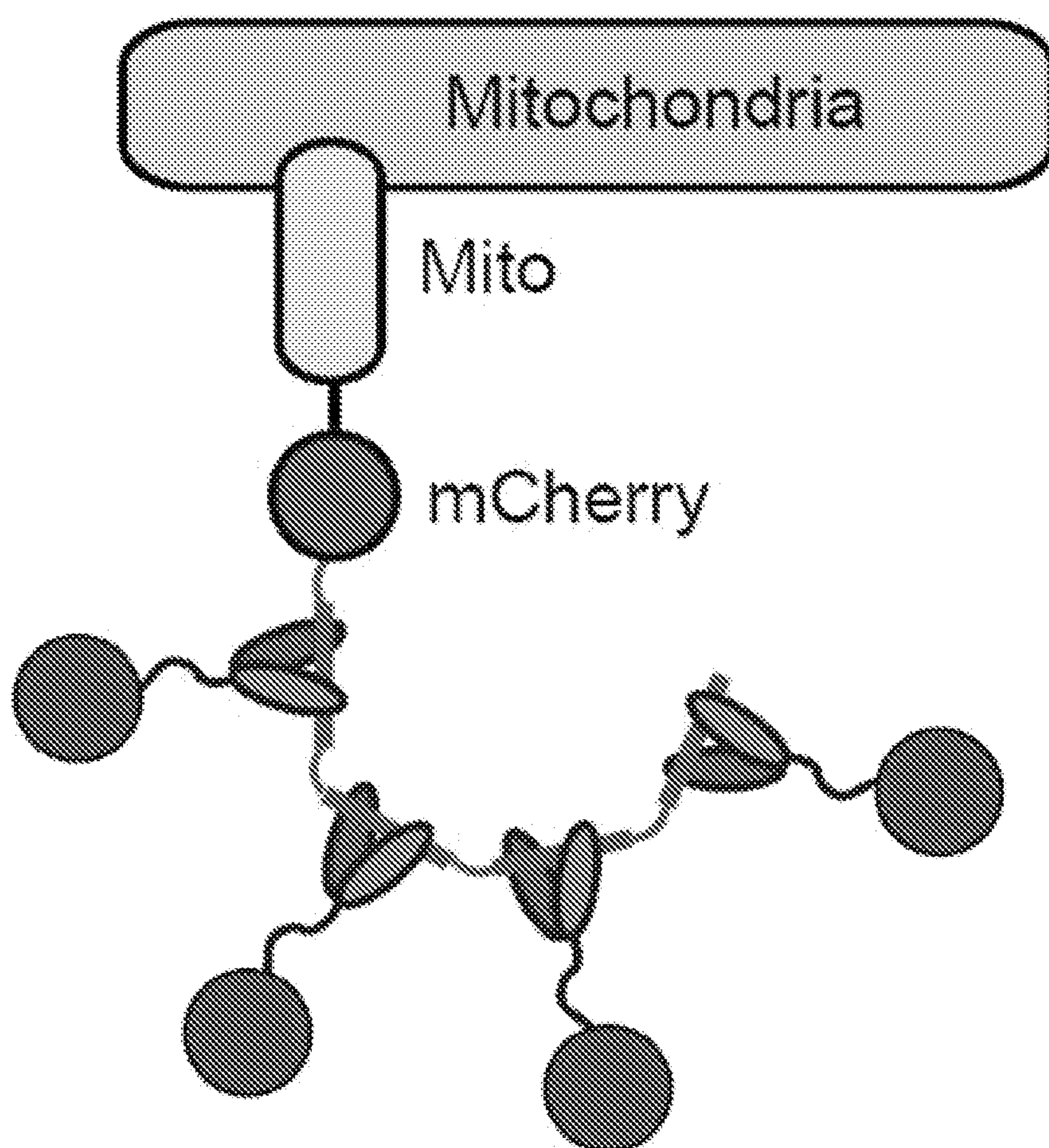


FIG. 1B

3 / 22

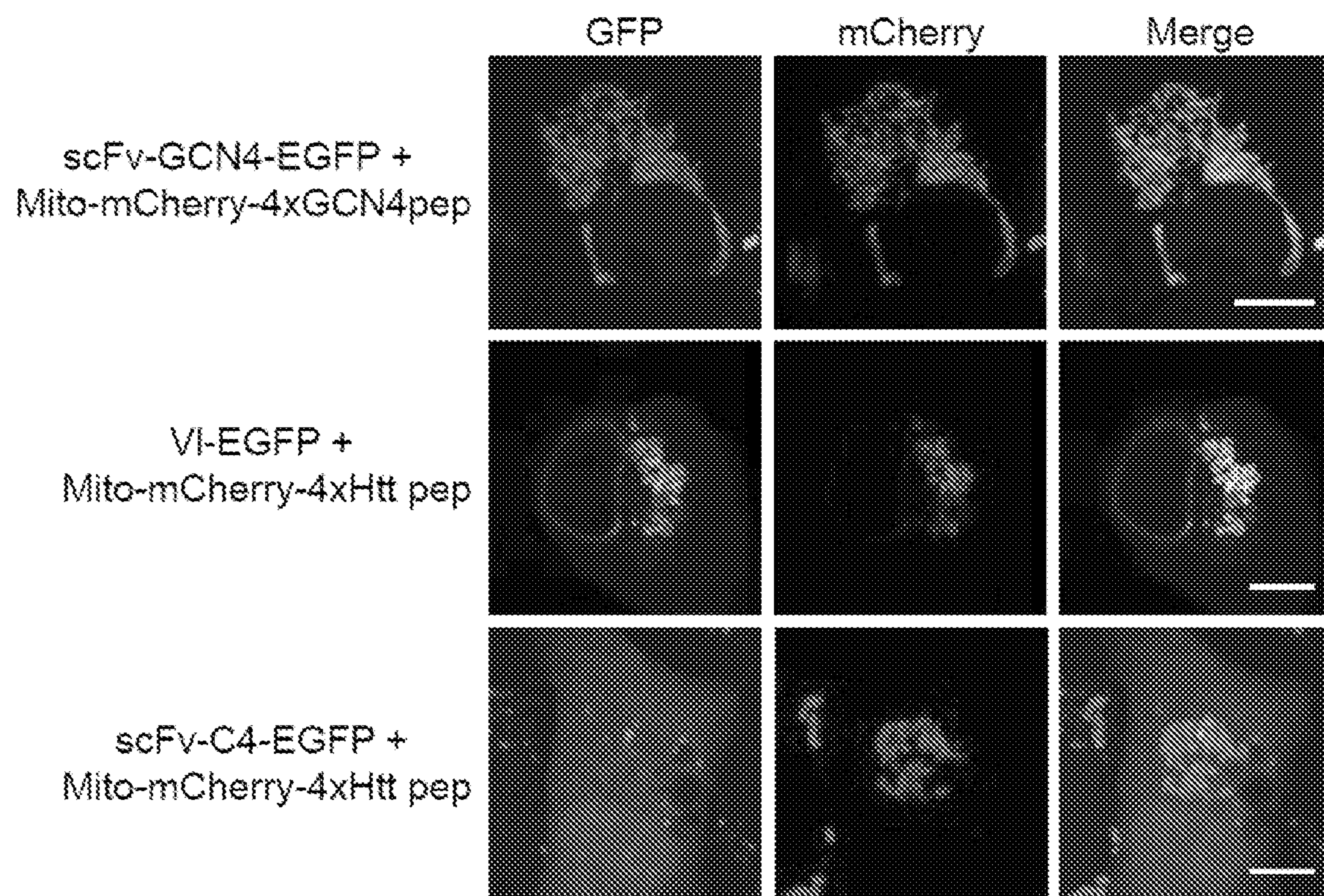


FIG. 1C

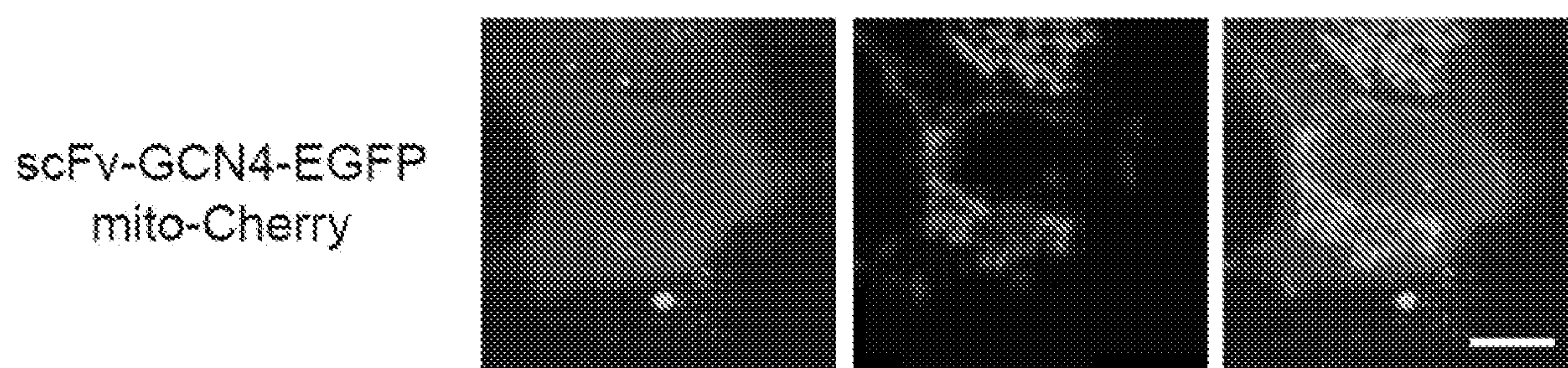


FIG. 1D

4 / 22

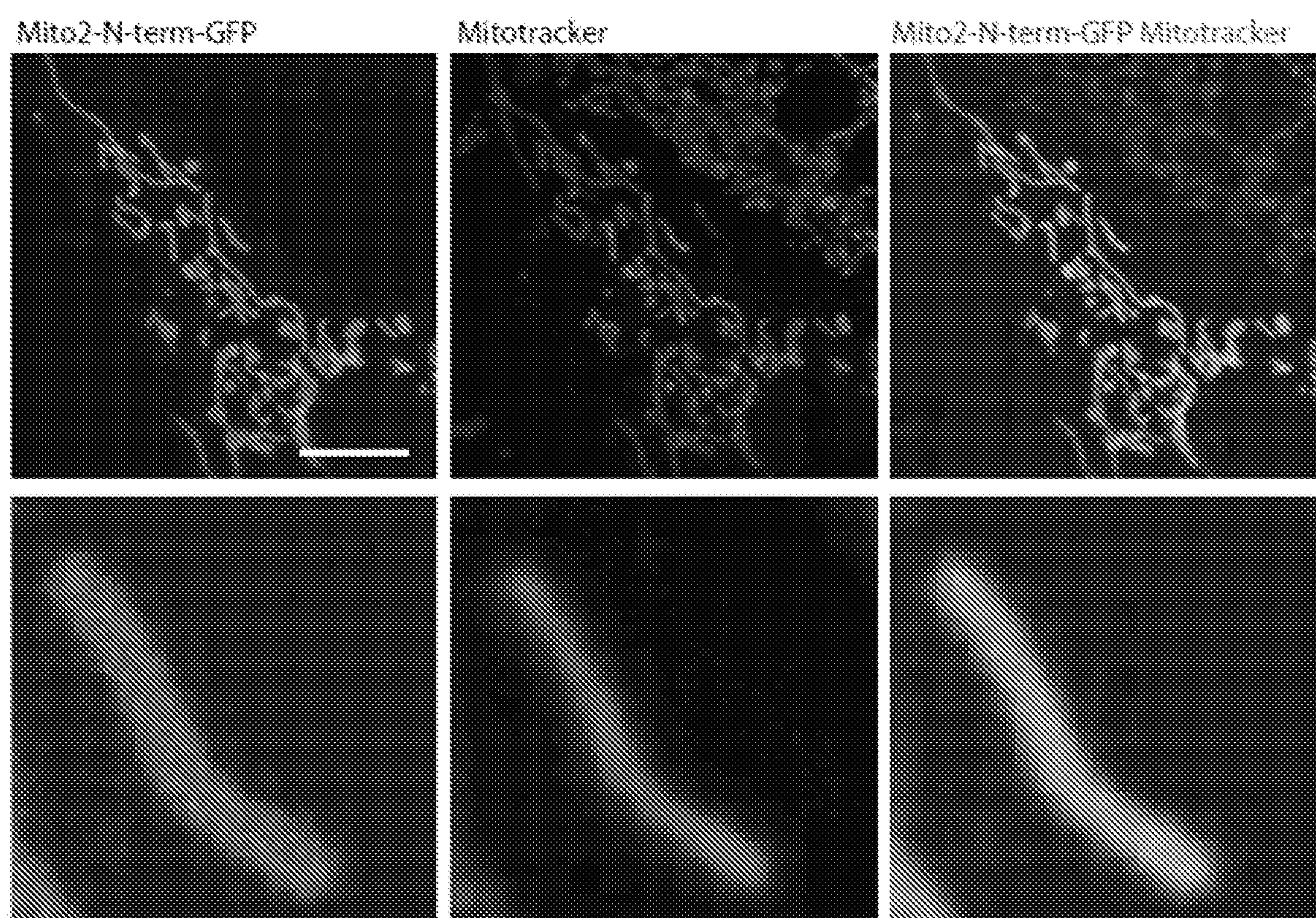


FIG. 2

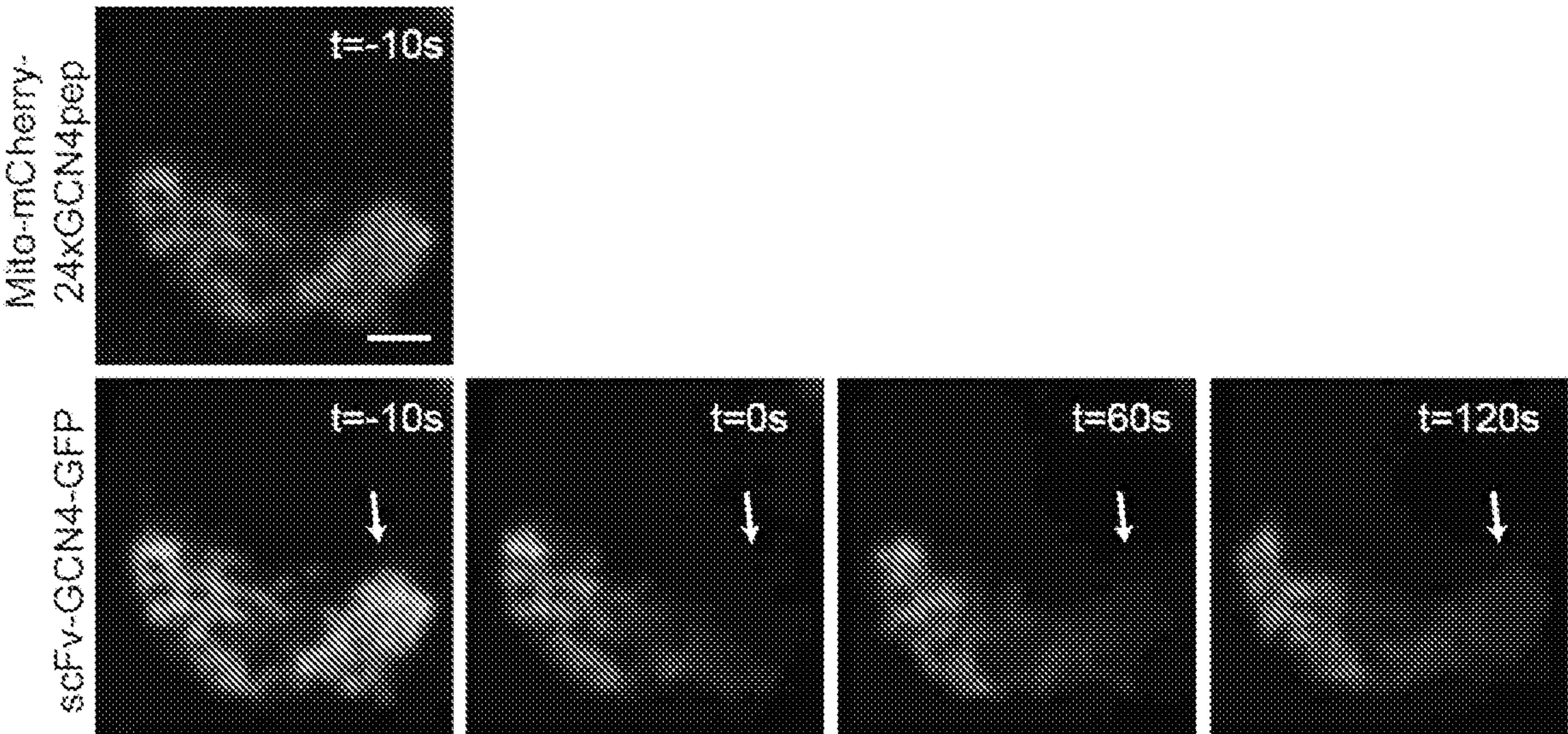


FIG. 3A

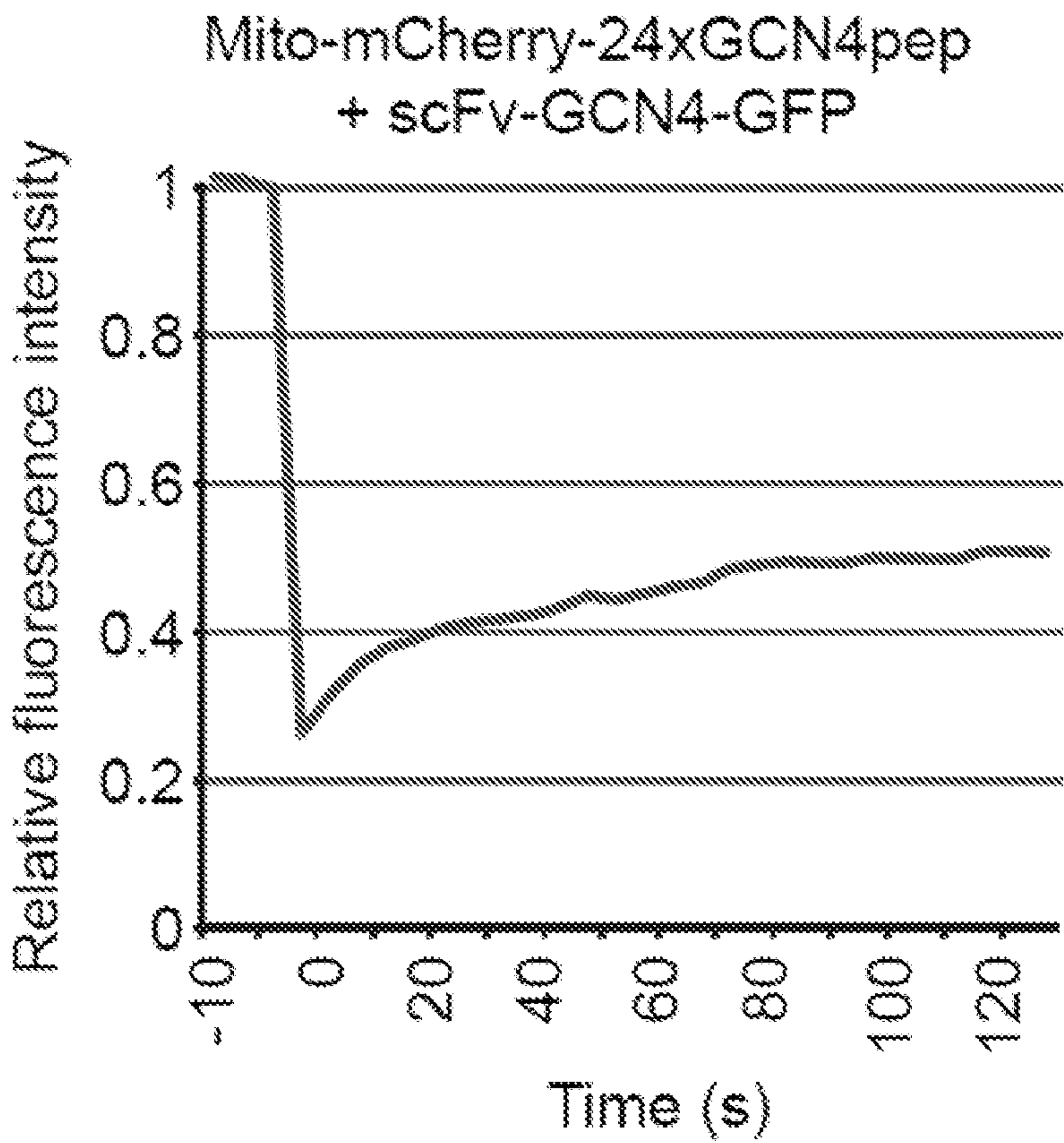


FIG. 3B

6 / 22

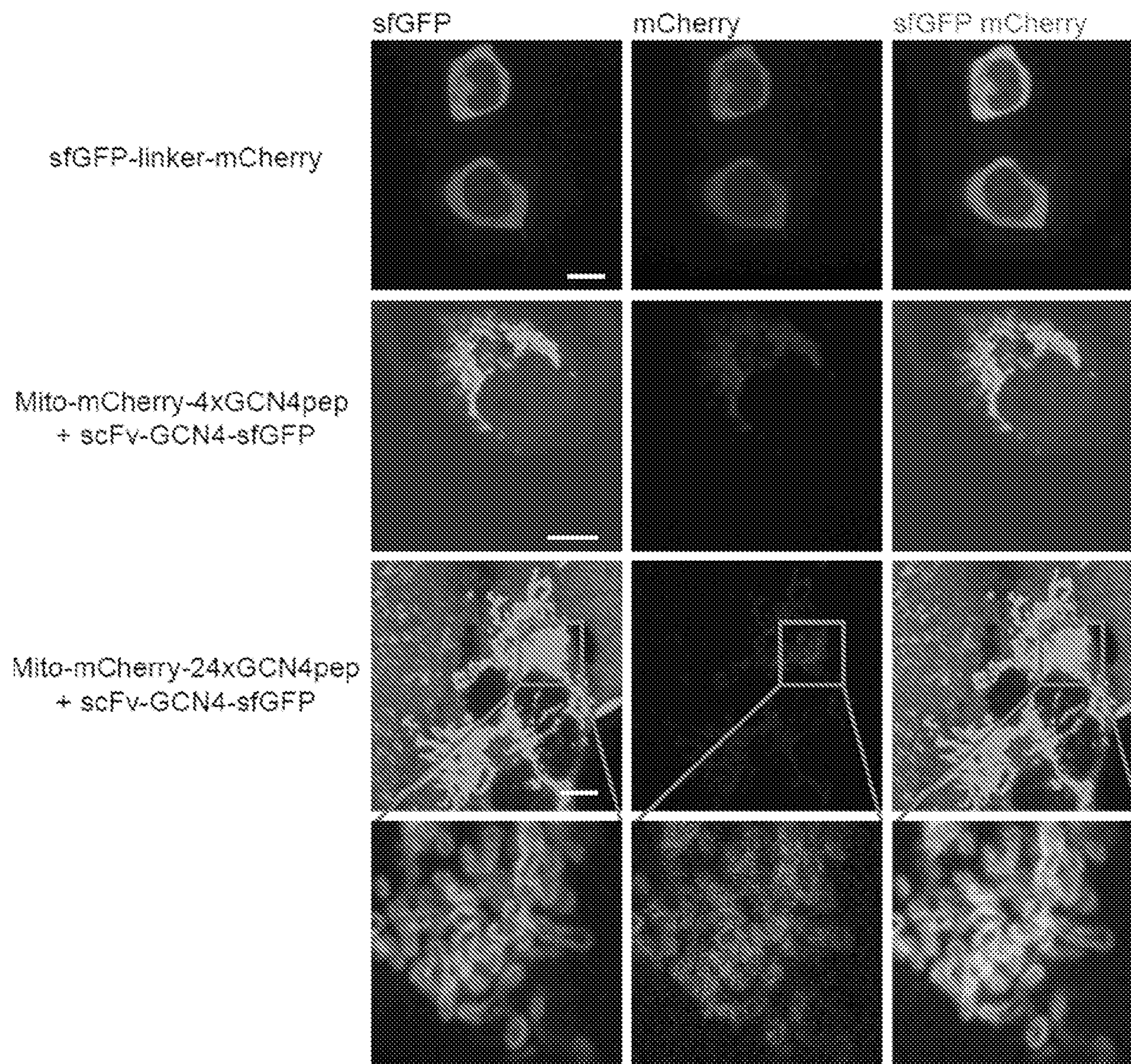


FIG. 3C

7 / 22

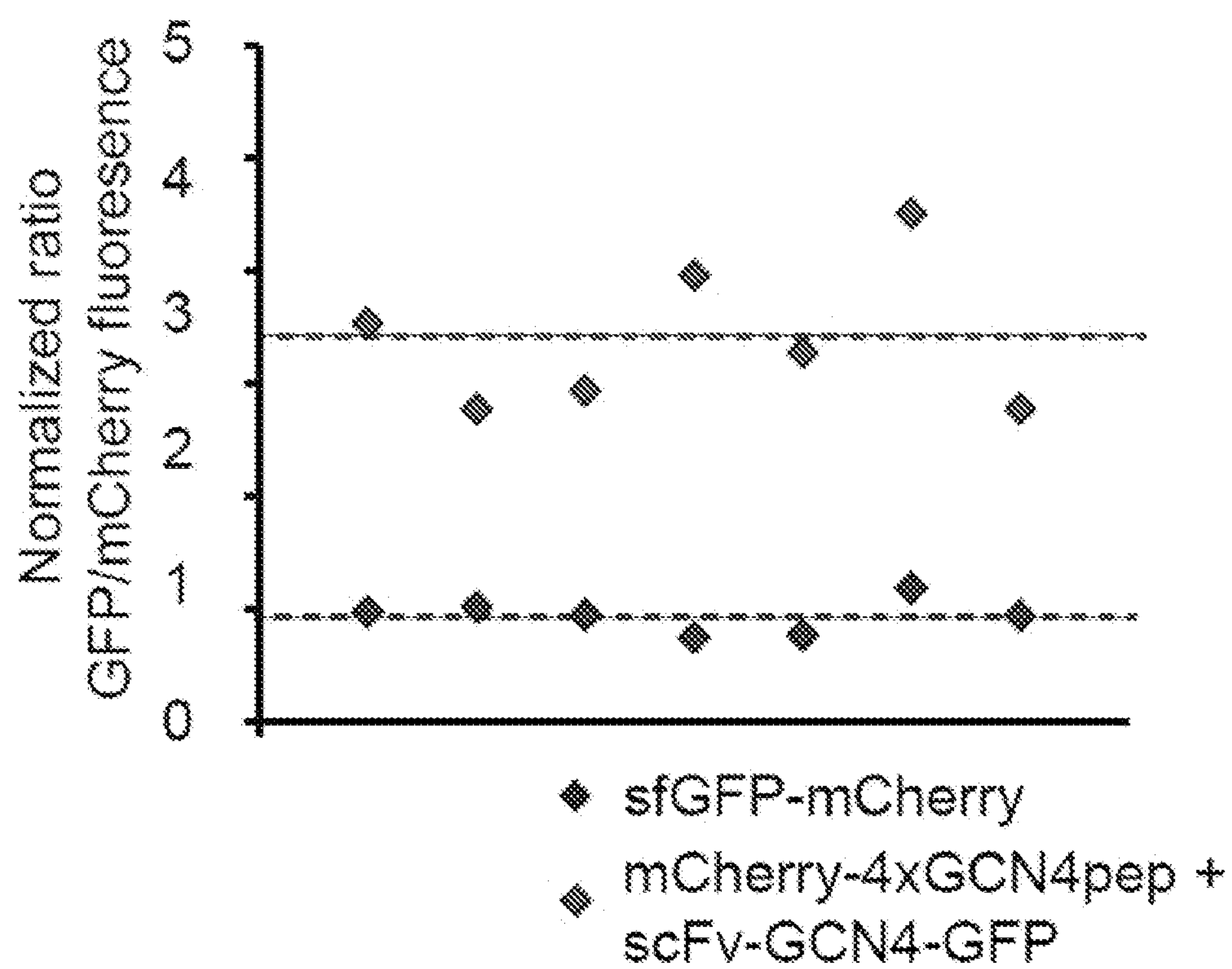


FIG. 3D

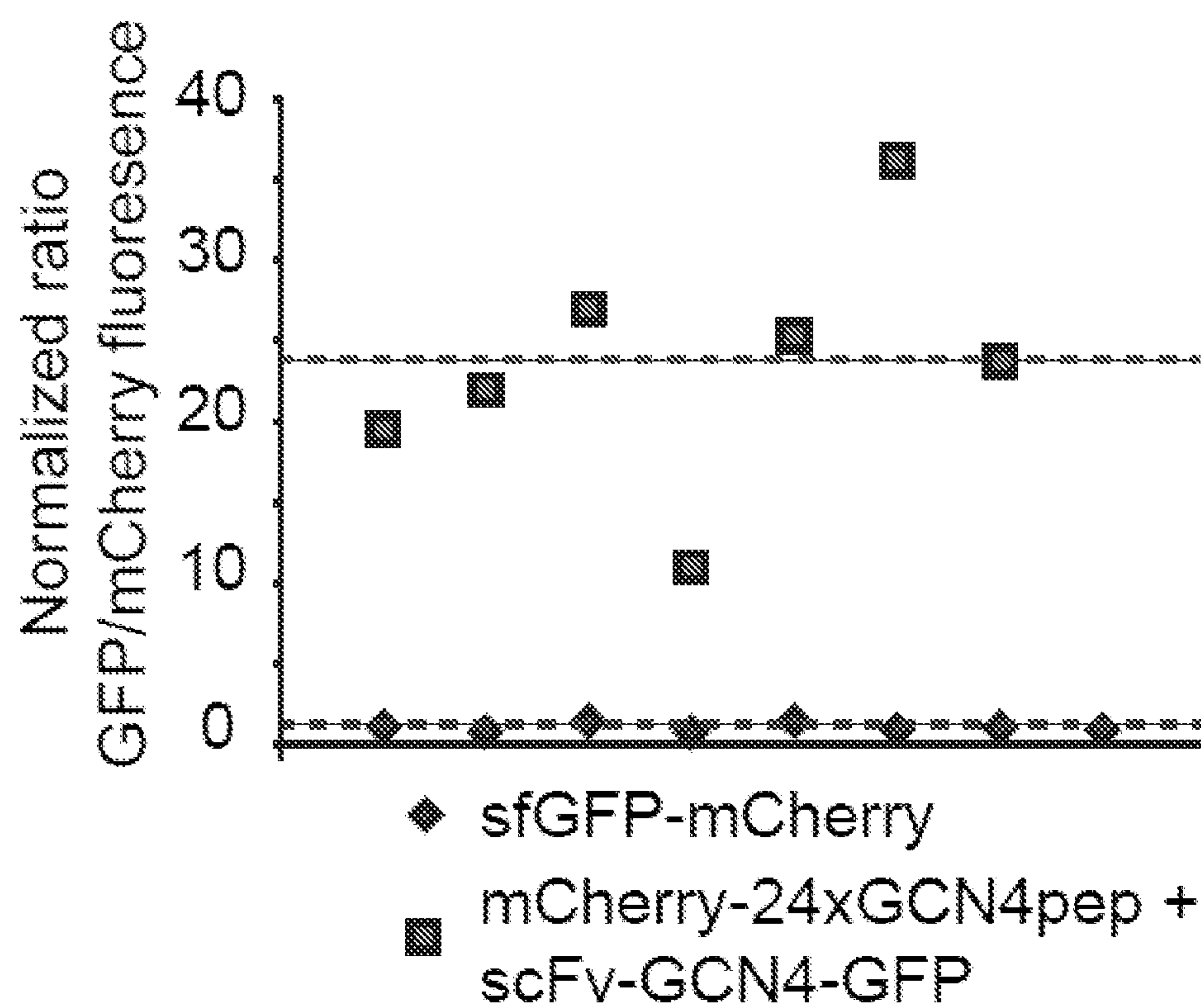


FIG. 3E

8 / 22

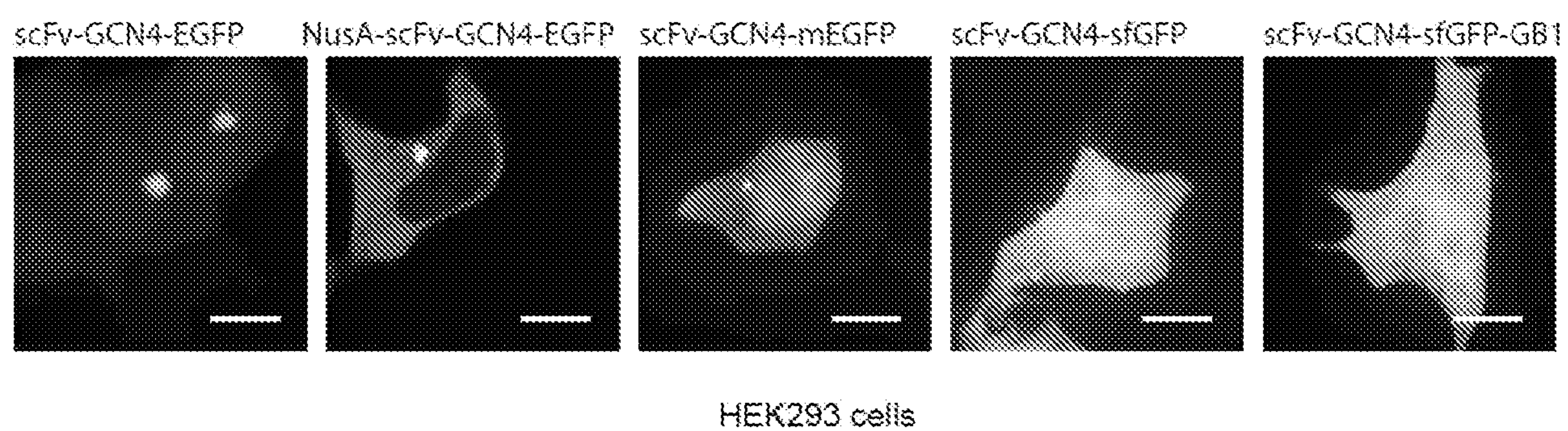


FIG. 4A

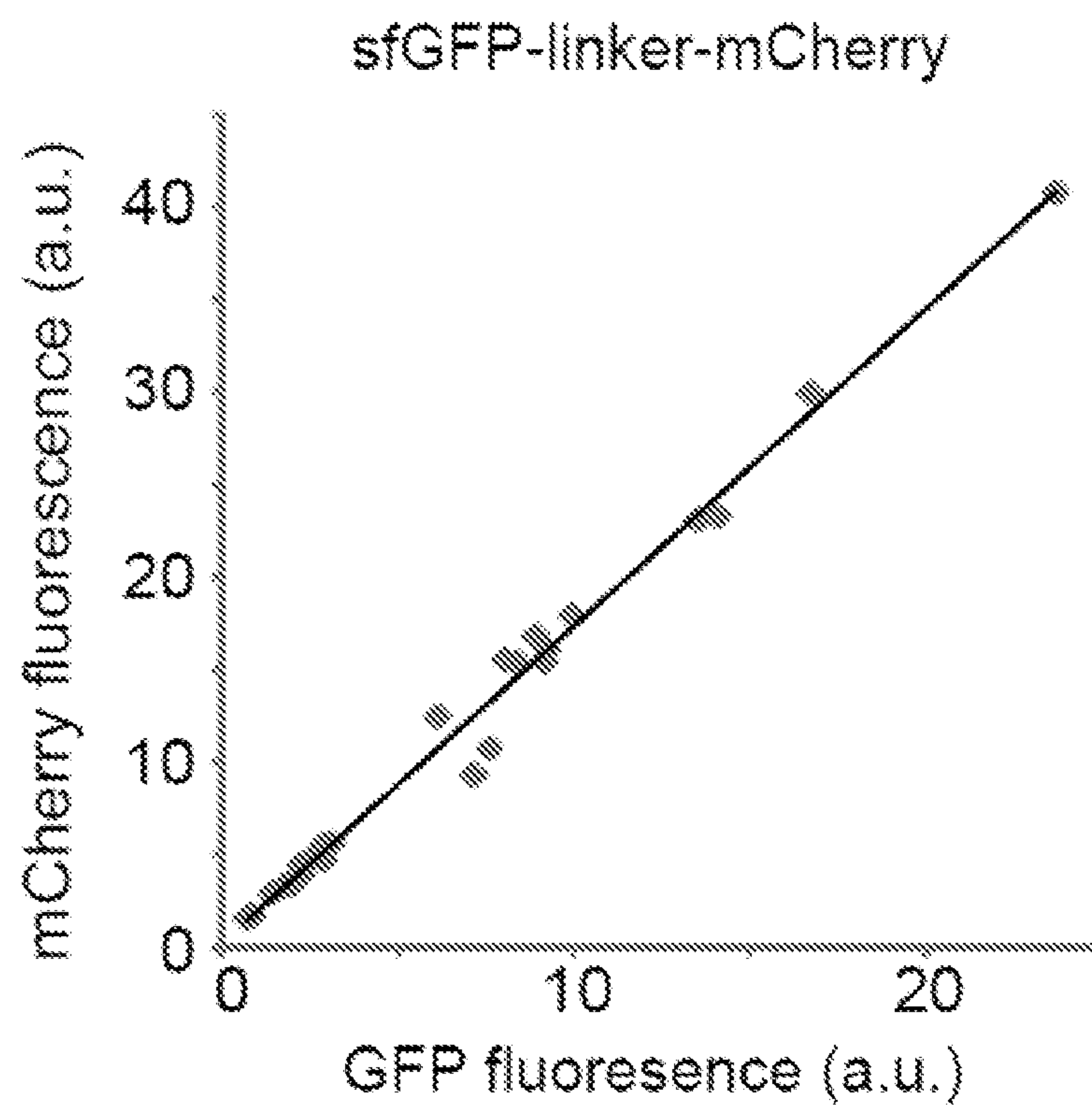


FIG. 4B

9 / 22

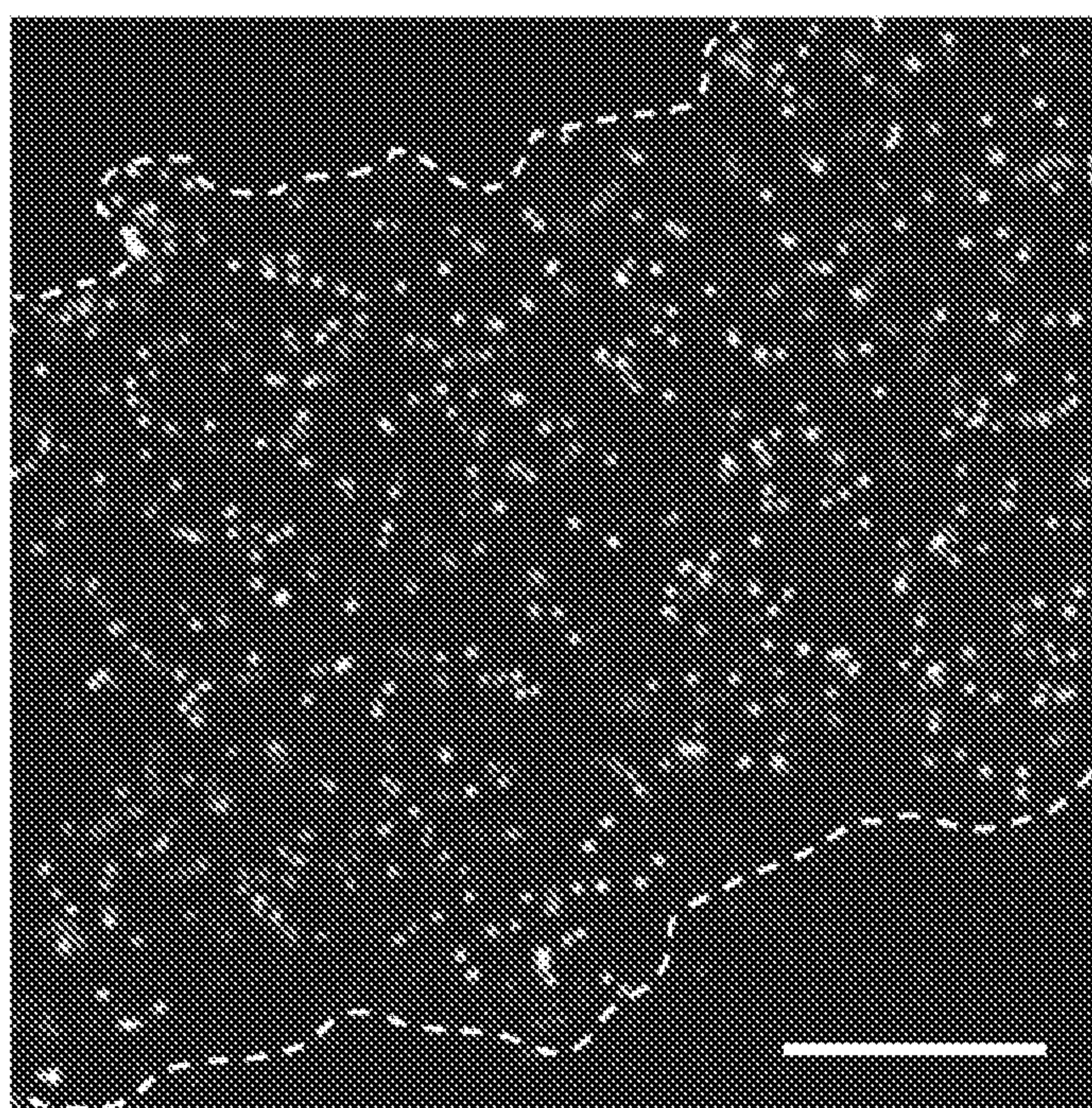
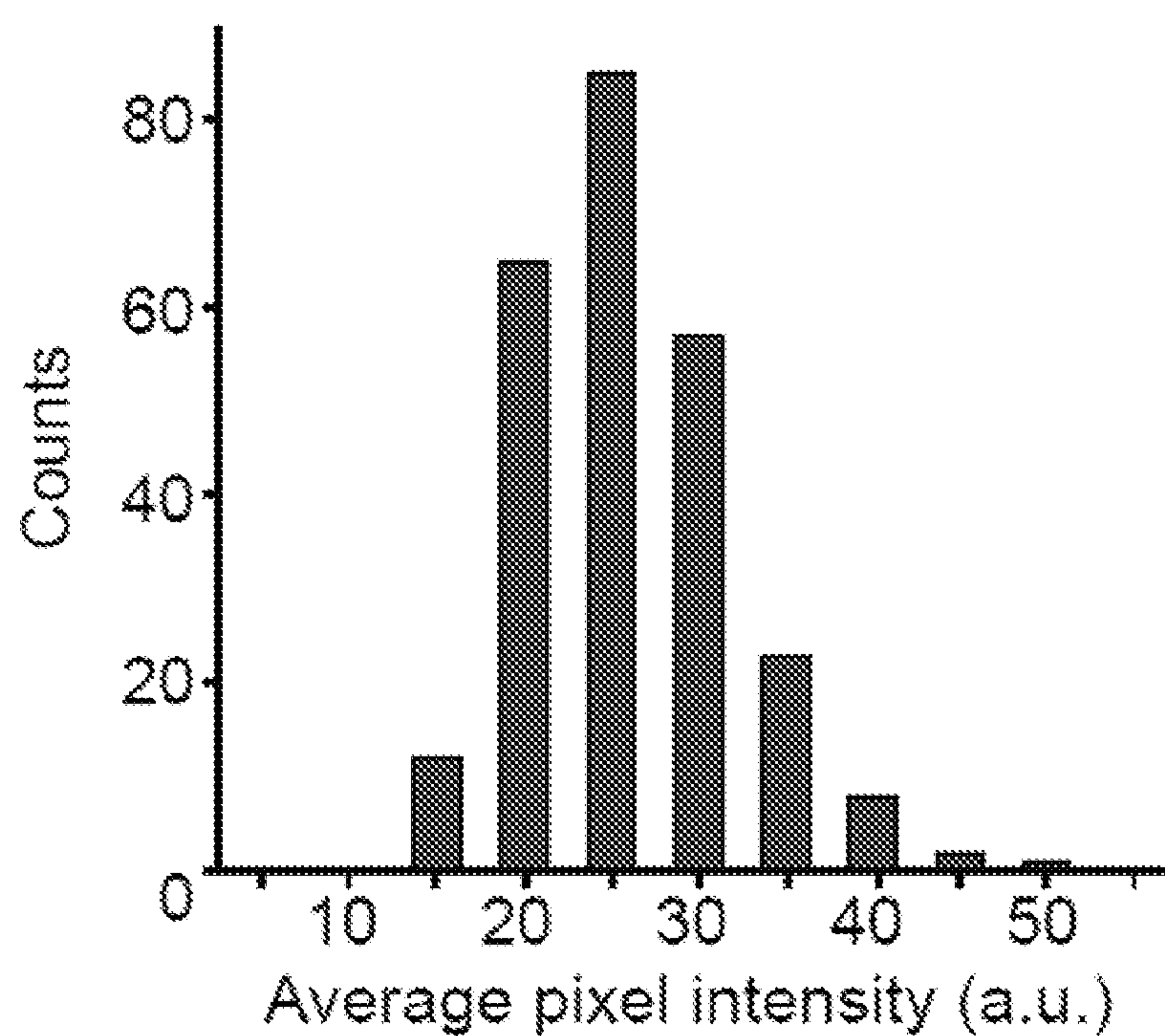
IFP-SunTag_{24X}-CAAX-GFPFluorescence intensity of
single SunTag-GFP foci

FIG. 5A

10 / 22

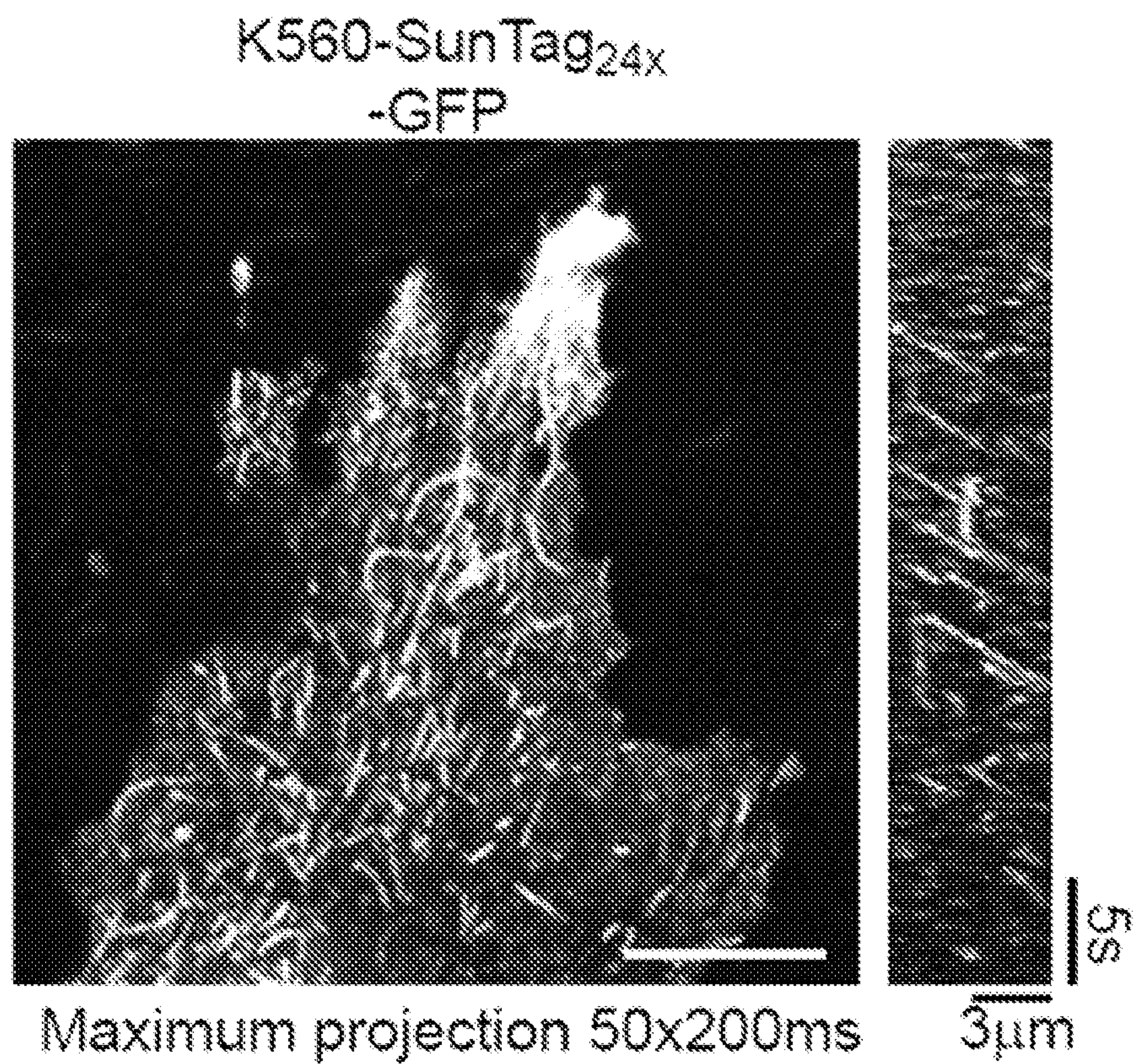


FIG. 5B

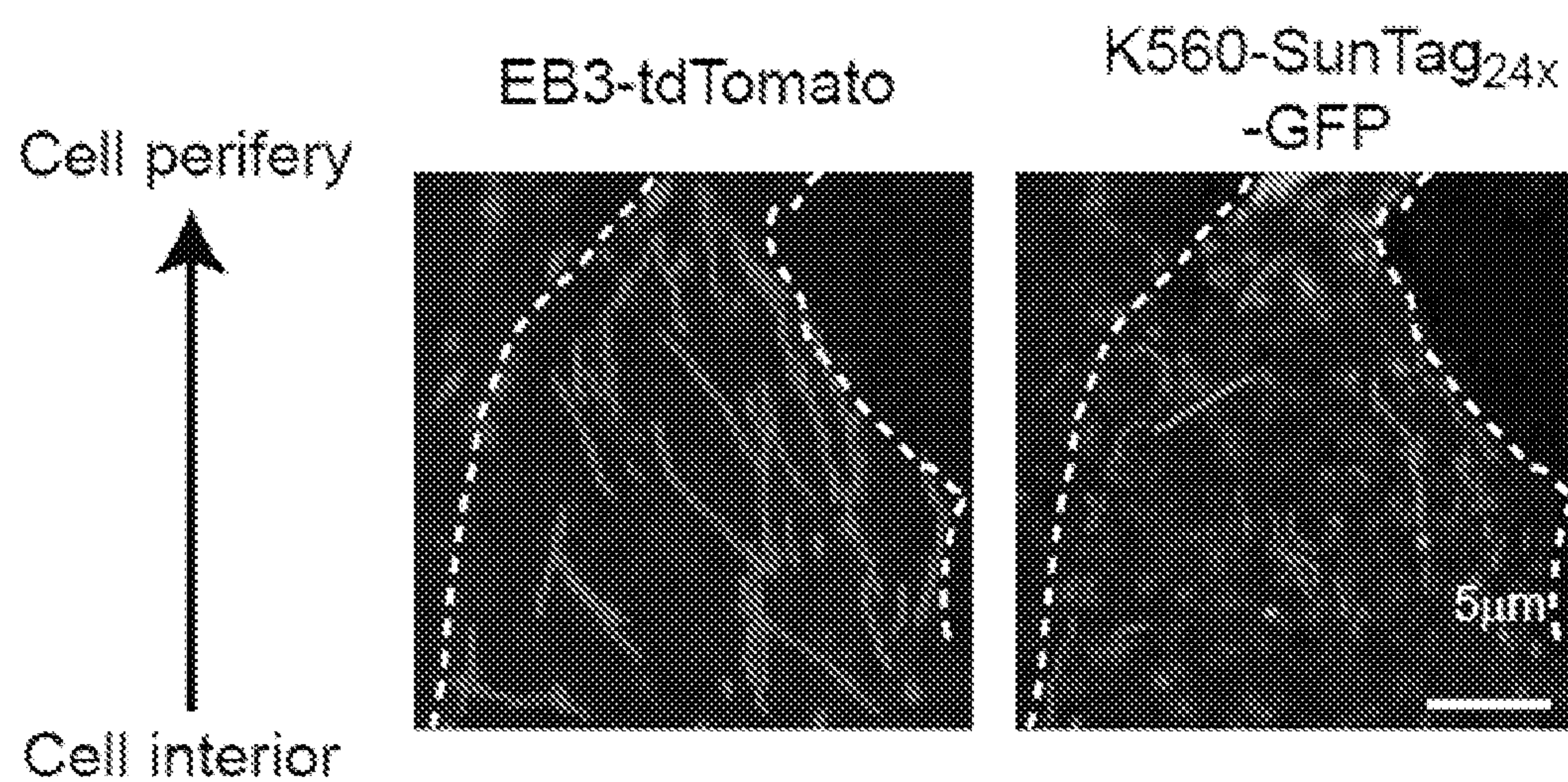


FIG. 5C

11 / 22

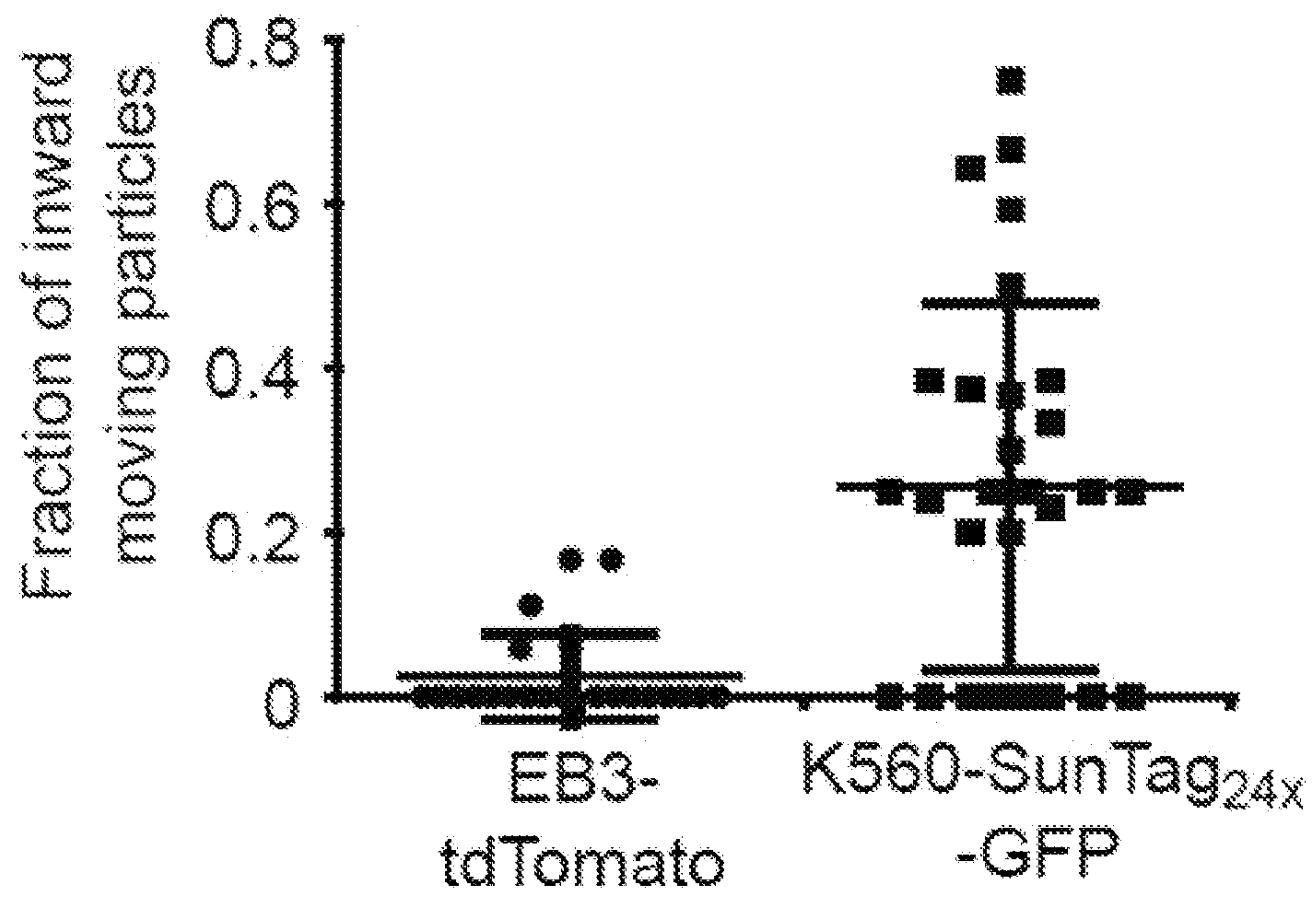


FIG. 5D

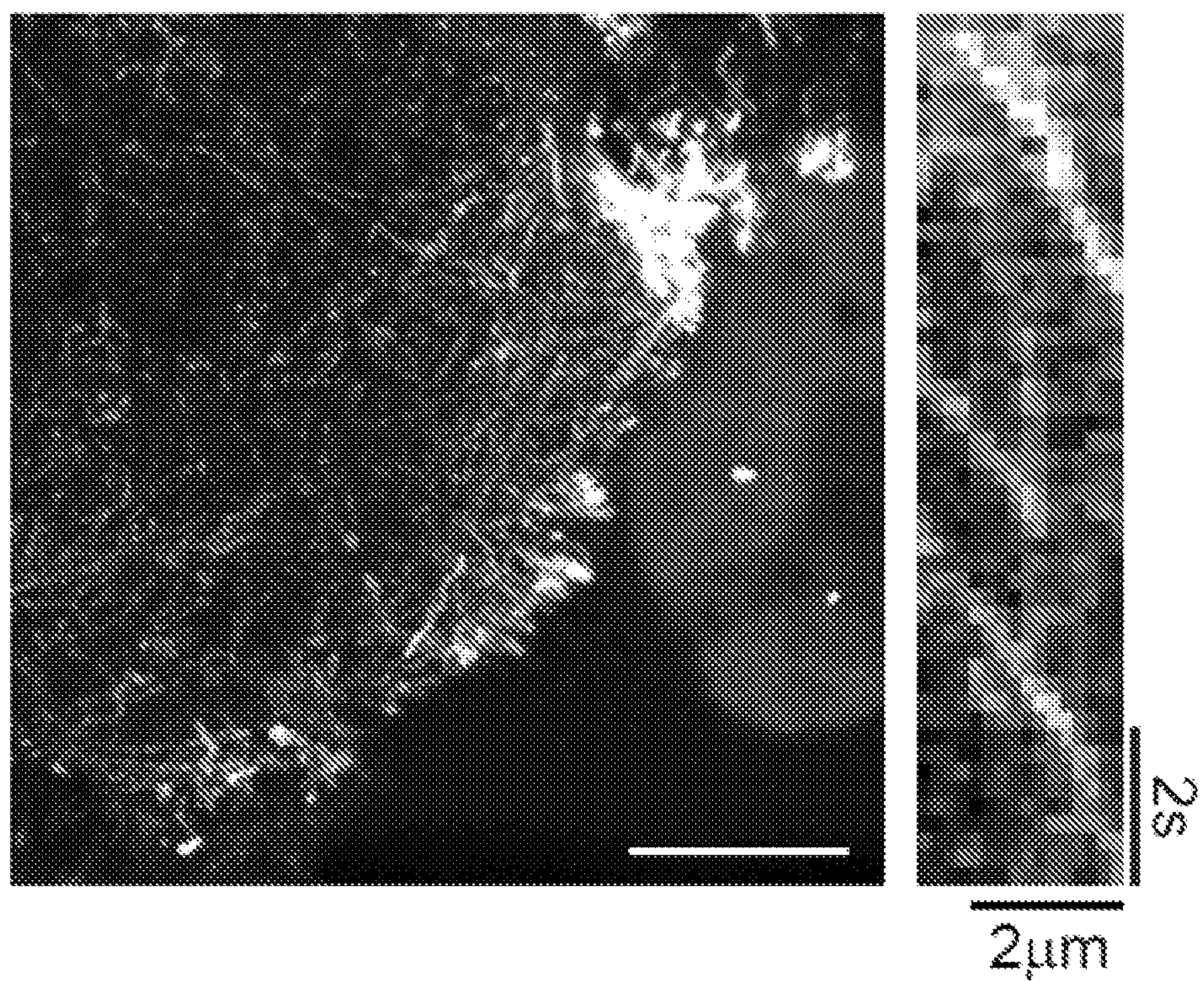
Kif18b-SunTag_{24x}-GFP

FIG. 5E

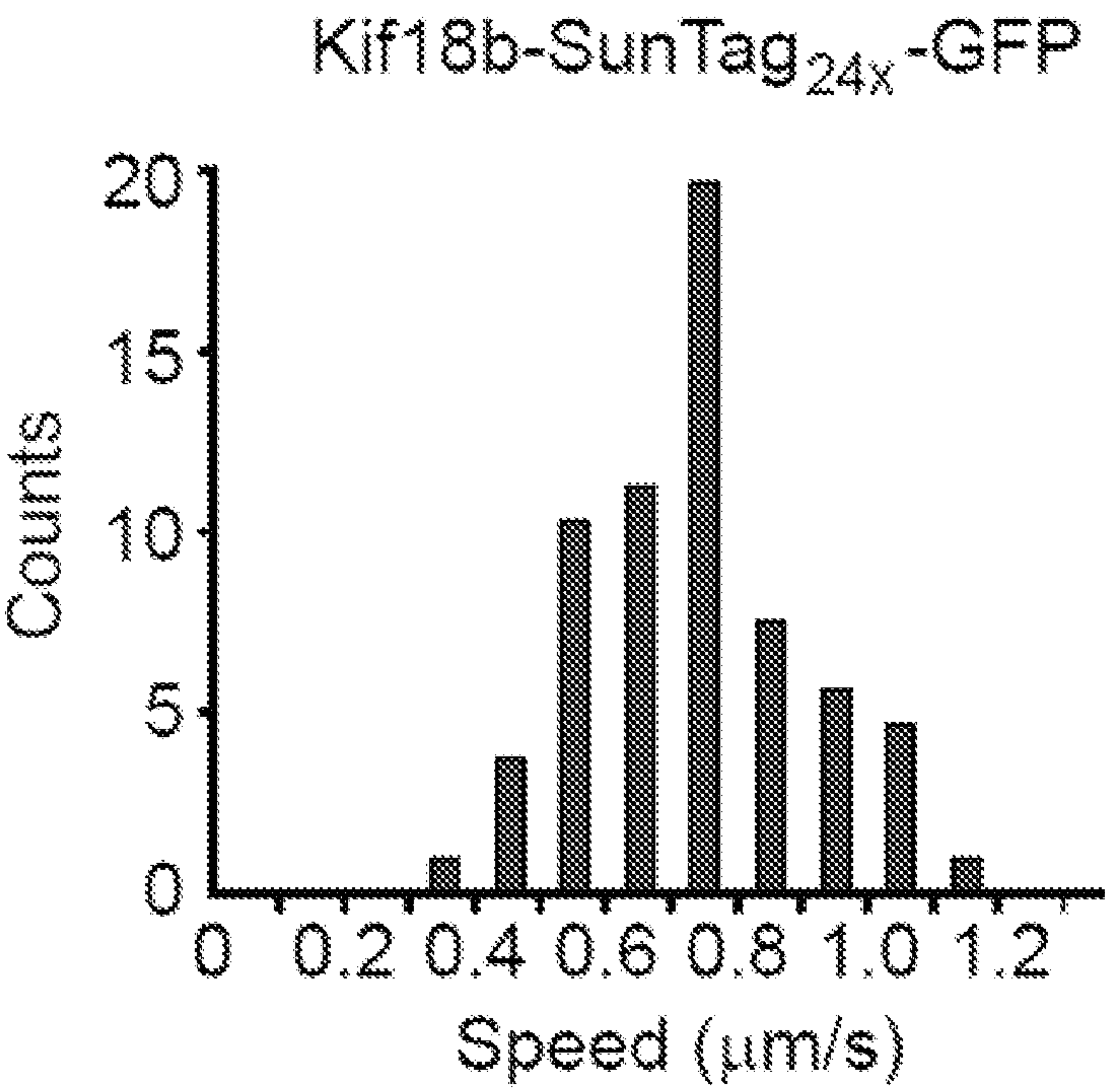
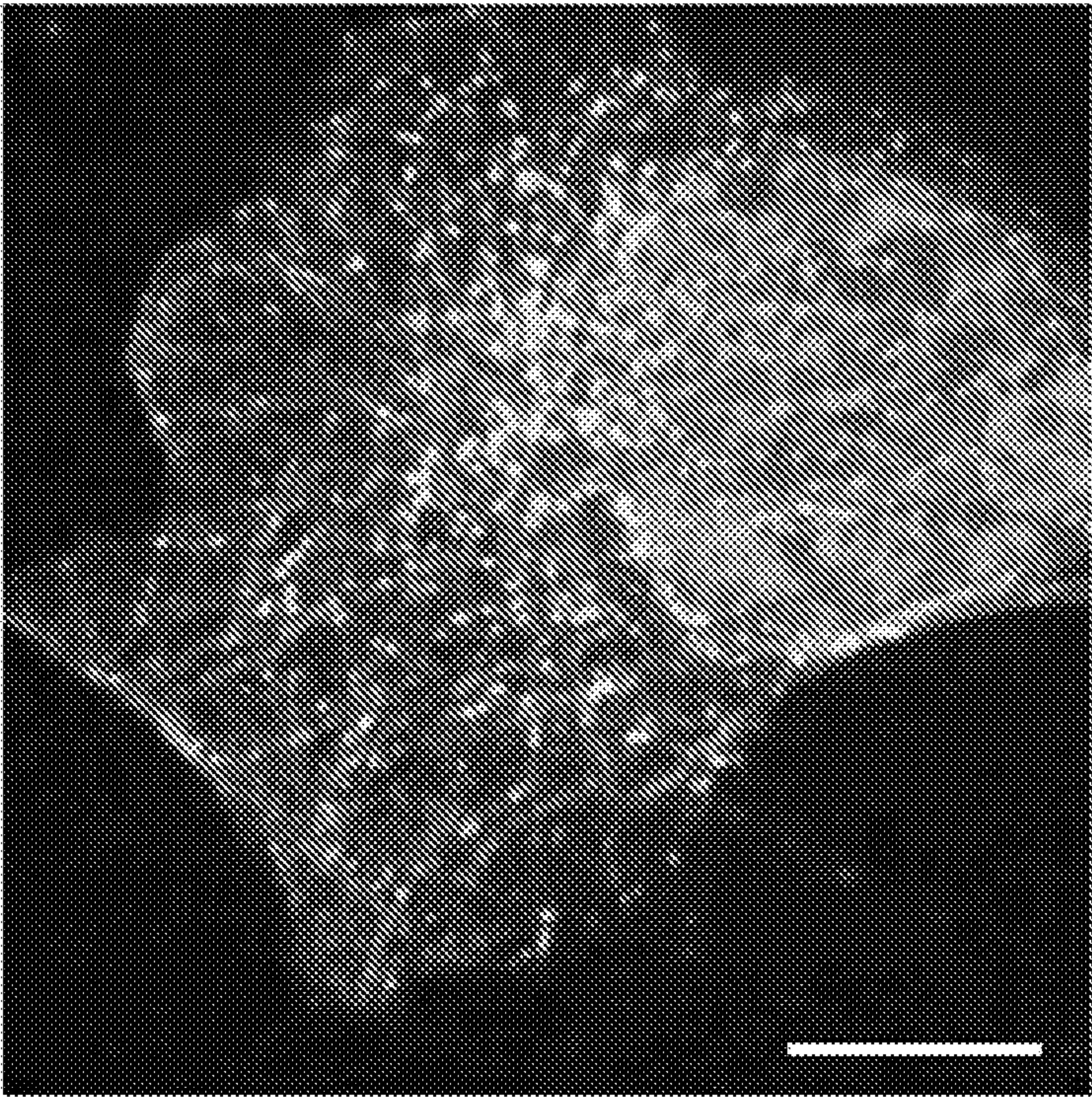


FIG. 5F



mCherry- α -tubulin
K560rig-SunTag_{24x}-GFP

FIG. 5G

13 / 22

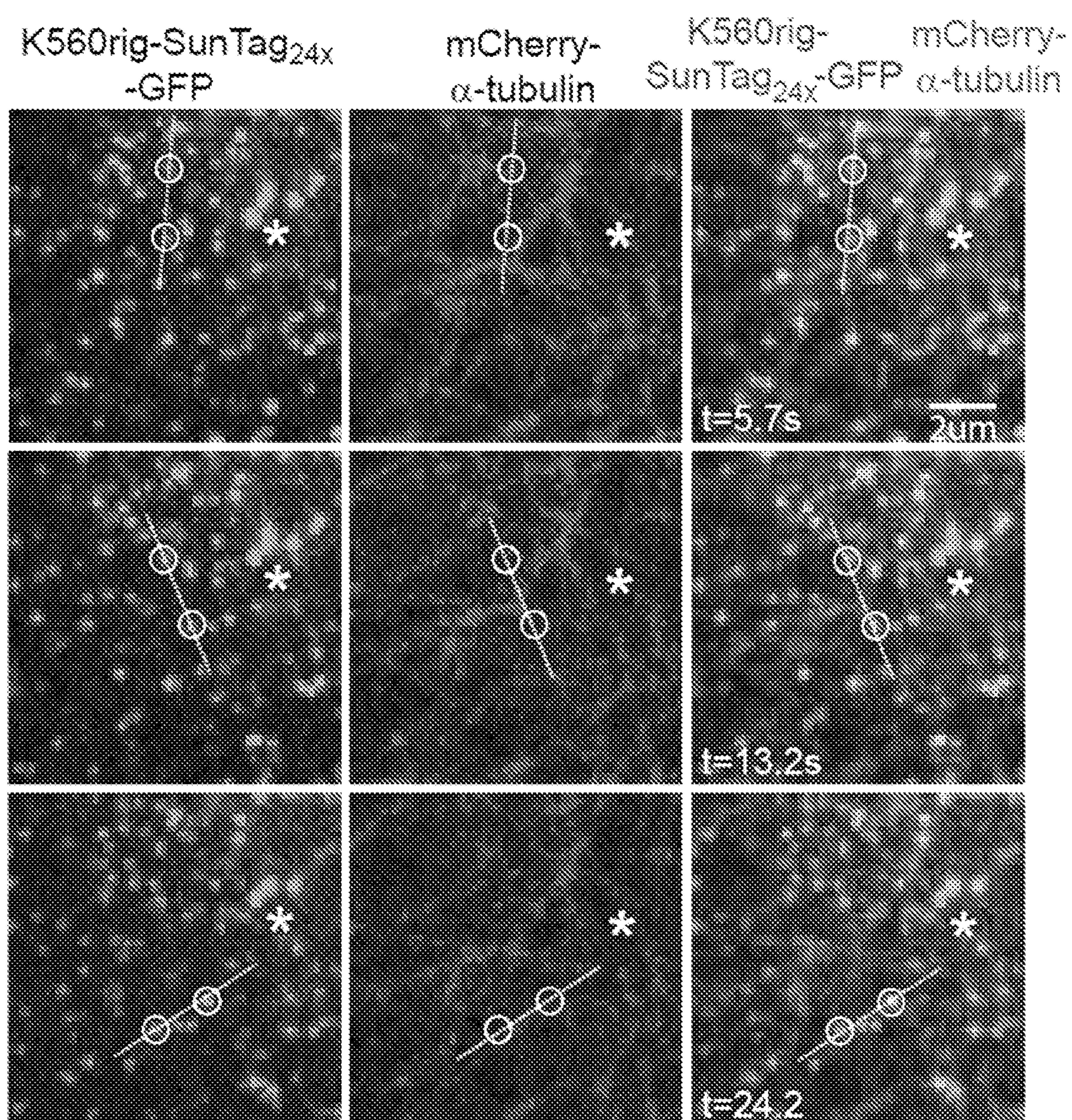


FIG. 5H

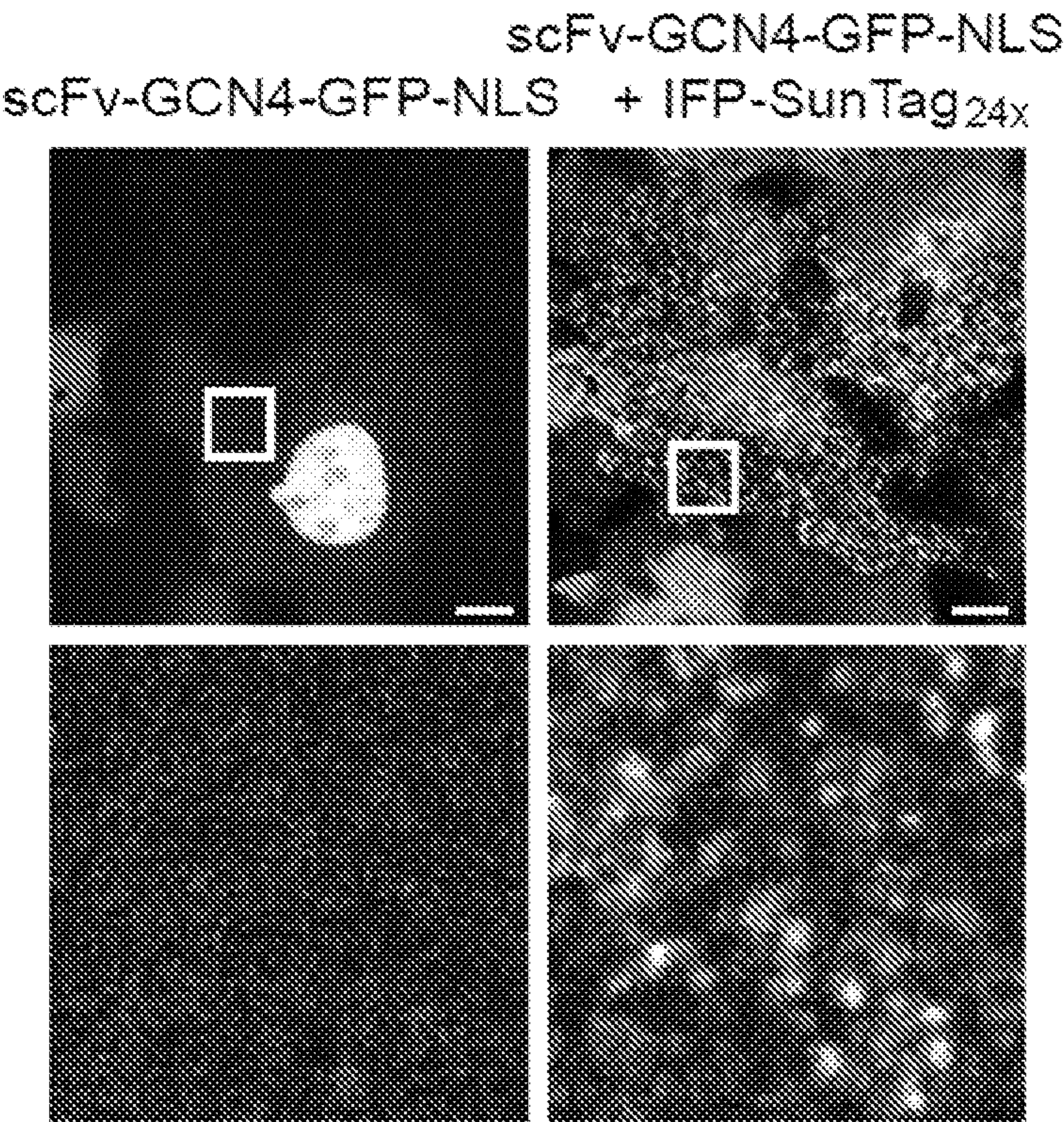


FIG. 6A

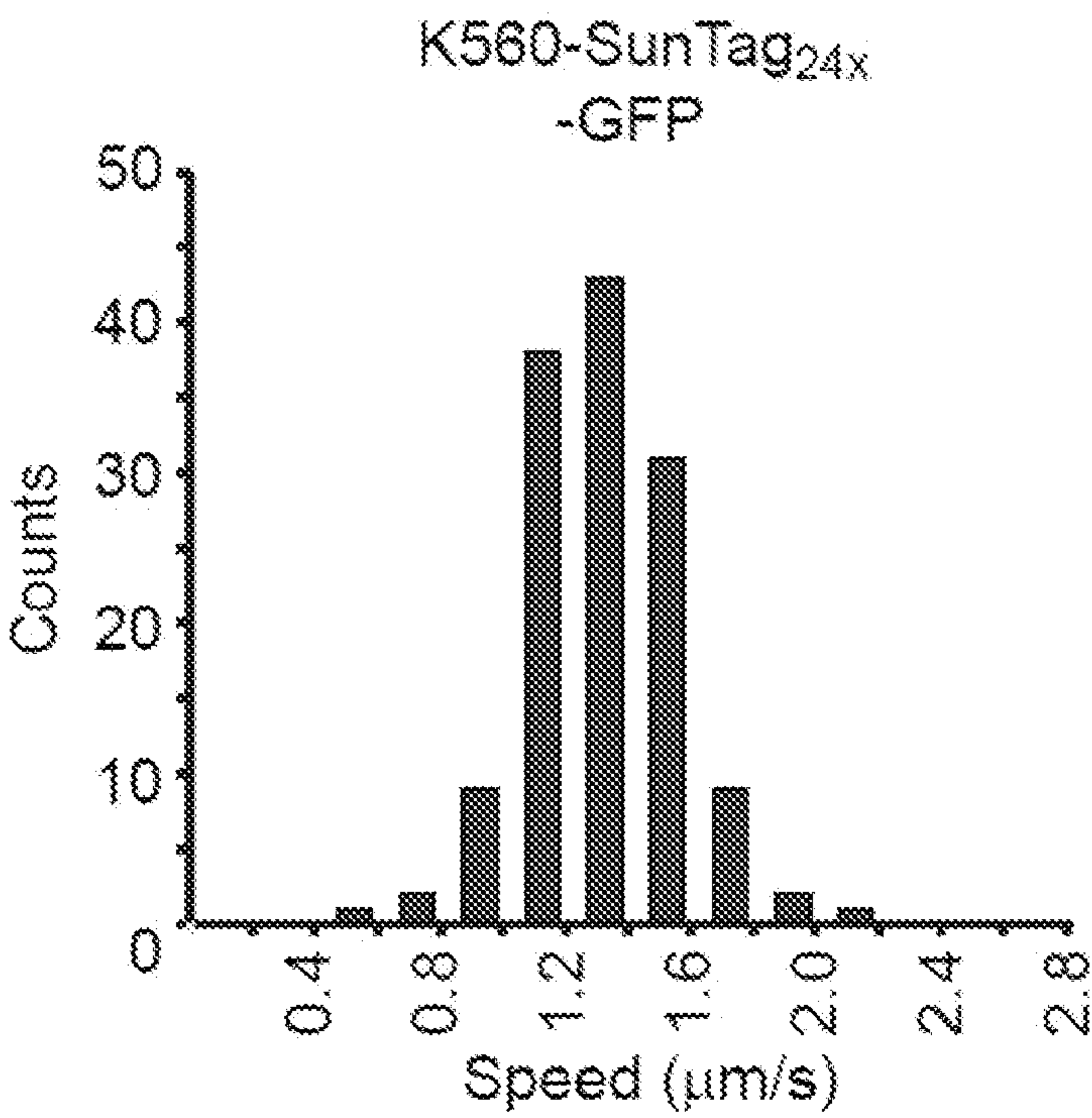


FIG. 6B

15 / 22

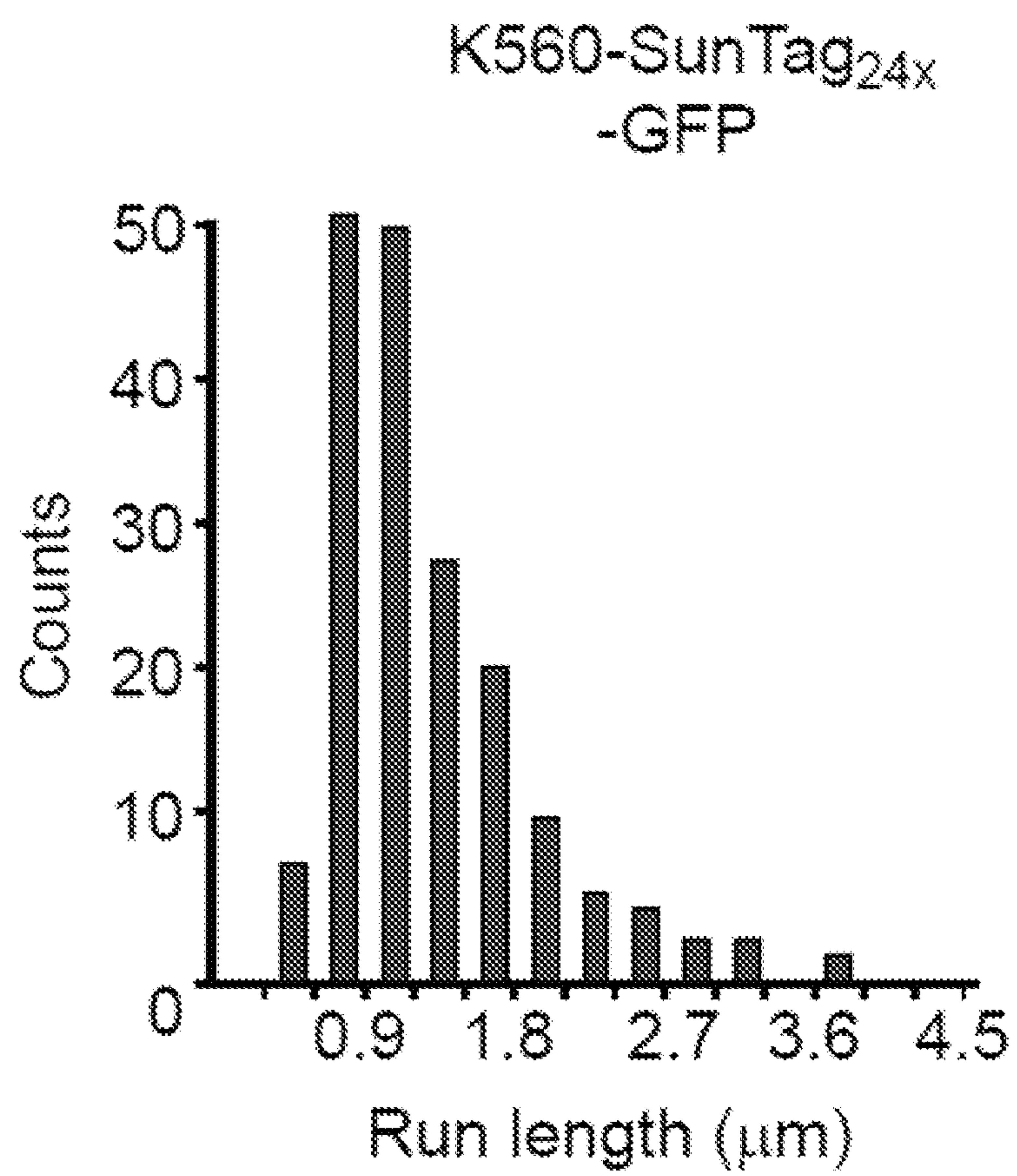


FIG. 6C

16 / 22

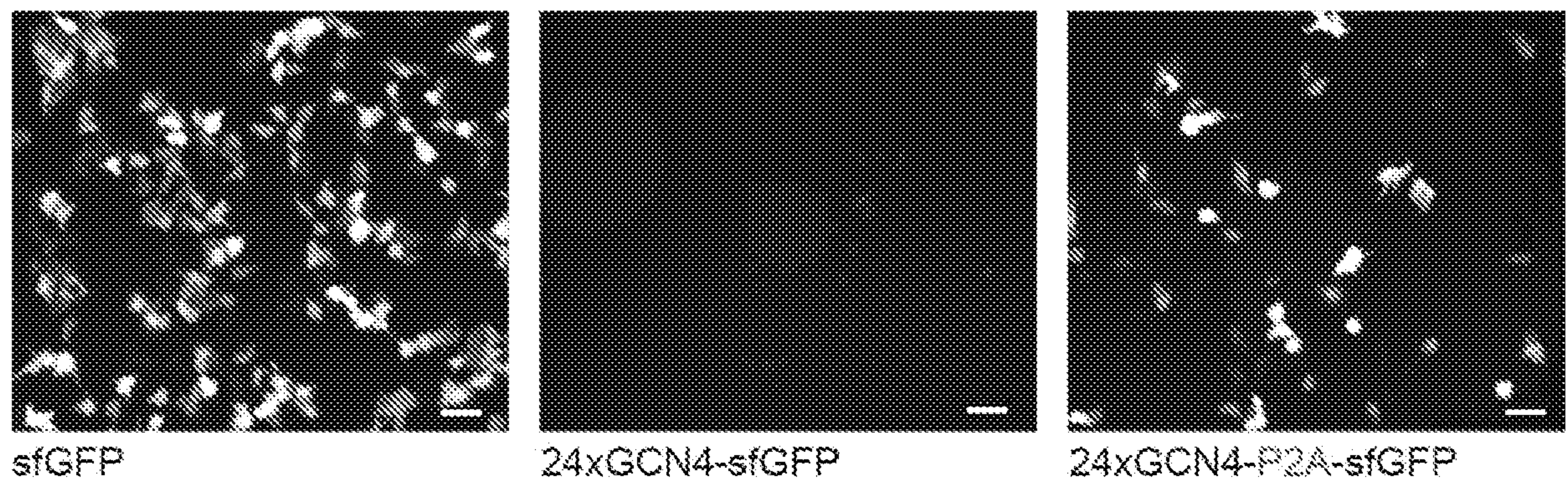


FIG. 7A

Peptide v1

LLPKNYHLENEVARLKKLVGERGGSGG

Peptide v4

EELSKNYHLENEVARLKKGSGSG

FIG. 7B

17 / 22

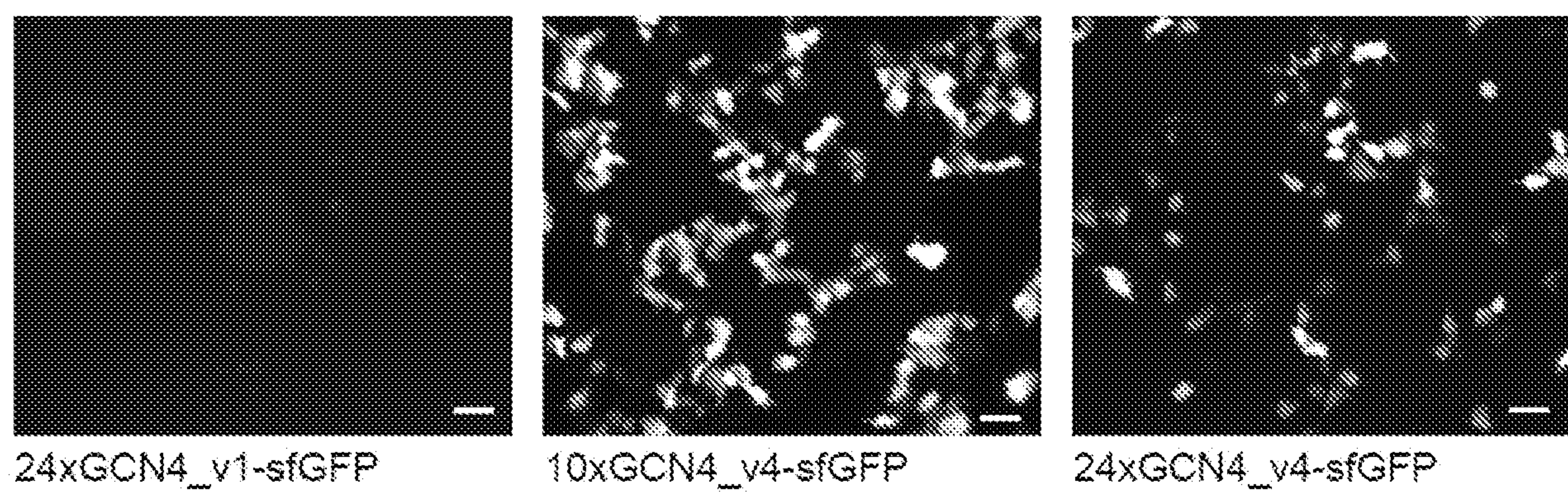


FIG. 7C

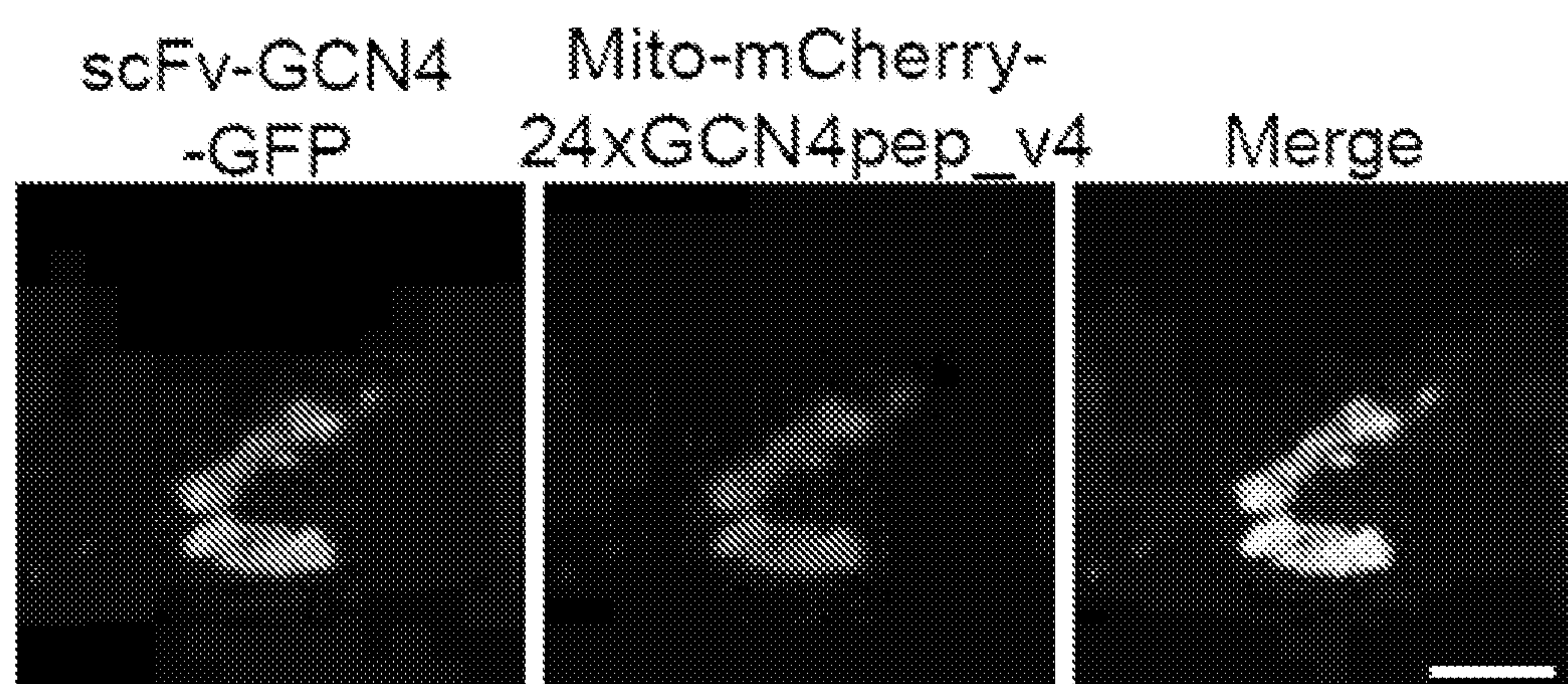


FIG. 7D

18 / 22

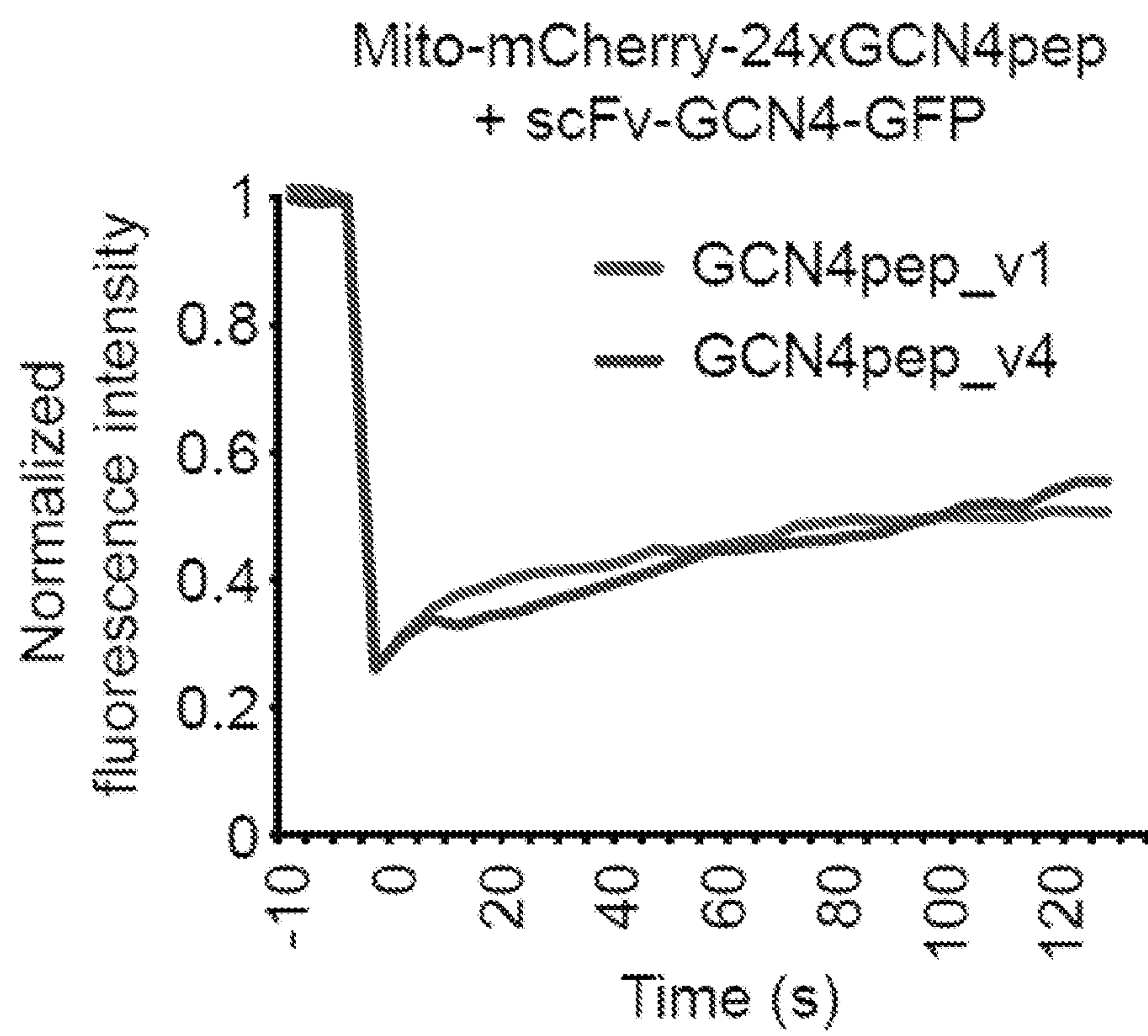


FIG. 7E

19 / 22

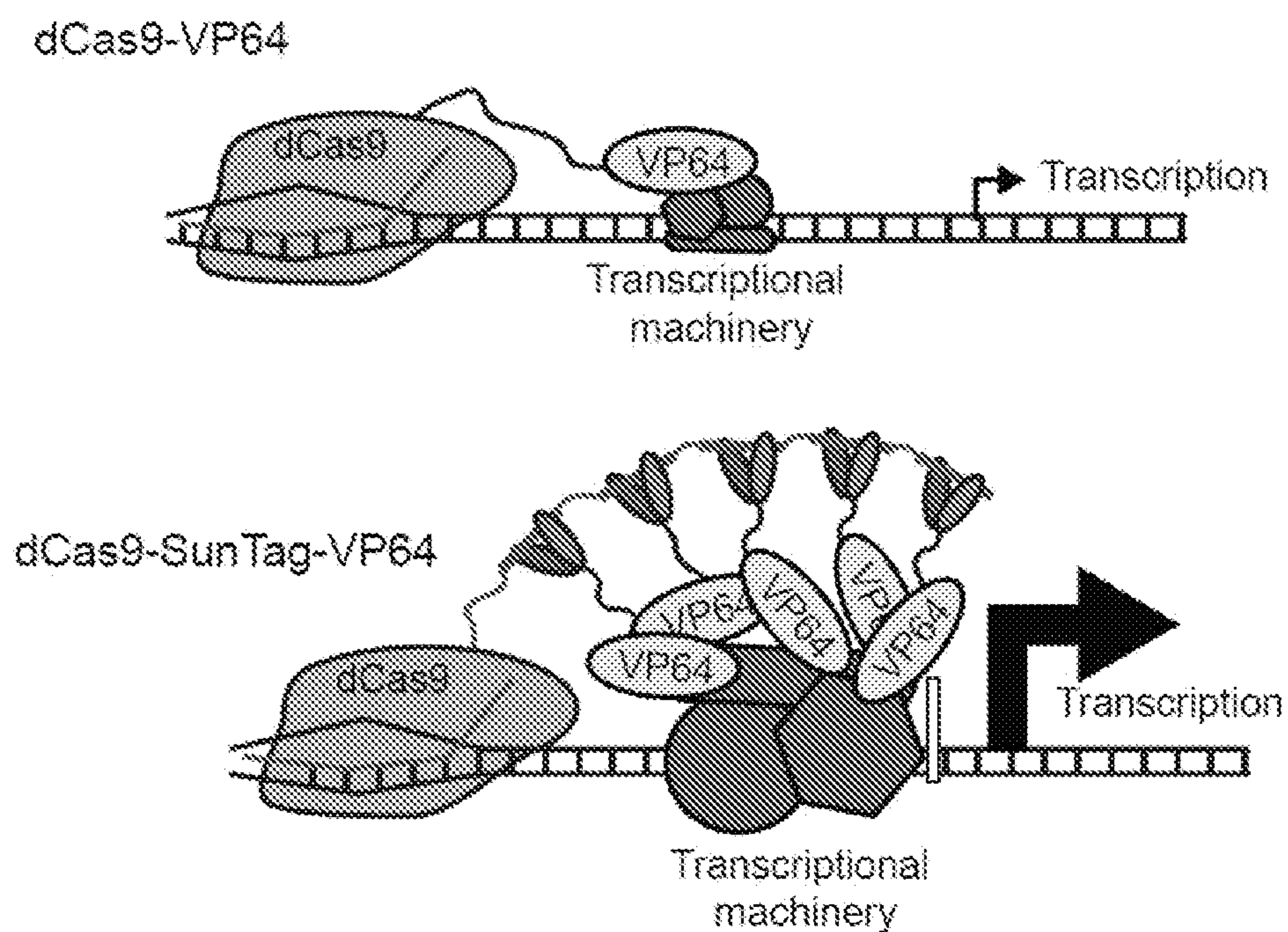


FIG. 8A

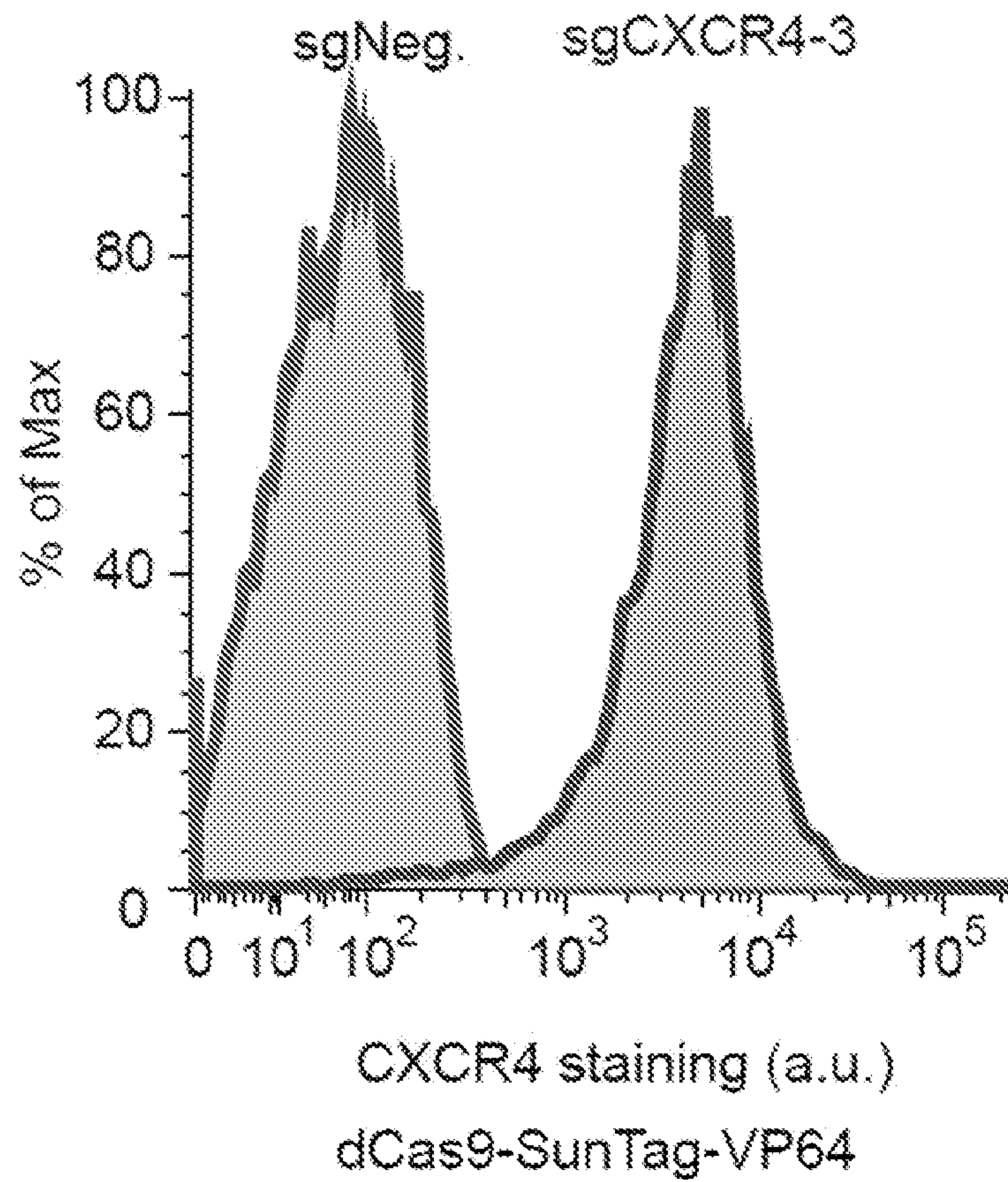


FIG. 8B

20 / 22

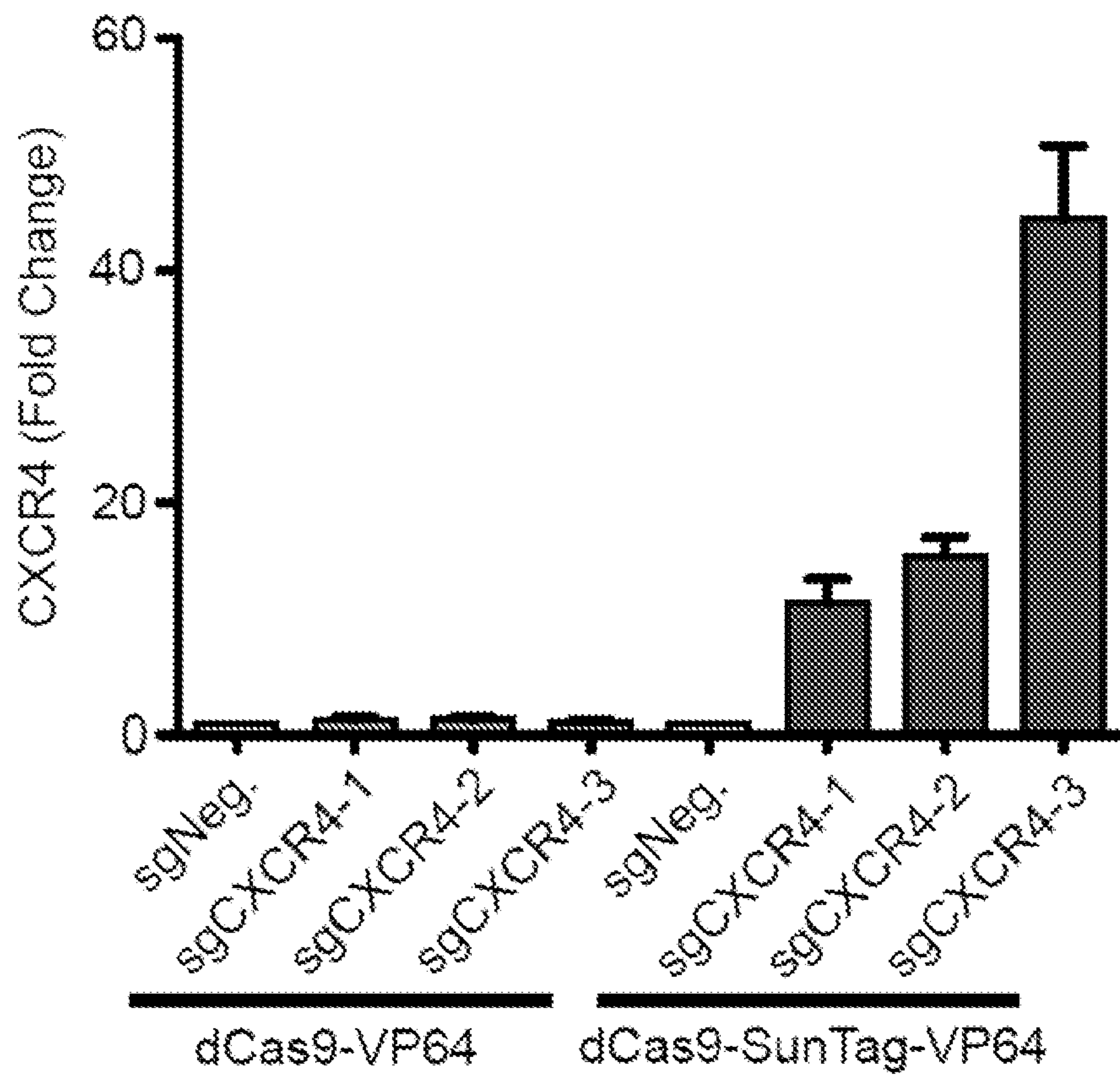


FIG. 8C

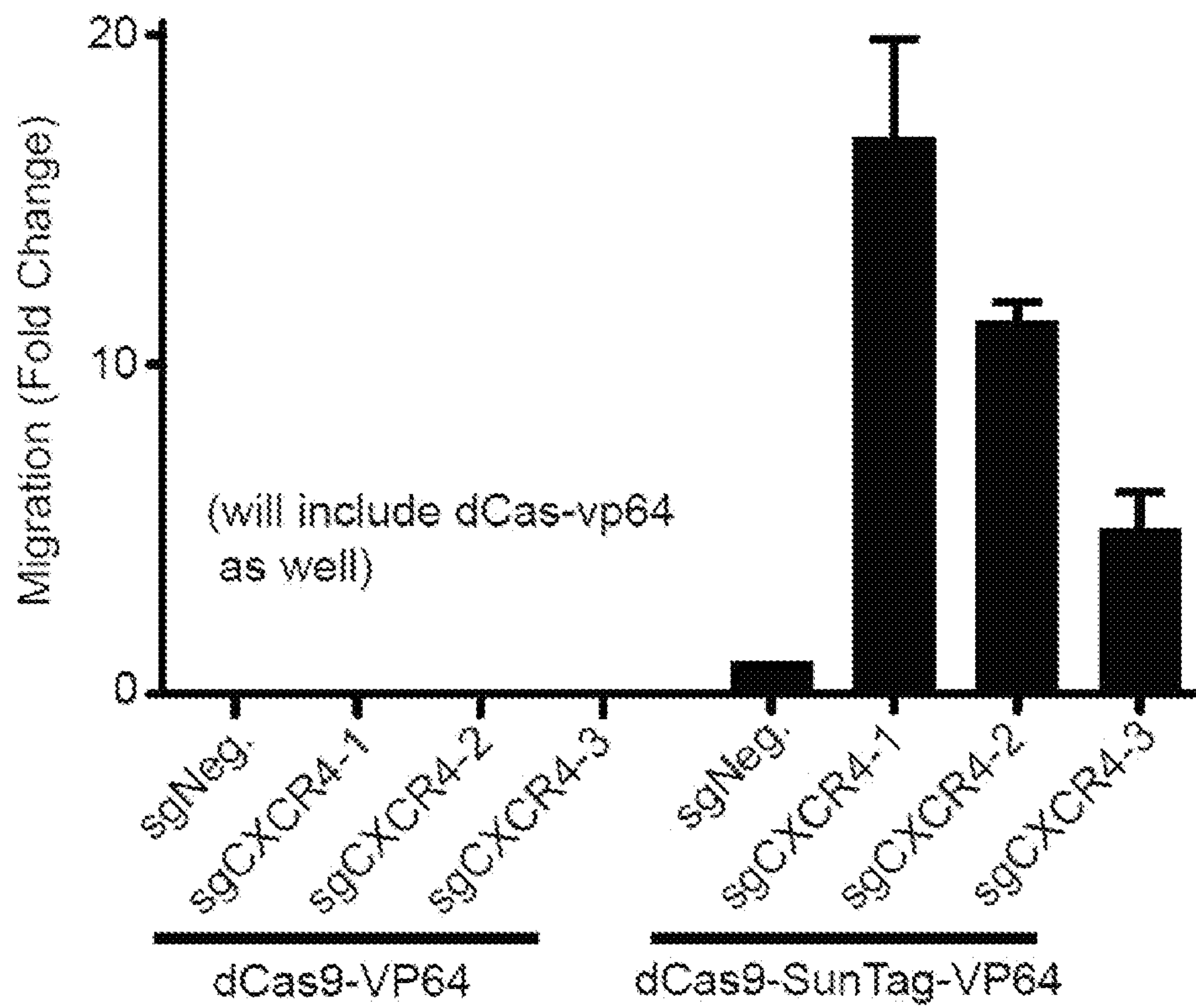


FIG. 8D

21 / 22

mRNA levels

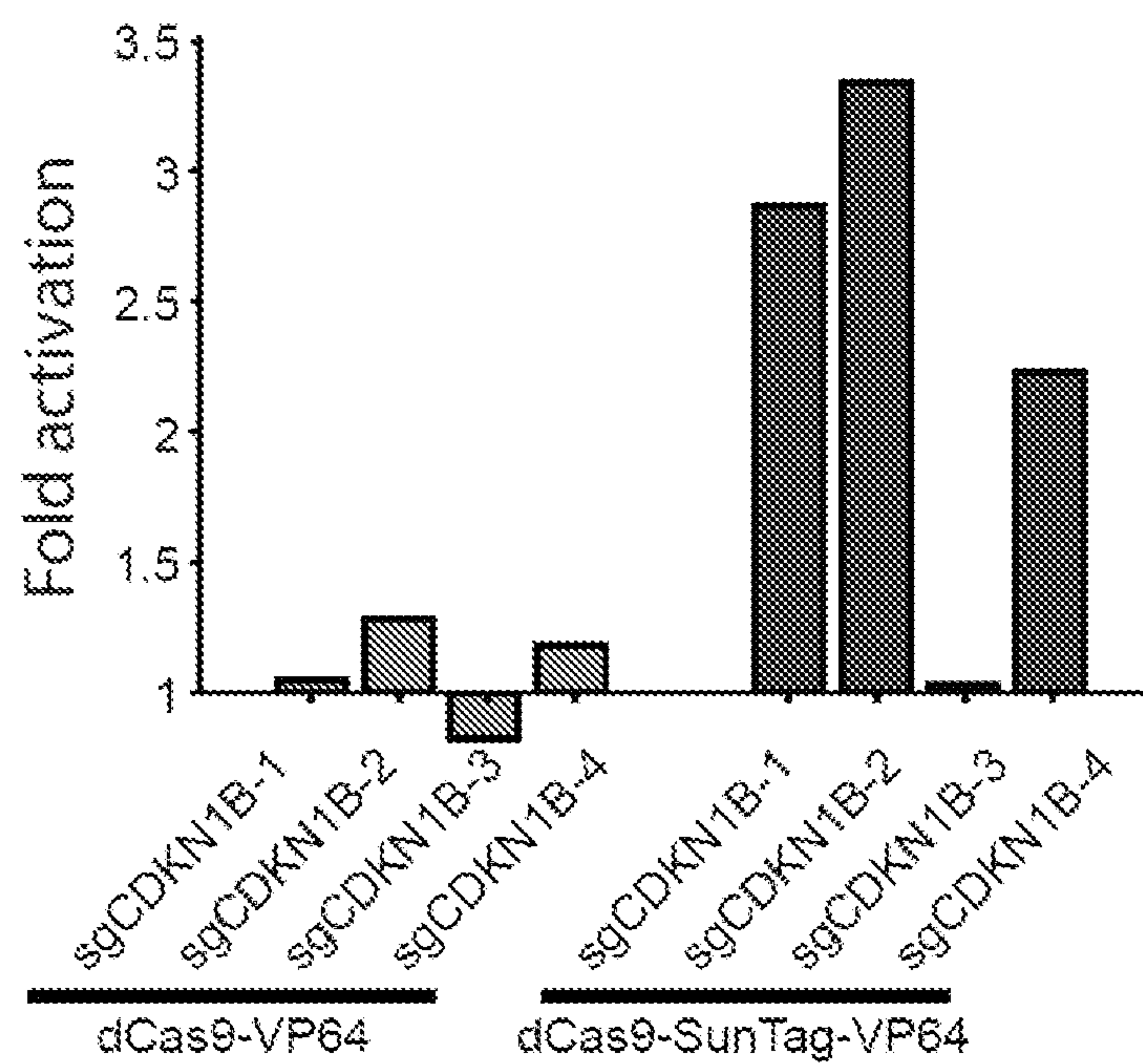


FIG. 8E

Cell Growth

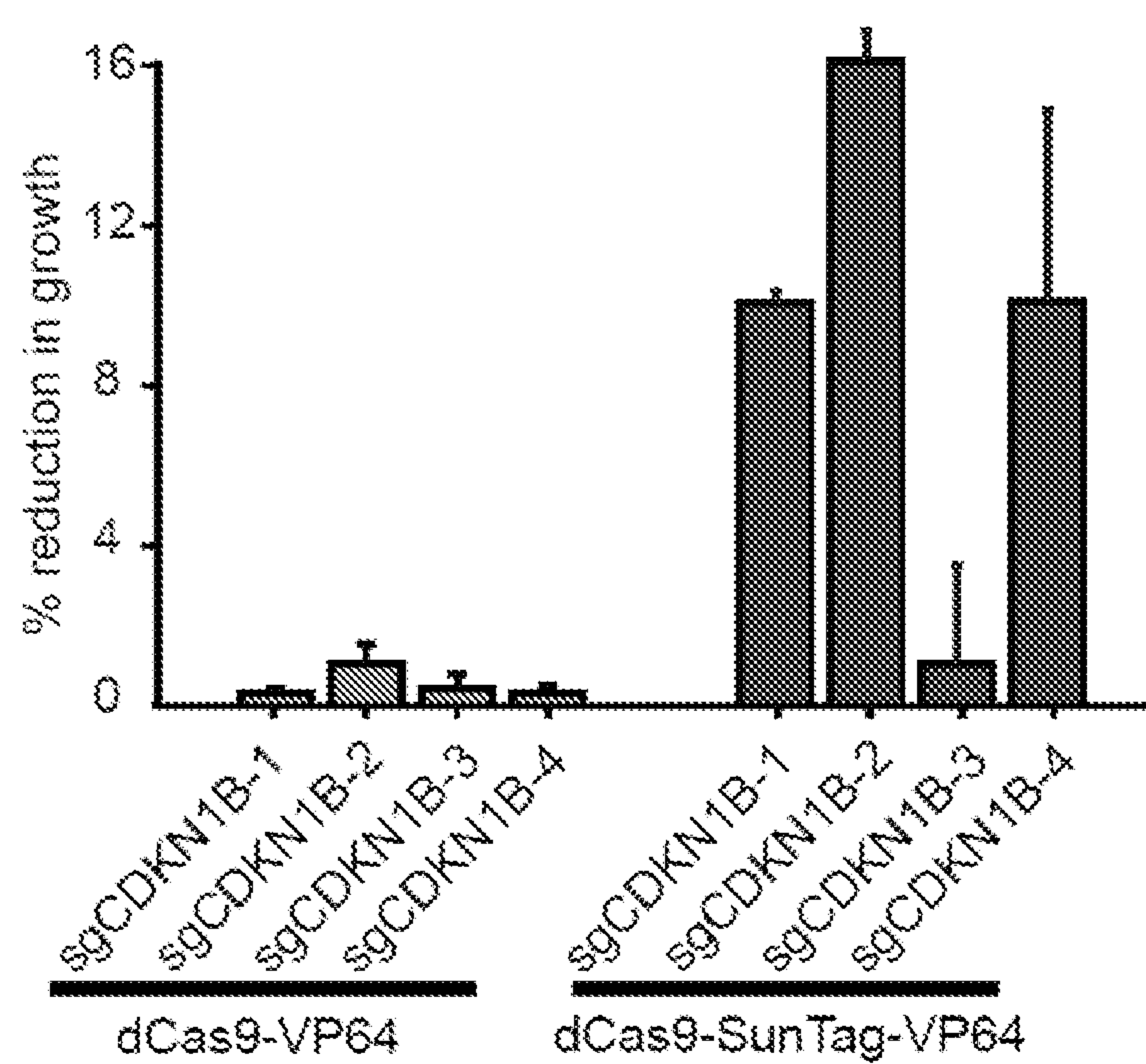


FIG. 8F

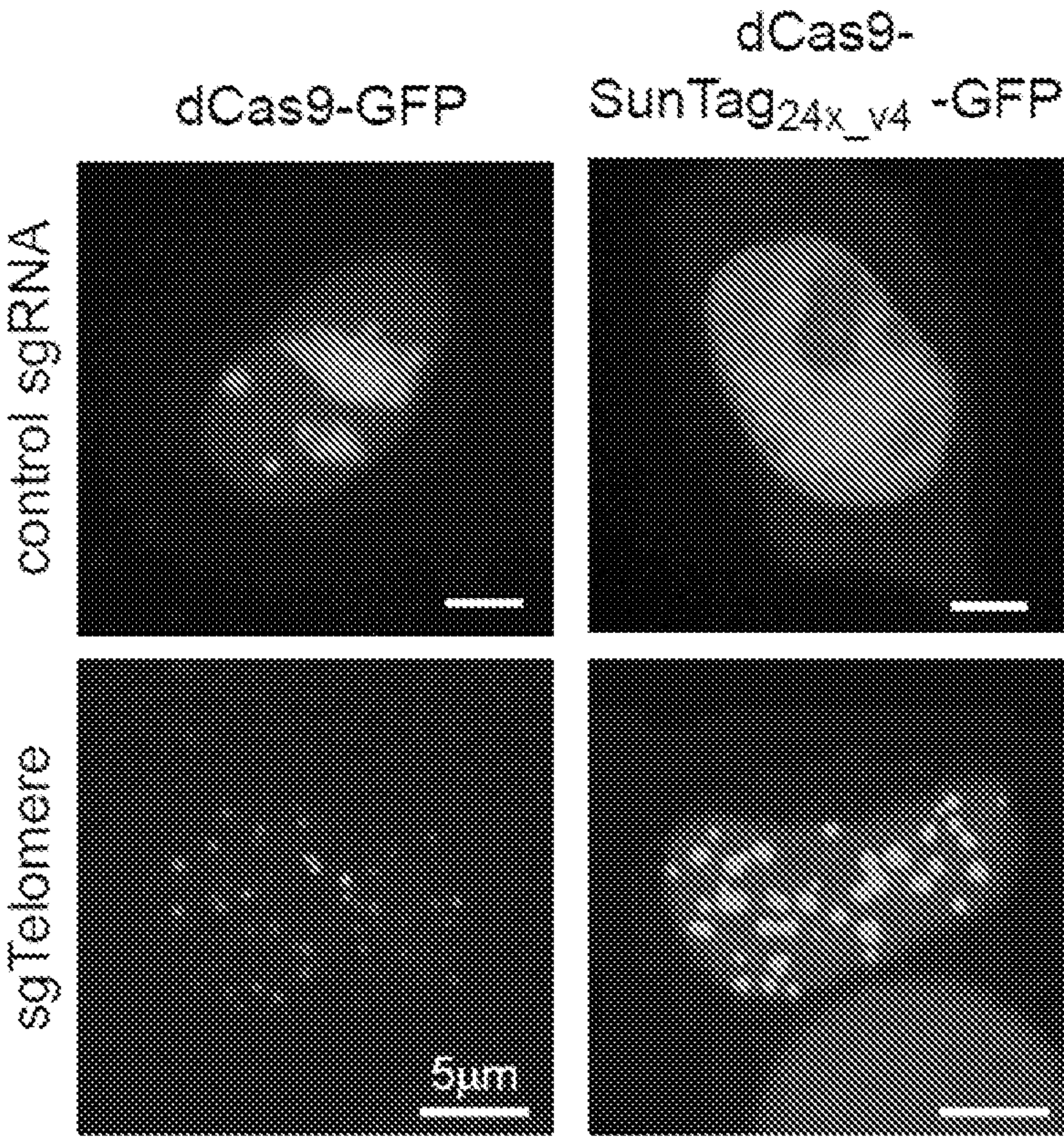


FIG. 9A

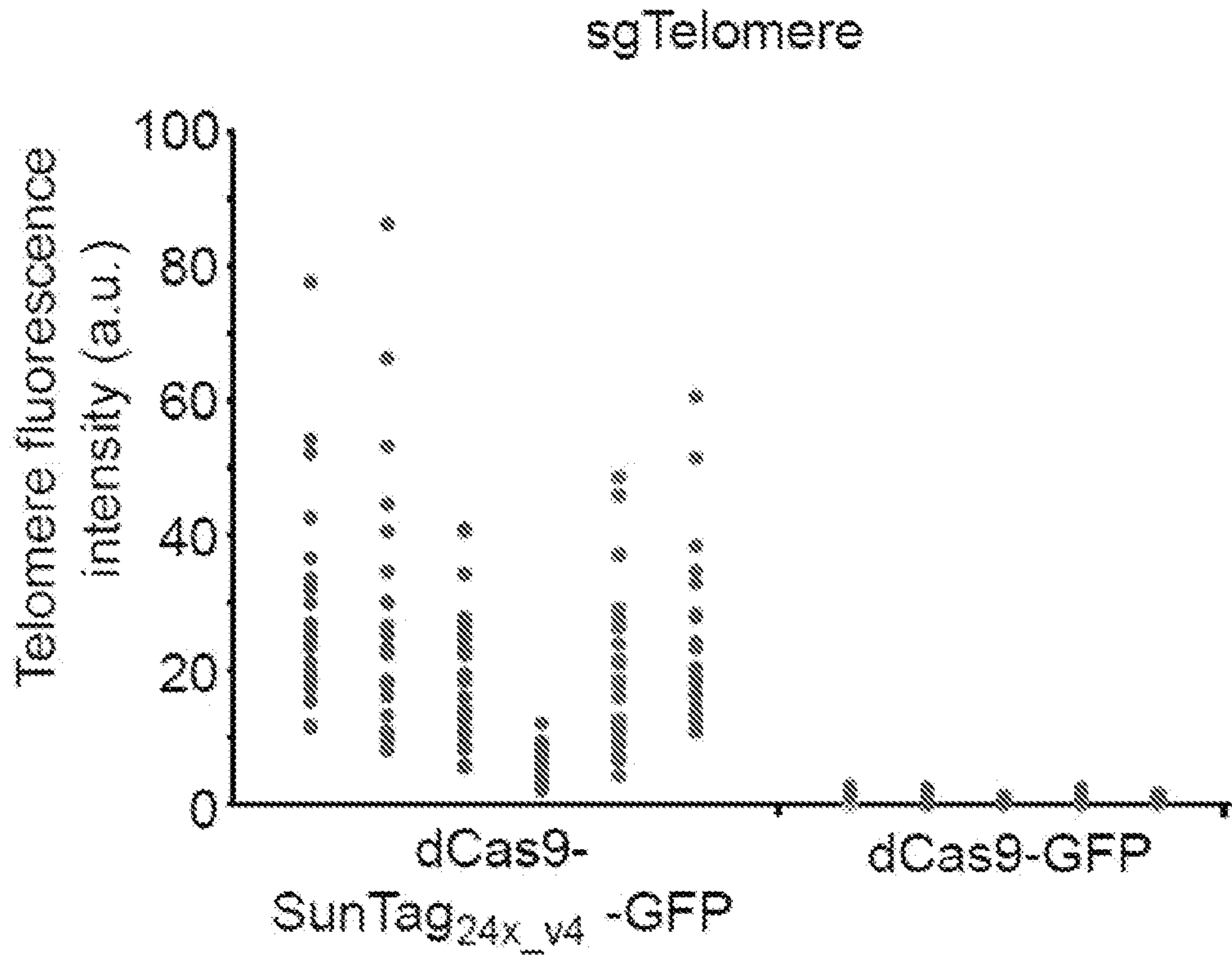


FIG. 9B

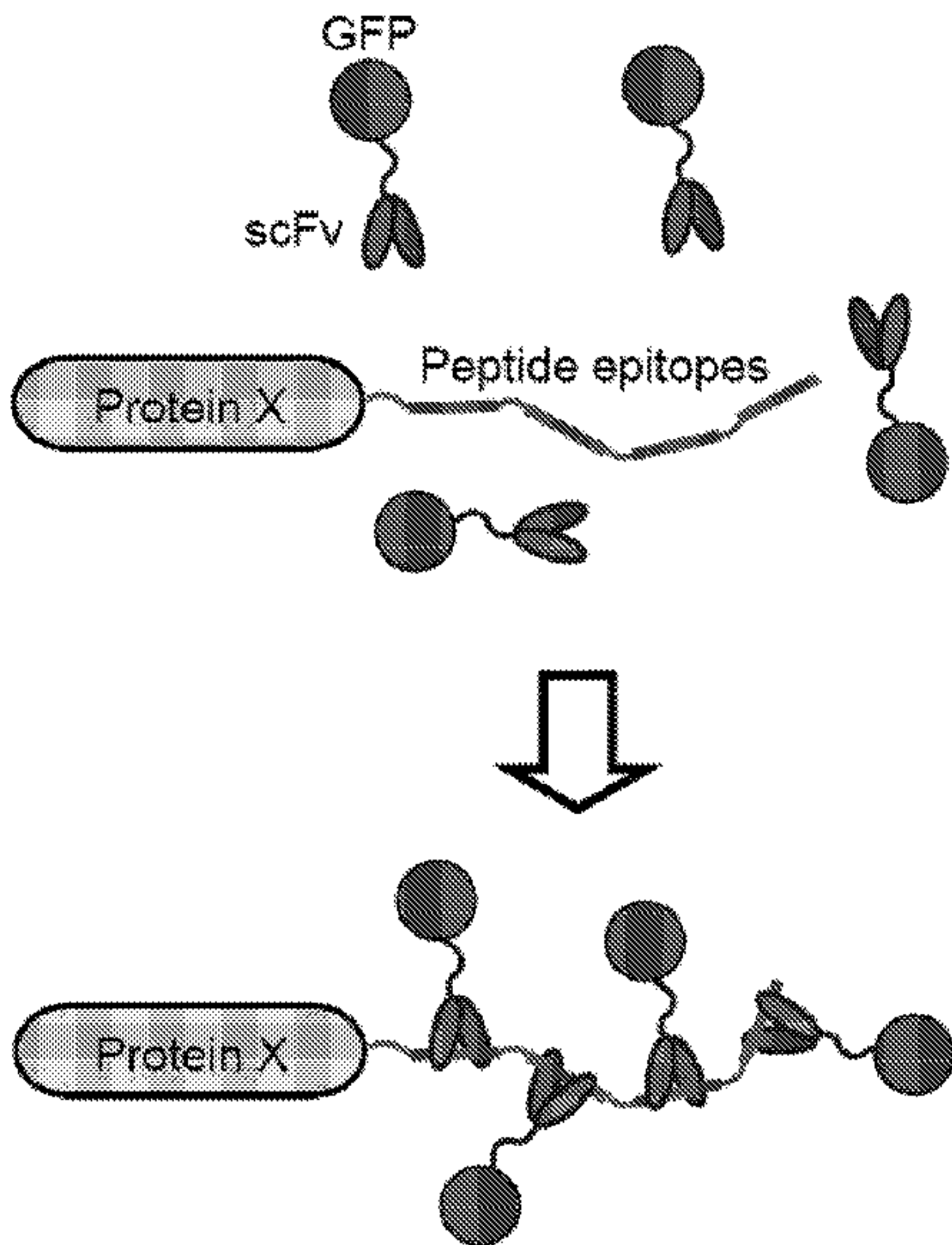


FIG. 1A