



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION
(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2018/01/19 (87) Date publication PCT/PCT Publication Date: 2018/07/26 (85) Entrée phase nationale/National Entry: 2019/06/17 (86) N° demande PCT/PCT Application No.: US 2018/014499 (87) N° publication PCT/PCT Publication No.: 2018/136783 (30) Priorité/Priority: 2017/01/20 (US62/448,841)	(51) Cl.Int./Int.Cl. <i>C12N 15/82</i> (2006.01), <i>C12N 15/113</i> (2010.01), <i>C12N 15/63</i> (2006.01), <i>C12N 9/22</i> (2006.01) (71) Demandeur/Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, US (72) Inventeurs/Inventors: JACOBSEN, STEVE E., US; PAPIKIAN, ASHOT, US (74) Agent: GOWLING WLG (CANADA) LLP
---	---

(54) Titre : ACTIVATION CIBLEE DE GENES DANS DES PLANTES
(54) Title: TARGETED GENE ACTIVATION IN PLANTS

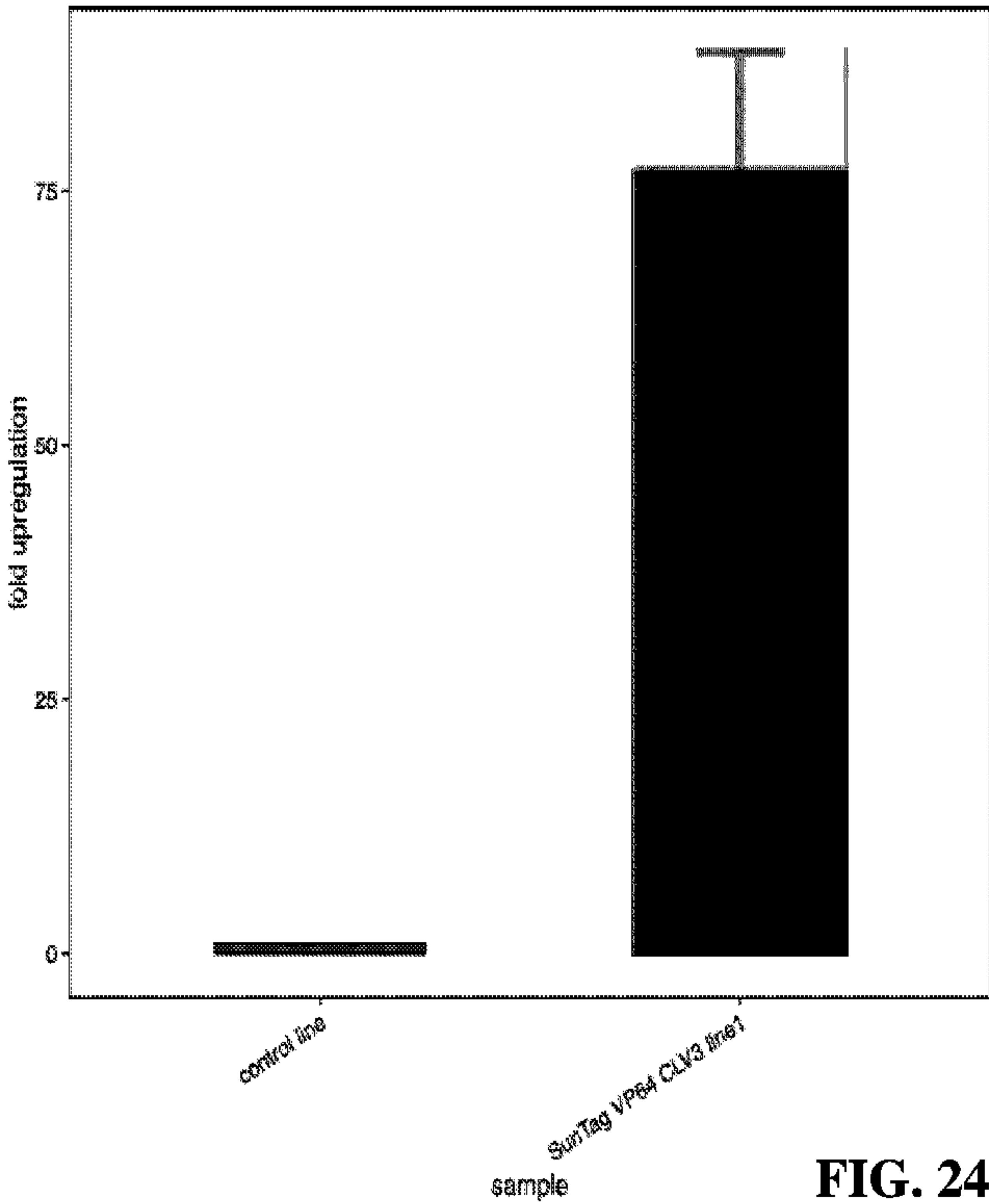


FIG. 24

(57) **Abrégé/Abstract:**
The present disclosure relates to the targeting of transcriptional activators to specific loci in plants to activate transcription of the targeted loci. Specifically, the present disclosure provides methods and compositions for using RNA-guided transcriptional activators to activate transcription of specific loci in plants.

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
26 July 2018 (26.07.2018)



(10) International Publication Number
WO 2018/136783 A1

(51) International Patent Classification:

C12N 15/82 (2006.01) C12N 15/113 (2010.01)
C12N 9/22 (2006.01) C12N 15/63 (2006.01)

(21) International Application Number:

PCT/US2018/014499

(22) International Filing Date:

19 January 2018 (19.01.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/448,841 20 January 2017 (20.01.2017) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: JACOBSEN, Steve, E.; 29087 Saddlebrook Drive, Agoura Hills, CA 91301 (US). PAPIKIAN, Ashot; 113B Magnolia Ave., Glendale, CA 91204 (US).

(74) Agent: ELLISON, Cory et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (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, DJ, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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.

(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))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: TARGETED GENE ACTIVATION IN PLANTS

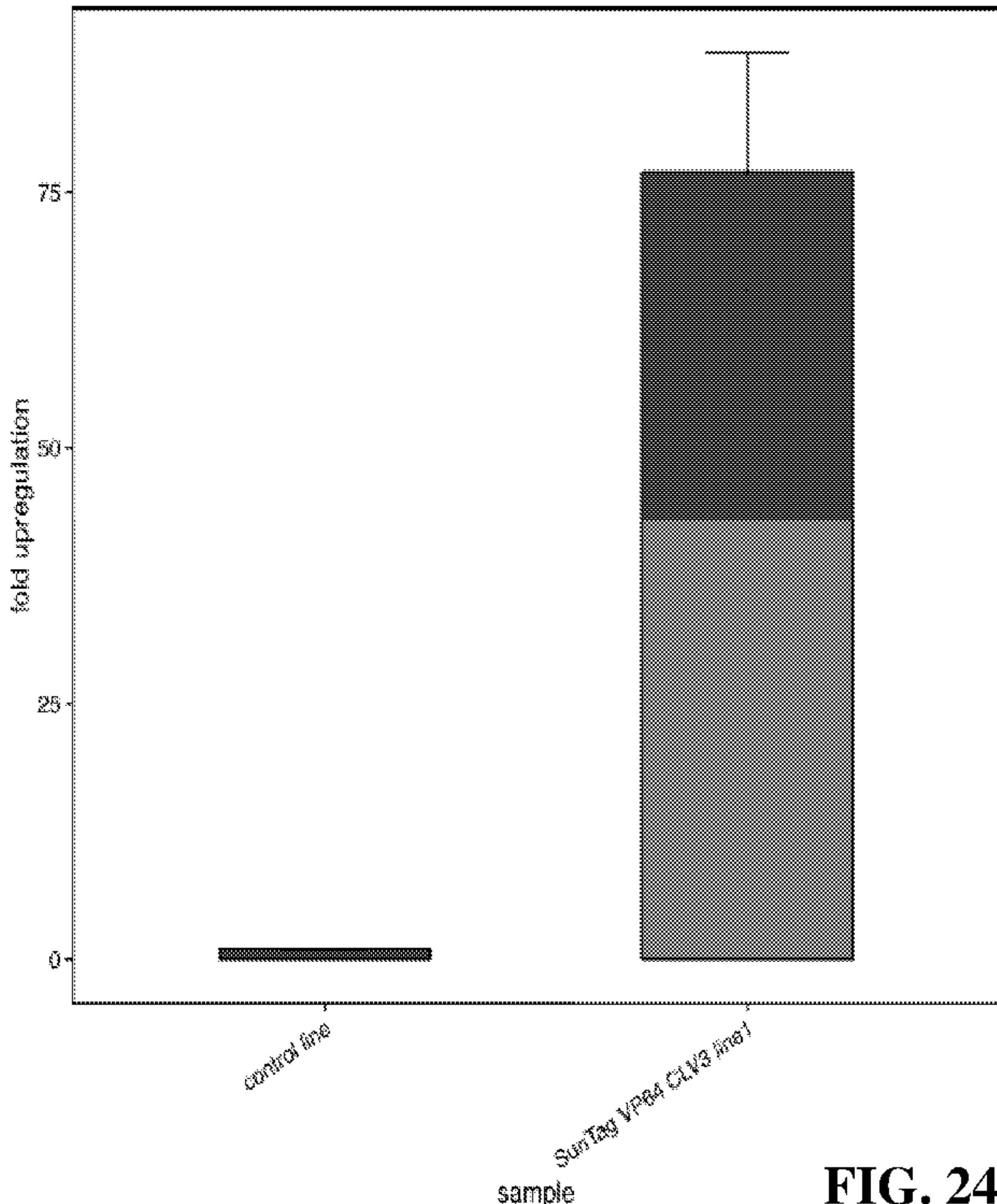


FIG. 24

(57) Abstract: The present disclosure relates to the targeting of transcriptional activators to specific loci in plants to activate transcription of the targeted loci. Specifically, the present disclosure provides methods and compositions for using RNA-guided transcriptional activators to activate transcription of specific loci in plants.

WO 2018/136783 A1

TARGETED GENE ACTIVATION IN PLANTS**CROSS-REFERENCE TO RELATED APPLICATION**

[0001] This application claims the benefit of U.S. Provisional Application No. 62/448,841, filed on January 20, 2017, the disclosure of which is incorporated herein by reference in its entirety.

SUBMISSION OF SEQUENCE LISTING ON ASCII TEXT FILE

[0002] The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 262232001440SEQLIST.txt, date recorded: January 16, 2018, size: 122 KB).

FIELD

[0003] The present disclosure relates to the targeting of transcriptional activators to specific loci in plants to activate transcription of the targeted loci. Specifically, the present disclosure provides methods and compositions for using RNA-guided transcriptional activators to activate transcription of specific loci in plants.

BACKGROUND

[0004] Transcriptional regulation is a key aspect of the growth and development of many organisms. In plants, transcriptional regulation plays a pivotal role in growth and development, as well as a multitude of biological pathways and processes. Indeed, the manipulation of gene expression in plants, such as the activation of a gene of interest, can have profound phenotypic impacts. In addition to influencing a phenotype, the activation of gene expression or transcriptional activation of a locus of interest can be useful for a wide variety of research purposes.

[0005] There is currently no robust method for selectively activating the expression of plant genes or other plant loci of interest. Accordingly, a need exists for methods of inducing transcriptional activation of specific loci in plants.

BRIEF SUMMARY

[0006] In one aspect, the present disclosure provides a method for activating expression of a target nucleic acid in a plant, including: (a) providing a plant including: a first

recombinant polypeptide including a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope; a second recombinant polypeptide including a transcriptional activator and an affinity polypeptide that specifically binds to the epitope; a crRNA and a tracrRNA, or fusions thereof; and (b) growing the plant under conditions whereby the first and second recombinant polypeptides are targeted to the target nucleic acid, thereby activating expression of the target nucleic acid. In some embodiments, the dCAS9 polypeptide has an amino acid sequence that is at least 80% identical to SEQ ID NO: 12. In some embodiments that may be combined with any of the preceding embodiments, the multimerized epitope includes a GCN4 epitope. In some embodiments, the multimerized epitope includes about 2 to about 10 copies of a GCN4 epitope. In some embodiments that may be combined with any of the preceding embodiments, the first polypeptide includes one or more linkers that link polypeptide units in the recombinant polypeptide. In some embodiments that may be combined with any of the preceding embodiments, the first polypeptide includes a nuclear localization signal (NLS). In some embodiments that may be combined with any of the preceding embodiments, the transcriptional activator is a VP64 polypeptide. In some embodiments, the VP64 polypeptide includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 31. In some embodiments that may be combined with any of the preceding embodiments, the affinity polypeptide is an antibody. In some embodiments, the antibody is an scFv antibody. In some embodiments, the antibody includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 28. In some embodiments that may be combined with any of the preceding embodiments, the second polypeptide includes one or more linkers that link polypeptide units in the recombinant polypeptide. In some embodiments that may be combined with any of the preceding embodiments, the second polypeptide includes an SV40-type NLS. In some embodiments, the SV40-type NLS includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 32. In some embodiments that may be combined with any of the preceding embodiments, the crRNA and the tracrRNA are fused together, thereby forming a guide RNA (gRNA). In some embodiments that may be combined with any of the preceding embodiments, expression of the activated nucleic acid is increased in the range of about 100-fold to about 10,000-fold as compared to a corresponding control.

[0007] In another aspect, the present disclosure provides a recombinant vector including: a first nucleic acid sequence including a plant promoter and that encodes a recombinant polypeptide including a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof

and a multimerized epitope; a second nucleic acid sequence including a plant promoter and that encodes a recombinant polypeptide including a transcriptional activator and an affinity polypeptide that specifically binds to the epitope; and a third nucleic acid sequence including a promoter and that encodes a crRNA and a tracrRNA, or fusions thereof. In some embodiments, the plant promoter in the first nucleic acid sequence is a UBQ10 promoter. In some embodiments, the UBQ10 promoter includes a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 2. In some embodiments that may be combined with any of the preceding embodiments, the first nucleic acid sequence includes a terminator sequence. In some embodiments, the terminator is an OCS terminator. In some embodiments, the OCS terminator includes a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 9. In some embodiments that may be combined with any of the preceding embodiments, the dCAS9 polypeptide includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 12. In some embodiments that may be combined with any of the preceding embodiments, the multimerized epitope includes a GCN4 epitope. In some embodiments, the multimerized epitope includes about 2 to about 10 copies of a GCN4 epitope. In some embodiments that may be combined with any of the preceding embodiments, the first polypeptide includes one or more linkers that link polypeptide units in the recombinant polypeptide. In some embodiments that may be combined with any of the preceding embodiments, the first polypeptide includes a nuclear localization signal (NLS). In some embodiments that may be combined with any of the preceding embodiments, the plant promoter in the second nucleic acid sequence is a UBQ10 promoter. In some embodiments, the UBQ10 promoter includes a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 2. In some embodiments that may be combined with any of the preceding embodiments, the second nucleic acid sequence includes a terminator sequence. In some embodiments, the terminator is a NOS terminator. In some embodiments, the NOS terminator includes a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 26. In some embodiments that may be combined with any of the preceding embodiments, the transcriptional activator is a VP64 polypeptide. In some embodiments, the VP64 polypeptide includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 31. In some embodiments that may be combined with any of the preceding embodiments, the affinity polypeptide is an antibody. In some embodiments, the antibody is an scFv antibody. In some embodiments, the antibody includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 28. In some embodiments that may be combined with any of the preceding embodiments, the second polypeptide includes one or more linkers that link polypeptide units

in the recombinant polypeptide. In some embodiments that may be combined with any of the preceding embodiments, the second polypeptide includes an SV40-type NLS. In some embodiments, the SV40-type NLS includes an amino acid sequence that is at least 80% identical to SEQ ID NO: 32. In some embodiments that may be combined with any of the preceding embodiments, the crRNA and the tracrRNA are fused together, thereby forming a guide RNA (gRNA). In some embodiments that may be combined with any of the preceding embodiments, the first and second nucleic acids are separated by a TBS insulator. In some embodiments, the TBS insulator includes a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 10.

[0008] In another aspect, the present disclosure provides a plant or plant cell including the vector of any of the preceding embodiments.

[0009] In another aspect, the present disclosure provides a method for activating expression of a target nucleic acid in a plant, including: (a) providing a plant including a vector of any one of the preceding embodiments; and (b) growing the plant under conditions whereby the first, second, and third nucleic acids in the vector are expressed and the resulting polypeptides are targeted to the target nucleic acid, thereby activating expression of the target nucleic acid. In some embodiments, expression of the activated nucleic acid is increased in the range of about 100-fold to about 10,000-fold as compared to a corresponding control.

[0010] In another aspect, the present disclosure provides a plant or plant cell including: a) a first recombinant polypeptide including a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope, b) a second recombinant polypeptide including a transcriptional activator and an affinity polypeptide that specifically binds to the epitope, and c) a crRNA and a tracrRNA, or fusions thereof. In some embodiments, the plant or plant cell includes a nucleic acid that has increased expression as compared to a corresponding control.

[0011] In another aspect, the present disclosure provides a plant or plant cell including: a) a first nucleic acid including a plant promoter and that encodes a recombinant polypeptide comprising a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope, b) a second nucleic acid including a plant promoter and that encodes a recombinant polypeptide comprising a transcriptional activator and an affinity polypeptide that specifically binds to the epitope, and c) a third nucleic acid including a promoter and that

encodes a crRNA and a tracrRNA, or fusions thereof. In some embodiments, the plant or plant cell includes a nucleic acid that has increased expression as compared to a corresponding control.

DESCRIPTION OF THE FIGURES

[0012] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the office upon request and payment of the necessary fee.

[0013] **FIG. 1** illustrates a schematic of the expression cassettes present in the vector housing the SunTag VP64 expression system.

[0014] **FIG. 2A-FIG. 2B** illustrates expression levels of sgRNA4 (**FIG. 2A**) and dCAS9 (**FIG. 2B**) in two independent T1 lines housing the SunTag VP64 construct, as well as in wild-type plants.

[0015] **FIG. 3** illustrates fluorescence microscopy of *N. benthamiana* plants transfected with the iteration of the SunTag VP64 vector where VP64 was fused to the NLS from Tanenbaum et al, 2014.

[0016] **FIG. 4** illustrates fluorescence microscopy of *N. benthamiana* plants transfected with the iteration of the SunTag VP64 vector where VP64 was fused to an SV40-type NLS.

[0017] **FIG. 5** illustrates fluorescence microscopy of T2 *A. thaliana* plants transformed with the iteration of the SunTag VP64 vector where VP64 was fused to an SV40-type NLS. Tissue shown is root tissue.

[0018] **FIG. 6A-FIG. 6B** illustrates *FWA* expression levels as determined by qRT-PCR in various lines. **FIG. 6A** illustrates *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*, *fwa* mutants, and two independent T1 lines housing the SunTag VP64 construct that contains gRNA4. **FIG. 6B** illustrates *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*, a T1 line housing the SunTag VP64 construct that does not contain any gRNA, and two independent T2 lines housing the SunTag VP64 construct that contains gRNA4.

[0019] **FIG. 7** illustrates *FWA* expression levels as determined by qRT-PCR in various lines. Shown is *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*,

fwa mutant, T1 lines housing the SunTag VP64 construct that does not contain any gRNA (lines 2, 3, and 4), and T1 lines housing the SunTag VP64 construct that contains gRNA4 (lines 10, 11, and 12).

[0020] **FIG. 8** illustrates *FWA* expression levels as determined by qRT-PCR in various lines. Shown is *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*, *fwa* mutants, T1 lines housing the SunTag VP64 construct that does not contain any gRNA (lines 2, 3, and 4), and T2 lines housing the SunTag VP64 construct that contains gRNA4 (lines 8, 9, 11, and 12).

[0021] **FIG. 9** illustrates *FWA* expression levels as determined by qRT-PCR in old and young leaf tissue from various Arabidopsis lines. The line samples include two wild-type *Col-0* samples, two *fwa* mutant samples, a T1 line housing the SunTag VP64 construct that does not contain any gRNA, and multiple T2 lines housing the SunTag VP64 construct that contains gRNA4.

[0022] **FIG. 10** illustrates flowering time in a SunTag VP64 + gRNA4 line and a control line that does not contain any guide RNA.

[0023] **FIG. 11** illustrates *FWA* expression levels as determined by qRT-PCR in various lines. Shown is *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*, a line housing the SunTag VP64 construct that contains gRNA17, and samples from lines housing the SunTag VP64 construct that contains gRNA4.

[0024] **FIG. 12** illustrates *FWA* expression levels as determined by qRT-PCR in various lines. Shown is *FWA* expression in the following Arabidopsis backgrounds: wild-type *Col-0*, a control line housing the SunTag VP64 construct that does not contain any gRNA, samples from lines housing the SunTag VP64 construct that contains gRNA17, and samples from lines housing the SunTag VP64 construct that contains gRNA4.

[0025] **FIG. 13** illustrates methylation analysis of the *FWA* promoter in various lines: wild-type *Col-0*, *fwa* mutants, and T2 lines (lines 1 and 2) housing the SunTag VP64 construct that contains gRNA4.

[0026] **FIG. 14** illustrates methylation analysis of the *FWA* promoter in various lines: wild-type *Col-0*, *fwa* mutants, and two lines (lines 11 and 12) housing the SunTag VP64 construct that contains gRNA4.

[0027] **FIG. 15** illustrates ChIP-qPCR results in wild-type plants and T2 lines housing the SunTag VP64 construct that contains gRNA4.

[0028] **FIG. 16** illustrates a screenshot of the genome browser analyzing the *FWA* promoter region in Cas9 ChIP samples from the SunTag VP64 + gRNA4 lines.

[0029] **FIG. 17** illustrates a screenshot of the genome browser analyzing a region of the genome (SEQ ID NO: 67) in Cas9 ChIP samples from the SunTag VP64 + gRNA4 lines that was bound by Cas9.

[0030] **FIG. 18** illustrates *GIS* expression levels as determined by qRT-PCR in various lines. Shown is *GIS* expression in the following Arabidopsis backgrounds: a T2 SunTag VP64 line that did not contain any gRNA, and a T1 line housing the SunTag VP64 construct that contains the tRNA:gRNA that targets *GIS*.

[0031] **FIG. 19** illustrates qRT-PCR data of the *FWA* locus. One SunTag no guide negative control line, one SunTag VP64 sgRNA4 positive control line, and four independent T1 SunTag no VP64 sgRNA4 lines are shown. Fold upregulation is plotted and a housekeeping gene, *IPP2*, was used as an internal control. Error bars indicate standard error of the mean of two replicates.

[0032] **FIG. 20A, FIG. 20B, and FIG. 20C** illustrate qRT-PCR data for *EVD* in T1 plants. Each of **FIG. 20A, FIG. 20B, and FIG. 20C** represent a separate experiment screening different T1 plants. Each T1 plant contains two guides targeting *EVD*. Col (wild type) and no guide samples are included as negative controls. “Control line” indicates a control with guides targeting the unrelated *SUPERMAN* gene. Fold upregulation is plotted and a housekeeping gene, *IPP2*, was used as an internal control. Error bars indicate standard error of the mean of two replicates.

[0033] **FIG. 21** illustrates qRT-PCR data for *EVD*. Expression data is from T2 plants from three independent lines with two guides targeting *EVD*. Three control lines are included as negative controls. The first one corresponds to a no guide control, while the next two

correspond to plants expressing gRNAs targeting the unrelated *SUPERMAN* gene. Fold upregulation is plotted and a housekeeping gene, *IPP2*, was used as an internal control. Error bars indicate standard error of the mean of two replicates.

[0034] **FIG. 22A – FIG. 22B** illustrate scatterplots showing significantly differentially expressed genes (DEGs) in T2 plants targeting *EVD* with two guides. Results for line 3 (**FIG. 22A**) and line 10 (**FIG. 22B**) are shown. Genes which showed at least a four-fold change are shown. Both upregulated copies of the *EVD* locus are labeled.

[0035] **FIG. 23** illustrates qRT-PCR data for *AP3*. Expression data is from two independent T1 lines with two guides targeting the promoter region of *AP3*. A control line expressing guides targeting the unrelated *EVD* gene is included as a negative control. Fold upregulation is plotted and a housekeeping gene, *IPP2*, was used as an internal control. Error bars indicate standard error of the mean of two replicates.

[0036] **FIG. 24** illustrates qRT-PCR data for *CLV3*. Expression data is from one T1 line with two guides targeting the promoter region of *CLV3*. A wild type sample is included as a negative control. Fold upregulation is plotted and a housekeeping gene, *IPP2*, was used as an internal control. Error bars indicate standard error of the mean of two replicates.

DETAILED DESCRIPTION

[0037] The following description is presented to enable a person of ordinary skill in the art to make and use the various embodiments. Descriptions of specific devices, techniques, methods, and applications are provided only as examples. Various modifications to the examples described herein will be readily apparent to those of ordinary skill in the art, and the general principles defined herein may be applied to other examples and applications without departing from the spirit and scope of the various embodiments. Thus, the various embodiments are not intended to be limited to the examples described herein and shown, but are to be accorded the scope consistent with the claims.

[0038] The present disclosure relates to the targeting of transcriptional activators to specific loci in plants to activate transcription of the targeted loci. Specifically, the present disclosure provides methods and compositions for using RNA-guided transcriptional activators to activate transcription of specific loci in plants.

[0039] The present disclosure relates to the amplification of transcriptional activation at a target nucleic acid. Recruitment of multiple copies of a protein to a target substrate (e.g. DNA, RNA, or protein) may amplify signals in biological systems. When the protein is a transcriptional activator, providing multiple copies of that transcriptional activator may result in amplification of the expression of that nucleic acid.

[0040] A synthetic system was previously developed for use in mammals for recruiting multiple copies of a protein to a target polypeptide chain, and this system was called a SunTag system (Tanenbaum et al., 2014)(WO2016011070). This system was also adapted so that the multiple copies of the protein using the SunTag system could be targeted to a nucleic acid using the CRISPR-Cas9 system (Tanenbaum et al., 2014). However, this system was developed for use in mammals.

[0041] The present disclosure is based, at least in part, on Applicant's development of a SunTag gene activation system that is functional in plants. The present disclosure provides the successful construction of a SunTag system that is operable in plants and uses CRISPR-based targeting to target a transcriptional activator to specific nucleic acids. This SunTag system was able to substantially increase expression of targeted genes as compared to corresponding controls. In some instances, many thousands-fold increases in gene expression were observed. This work presents the opportunity for robust and selective activation of plant genes or other nucleic acids in plants, which may serve both research purposes as well as be used in applications for crop improvement.

[0042] Accordingly, the present disclosure provides methods and compositions for the recruitment of multiple copies of a transcriptional activator to a target nucleic acid in plants via CRISPR-based targeting in a manner that allows for transcriptional activation of the target nucleic acid. In certain aspects, this specific targeting involves the use of a system that includes (1) a nuclease-deficient CAS9 polypeptide that is recombinantly fused to a multimerized epitope, (2) a transcriptional activator polypeptide that is recombinantly fused to an affinity polypeptide, and (3) a guide RNA (gRNA). In this aspect, the dCAS9 portion of the dCAS9-multimerized epitope fusion protein is involved with targeting a target nucleic acid as directed by the guide RNA. The multimerized epitope portion of the dCAS9-multimerized epitope fusion protein is involved with binding to the affinity polypeptide (which is recombinantly fused to a transcriptional activator). The affinity polypeptide portion of the transcriptional activator-affinity polypeptide fusion protein is involved with binding to

the multimerized epitope so that the transcriptional activator can be in association with dCAS9. The transcriptional activator portion of the transcriptional activator-affinity polypeptide fusion protein is involved with activating transcription of a target nucleic acid, once the complex has been targeted to a target nucleic acid via the guide RNA.

[0043] As described above, certain aspects of the present disclosure involve CRISPR-based targeting of a target nucleic acid, which involves use of a CRISPR-CAS9 targeting system. CRISPR-CAS9 systems involve the use of a CRISPR RNA (crRNA), a trans-activating CRISPR RNA (tracrRNA), and a CAS9 protein. The crRNA and tracrRNA aid in directing the CAS9 protein to a target nucleic acid sequence, and these RNA molecules can be specifically engineered to target specific nucleic acid sequences. In particular, certain aspects of the present disclosure involve the use of a single guide RNA (gRNA) that reconstitutes the function of the crRNA and the tracrRNA. Further, certain aspects of the present disclosure involve a CAS9 protein that does not exhibit DNA cleavage activity (dCAS9). As disclosed herein, gRNA molecules may be used to direct a dCAS9 protein to a target nucleic acid sequence.

[0044] The use of the terms “a,” “an,” and “the,” and similar referents in the context of describing the disclosure (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms “comprising,” “having,” “including,” and “containing” are to be construed as open-ended terms (*i.e.*, meaning “including, but not limited to,”) unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. For example, if the range 10-15 is disclosed, then 11, 12, 13, and 14 are also disclosed. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (*e.g.*, “such as”) provided herein, is intended merely to better illuminate the embodiments of the disclosure and does not pose a limitation on the scope of the disclosure unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the embodiments of the disclosure.

[0045] Reference to “about” a value or parameter herein refers to the usual error range for the respective value readily known to the skilled person in this technical field. Reference to “about” a value or parameter herein includes (and describes) aspects that are directed to that value or parameter *per se*. For example, description referring to “about X” includes description of “X.”

[0046] It is understood that aspects and embodiments of the present disclosure described herein include “comprising,” “consisting,” and “consisting essentially of” aspects and embodiments.

[0047] It is to be understood that one, some, or all of the properties of the various embodiments described herein may be combined to form other embodiments of the present disclosure. These and other aspects of the present disclosure will become apparent to one of skill in the art. These and other embodiments of the present disclosure are further described by the detailed description that follows.

[0048] The terms “isolated” and “purified” as used herein refers to a material that is removed from at least one component with which it is naturally associated (e.g., removed from its original environment). The term “isolated,” when used in reference to an isolated protein, refers to a protein that has been removed from the culture medium of the host cell that expressed the protein. As such an isolated protein is free of extraneous or unwanted compounds (e.g., nucleic acids, native bacterial or other proteins, etc.).

Recombinant Polypeptides

[0049] The present disclosure relates to the use of recombinant polypeptides to activate expression of a target nucleic acid. In certain aspects, the targeting involves the use of a nuclease-deficient CAS9 polypeptide that is recombinantly fused to a multimerized epitope. In certain aspects, the targeting involves the use of a transcriptional activator polypeptide that is recombinantly fused to an affinity polypeptide.

[0050] As used herein, a “polypeptide” is an amino acid sequence including a plurality of consecutive polymerized amino acid residues (*e.g.*, at least about 15 consecutive polymerized amino acid residues). “Polypeptide” refers to an amino acid sequence, oligopeptide, peptide, protein, or portions thereof, and the terms “polypeptide” and “protein” are used interchangeably.

[0051] Polypeptides as described herein also include polypeptides having various amino acid additions, deletions, or substitutions relative to the native amino acid sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain non-conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure, and thus may be referred to as conservatively modified variants. A conservatively modified variant may include individual substitutions, deletions or additions to a polypeptide sequence which result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables 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 disclosure. The following eight groups contain amino acids that are conservative 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 8) Cysteine (C), Methionine (M) (see, *e.g.*, Creighton, *Proteins* (1984)). A modification of an amino acid to produce a chemically similar amino acid may be referred to as an analogous amino acid.

[0052] Recombinant polypeptides of the present disclosure that are composed of individual polypeptide domains may be described based on the individual polypeptide domains of the overall recombinant polypeptide. A domain in such a recombinant polypeptide refers to the particular stretches of contiguous amino acid sequences with a particular function or activity. For example, a recombinant polypeptide that is a fusion of a transcriptional activator polypeptide and an affinity polypeptide, the contiguous amino acids that encode the transcriptional activator polypeptide may be described as the transcriptional activator domain in the overall recombinant polypeptide, and the contiguous amino acids that encode the affinity polypeptide may be described as the affinity domain in the overall recombinant polypeptide. Individual domains in an overall recombinant protein may also be referred to as units of the recombinant protein. Recombinant polypeptides that are composed of individual polypeptide domains may also be referred to as fusion polypeptides.

[0053] Certain aspects of the present disclosure relate to a nuclease-deficient CAS9 polypeptide that is recombinantly fused to a multimerized epitope (*e.g.* dCAS9-multimerized epitope fusion protein). The dCAS9 polypeptide domain of a dCAS9-multimerized epitope fusion protein may be in an N-terminal orientation or a C-terminal orientation relative to the multimerized epitope domain. The multimerized epitope domain of a dCAS9-multimerized epitope fusion protein may be in an N-terminal orientation or a C-terminal orientation relative to the dCAS9 polypeptide domain. In some embodiments, a dCAS9-multimerized epitope fusion protein may be a direct fusion of a dCAS9 polypeptide domain and a multimerized epitope domain. In some embodiments, a dCAS9-multimerized epitope fusion protein may be an indirect fusion of a dCAS9 polypeptide domain and a multimerized epitope domain. In embodiments where the fusion is indirect, a linker domain or other contiguous amino acid sequence may separate the dCAS9 polypeptide domain and the multimerized epitope domain.

[0054] Certain aspects of the present disclosure relate to a transcriptional activator polypeptide that is recombinantly fused to an affinity polypeptide (*e.g.* transcriptional activator-affinity polypeptide fusion protein). The transcriptional activator polypeptide domain of a transcriptional activator-affinity polypeptide fusion protein may be in an N-terminal orientation or a C-terminal orientation relative to the affinity polypeptide. The affinity polypeptide domain of a transcriptional activator-affinity polypeptide fusion protein may be in an N-terminal orientation or a C-terminal orientation relative to the transcriptional activator polypeptide domain. In some embodiments, a transcriptional activator-affinity polypeptide fusion protein may be a direct fusion of a transcriptional activator polypeptide domain and an affinity polypeptide domain. In some embodiments, a transcriptional activator-affinity polypeptide fusion protein may be an indirect fusion of a transcriptional activator polypeptide domain and an affinity polypeptide domain. In embodiments where the fusion is indirect, a linker domain or other contiguous amino acid sequence may separate the transcriptional activator polypeptide domain and the affinity polypeptide domain.

Linkers

[0055] Various linkers may be used in the construction of recombinant proteins as described herein. In general, linkers are short peptides that separate the different domains in a multi-domain protein. They may play an important role in fusion proteins, affecting the crosstalk between the different domains, the yield of protein production, and the stability and/or the activity of the fusion proteins. Linkers are generally classified into 2 major

categories: flexible or rigid. Flexible linkers are typically used when the fused domains require a certain degree of movement or interaction, and these linkers are usually composed of small amino acids such as, for example, glycine (G), serine (S) or proline (P).

[0056] The certain degree of movement between domains allowed by flexible linkers is an advantage in some fusion proteins. However, it has been reported that flexible linkers can sometimes reduce protein activity due to an inefficient separation of the two domains. In this case, rigid linkers may be used since they enforce a fixed distance between domains and promote their independent functions. A thorough description of several linkers has been provided in Chen X et al., 2013, *Advanced Drug Delivery Reviews* 65 (2013) 1357–1369).

[0057] Various linkers may be used in, for example, the construction of recombinant polypeptides as described herein. Linkers may be used in *e.g.* dCAS9-multimerized epitope fusion proteins as described herein to separate the coding sequences of the dCAS9 polypeptide and the multimerized epitope polypeptide. Linkers may be used in *e.g.* transcriptional activator-affinity polypeptide fusion proteins as described herein to separate the coding sequences of the transcriptional activator polypeptide and the affinity polypeptide. For example, a variety of wiggly/flexible linkers, stiff/rigid linkers, short linkers, and long linkers may be used as described herein. Various linkers as described herein may be used in the construction of recombinant proteins as described herein.

[0058] A variety of shorter or longer linker regions are known in the art, for example corresponding to a series of glycine residues, a series of adjacent glycine-serine dipeptides, a series of adjacent glycine-glycine-serine tripeptides, or known linkers from other proteins. A flexible linker may include, for example, the amino acid sequence: SSGPPPGTG (SEQ ID NO: 64) and variants thereof. A rigid linker may include, for example, the amino acid sequence: AEAAAKEAAKA (SEQ ID NO: 65) and variants thereof. The XTEN linker, SGSETPGTSESATPES (SEQ ID NO: 66), and variants thereof, described in Guilinget et al, 2014 (*Nature Biotechnology* 32, 577–582), may also be used. This particular linker was previously shown to produce the best results among other linkers in a protein fusion between dCAS9 and the nuclease FokI.

[0059] Recombinant polypeptides of the present disclosure may contain one or more linkers that contain an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about

60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of any one of SEQ ID NO: 15 and/or SEQ ID NO: 30.

Nuclear Localization Signals (NLS)

[0060] Recombinant polypeptides of the present disclosure may contain one or more nuclear localization signals (NLS). Nuclear localization signals may also be referred to as nuclear localization sequences, domains, peptides, or other terms readily apparent to those of skill in the art. Nuclear localization signals are a translocation sequence that, when present in a polypeptide, direct that polypeptide to localize to the nucleus of a eukaryotic cell.

[0061] Various nuclear localization signals may be used in recombinant polypeptides of the present disclosure. For example, one or more SV40-type NLS or one or more REX NLS may be used in recombinant polypeptides. Recombinant polypeptides may also contain two or more tandem copies of a nuclear localization signal. For example, recombinant polypeptides may contain at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten copies, either tandem or not, of a nuclear localization signal.

[0062] Recombinant polypeptides of the present disclosure may contain one or more nuclear localization signals that contain an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of any one of SEQ ID NO: 14, SEQ ID NO: 32, and/or SEQ ID NO: 34.

Tags, Reporters, and Other Features

[0063] Recombinant polypeptides of the present disclosure may contain one or more tags that allow for *e.g.* purification and/or detection of the recombinant polypeptide. Various tags may be used herein and are well-known to those of skill in the art. Exemplary tags may

include HA, GST, FLAG, MBP, etc., and multiple copies of one or more tags may be present in a recombinant polypeptide.

[0064] Recombinant polypeptides of the present disclosure may contain one or more reporters that allow for *e.g.* visualization and/or detection of the recombinant polypeptide. A reporter polypeptide encodes a protein that may be readily detectable due to its biochemical characteristics such as, for example, enzymatic activity or chemifluorescent features. Reporter polypeptides may be detected in a number of ways depending on the characteristics of the particular reporter. For example, a reporter polypeptide may be detected by its ability to generate a detectable signal (*e.g.* fluorescence), by its ability to form a detectable product, etc. Various reporters may be used herein and are well-known to those of skill in the art. Exemplary reporters may include GFP, GUS, mCherry, luciferase, etc., and multiple copies of one or more tags may be present in a recombinant polypeptide.

[0065] Recombinant polypeptides of the present disclosure may contain one or more polypeptide domains that serve a particular purpose depending on the particular goal/need. For example, recombinant polypeptides may contain a GB1 polypeptide. Recombinant polypeptides may contain translocation sequences that target the polypeptide to a particular cellular compartment or area. Suitable features will be readily apparent to those of skill in the art.

Transcriptional Activators

[0066] Certain aspects of the present disclosure involve targeting a transcriptional activator to a target nucleic acid such that the transcriptional activator activates the expression/transcription of the target nucleic acid. In some embodiments, a transcriptional activator is present in a recombinant polypeptide that contains a transcriptional activator polypeptide and an affinity polypeptide.

[0067] Transcriptional activators are polypeptides that facilitate the activation of transcription/expressions of a nucleic acid (*e.g.* a gene). Transcriptional activators may be DNA-binding proteins that bind to enhancers, promoters, or other regulatory elements of a nucleic acid, which then promotes expression of the nucleic acid. Transcriptional activators may interact with proteins that are components of transcriptional machinery or other proteins that are involved in regulation of transcription in a manner that promotes expression of the nucleic acid.

[0068] Transcriptional activators of the present disclosure may be endogenous to the host plant, or they may be exogenous/heterologous to the host plant. In some embodiments, the transcriptional activator is a viral transcriptional activator. In some embodiments, the transcriptional activator is derived from Herpes Simplex Virus. For example, one or more copies of a Herpes Simplex Virus Viral Protein 16 (VP16) domain may be used herein. In some embodiments, at least two, at least three, or at least four or more copies of a VP16 domain may be used as a transcriptional activator. A polypeptide containing 4 copies of the Herpes Simplex Virus Viral Protein 16 (VP16) domain is known as a VP64 domain.

[0069] In some embodiments, the transcriptional activator is a VP64 polypeptide. A VP64 polypeptide of the present disclosure may contain an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of SEQ ID NO: 31.

[0070] Other exemplary transcriptional activators include, for example, the the EDLL motif present in the ERF/EREBP family of transcriptional regulators in plants, activation domains of or full-length transcription factors, the TAL activation domain derived from the transcription activator-like effector (TALE) proteins from the plant pathogen *Xanthomonas*, plant endogenous and exogenous histone acetylases (*e.g.* p300 from mammals), histone methylases (*e.g.* H3K4 methylation depositors (*SDG2*)), histone demethylases (*e.g.* H3K9 demethylases (*IBM1*)), Polymerase II subunits, and various combinations of the above mentioned transcriptional activators. For example, VP64 and EDLL may each be fused to an scFv antibody in the SunTag system and co-expressed for targeted activation. In the latter case, each fusion would bind to the epitope tail fused to dCas9.

[0071] Additional transcriptional activators that may be used in the methods and compositions described herein will be readily apparent to those of skill in the art.

Affinity Polypeptides

[0072] Certain aspects of the present disclosure relate to recombinant polypeptides that contain an affinity polypeptide. Affinity polypeptides of the present disclosure may bind to

one or more epitopes (*e.g.* a multimerized epitope). In some embodiments, an affinity polypeptide is present in a recombinant polypeptide that contains a transcriptional activator polypeptide and an affinity polypeptide.

[0073] A variety of affinity polypeptides are known in the art and may be used herein. Generally, the affinity polypeptide should be stable in the conditions present in the intracellular environment of a plant cell. Additionally, the affinity polypeptide should specifically bind to its corresponding epitope with minimal cross-reactivity.

[0074] The affinity polypeptide may be an antibody such as, for example, an scFv. The antibody may be optimized for stability in the plant intracellular environment. When a GCN4 epitope is used in the methods described herein, a suitable affinity polypeptide that is an antibody may contain an anti-GCN4 scFv domain.

[0075] In embodiments where the affinity polypeptide is an scFv antibody, the polypeptide may contain an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of SEQ ID NO: 28.

[0076] Other exemplary affinity polypeptides include, for example, proteins with SH2 domains or the domain itself, 14-3-3 proteins, proteins with SH3 domains or the domain itself, the Alpha-Syntrophin PDZ protein interaction domain, the PDZ signal sequence, or proteins from plants which can recognize AGO hook motifs (*e.g.* AGO4 from *Arabidopsis thaliana*).

[0077] Additional affinity polypeptides that may be used in the methods and compositions described herein will be readily apparent to those of skill in the art.

Epitopes and Multimerized Epitopes

[0078] Certain aspects of the present disclosure relate to recombinant polypeptides that contain an epitope or a multimerized epitope. Epitopes of the present disclosure may bind to

an affinity polypeptide. In some embodiments, an epitope or multimerized epitope is present in a recombinant polypeptide that contains dCAS9 polypeptide.

[0079] Epitopes of the present disclosure may be used for recruiting affinity polypeptides (and any polypeptides they may be recombinantly fused to) to a dCAS9 polypeptide. In embodiments where a dCAS9 polypeptide is fused to an epitope or a multimerized epitope, the dCAS9 polypeptide may be fused to one copy of 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.

[0080] A variety of epitopes and multimerized epitopes are known in the art and may be used herein. In general, the epitope or multimerized epitope may be any polypeptide sequence that is specifically recognized by an affinity polypeptide of the present disclosure. Exemplary epitopes may include a 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 FLAG octapeptide, a strep tag or strep tag II, a V5 tag, a VSV-G epitope, and a GCN4 epitope.

[0081] Other exemplary amino acid sequences that may serve as epitopes and multimerized epitopes include, for example, phosphorylated tyrosines in specific sequence contexts recognized by SH2 domains, characteristic consensus sequences containing phosphoserines recognized by 14-3-3 proteins, proline rich peptide motifs recognized by SH3 domains, the PDZ protein interaction domain or the PDZ signal sequence, and the AGO hook motif from plants.

[0082] Epitopes described herein may also be multimerized. Multimerized epitopes may include at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, or at least 24 or more copies of an epitope.

[0083] Multimerized epitopes may be present as tandem copies of an epitope, or each individual epitope may be separated from another epitope in the multimerized epitope by a linker or other amino acid sequence. Suitable linker regions are known in the art and are described herein. The linker may be configured to allow the binding of affinity polypeptides to adjacent epitopes without, or without substantial, steric hindrance. Linker sequences may also be configured to provide an unstructured or linear region of the polypeptide to which

they are recombinantly fused. The linker sequence may comprise *e.g.* one or more glycines and/or serines. The linker sequences may be *e.g.* at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 or more amino acids in length.

[0084] In some embodiments, the epitope is a GCN4 epitope (SEQ ID NO: 47). In some embodiments, the multimerized epitope contains at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, or at least 24 copies of a GCN4 epitope (SEQ ID NO: 47). In some embodiments, the multimerized epitope contains 10 copies of a GCN4 epitope (SEQ ID NO: 16).

[0085] Additional epitopes and multimerized epitopes that may be used in the methods and compositions described herein will be readily apparent to those of skill in the art.

CRISPR-CAS9

[0086] Certain aspects of the present disclosure involve CRISPR-based targeting of a transcriptional activator to a target nucleic acid, which involves use of a CRISPR-CAS9 targeting system. In some embodiments, an epitope or multimerized epitope of the present disclosure is present in a recombinant polypeptide that contains dCAS9 polypeptide.

[0087] CRISPR systems naturally use small base-pairing guide RNAs to target and cleave foreign DNA elements in a sequence-specific manner (Wiedenheft *et al.*, 2012). There are diverse CRISPR systems in different organisms that may be used to target proteins of the present disclosure to a target nucleic acid. One of the simplest systems is the type II CRISPR system from *Streptococcus pyogenes*. Only a single gene encoding the CAS9 protein and two RNAs, a mature CRISPR RNA (crRNA) and a partially complementary trans-acting RNA (tracrRNA), are necessary and sufficient for RNA-guided silencing of foreign DNAs (Jinek *et al.*, 2012). Maturation of crRNA requires tracrRNA and RNase III (Deltcheva *et al.*, 2011). However, this requirement can be bypassed by using an engineered small guide RNA (gRNA) containing a designed hairpin that mimics the tracrRNA-crRNA complex (Jinek *et al.*, 2012). Base pairing between the gRNA and target DNA normally causes double-strand breaks (DSBs) due to the endonuclease activity of CAS9.

[0088] It is known that the endonuclease domains of the CAS9 protein can be mutated to create a programmable RNA-dependent DNA-binding protein (dCAS9) (Qi *et al.*, 2013). The fact that duplex gRNA-dCAS9 binds target sequences without endonuclease activity has been used to tether regulatory proteins, such as transcriptional activators or repressors, to promoter regions in order to modify gene expression (Gilbert *et al.*, 2013), and CAS9 transcriptional activators have been used for target specificity screening and paired nickases for cooperative genome engineering (Mali *et al.*, 2013, Nature Biotechnology 31:833-838). Thus, dCAS9 may be used as a modular RNA-guided platform to recruit different proteins to DNA in a highly specific manner. One of skill in the art would recognize other RNA-guided DNA binding protein/RNA complexes that can be used equivalently to CRISPR-CAS9.

CAS9 Proteins

[0089] A variety of CAS9 proteins may be used in the methods of the present disclosure. There are several CAS9 genes present in different bacteria species (Esvelt, K et al, 2013, Nature Methods). One of the most characterized CAS9 proteins is the CAS9 protein from *S. pyogenes* that, in order to be active, needs to bind a gRNA with a specific sequence and the presence of a PAM motif (NGG, where N is any nucleotide) at the 3' end of the target locus. However, other CAS9 proteins from different bacterial species show differences in 1) the sequence of the gRNA they can bind and 2) the sequence of the PAM motif. Therefore, it is possible that other CAS9 proteins such as, for example, those from *Streptococcus thermophilus* or *N. meningitidis* may also be utilized herein. Indeed, these two CAS9 proteins have a smaller size (around 1100 amino acids) as compared to *S. pyogenes* CAS9 (1400 amino acids), which may confer some advantages during cloning or protein expression.

[0090] CAS9 proteins from a variety of bacteria have been used successfully in engineered CRISPR-CAS9 systems. There are also versions of CAS9 proteins available in which the codon usage has been more highly optimized for expression in eukaryotic systems, such as human codon optimized CAS9 (Cell, 152:1173-1183) and plant optimized CAS9 (Nature Biotechnology, 31:688-691).

[0091] CAS9 proteins may also be modified for various purposes. For example, CAS9 proteins may be engineered to contain a nuclear-localization sequence (NLS). CAS9 proteins may be engineered to contain an NLS at the N-terminus of the protein, at the C-terminus of the protein, or at both the N- and C-terminus of the protein. Engineering a CAS9 protein to

contain an NLS may assist with directing the protein to the nucleus of a host cell. One of skill in the art would be able to readily identify a suitable CAS9 protein for use in the methods and compositions of the present disclosure.

[0092] Exemplary CAS9 proteins that may be used in the methods and compositions of the present disclosure may include, for example, a CAS9 protein having the amino acid sequence of any one of SEQ ID NO: 48, SEQ ID NO: 49, and/or SEQ ID NO: 50, homologs thereof, and fragments thereof. In some embodiments, the CAS9 polypeptide is a dCAS9 polypeptide. dCAS9 polypeptides may contain an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of SEQ ID NO: 12.

CRISPR RNAs

[0093] The CRISPR RNA (crRNA) of the present disclosure may take a variety of forms. As described above, the sequence of the crRNA is involved in conferring specificity to targeting a specific nucleic acid.

[0094] Many different crRNA molecules can be designed to target many different sequences. With respect to targeting, target nucleic acids generally require the PAM sequence, NGG, at the end of the 20 base pair target sequence. crRNAs of the present disclosure may be expressed as a single crRNA molecule, or they may be expressed in the form of a crRNA/tracrRNA hybrid molecule where the crRNA and the tracrRNA have been fused together, forming a guide RNA (gRNA). crRNA molecules and/or guide RNA molecules may be extended to include sites for the binding of RNA binding proteins.

[0095] Multiple crRNAs and/or guide RNAs can be encoded into a single CRISPR array to enable simultaneous targeting to several sites (Science 2013: Vol. pp. 819-823). For example, the tracrRNA may be expressed separately, and two adjacent target sequences may be encoded in a pre-crRNA array interspaced with repeats.

[0096] A variety of promoters may be used to drive expression of the crRNA and/or the guide RNA. crRNAs and/or guide RNAs may be expressed using a Pol III promoter such as, for example, the U6 promoter or the H1 promoter (eLife 2013 2:e00471). For example, an approach in plants has been described using three different Pol III promoters from three different Arabidopsis U6 genes, and their corresponding gene terminators (BMC Plant Biology 2014 14:327). One skilled in the art would readily understand that many additional Pol III promoters could be utilized to simultaneously express many crRNAs and/or guide RNAs to many different locations in the genome simultaneously. The use of different Pol III promoters for each crRNA and/or gRNA expression cassette may be desirable to reduce the chances of natural gene silencing that can occur when multiple copies of identical sequences are expressed in plants. In addition, crRNAs and/or guide RNAs can be modified to improve the efficiency of their function in guiding CAS9 to a target nucleic acid. For example, it has been shown that adding either 8 or 20 additional nucleotides to the gRNA in order to extend the hairpin by 4 or 10 base pairs resulted in more efficient CAS9 activity (eLife 2013 2:e00471).

[0097] In some embodiments, the guide RNA is driven by a U6 promoter. In some embodiments, the guide RNA is driven by a promoter having a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 36.

[0098] Alternatively, a tRNA-gRNA expression cassette (Xie, X et al, 2015, Proc Natl Acad Sci U S A. 2015 Mar 17;112(11):3570-5) may be used to deliver multiple gRNAs simultaneously with high expression levels. In such an embodiment, a tRNA in such a cassette may have a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about

98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 43.

Trans-activating CRISPR RNAs

[0099] The trans-activating CRISPR RNA (tracrRNA) of the present disclosure may take a variety of forms, as will be readily understood by one of skill in the art. As described above, tracrRNAs are involved in the maturation of a crRNA. tracrRNAs of the present disclosure may be expressed as a single tracrRNA molecule, or they may be expressed in the form of a crRNA/tracrRNA hybrid molecule where the crRNA and the tracrRNA have been fused together, forming a guide RNA (gRNA). tracrRNA molecules and/or guide RNA molecules may be extended to include sites for the binding of RNA binding proteins.

[0100] As CRISPR systems naturally exist in a variety of bacteria, the framework of the crRNA and tracrRNA in these bacteria may be adapted for use in the methods and compositions described herein. crRNAs, tracrRNAs, and/or guide RNAs of the present disclosure may be constructed based on the framework of one or more of these molecules in, for example, *S.pyogenes*, *Streptococcus thermophilus*, and/or *N. meningitidis*. For example, a guide RNA of the present disclosure may be constructed based on the framework of the crRNA and tracrRNA from *S.pyogenes* (SEQ ID NO: 51), *Streptococcus thermophilus* (SEQ ID NO: 52), and/or *N. meningitidis* (SEQ ID NO: 53). In these exemplary frameworks, the 5' end of the sequence contains 20 generic nucleotides (N) that correspond to the crRNA targeting sequence. This sequence will vary depending on the sequence of the particular nucleic acid being targeted.

[0101] In some embodiments, the tracrRNA component may have a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 38.

Recombinant Nucleic Acids Encoding Recombinant Proteins

[0102] Certain aspects of the present disclosure relate to recombinant nucleic acids encoding recombinant proteins of the present disclosure. Certain aspects of the present disclosure relate to recombinant nucleic acids encoding various portions/domains of recombinant proteins of the present disclosure.

[0103] As used herein, the terms “polynucleotide,” “nucleic acid,” and variations thereof shall be generic to polydeoxyribonucleotides (containing 2-deoxy-D-ribose), to polyribonucleotides (containing D-ribose), to any other type of polynucleotide that is an N-glycoside of a purine or pyrimidine base, and to other polymers containing non-nucleotidic backbones, provided that the polymers contain nucleobases in a configuration that allows for base pairing and base stacking, as found in DNA and RNA. Thus, these terms include known types of nucleic acid sequence modifications, for example, substitution of one or more of the naturally occurring nucleotides with an analog, and inter-nucleotide modifications. As used herein, the symbols for nucleotides and polynucleotides are those recommended by the IUPAC-IUB Commission of Biochemical Nomenclature.

[0104] In some embodiments, a recombinant nucleic acid is provided containing a plant promoter and that encodes a recombinant polypeptide containing a nuclease-deficient CAS9 polypeptide (dCAS9) and a multimerized epitope. This recombinant nucleic acid may encode a recombinant polypeptide having an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of SEQ ID NO: 11.

[0105] In some embodiments, a recombinant nucleic acid is provided containing a plant promoter and that encodes recombinant polypeptide containing a transcriptional activator and an affinity polypeptide. This recombinant nucleic acid may encode a recombinant polypeptide having an amino acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least

about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% amino acid identity to the amino acid sequence of SEQ ID NO: 27.

[0106] Recombinant nucleic acids are also provided that have a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, and SEQ ID NO: 45.

[0107] Sequences of the polynucleotides of the present disclosure may be prepared by various suitable methods known in the art, including, for example, direct chemical synthesis or cloning. For direct chemical synthesis, formation of a polymer of nucleic acids typically involves sequential addition of 3'-blocked and 5'-blocked nucleotide monomers to the terminal 5'-hydroxyl group of a growing nucleotide chain, wherein each addition is effected by nucleophilic attack of the terminal 5'-hydroxyl group of the growing chain on the 3'-position of the added monomer, which is typically a phosphorus derivative, such as a phosphotriester, phosphoramidite, or the like. Such methodology is known to those of ordinary skill in the art and is described in the pertinent texts and literature (*e.g.*, in Matteucci *et al.*, (1980) Tetrahedron Lett 21:719-722; U.S. Pat. Nos. 4,500,707; 5,436,327; and 5,700,637). In addition, the desired sequences may be isolated from natural sources by splitting DNA using appropriate restriction enzymes, separating the fragments using gel electrophoresis, and thereafter, recovering the desired polynucleotide sequence from the gel via techniques known to those of ordinary skill in the art, such as utilization of polymerase chain reactions (PCR; *e.g.*, U.S. Pat. No. 4,683,195).

[0108] The nucleic acids employed in the methods and compositions described herein may be codon optimized relative to a parental template for expression in a particular host cell. Cells differ in their usage of particular codons, and codon bias corresponds to relative abundance of particular tRNAs in a given cell type. By altering codons in a sequence so that they are tailored to match with the relative abundance of corresponding tRNAs, it is possible to increase expression of a product (*e.g.* a polypeptide) from a nucleic acid. Similarly, it is possible to decrease expression by deliberately choosing codons corresponding to rare tRNAs. Thus, codon optimization/deoptimization can provide control over nucleic acid expression in a particular cell type (*e.g.* bacterial cell, plant cell, mammalian cell, etc.). Methods of codon optimizing a nucleic acid for tailored expression in a particular cell type are well-known to those of skill in the art.

Methods of Identifying Sequence Similarity

[0109] Various methods are known to those of skill in the art for identifying similar (*e.g.* homologs, orthologs, paralogs, etc.) polypeptide and/or polynucleotide sequences, including phylogenetic methods, sequence similarity analysis, and hybridization methods.

[0110] Phylogenetic trees may be created for a gene family by using a program such as CLUSTAL (Thompson et al. *Nucleic Acids Res.* 22: 4673-4680 (1994); Higgins et al. *Methods Enzymol* 266: 383-402 (1996)) or MEGA (Tamura et al. *Mol. Biol. & Evo.* 24:1596-1599 (2007)). Once an initial tree for genes from one species is created, potential orthologous sequences can be placed in the phylogenetic tree and their relationships to genes from the species of interest can be determined. Evolutionary relationships may also be inferred using the Neighbor-Joining method (Saitou and Nei, *Mol. Biol. & Evo.* 4:406-425 (1987)). Homologous sequences may also be identified by a reciprocal BLAST strategy. Evolutionary distances may be computed using the Poisson correction method (Zuckerkandl and Pauling, pp. 97-166 in *Evolving Genes and Proteins*, edited by V. Bryson and H.J. Vogel. Academic Press, New York (1965)).

[0111] In addition, evolutionary information may be used to predict gene function. Functional predictions of genes can be greatly improved by focusing on how genes became similar in sequence (*i.e.* by evolutionary processes) rather than on the sequence similarity itself (Eisen, *Genome Res.* 8: 163-167 (1998)). Many specific examples exist in which gene function has been shown to correlate well with gene phylogeny (Eisen, *Genome Res.* 8: 163-

167 (1998)). By using a phylogenetic analysis, one skilled in the art would recognize that the ability to deduce similar functions conferred by closely-related polypeptides is predictable.

[0112] When a group of related sequences are analyzed using a phylogenetic program such as CLUSTAL, closely related sequences typically cluster together or in the same clade (a group of similar genes). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle, *J. Mol. Evol.* 25: 351-360 (1987)). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous sequences, or orthologous sequences that share the same function (see also, for example, Mount, *Bioinformatics: Sequence and Genome Analysis* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., page 543 (2001)).

[0113] To find sequences that are homologous to a reference sequence, BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the disclosure. BLAST protein searches can be performed with the BLASTX program, score=50, wordlength=3, to obtain amino acid sequences homologous to a protein or polypeptide of the disclosure. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al. (1997) *supra*. When utilizing BLAST, Gapped BLAST, or PSI-BLAST, the default parameters of the respective programs (*e.g.*, BLASTN for nucleotide sequences, BLASTX for proteins) can be used.

[0114] Methods for the alignment of sequences and for the analysis of similarity and identity of polypeptide and polynucleotide sequences are well-known in the art.

[0115] As used herein “sequence identity” refers to the percentage of residues that are identical in the same positions in the sequences being analyzed. As used herein “sequence similarity” refers to the percentage of residues that have similar biophysical / biochemical

characteristics in the same positions (e.g. charge, size, hydrophobicity) in the sequences being analyzed.

[0116] Methods of alignment of sequences for comparison are well-known in the art, including manual alignment and computer assisted sequence alignment and analysis. This latter approach is a preferred approach in the present disclosure, due to the increased throughput afforded by computer assisted methods. As noted below, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

[0117] The determination of percent sequence identity and/or similarity between any two sequences can be accomplished using a mathematical algorithm. Examples of such mathematical algorithms are the algorithm of Myers and Miller, CABIOS 4:11-17 (1988); the local homology algorithm of Smith et al., Adv. Appl. Math. 2:482 (1981); the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); the search-for-similarity-method of Pearson and Lipman, Proc. Natl. Acad. Sci. 85:2444-2448 (1988); the algorithm of Karlin and Altschul, Proc. Natl. Acad. Sci. USA 87:2264-2268 (1990), modified as in Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5877 (1993).

[0118] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity and/or similarity. Such implementations include, for example: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the AlignX program, version 10.3.0 (Invitrogen, Carlsbad, CA) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. Gene 73:237-244 (1988); Higgins et al. CABIOS 5:151-153 (1989); Corpet et al., Nucleic Acids Res. 16:10881-90 (1988); Huang et al. CABIOS 8:155-65 (1992); and Pearson et al., Meth. Mol. Biol. 24:307-331 (1994). The BLAST programs of Altschul et al. J. Mol. Biol. 215:403-410 (1990) are based on the algorithm of Karlin and Altschul (1990) supra.

[0119] Polynucleotides homologous to a reference sequence can be identified by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well

characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in references cited below (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. ("Sambrook") (1989); Berger and Kimmel, *Guide to Molecular Cloning Techniques*, *Methods in Enzymology*, vol. 152 Academic Press, Inc., San Diego, Calif. ("Berger and Kimmel") (1987); and Anderson and Young, "Quantitative Filter Hybridisation." In: Hames and Higgins, ed., *Nucleic Acid Hybridisation, A Practical Approach*. Oxford, TRL Press, 73-111 (1985)).

[0120] Encompassed by the disclosure are polynucleotide sequences that are capable of hybridizing to the disclosed polynucleotide sequences and fragments thereof under various conditions of stringency (see, for example, Wahl and Berger, *Methods Enzymol.* 152: 399-407 (1987); and Kimmel, *Methods Enzymo.* 152: 507-511, (1987)). Full length cDNA, homologs, orthologs, and paralog of polynucleotides of the present disclosure may be identified and isolated using well-known polynucleotide hybridization methods.

[0121] With regard to hybridization, conditions that are highly stringent, and means for achieving them, are well known in the art. See, for example, Sambrook et al. (1989) (*supra*); Berger and Kimmel (1987) pp. 467-469 (*supra*); and Anderson and Young (1985)(*supra*).

[0122] Hybridization experiments are generally conducted in a buffer of pH between 6.8 to 7.4, although the rate of hybridization is nearly independent of pH at ionic strengths likely to be used in the hybridization buffer (Anderson and Young (1985)(*supra*)). In addition, one or more of the following may be used to reduce non-specific hybridization: sonicated salmon sperm DNA or another non-complementary DNA, bovine serum albumin, sodium pyrophosphate, sodium dodecylsulfate (SDS), polyvinyl-pyrrolidone, ficoll and Denhardt's solution. Dextran sulfate and polyethylene glycol 6000 act to exclude DNA from solution, thus raising the effective probe DNA concentration and the hybridization signal within a given unit of time. In some instances, conditions of even greater stringency may be desirable or required to reduce non-specific and/or background hybridization. These conditions may

be created with the use of higher temperature, lower ionic strength and higher concentration of a denaturing agent such as formamide.

[0123] Stringency conditions can be adjusted to screen for moderately similar fragments such as homologous sequences from distantly related organisms, or to highly similar fragments such as genes that duplicate functional enzymes from closely related organisms. The stringency can be adjusted either during the hybridization step or in the post-hybridization washes. Salt concentration, formamide concentration, hybridization temperature and probe lengths are variables that can be used to alter stringency. As a general guideline, high stringency is typically performed at $T_m-5^{\circ}\text{C}$ to $T_m-20^{\circ}\text{C}$, moderate stringency at $T_m-20^{\circ}\text{C}$ to $T_m-35^{\circ}\text{C}$ and low stringency at $T_m-35^{\circ}\text{C}$ to $T_m-50^{\circ}\text{C}$ for duplex >150 base pairs. Hybridization may be performed at low to moderate stringency ($25-50^{\circ}\text{C}$ below T_m), followed by post-hybridization washes at increasing stringencies. Maximum rates of hybridization in solution are determined empirically to occur at $T_m-25^{\circ}\text{C}$ for DNA-DNA duplex and $T_m-15^{\circ}\text{C}$ for RNA-DNA duplex. Optionally, the degree of dissociation may be assessed after each wash step to determine the need for subsequent, higher stringency wash steps.

[0124] High stringency conditions may be used to select for nucleic acid sequences with high degrees of identity to the disclosed sequences. An example of stringent hybridization conditions obtained in a filter-based method such as a Southern or northern blot for hybridization of complementary nucleic acids that have more than 100 complementary residues is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

[0125] Hybridization and wash conditions that may be used to bind and remove polynucleotides with less than the desired homology to the nucleic acid sequences or their complements of the present disclosure include, for example: 6X SSC and 1% SDS at 65°C ; 50% formamide, 4X SSC at 42°C ; 0.5X SSC to 2.0 X SSC, 0.1% SDS at 50°C to 65°C ; or 0.1X SSC to 2X SSC, 0.1% SDS at 50°C - 65°C ; with a first wash step of, for example, 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with, for example, a subsequent wash step with 0.2 X SSC and 0.1% SDS at 65°C for 10, 20 or 30 minutes.

[0126] For identification of less closely related homologs, wash steps may be performed at a lower temperature, e.g., 50°C . An example of a low stringency wash step employs a

solution and conditions of at least 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS over 30 min. Greater stringency may be obtained at 42°C in 15 mM NaCl, with 1.5 mM trisodium citrate, and 0.1% SDS over 30 min. Wash procedures will generally employ at least two final wash steps. Additional variations on these conditions will be readily apparent to those skilled in the art (see, for example, US Patent Application No. 20010010913).

[0127] If desired, one may employ wash steps of even greater stringency, including conditions of 65°C -68°C in a solution of 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS, or about 0.2X SSC, 0.1% SDS at 65° C and washing twice, each wash step of 10, 20 or 30 min in duration, or about 0.1 X SSC, 0.1% SDS at 65° C and washing twice for 10, 20 or 30 min. Hybridization stringency may be increased further by using the same conditions as in the hybridization steps, with the wash temperature raised about 3°C to about 5°C, and stringency may be increased even further by using the same conditions except the wash temperature is raised about 6°C to about 9°C.

Target Nucleic Acids of the Present Disclosure

[0128] Transcriptional activators of the present disclosure may be targeted to specific target nucleic acids to induce activation/expression of the target nucleic acid. In some embodiments, transcriptional activators activate expression of a target nucleic acid by being targeted to the nucleic acid with the assistance of a guide RNA (CRISPR-based targeting). In this sense, a target nucleic acid of the present disclosure is targeted based on the particular nucleotide sequence in the target nucleic acid that is recognized by the targeting portion of the crRNA or guide RNA that is used according to the methods of the present disclosure.

[0129] Various types of nucleic acids may be targeted for activation of expression, as will be readily apparent to one of skill in the art. The target nucleic acid may be located within the coding region of a target gene or upstream or downstream thereof. Moreover, the target nucleic acid may reside endogenously in a target gene or may be inserted into the gene, e.g., heterologous, for example, using techniques such as homologous recombination. For example, a target gene of the present disclosure can be operably linked to a control region, such as a promoter, that contains a sequence that can be recognized by e.g. a crRNA/tracrRNA and/or a guide RNA of the present disclosure such that a transcriptional activator of the present disclosure may be targeted to that sequence.

[0130] In some embodiments, the target nucleic acid is endogenous to the plant where the expression of one or more genes is activated according to the methods described herein. In some embodiments, the target nucleic acid is a transgene of interest that has been inserted into a plant. Methods of introducing transgenes into plants are well known in the art. Transgenes may be inserted into plants in order to provide a production system for a desired protein, or may be added to the genetic complement in order to modulate the metabolism of a plant.

[0131] Suitable target nucleic acids will be readily apparent to one of skill in the art depending on the particular need or outcome. The target nucleic acid may be in e.g. a region of euchromatin (e.g. highly expressed gene), or the target nucleic acid may be in a region of heterochromatin (e.g. centromere DNA). Use of transcriptional activators according to the methods described herein to induce transcriptional activation in a region of heterochromatin or other highly methylated region of a plant genome may be especially useful in certain research embodiments. For example, activation of a retrotransposon in a plant genome may find use in inducing mutagenesis of other genomic regions in that genome. A target nucleic acid of the present disclosure may be methylated or it may be unmethylated.

[0132] Exemplary target genes for e.g. research or other purposes may include, for example, AS1, PHYB, DWF4, CLV3, and API. The methods of the present disclosure may also provide a quantitative approach to comparing guide RNA efficiency at activating expression of target genes using plant-based SunTag expression systems.

Plants of the Present Disclosure

[0133] Certain aspects of the present disclosure relate to plants containing transcriptional activators that are targeted to one or more target nucleic acids in the plant in order to activate transcription/expression of the target nucleic acid.

[0134] As used herein, a “plant” refers to any of various photosynthetic, eukaryotic multicellular organisms of the kingdom Plantae, characteristically producing embryos, containing chloroplasts, having cellulose cell walls and lacking locomotion. As used herein, a “plant” includes any plant or part of a plant at any stage of development, including seeds, suspension cultures, plant cells, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, microspores, and progeny thereof. Also included are cuttings, and cell or tissue cultures. As used in conjunction with the present disclosure, plant

tissue includes, for example, whole plants, plant cells, plant organs, e.g., leafs, stems, roots, meristems, plant seeds, protoplasts, callus, cell cultures, and any groups of plant cells organized into structural and/or functional units.

[0135] Any plant cell may be used in the present disclosure so long as it remains viable after being transformed with a sequence of nucleic acids. Preferably, the plant cell is not adversely affected by the transduction of the necessary nucleic acid sequences, the subsequent expression of the proteins or the resulting intermediates.

[0136] As disclosed herein, a broad range of plant types may be modified to incorporate recombinant polypeptides and/or polynucleotides of the present disclosure. Suitable plants that may be modified include both monocotyledonous (monocot) plants and dicotyledonous (dicot) plants.

[0137] Examples of suitable plants may include, for example, species of the Family Gramineae, including *Sorghum bicolor* and *Zea mays*; species of the genera: *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Pisum*, *Phaseolus*, *Lolium*, *Oryza*, *Avena*, *Hordeum*, *Secale*, and *Triticum*.

[0138] In some embodiments, plant cells may include, for example, those from corn (*Zea mays*), canola (*Brassica napus*, *Brassica rapa* ssp.), *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), duckweed (*Lemna*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucijra*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea*

americana), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia* spp.), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

[0139] Examples of suitable vegetables plants may include, for example, tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*).

[0140] Examples of suitable ornamental plants may include, for example, azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbiapulcherrima*), and chrysanthemum.

[0141] Examples of suitable conifer plants may include, for example, loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*), Western hemlock (*Isuga canadensis*), Sitka spruce (*Picea glauca*), redwood (*Sequoia sempervirens*), silver fir (*Abies amabilis*), balsam fir (*Abies balsamea*), Western red cedar (*Thuja plicata*), and Alaska yellow-cedar (*Chamaecyparis nootkatensis*).

[0142] Examples of suitable leguminous plants may include, for example, guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, peanuts (*Arachis* sp.), crown vetch (*Vicia* sp.), hairy vetch, adzuki bean, lupine (*Lupinus* sp.), trifolium, common bean (*Phaseolus* sp.), field bean (*Pisum* sp.), clover (*Melilotus* sp.) Lotus, trefoil, lens, and false indigo.

[0143] Examples of suitable forage and turf grass may include, for example, alfalfa (*Medicago* s sp.), orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop.

[0144] Examples of suitable crop plants and model plants may include, for example, *Arabidopsis*, corn, rice, alfalfa, sunflower, canola, soybean, cotton, peanut, sorghum, wheat, tobacco, and lemna.

[0145] The plants of the present disclosure may be genetically modified in that recombinant nucleic acids have been introduced into the plants, and as such the genetically modified plants do not occur in nature. A suitable plant of the present disclosure is one capable of expressing one or more nucleic acid constructs encoding one or more recombinant proteins. The recombinant proteins encoded by the nucleic acids may *e.g.* recombinant polypeptides containing a nuclease-deficient CAS9 polypeptide (dCAS9) and a multimerized epitope, as well as recombinant polypeptides containing a transcriptional activator and an affinity polypeptide.

[0146] As used herein, the terms “transgenic plant” and “genetically modified plant” are used interchangeably and refer to a plant which contains within its genome a recombinant nucleic acid. Generally, the recombinant nucleic acid is stably integrated within the genome such that the polynucleotide is passed on to successive generations. However, in certain embodiments, the recombinant nucleic acid is transiently expressed in the plant. The recombinant nucleic acid may be integrated into the genome alone or as part of a recombinant expression cassette. “Transgenic” is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of exogenous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic.

[0147] “Recombinant nucleic acid” or “heterologous nucleic acid” or “recombinant polynucleotide” as used herein refers to a polymer of nucleic acids wherein at least one of the following is true: (a) the sequence of nucleic acids is foreign to (*i.e.*, not naturally found in) a given host cell; (b) the sequence may be naturally found in a given host cell, but in an unnatural (*e.g.*, greater than expected) amount; or (c) the sequence of nucleic acids contains two or more subsequences that are not found in the same relationship to each other in nature. For example, regarding instance (c), a recombinant nucleic acid sequence will have two or more sequences from unrelated genes arranged to make a new functional nucleic acid. Specifically, the present disclosure describes the introduction of an expression vector into a plant cell, where the expression vector contains a nucleic acid sequence coding for a protein that is not normally found in a plant cell or contains a nucleic acid coding for a protein that is normally found in a plant cell but is under the control of different regulatory sequences. With reference to the plant cell’s genome, then, the nucleic acid sequence that codes for the protein is recombinant. A protein that is referred to as recombinant generally implies that it is

encoded by a recombinant nucleic acid sequence which may be present in the plant cell. Recombinant proteins of the present disclosure may also be exogenously supplied directly to host cells (*e.g.* plant cells).

[0148] A “recombinant” polypeptide, protein, or enzyme of the present disclosure, is a polypeptide, protein, or enzyme that is encoded by a “recombinant nucleic acid” or “heterologous nucleic acid” or “recombinant polynucleotide.”

[0149] In some embodiments, the genes encoding the recombinant proteins in the plant cell may be heterologous to the plant cell. In certain embodiments, the plant cell does not naturally produce one or more polypeptides of the present disclosure, and contains heterologous nucleic acid constructs capable of expressing one or more genes necessary for producing those molecules. In certain embodiments, the plant cell does not naturally produce one or more polypeptides of the present disclosure, and is provided the one or more polypeptides through exogenous delivery of the polypeptides directly to the plant cell without the need to express a recombinant nucleic acid encoding the recombinant polypeptide in the plant cell.

[0150] Recombinant nucleic acids and/or recombinant proteins of the present disclosure may be present in host cells (*e.g.* plant cells). In some embodiments, recombinant nucleic acids are present in an expression vector, and the expression vector may be present in host cells (*e.g.* plant cells).

Expression of Recombinant Proteins in Plants

[0151] Recombinant polypeptides of the present disclosure may be introduced into plant cells via any suitable methods known in the art. For example, a recombinant polypeptide can be exogenously added to plant cells and the plant cells are maintained under conditions such that the recombinant polypeptide is involved with targeting one or more target nucleic acids to activate the expression of the target nucleic acids in the plant cells. Alternatively, a recombinant nucleic acid encoding a recombinant polypeptide of the present disclosure can be expressed in plant cells and the plant cells are maintained under conditions such that the recombinant polypeptides of the present disclosure are targeted to one or more target nucleic acids and activate the expression of the target gene in the plant cells. Additionally, in some embodiments, a recombinant polypeptide of the present disclosure may be transiently expressed in a plant via viral infection of the plant, or by introducing a recombinant

polypeptide-encoding RNA into a plant to activate the expression of a target nucleic acid of interest. Methods of introducing recombinant proteins via viral infection or via the introduction of RNAs into plants are well known in the art. For example, Tobacco rattle virus (TRV) has been successfully used to introduce zinc finger nucleases in plants to cause genome modification (“Nontransgenic Genome Modification in Plant Cells”, *Plant Physiology* 154:1079-1087 (2010)).

[0152] A recombinant nucleic acid encoding a recombinant polypeptide of the present disclosure can be expressed in a plant with any suitable plant expression vector. Typical vectors useful for expression of recombinant nucleic acids in higher plants are well known in the art and include, for example, vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* (e.g., see Rogers et al., *Meth. in Enzymol.* (1987) 153:253-277). These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 (e.g., see of Schardl et al., *Gene* (1987) 61:1-11; and Berger et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:8402-8406); and plasmid pBI 101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

[0153] In addition to regulatory domains, recombinant polypeptides of the present disclosure can be expressed as a fusion protein that is coupled to, for example, a maltose binding protein ("MBP"), glutathione S transferase (GST), hexahistidine, c-myc, or the FLAG epitope for ease of purification, monitoring expression, or monitoring cellular and subcellular localization.

[0154] Moreover, a recombinant nucleic acid encoding a recombinant polypeptide of the present disclosure can be modified to improve expression of the recombinant protein in plants by using codon preference. When the recombinant nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended plant host where the nucleic acid is to be expressed. For example, recombinant nucleic acids of the present disclosure can be modified to account for the specific codon preferences and GC content preferences of monocotyledons and dicotyledons, as these preferences have been shown to differ (Murray et al., *Nucl. Acids Res.* (1989) 17: 477-498).

[0155] In some embodiments, recombinant polypeptides of the present disclosure can be used to create functional “overexpression” mutations in a plant by releasing repression of the

target gene expression as a consequence of transcriptional activation of the target nucleic acid. Release of gene expression repression, which may lead to activation of gene expression, may be of a structural gene, e.g., one encoding a protein having for example enzymatic activity, or of a regulatory gene, e.g., one encoding a protein that in turn regulates expression of a structural gene.

[0156] The present disclosure further provides expression vectors encoding recombinant polypeptides of the present disclosure. A nucleic acid sequence coding for the desired recombinant nucleic acid of the present disclosure can be used to construct a recombinant expression vector which can be introduced into the desired host cell. A recombinant expression vector will typically contain a nucleic acid encoding a recombinant protein of the present disclosure, operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the nucleic acid in the intended host cell, such as tissues of a transformed plant.

[0157] Recombinant nucleic acids *e.g.* encoding recombinant polypeptides of the present disclosure may be expressed on multiple expression vectors or they may be expressed on a single expression vector. In some embodiments, recombinant nucleic acids encoding (1) recombinant polypeptides containing a nuclease-deficient CAS9 polypeptide (dCAS9) and a multimerized epitope, (2) recombinant polypeptides containing a transcriptional activator and an affinity polypeptide, and (3) a crRNA and a tracrRNA, or fusions thereof (guide RNA), are all expressed on a single vector.

[0158] For example, plant expression vectors may include (1) a cloned gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

[0159] In some embodiments, an expression vector containing recombinant nucleic acids of the present disclosure may contain a plant-specific TBS insulator sequence having a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at

least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 10.

[0160] A plant promoter, or functional fragment thereof, can be employed to control the expression of a recombinant nucleic acid of the present disclosure in regenerated plants. The selection of the promoter used in expression vectors will determine the spatial and temporal expression pattern of the recombinant nucleic acid in the modified plant, e.g., the nucleic acid encoding the recombinant polypeptide of the present disclosure is only expressed in the desired tissue or at a certain time in plant development or growth. Certain promoters will express recombinant nucleic acids in all plant tissues and are active under most environmental conditions and states of development or cell differentiation (i.e., constitutive promoters). Other promoters will express recombinant nucleic acids in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and the selection will reflect the desired location of accumulation of the gene product. Alternatively, the selected promoter may drive expression of the recombinant nucleic acid under various inducing conditions.

[0161] Examples of suitable constitutive promoters may include, for example, the core promoter of the Rsyn7, the core CaMV 35S promoter (Odell et al., *Nature* (1985) 313:810-812), CaMV 19S (Lawton et al., 1987), rice actin (Wang et al., 1992; U.S. Pat. No. 5,641,876; and McElroy et al., *Plant Cell* (1985) 2:163-171); ubiquitin (Christensen et al., *Plant Mol. Biol.* (1989) 12:619-632; and Christensen et al., *Plant Mol. Biol.* (1992) 18:675-689), pEMU (Last et al., *Theor. Appl. Genet.* (1991) 81:581-588), MAS (Velten et al., *EMBO J.* (1984) 3:2723-2730), nos (Ebert et al., 1987), Adh (Walker et al., 1987), the P- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Pat. No. 5,683,439), the Nos promoter, the pEmu promoter, the rubisco promoter, the GRP 1 - 8 promoter, and other transcription initiation regions from various plant genes known to those of skilled artisans, and constitutive promoters described in, for example, U.S. Pat. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5, 608,142.

[0162] In some embodiments, expression of a nucleic acid of the present disclosure may be driven (in operable linkage) with a UBQ10 promoter. In some embodiments, expression of a nucleic acid of the present disclosure may be driven (in operable linkage) with a promoter having a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 2.

[0163] Examples of suitable tissue specific promoters may include, for example, the lectin promoter (Vodkin et al., 1983; Lindstrom et al., 1990), the corn alcohol dehydrogenase 1 promoter (Vogel et al., 1989; Dennis et al., 1984), the corn light harvesting complex promoter (Simpson, 1986; Bansal et al., 1992), the corn heat shock protein promoter (Odell et al., Nature (1985) 313:810-812; Rochester et al., 1986), the pea small subunit RuBP carboxylase promoter (Poulsen et al., 1986; Cashmore et al., 1983), the Ti plasmid mannopine synthase promoter (Langridge et al., 1989), the Ti plasmid nopaline synthase promoter (Langridge et al., 1989), the petunia chalcone isomerase promoter (Van Tunen et al., 1988), the bean glycine rich protein 1 promoter (Keller et al., 1989), the truncated CaMV 35s promoter (Odell et al., Nature (1985) 313:810-812), the potato patatin promoter (Wenzler et al., 1989), the root cell promoter (Conkling et al., 1990), the maize zein promoter (Reina et al., 1990; Kriz et al., 1987; Wandelt and Feix, 1989; Langridge and Feix, 1983; Reina et al., 1990), the globulin-1 promoter (Belanger and Kriz et al., 1991), the α -tubulin promoter, the cab promoter (Sullivan et al., 1989), the PEPCase promoter (Hudspeth & Grula, 1989), the R gene complex-associated promoters (Chandler et al., 1989), and the chalcone synthase promoters (Franken et al., 1991).

[0164] Alternatively, the plant promoter can direct expression of a recombinant nucleic acid of the present disclosure in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as “inducible” promoters. Environmental conditions that may affect transcription by inducible promoters include, for example, pathogen attack, anaerobic conditions, or the presence of light. Examples of inducible promoters include, for example, the AdhI promoter which is inducible

by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, and the PPKK promoter which is inducible by light. Examples of promoters under developmental control include, for example, promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. An exemplary promoter is the anther specific promoter 5126 (U.S. Pat. Nos. 5,689,049 and 5,689,051). The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

[0165] Moreover, any combination of a constitutive or inducible promoter, and a non-tissue specific or tissue specific promoter may be used to control the expression of various recombinant polypeptides of the present disclosure.

[0166] The recombinant nucleic acids of the present disclosure and/or a vector housing a recombinant nucleic acid of the present disclosure, may also contain a regulatory sequence that serves as a 3' terminator sequence. One of skill in the art would readily recognize a variety of terminators that may be used in the recombinant nucleic acids of the present disclosure. For example, a recombinant nucleic acid of the present disclosure may contain a 3' NOS terminator.

[0167] In some embodiments, recombinant nucleic acids of the present disclosure contain a transcriptional termination site. Transcription termination sites may include, for example, OCS terminators and NOS terminators.

[0168] In some embodiments, recombinant nucleic acids of the present disclosure contain a transcriptional termination site having a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 9.

[0169] In some embodiments, recombinant nucleic acids of the present disclosure contain a transcriptional termination site having a nucleic acid sequence with at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least

about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% nucleic acid sequence identity to the nucleic acid sequence of SEQ ID NO: 26.

[0170] Plant transformation protocols as well as protocols for introducing recombinant nucleic acids of the present disclosure into plants may vary depending on the type of plant or plant cell, e.g., monocot or dicot, targeted for transformation. Suitable methods of introducing recombinant nucleic acids of the present disclosure into plant cells and subsequent insertion into the plant genome include, for example, microinjection (Crossway et al., *Biotechniques* (1986) 4:320-334), electroporation (Riggs et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:5602-5606), *Agrobacterium*-mediated transformation (U.S. Pat. No. 5,563,055), direct gene transfer (Paszkowski et al., *EMBO J.* (1984) 3:2717-2722), and ballistic particle acceleration (U.S. Pat. No. 4,945,050; Tomes et al. (1995). "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe et al., *Biotechnology* (1988) 6:923-926).

[0171] Additionally, recombinant polypeptides of the present disclosure can be targeted to a specific organelle within a plant cell. Targeting can be achieved by providing the recombinant protein with an appropriate targeting peptide sequence. Examples of such targeting peptides include, for example, secretory signal peptides (for secretion or cell wall or membrane targeting), plastid transit peptides, chloroplast transit peptides, mitochondrial target peptides, vacuole targeting peptides, nuclear targeting peptides, and the like (e.g., see Reiss et al., *Mol. Gen. Genet.* (1987) 209(1):116-121; Settles and Martienssen, *Trends Cell Biol* (1998) 12:494-501; Scott et al., *J Biol Chem* (2000) 10:1074; and Luque and Correias, *J Cell Sci* (2000) 113:2485-2495).

[0172] The modified plant may be grown into plants in accordance with conventional ways (e.g., see McCormick et al., *Plant Cell. Reports* (1986) 81-84.). These plants may then be grown, and pollinated with either the same transformed strain or different strains, with the resulting hybrid having the desired phenotypic characteristic. Two or more generations may be grown to ensure that the subject phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure the desired phenotype or other property has been achieved.

Methods of Activating Expression of a Target Nucleic Acid in Plants

[0173] Growing conditions sufficient for the recombinant polypeptides of the present disclosure to be expressed in the plant to be targeted to and activate the expression of one or more target nucleic acids of the present disclosure are well known in the art and include any suitable growing conditions disclosed herein. Typically, the plant is grown under conditions sufficient to express a recombinant polypeptide of the present disclosure, and for the expressed recombinant polypeptides to be localized to the nucleus of cells of the plant in order to be targeted to and activate the expression of the target nucleic acids (if those targets are present in the nucleus). Generally, the conditions sufficient for the expression of the recombinant polypeptide will depend on the promoter used to control the expression of the recombinant polypeptide. For example, if an inducible promoter is utilized, expression of the recombinant polypeptide in a plant will require that the plant to be grown in the presence of the inducer.

[0174] As noted above, growing conditions sufficient for the recombinant polypeptides of the present disclosure to be expressed in the plant to be targeted to and activate the expression of one or more target nucleic acids may vary depending on a number of factors (*e.g.* species of plant, use of inducible promoter, etc.). Suitable growing conditions may include, for example, ambient environmental conditions, standard greenhouse conditions, growth in long days under standard environmental conditions (*e.g.* 16 hours of light, 8 hours of dark), growth in 12 hour light : 12 hour dark day/night cycles, etc.

[0175] Various time frames may be used to observe activation in expression of a target nucleic acid according to the methods of the present disclosure. Plants may be observed/assayed for activation in expression of a target nucleic acid after, for example, about 5 days of growth, about 10 days of growth, about 15 days after growth, about 20 days after growth, about 25 days after growth, about 30 days after growth, about 35 days after growth, about 40 days after growth, about 50 days after growth, or 55 days or more of growth.

[0176] A target nucleic acid of the present disclosure in a plant cell housing recombinant polypeptides of the present disclosure may have its expression increased/upregulated by at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 55%, at least about

60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% as compared to a corresponding control. Various controls will be readily apparent to one of skill in the art. For example, a control may be a corresponding plant or plant cell that does not contain recombinant polypeptides of the present disclosure (*e.g.* wild-type plant or plant cell).

[0177] A target nucleic acid of the present disclosure may have its expression upregulated/activated as compared to a corresponding control nucleic acid. A target nucleic acid may have its expression upregulated at least about 1-fold, at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 10-fold, at least about 15-fold, at least about 20-fold, at least about 25-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 150-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, at least about 600-fold, at least about 700-fold, at least about 800-fold, at least about 900-fold, at least about 1,000-fold, at least about 1,250-fold, at least about 1,500-fold, at least about 1,750-fold, at least about 2,000-fold, at least about 2,500-fold, at least about 3,000-fold, at least about 3,500-fold, at least about 4,000-fold, at least about 4,500-fold, at least about 5,000-fold, at least about 5,500-fold, at least about 6,000-fold, at least about 6,500-fold, at least about 7,000-fold, at least about 7,500-fold, at least about 8,000-fold, at least about 8,500-fold, at least about 9,000-fold, at least about 9,500-fold, at least about 10,000-fold, at least about 12,000-fold, at least about 14,00-fold, at least about 16,000-fold, at least about 18,000-fold, or at least about 20,000-fold or more as compared to a corresponding control nucleic acid. In some embodiments, a target nucleic acid may have its expression upregulated in the range of about 1,000-fold to about 10,000-fold as compared to a corresponding control nucleic acid. As stated above, various controls will be readily apparent to one of skill in the art. For example, a control nucleic acid may be a corresponding nucleic acid from a plant or plant cell that does not contain a nucleic acid encoding a recombinant polypeptide of the present disclosure.

[0178] Comparisons in the present disclosure may also be in reference to corresponding control plants. Various control plants will be readily apparent to one of skill in the art. For example, a control plant may be a plant that does not contain one or more of: (1) a

recombinant polypeptide including a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope, (2) a recombinant polypeptide including a transcriptional activator and an affinity polypeptide, and/or (3) a crRNA and a tracrRNA, or fusions thereof.

[0179] Methods of probing the expression level of a nucleic acid are well-known to those of skill in the art. For example, qRT-PCR analysis may be used to determine the expression level of a population of nucleic acids isolated from a nucleic acid-containing sample (*e.g.* plants, plant tissues, or plant cells).

[0180] It is to be understood that while the present disclosure has been described in conjunction with the preferred specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the present disclosure. Other aspects, advantages, and modifications within the scope of the present disclosure will be apparent to those skilled in the art to which the present disclosure pertains.

EXAMPLES

[0181] The following examples are offered to illustrate provided embodiments and are not intended to limit the scope of the present disclosure.

Example 1: SunTag VP64 System for Targeted Gene Activation in Plants

[0182] This Example demonstrates the targeting of the VP64 transcriptional activator, using a SunTag system, to specific loci in plants and the subsequent transcriptional activation of the targeted loci.

Introduction

[0183] Recently, a technique called SunTag was developed to recruit many effector proteins simultaneously to a location via one dCAS9 protein. In this way, there is an amplification of the effect of targeting, and improved magnitude of gene regulation (Tanenbaum et al, 2014). Tanenbaum et al. described that a dCas9 protein was fused to an unstructured peptide that contains up to 24 copies of the GCN4 epitope. A single chain antibody, scFV, designed to bind this peptide sequence with high affinity and specificity, was fused to an effector protein for gene regulation. Co-expression of the two components allows binding of up to 24 copies of the antibody-fused effector protein to each CAS9-GCN4 fusion

protein. In the case of VP64 as an effector protein, this procedure resulted in very high activation of gene expression compared to simple CAS9-VP64 fusion proteins.

[0184] Although the SunTag system described above (Tanenbaum et al, 2014) resulted in targeted activation of gene expression, this system was designed for and tested in mammalian cells, and such a method has not been shown to work in plants. Moreover, given the multitude of differences in the cellular environment between plant and animal cells, as well as the differences in nucleic acid structure and function between plant and animal cells, it was not known that such a method could even work in plants.

[0185] The present Example describes Applicant's development of a SunTag system capable of specifically activating target gene expression in plants.

Materials and Methods

Plasmid Construction

[0186] The SunTag system described in Tanenbaum et al, 2014 was designed to upregulate/activate genes in mammalian systems. Therefore, transferring this system directly into plants would fail to upregulate/activate selected genomic targets. New SunTag constructs thus needed to be constructed and tested in plants as a first step to seeing if this system could be used to activate gene expression of a specific locus in plants.

[0187] The SunTag VP64 constructs as described in Tanenbaum et al, 2014 were ordered from Addgene (pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP and pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS). In order for the SunTag system to successfully be expressed and work in plants, various components of the system needed to be modified and adapted for use in plants.

[0188] Plant-specific promoters and transcriptional terminators were used in the new construct, although a human codon-optimized, nuclease-deficient (hdCAS9) was also used. Human codon optimized dCas9 expression, which is fused to one HA tag, two nuclear localization signals, and a linker followed by a 10x epitope tail (10xGCN4), was driven by the plant UBIQUITIN10 (UBQ10) promoter, which is ubiquitously expressed in Arabidopsis. The UBQ10 promoter preceding dCas9-10xGCN4 was followed by an Omega translational enhancer sequence. The single chain antibody (scFV) portion of the system, which was also

driven by the UBQ10 promoter, was fused to superfolder GFP, followed by a linker, VP64, another linker, an NLS that was added for plant nuclear localization, GB1, and a REX NLS. The dCas9-10xGCN4 and scFv-VP64 cassettes were separated by a plant-specific TBS insulator sequence (SEQ ID NO: 10). gRNA expression was controlled by the Pol III specific U6 promoter and termination was controlled by the Pol III termination sequence.

[0189] All features of the constructed SunTag VP64 system were present on a single vector. The dCAS9-10xGCN4 cassette, scFv-VP64 cassette, and respective gRNA cassette were cloned into a binary vector using In-Fusion cloning. Only one respective gRNA cassette was present in the SunTag vector transformed into plants. For evaluating the different gRNA cassettes, different independent SunTag vectors were constructed, each housing a respective gRNA cassette. A schematic of the expression cassettes for the SunTag VP64 system is presented in **FIG. 1**.

Construction of dCAS9-10xGCN4 Cassette

[0190] The dCAS9-10xGCN4 portion of the SunTag VP64 vector that was constructed is contained in expression cassette pUBQ10_Omega RBC_dCas9_1xHA_2xNLS_flexible linker_10xGCN4 (nucleic acid sequence presented in SEQ ID NO: 1). This cassette contains the following features and nucleic acid sequences are provided: UBQ10 promoter (SEQ ID NO: 2), Omega RBC translation enhancer (SEQ ID NO: 3), dCas9 (SEQ ID NO: 4), 1xHA (SEQ ID NO: 5), 2xNLS (SEQ ID NO: 6), flexible linker (SEQ ID NO: 7), 10xGCN4 (SEQ ID NO: 8). The expression cassette further included an OCS terminator (SEQ ID NO: 9).

[0191] This expression cassette produces a recombinant dCas9-10xGCN4 fusion protein (SEQ ID NO: 11): dCAS9-1xHA-2xNLS-flexible linker-10xGCN4. The amino acid sequences of features present in the recombinant fusion protein expressed from this expression cassette are: dCAS9 (SEQ ID NO: 12), 1xHA (SEQ ID NO: 13), 2x NLS (SEQ ID NO: 14), flexible linker (SEQ ID NO: 15), and 10xGCN4 (SEQ ID NO: 16).

Construction of scFv-VP64 Cassette

[0192] The scFv-VP64 portion of the SunTag VP64 vector that was constructed is contained in expression cassette pUBQ10-scFv-sfGFP-glycine linker-VP64-glycine linker-SV40 type NLS-GB1-REX NLS-NOS terminator (nucleic acid sequence presented in SEQ ID NO: 17). This cassette contains the following features and nucleic acid sequences are

provided: UBQ10 promoter (SEQ ID NO: 18), scFv single chain antibody (SEQ ID NO: 19), sfGFP (SEQ ID NO: 20), glycine linker (SEQ ID NO: 21), VP64 (SEQ ID NO: 22), glycine linker (SEQ ID NO: 21), SV40 type NLS (SEQ ID NO: 23), GB1 (SEQ ID NO: 24), REX NLS (SEQ ID NO: 25), and NOS terminator (SEQ ID NO: 26).

[0193] This expression cassette produces a recombinant scFv-VP64 fusion protein (SEQ ID NO: 27): scFv-sfGFP-glycine linker-VP64-glycine linker-SV40 type NLS-GB1-REX NLS. The amino acid sequences of features present in the recombinant fusion protein expressed from this expression cassette are: scFv (SEQ ID NO: 28), sfGFP (SEQ ID NO: 29), glycine linker (SEQ ID NO: 30), VP64 (SEQ ID NO: 31), SV40-type NLS (SEQ ID NO: 32), GB1 (SEQ ID NO: 33), and REX NLS (SEQ ID NO: 34).

Construction of gRNA Cassettes

[0194] For targeting the *FWA* gene promoter, a number of different gRNA expression cassettes were constructed. One such expression cassette was U6:gRNA4 (nucleic acid sequence presented in SEQ ID NO: 35). This cassette contains the following features and nucleic acid sequences are provided: U6 promoter (SEQ ID NO: 36), protospacer #4 (SEQ ID NO: 37), gRNA backbone (SEQ ID NO: 38), and PolIII terminator (SEQ ID NO: 39).

[0195] A similar expression cassette that was constructed was U6:gRNA17 (nucleic acid sequence presented in SEQ ID NO: 40). This cassette contains the following features and nucleic acid sequences are provided: U6 promoter (SEQ ID NO: 36), protospacer #17 (SEQ ID NO: 41), gRNA backbone (SEQ ID NO: 38), and Pol III terminator (SEQ ID NO: 39).

[0196] For targeting the *GIS* gene promoter, a tRNA:gRNA expression cassette was constructed. This cassette contained two different gRNA molecules targeting different regions of the *GIS* promoter. The nucleic acid sequence for the *GIS* tRNA:gRNA expression cassette is presented in SEQ ID NO: 42. The structure of this cassette is as follows: U6 promoter-tRNA-protospacer #1-gRNA backbone-tRNA-protospacer #2-gRNA backbone-Pol III terminator. Nucleic acid sequences of the features include U6 promoter (SEQ ID NO: 36), tRNA (SEQ ID NO: 43), protospacer #1 (SEQ ID NO: 44), gRNA backbone (SEQ ID NO: 38), protospacer #2 (SEQ ID NO: 45), and Pol III terminator (SEQ ID NO: 39).

Design of tRNA:gRNA Cassette for Targeting the FWA Promoter

[0197] A tRNA:gRNA expression cassette was designed for targeting the *FWA* promoter. This cassette has a similar structure as the tRNA:gRNA cassette described above for the *GIS* promoter. This cassette for targeting *FWA* includes two different gRNA molecules and uses protospacer #4 and protospacer #17 as described above. The sequence of this cassette is presented in SEQ ID NO: 46.

Construct Transformation into Arabidopsis

[0198] The vector described above housing the SunTag VP64 expression system was transformed into *Agrobacterium*. The vector was then introduced into *Col-0* wild-type *Arabidopsis thaliana* plants using *Agrobacterium*-mediated transformation via the floral dip method. T1 transgenic plants were selected based on their resistance to Hygromycin.

Flowering Time Assays

[0199] Flowering time in plants was scored by measuring the number of rosette and caulinar leaves.

Fluorescent Microscopy

[0200] Visualization of sfGFP fluorescence was performed using a Zeiss confocal microscope and recommended wavelengths to visualize GFP fluorescence. Leaf sections were taken from transgenic SunTag VP64 plants and placed on microscope slides for visualization. Other tissues imaged included roots. In the latter case, seeds from transgenic plants were plated on plates containing Murashige and Skoog (MS) media and Hygromycin, and plates were then grown vertically to allow roots to extend. Once grown, seedlings were placed on microscope slides and GFP was visualized using a confocal microscope.

qRT-PCR

[0201] qRT-PCR assays were conducted according to standard methods and the manufacturer's protocol. The Superscript III First-Strand synthesis kit (Invitrogen) was used for these assays.

Bisulfite Sequencing

[0202] BS-Seq libraries were generated as previously reported (Cokus *et al.*, 2008) and all libraries were sequenced using the HiSeq 2000 platform following manufacturer instructions (Illumina) at a length of 50 bp. Bisulfite-Seq (BS-Seq) reads were aligned to the TAIR10 version of the *Arabidopsis thaliana* reference genome using BS-seeker. For BS-Seq, up to 2 mismatches were allowed and only uniquely mapped reads were used.

Chromatin Immunoprecipitation (ChIP) and ChIP-Seq

[0203] Transgenic SunTag VP64 seeds were plated on MS media and grown. Tissue was collected and two grams were used to grind the tissue. Nuclear Isolation Buffer, protease inhibitors, and 1% formaldehyde was then added to the powder. This solution was incubated at room temperature on a rotator for 10 minutes. Glycine was then added to stop cross-linking. The solution was filtered, spun down, and the resulting pellet was resuspended with extraction buffer 2+inhibitors. This was spun down, and the resulting pellet was resuspended with extraction buffer 3+inhibitors. This was spun and resuspended with Nuclear Lysis Buffer. The solution was moved to a new tube and diluted with ChIP dilution buffer. Samples were then sonicated (30 seconds on, 30 seconds off at maximum power for 15 minutes). dCas9 and the SunTag system were then immunoprecipitated using an anti-HA antibody. Samples were then washed and eluted. DNA was then extracted using phenol-chloroform and libraries were then made for sequencing by following the procedures recommended by the NuGEN kit used. Sequencing reads were then aligned using bowtie2.

Results*Evaluating Component Expression*

[0204] qRT-PCR of T1 plants housing the SunTag VP64 expression system was performed to determine if the various components were being properly expressed. Specifically, expression of the guide RNAs and dCAS9 was evaluated as a proxy for expression of the system. As shown in **FIG. 2A** and **FIG. 2B**, it was found that both sgRNA4 and dCAS9 in the T1 plants were being expressed.

Evaluating Component Nuclear Localization

[0205] After determining that the components of the SunTag VP64 expression system were being expressed in T1 plants as described above, T1 plants were evaluated using fluorescent microscopy to determine if the scFv-VP64 fusion protein was being targeted to the nucleus (as well as to probe proper production of the fusion protein). Nuclear localization was evaluated using transient expression assays and fluorescent microscopy analysis in *N. benthamiana* plants transfected with the SunTag VP64 vector. Because the scFv-VP64 fusion protein also contained sfGFP, probing for GFP nuclear localization provides a read-out of nuclear localization of the fusion protein.

[0206] In an earlier iteration of the SunTag VP64 vector that was constructed, the first nuclear localization signal (NLS) following VP64 in the scFv-VP64 cassette (added to C-terminus of VP64) was the same as that used in Tanenbaum et al, 2014. However, with that NLS, the scFv-VP64 fusion protein did not localize to the nucleus in plants evaluated (**FIG. 3**).

[0207] The earlier SunTag VP64 vector was thus re-designed to replace the failed NLS described above with a linker followed by a modified SV40-type NLS. This is the vector described above in the Materials and Methods. T1 plants housing this SunTag VP64 vector were similarly evaluated for nuclear localization of the scFv-VP64 fusion protein. As can be seen in **FIG. 4**, the SV40-type NLS was able to facilitate nuclear localization of the scFv-VP64 fusion protein.

[0208] Nuclear localization of the scFv-VP64 fusion protein containing the SV40-type NLS was also evaluated in T2 *A. thaliana* plants housing the SunTag VP64 construct. The roots of these plants were evaluated for nuclear localization. As can be seen in **FIG. 5**, the SV40-type NLS was able to facilitate nuclear localization of the scFv-VP64 fusion protein.

Activation of FWA Expression Using gRNA4

[0209] Following confirmation that the SunTag VP64 expression system components were being expressed and localized to the nucleus as described above, various plant lines were evaluated for whether this system could activate expression of a targeted gene. Various T1 and/or T2 lines housing the SunTag VP64 construct that contains gRNA4 (which targets the *FWA* promoter) were evaluated for expression levels of *FWA*.

[0210] As can be seen in **FIG. 6A**, two independent T1 lines housing the SunTag VP64 construct that contains gRNA4 showed substantially increased expression of *FWA* as compared to wild-type plants (Col). In wild-type plants, the promoter of *FWA* is methylated, which results in very low (if any) levels of expression of this gene. However, the SunTag VP64 expression system was able to activate *FWA* expression in an otherwise wild-type genetic background. In *fwa* mutants, an epimutation results in loss of methylation from the *FWA* promoter and subsequent high levels of expression of this gene, as was observed in **FIG. 6A**.

[0211] As can be seen in **FIG. 6B**, two independent T2 lines housing the SunTag VP64 construct that contains gRNA4 showed substantially increased expression of *FWA* as compared to wild-type plants (Col). *FWA* expression was also substantially increased in the T2 lines as compared to a T1 SunTag VP64 line that did not contain any gRNA. Additional control and experimental SunTag VP64 + gRNA lines were evaluated for *FWA* expression, which produced similar results as described above (**FIG. 7** and **FIG. 8**).

[0212] In a related assay, both old and young leaf tissue from various Arabidopsis lines was evaluated for *FWA* expression. As can be seen in **FIG. 9**, SunTag VP64 Arabidopsis lines containing gRNA4 showed upregulation of *FWA* expression as compared to wild-type plants and as compared to a T1 SunTag VP64 line that did not contain any gRNA. Further, the data demonstrates that older leaves have reduced levels of *FWA* RNA, which is consistent with previous reports.

[0213] Given that *FWA* is a repressor of flowering time, the timing of flowering could serve as a phenotypic illustration of activation of the *FWA* gene. Accordingly, flowering time was evaluated in SunTag VP64 Arabidopsis lines containing gRNA4, as well as in control lines that do not contain any gRNA. It was found that the SunTag VP64 Arabidopsis lines containing gRNA4 were slightly late flowering as compared to the no gRNA control lines (**FIG. 10**). This delayed flowering phenotype is consistent with activation of the *FWA* gene.

[0214] Overall, the results suggest that, in the SunTag VP64 lines containing a gRNA that targets the *FWA* promoter (gRNA4), the gRNA is able to successfully guide Cas9 to the *FWA* locus, and that VP64 is then able to activate expression of *FWA*.

Activation of FWA Expression Using gRNA17

[0215] The results described above indicate that SunTag VP64 constructs containing gRNA4 were able to successfully activate *FWA* expression. Another gRNA that targets the *FWA* promoter (gRNA17) was also tested in a SunTag VP64 construct to evaluate if lines containing this construct also exhibited activation of *FWA* expression.

[0216] As can be seen in **FIG. 11**, a line housing the SunTag VP64 construct that contains gRNA17 also showed substantially increased expression of *FWA* as compared to wild-type plants (Col). In this assay, lines containing a SunTag VP64 construct that contains gRNA4 showed greater activation of *FWA* than achieved in lines containing gRNA17, but all SunTag VP64 lines containing a gRNA that targets the *FWA* promoter showed substantially increased expression of *FWA* as compared to wild-type plants (Col). Additional control and experimental SunTag VP64 + gRNA lines were evaluated for *FWA* expression, which produced similar results as described above (**FIG. 12**).

[0217] Taken together, the results suggest that, in the SunTag VP64 lines containing a gRNA that targets the *FWA* promoter, the gRNA is able to successfully guide Cas9 to the *FWA* locus, and that VP64 is then able to activate expression of *FWA*.

Methylation Status of FWA Promoter in FWA-Activated Lines

[0218] The results above suggest that, in the SunTag VP64 lines containing a gRNA that targets the *FWA* promoter, the gRNA is able to successfully guide Cas9 to the *FWA* locus, and that VP64 is then able to activate expression of *FWA*. Also, as described above, the SunTag VP64 + gRNA lines are in an otherwise wild-type genetic background, and in wild-type *Arabidopsis* plants, the *FWA* promoter is highly methylated such that *FWA* expression is effectively silenced. To investigate the methylation status of the *FWA* promoter in SunTag VP64 + gRNA lines, bisulfite sequencing and analysis assays in various lines were conducted.

[0219] From **FIG. 13**, it was seen that two independent T2 SunTag VP64 + gRNA4 lines did not show differential methylation in the *FWA* promoter as compared to wild-type plants. However, as can be seen in **FIG. 14**, two different lines containing SunTag VP64 + gRNA4 did show a moderate level of decreased methylation in the *FWA* promoter as compared to wild-type plants. These results suggest that there is not a clear link between *FWA* activation

and methylation status of the *FWA* promoter in SunTag VP64 + gRNA lines. However, taken together, the data does indicate that the SunTag VP64 system described herein can activate expression of a methylated gene.

ChIP Analysis of Cas9-Bound Targets

[0220] As described above, the results suggest that, in the SunTag VP64 lines containing a gRNA that targets the *FWA* promoter, the gRNA is able to successfully guide Cas9 to the *FWA* locus, and that VP64 is then able to activate expression of *FWA*. To confirm that Cas9 was targeted to the *FWA* promoter in these lines (specifically the SunTag VP64 + gRNA4 line), ChIP-qPCR of Cas9 using an anti-HA antibody (Cas9 is 1xHA tagged) was performed. As can be seen in **FIG. 15**, ChIP-qPCR confirmed Cas9 binding to *FWA* via gRNA4.

[0221] ChIP samples were further subjected to ChIP-seq to analyze genome-wide binding of Cas9 to genomic regions. The results illustrated in **FIG. 16** demonstrate an enrichment of Cas9 over the *FWA* promoter. ChIP-seq analysis also revealed only one major off-target of gRNA4 (**FIG. 17**). This off-target contained a PAM sequence + 14 base pairs that were complementary to gRNA4.

[0222] These results demonstrate that, in the SunTag VP64 + gRNA system, Cas9 is able to be guided to its targets as specified by the gRNA, and that Cas9 is able to bind these targets. The results further suggest that this successful targeting is responsible for the activation of *FWA* by VP64.

Activation of GIS Expression Using tRNA:gRNA

[0223] The data described above indicates that the SunTag VP64 + gRNA construct that was designed to target the *FWA* promoter was successful in doing so, and also successful at activating expression of *FWA*. To evaluate nucleic acid targets other than *FWA*, a SunTag VP64 construct was designed that contained a tRNA:gRNA cassette that targeted the *GIS* locus. CRISPR-targeting technology involving tRNA-gRNA expression cassettes is described in Xie et al, PNAS (2015). This tRNA:gRNA system utilizes a plant's endogenous tRNA processing system to produce mature gRNAs from a single transcript, and allows for the delivery of multiple gRNAs simultaneously with high expression level. The tRNA:gRNA cassette that was designed to target *GIS* is further described in the Materials and Methods.

This SunTag VP64 construct was transformed into wild-type plants, and expression of *GIS* in these transformed plants was evaluated.

[0224] As can be seen in **FIG. 18**, a T1 line housing the SunTag VP64 construct that contains the tRNA:gRNA that targets *GIS* showed substantially increased expression of *GIS* as compared to a T2 SunTag VP64 line that did not contain any gRNA. As above, the results suggest that, in the SunTag VP64 lines containing a gRNA that targets *GIS*, the gRNA is able to successfully guide Cas9 to the *GIS* locus, and that VP64 is then able to activate expression of *GIS*.

Conclusion

[0225] The data presented in this Example demonstrates the successful construction of a SunTag VP64 + gRNA system that is operable in plants. The SunTag system was able to substantially increase expression of targeted genes as compared to corresponding controls. In some instances, many thousands-fold increases in gene expression were observed, and the system was also able to activate expression of a normally methylated gene (*FWA*). This work presents the opportunity for robust and selective activation of plant genes or other nucleic acids in plants, which may serve both research purposes as well as be used in applications for crop improvement.

Example 2: SunTag without VP64 does not activate the expression of *FWA*

[0226] This Example demonstrates that the targeting of SunTag to the *FWA* promoter with guide RNA 4, but without the VP64 transcriptional activator (no effector), does not lead to the activation of *FWA* expression.

Materials and Methods

[0227] Relevant and applicable Materials and Methods for this Example are as described in Example 1. One difference is that instead of fusing linker-VP64-linker-NLS to the antibody, only linker-NLS was fused to obtain a no effector control in this Example. DNA and protein sequence information for the antibody region without a VP64 fusion is provided below. Further, the guide RNA 4 previously described in Example 1 was used in this Example to evaluate the no effector control constructs.

[0228] The relevant control vector used in this Example contained the scFv antibody region without VP64. This vector contained the following features: UBQ10 promoter – scFv – sfGFP - unique BsiWI site for cloning effectors - glycine linker - NLS added for plant nuclear localization - unique BsiWI site for cloning effectors - GB1 - REX NLS - NOS terminator. The nucleotide sequence of this expression cassette is presented in SEQ ID NO: 54.

[0229] The fusion polypeptide encoded from the cassette described above contained the following features: scFv – sfGFP - glycine linker - NLS added for plant nuclear localization - GB1 - REX NLS. The amino acid sequence is presented in SEQ ID NO: 55.

Results

[0230] To confirm that the physical binding of Cas9 to the *FWA* promoter does not activate transcription, qRT-PCR was done with RNA from plants expressing SunTag guide 4 without an effector. Four T1 lines were screened for transcriptional activation. As shown in **FIG. 19**, expression of *FWA* was not activated when VP64 was omitted. A negative control with no guide RNA, and a positive control with VP64 and guide 4 are also plotted for comparison. These results show that the recruitment of VP64 is responsible for the ectopic expression of *FWA*.

Example 3: SunTag VP64-mediated activation of a diverse set of genomic targets

[0231] This Example demonstrates that SunTag VP64 is able to activate a methylated transposable element (*EVADÉ*), and two additional genes involved in development which have no promoter methylation.

Materials and Methods

[0232] Relevant and applicable Materials and Methods for this Example are as described in Example 1. A notable difference is that different guide RNAs are used to target each respective locus. All sgRNA expression is driven by the U6 promoter in each case. For each target, both guides are on the same binary vector and are cloned in tandem. Sequence information is provided below.

[0233] For the guide RNAs (sgRNA), each guide RNA was driven by the U6 promoter, the nucleotide sequence of which is presented in SEQ ID NO: 56. The sgRNA backbone

sequence used for each guide RNA is presented in SEQ ID NO: 57. The two spacer sequences used in the *EVD* guide RNAs are presented in SEQ ID NO: 58 (spacer 1 for *EVD*) and SEQ ID NO: 59 (spacer 2 for *EVD*). The two spacer sequences used in the *AP3* guide RNAs are presented in SEQ ID NO: 60 (spacer 1 for *AP3*) and SEQ ID NO: 61 (spacer 2 for *AP3*). The two spacer sequences used in the *CLV3* guide RNAs are presented in SEQ ID NO: 62 (spacer 1 for *CLV3*) and SEQ ID NO: 63 (spacer 2 for *CLV3*).

Results

[0234] SunTag VP64 was targeted to the retrotransposon *EVD*, which is normally DNA methylated and silenced. Two separate guides (both driven by the U6 promoter) were targeted simultaneously to the 5' end of the coding region. As shown in **FIG. 20A-20C**, 15 different T1 lines were screened by qRT-PCR for *EVD* transcripts. Numerous positive lines ectopically expressing *EVD* were found, indicating that SunTag VP64 can be utilized for targeted activation of transposons or transposon families. It also provides further evidence that methylated loci are amenable to VP64 mediated activation in plants.

[0235] qRT-PCR measurement of RNA expression of three independent *EVD* targeting SunTag VP64 T2 lines provides further confirmation of the activation of *EVD* (**FIG. 21**), and showed that the activation is stable over multiple generations.

[0236] RNA-seq was also performed to confirm the upregulation of *EVD*. There are two copies of the *EVD* transposon in the *Arabidopsis* genome that are predicted to be upregulated by the guide RNAs used, one that resides in a region of heterochromatin near the centromere on chromosome 1 (At1g34967), and one that resides in an area of the genome that is generally euchromatic on chromosome 5 (At5g17125). It was found that both copies were highly upregulated and genome-wide analyses indicated that the activation of *EVD* was highly specific, with very few other genes affected (**FIG. 22A – FIG. 22B**). These results show that SunTag VP64 can be used to specifically activate genes in both heterochromatin and euchromatin.

[0237] Two additional genes were targeted for activation using the SunTag VP64 activation system. One gene is *AP3*, which regulates floral development. It has no methylation present in its promoter. Two designed sgRNAs were targeted to its promoter and as shown by qRT-PCR in **FIG. 23**, two independent T1 lines displayed an upregulation of *AP3* transcription. *CLV3*, which controls the development of the shoot apical meristem

among other functions, is another gene that was targeted with two guides simultaneously. The *CLV3* locus is not methylated in its promoter. **FIG. 24** shows qRT-PCR expression data of one T1 SunTag VP64 line, and displays an upregulation of the *CLV3* transcript.

[0238] Overall, these results suggest that Applicant's SunTag VP64 system for use in plants is a highly robust activation tool that is able to access multiple chromatin contexts through sgRNA complementarity and can subsequently activate transcription of a diverse set of genomic targets. This tool can also be used to activate genes with promoter methylation as well as genes with no promoter methylation. The observation that SunTag VP64 can be used to transiently activate transposons raises the possibility of using this tool to create new transposition events in the genome, with very few off target effects and without the use of reagents that demethylate the entire genome.

References

Tanenbaum et al. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* 159, 635–646, October 23, 2014.

CLAIMS

What is claimed is:

1. A method for activating expression of a target nucleic acid in a plant, comprising:
 - (a) providing a plant comprising:
 - a first recombinant polypeptide comprising a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope;
 - a second recombinant polypeptide comprising a transcriptional activator and an affinity polypeptide that specifically binds to the epitope;
 - a crRNA and a tracrRNA, or fusions thereof; and
 - (b) growing the plant under conditions whereby the first and second recombinant polypeptides are targeted to the target nucleic acid, thereby activating expression of the target nucleic acid.
2. The method of claim 1, wherein the dCAS9 polypeptide has an amino acid sequence that is at least 80% identical to SEQ ID NO: 12.
3. The method of any one of claims 1-2, wherein the multimerized epitope comprises a GCN4 epitope.
4. The method of claim 3, wherein the multimerized epitope comprises about 2 to about 10 copies of a GCN4 epitope.
5. The method of any one of claims 1-4, wherein the first polypeptide comprises one or more linkers that link polypeptide units in the recombinant polypeptide.
6. The method of any one of claims 1-5, wherein the first polypeptide comprises a nuclear localization signal (NLS).
7. The method of any one of claims 1-6, wherein the transcriptional activator is a VP64 polypeptide.
8. The method of claim 7, wherein the VP64 polypeptide comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 31.

9. The method of any one of claims 1-8, wherein the affinity polypeptide is an antibody.
10. The method of claim 9, wherein the antibody is an scFv antibody.
11. The method of claim 10, wherein the antibody comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 28.
12. The method of any one of claims 1-11, wherein the second polypeptide comprises one or more linkers that link polypeptide units in the recombinant polypeptide.
13. The method of any one of claims 1-12, wherein the second polypeptide comprises an SV40-type NLS.
14. The method of claim 13, wherein the SV40-type NLS comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 32.
15. The method of any one of claims 1-14, wherein the crRNA and the tracrRNA are fused together, thereby forming a guide RNA (gRNA).
16. The method of any one of claims 1-15, wherein expression of the activated nucleic acid is increased in the range of about 100-fold to about 10,000-fold as compared to a corresponding control.
17. A recombinant vector comprising:
 - a first nucleic acid sequence comprising a plant promoter and that encodes a recombinant polypeptide comprising a nuclease-deficient CAS9 polypeptide (dCAS9) or fragment thereof and a multimerized epitope;
 - a second nucleic acid sequence comprising a plant promoter and that encodes a recombinant polypeptide comprising a transcriptional activator and an affinity polypeptide that specifically binds to the epitope; and
 - a third nucleic acid sequence comprising a promoter and that encodes a crRNA and a tracrRNA, or fusions thereof.
18. The vector of claim 17, wherein the plant promoter in the first nucleic acid sequence is a UBQ10 promoter.

19. The vector of claim 18, wherein the UBQ10 promoter comprises a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 2.
20. The vector of any one of claims 17-19, wherein the first nucleic acid sequence comprises a terminator sequence.
21. The vector of claim 20, wherein the terminator is an OCS terminator.
22. The vector of claim 21, wherein the OCS terminator comprises a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 9.
23. The vector of any one of claims 17-22, wherein the dCAS9 polypeptide comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 12.
24. The vector of any one of claims 17-23, wherein the multimerized epitope comprises a GCN4 epitope.
25. The vector of claim 24, wherein the multimerized epitope comprises about 2 to about 10 copies of a GCN4 epitope.
26. The vector of any one of claims 17-25, wherein the first polypeptide comprises one or more linkers that link polypeptide units in the recombinant polypeptide.
27. The vector of any one of claims 17-25, wherein the first polypeptide comprises a nuclear localization signal (NLS).
28. The vector of any one of claims 17-27, wherein the plant promoter in the second nucleic acid sequence is a UBQ10 promoter.
29. The vector of claim 28, wherein the UBQ10 promoter comprises a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 2.
30. The vector of any one of claims 17-29, wherein the second nucleic acid sequence comprises a terminator sequence.
31. The vector of claim 30, wherein the terminator is a NOS terminator.
32. The vector of claim 31, wherein the NOS terminator comprises a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 26.

33. The vector of any one of claims 17-32, wherein the transcriptional activator is a VP64 polypeptide.
34. The vector of claim 33, wherein the VP64 polypeptide comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 31.
35. The vector of any one of claims 17-34, wherein the affinity polypeptide is an antibody.
36. The vector of claim 35, wherein the antibody is an scFv antibody.
37. The vector of claim 36, wherein the antibody comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 28.
38. The vector of any one of claims 17-37, wherein the second polypeptide comprises one or more linkers that link polypeptide units in the recombinant polypeptide.
39. The vector of any one of claims 17-38, wherein the second polypeptide comprises an SV40-type NLS.
40. The vector of claim 39, wherein the SV40-type NLS comprises an amino acid sequence that is at least 80% identical to SEQ ID NO: 32.
41. The vector of any one of claims 17-40, wherein the crRNA and the tracrRNA are fused together, thereby forming a guide RNA (gRNA).
42. The vector of any one of claims 17-41, wherein the first and second nucleic acids are separated by a TBS insulator.
43. The vector of claim 42, wherein the TBS insulator comprises a nucleic acid sequence that is at least 80% identical to SEQ ID NO: 10.
44. A plant or plant cell comprising the vector of any one of claims 17-43.
45. A method for activating expression of a target nucleic acid in a plant, comprising:
 - (a) providing a plant comprising a vector of any one of claims 17-43;

(b) growing the plant under conditions whereby the first, second, and third nucleic acids in the vector are expressed and the resulting polypeptides are targeted to the target nucleic acid, thereby activating expression of the target nucleic acid.

46. The method of claim 45, wherein expression of the activated nucleic acid is increased in the range of about 100-fold to about 10,000-fold as compared to a corresponding control.

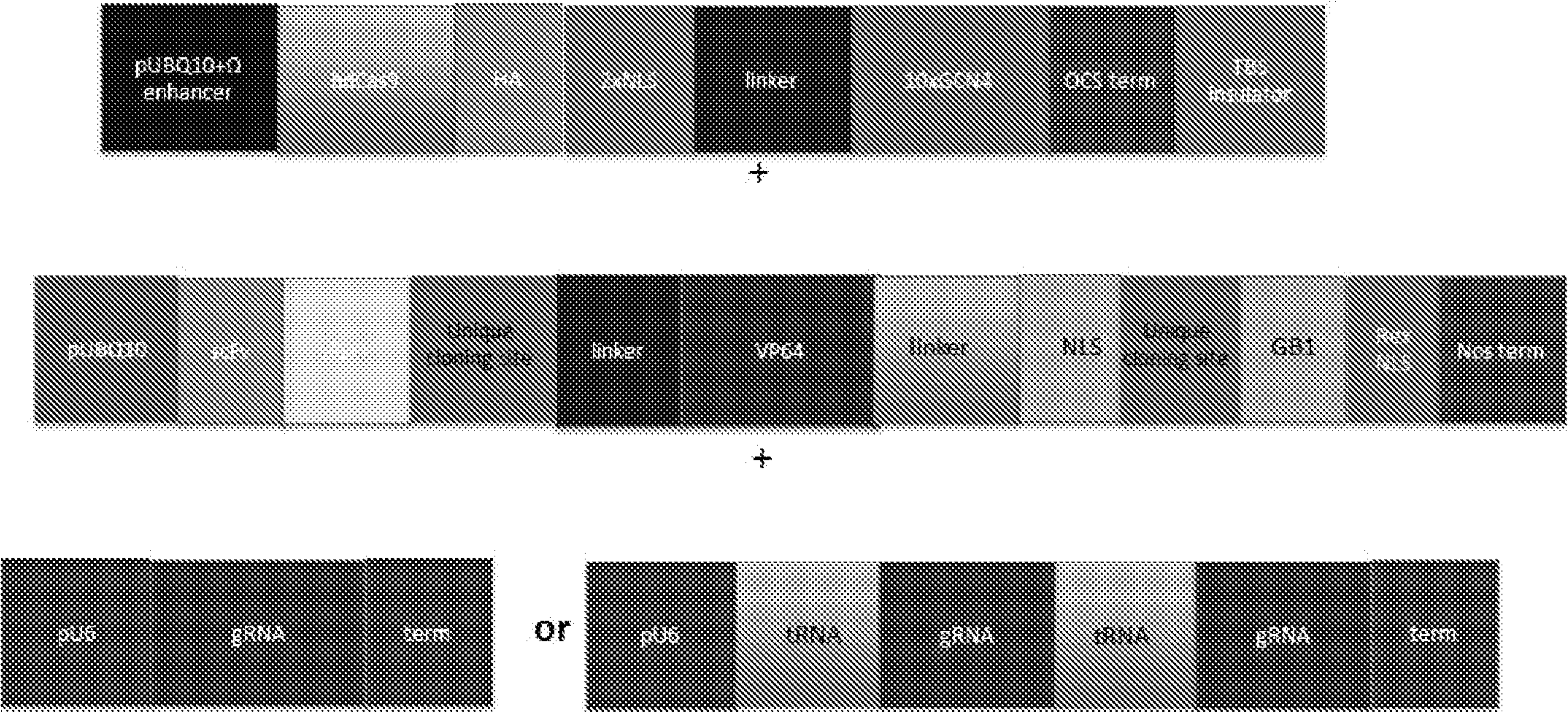


FIG. 1

FIG. 2A

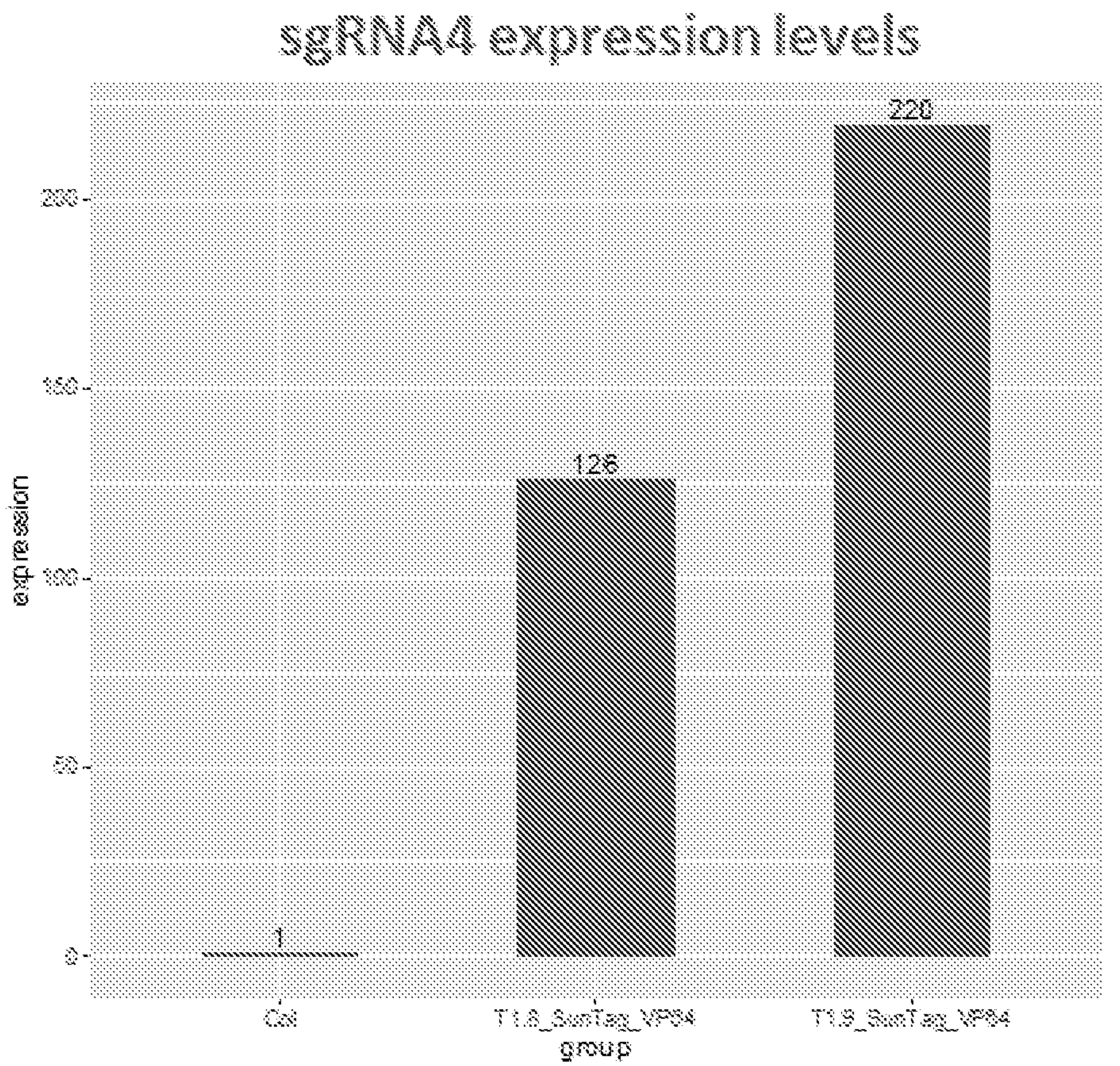
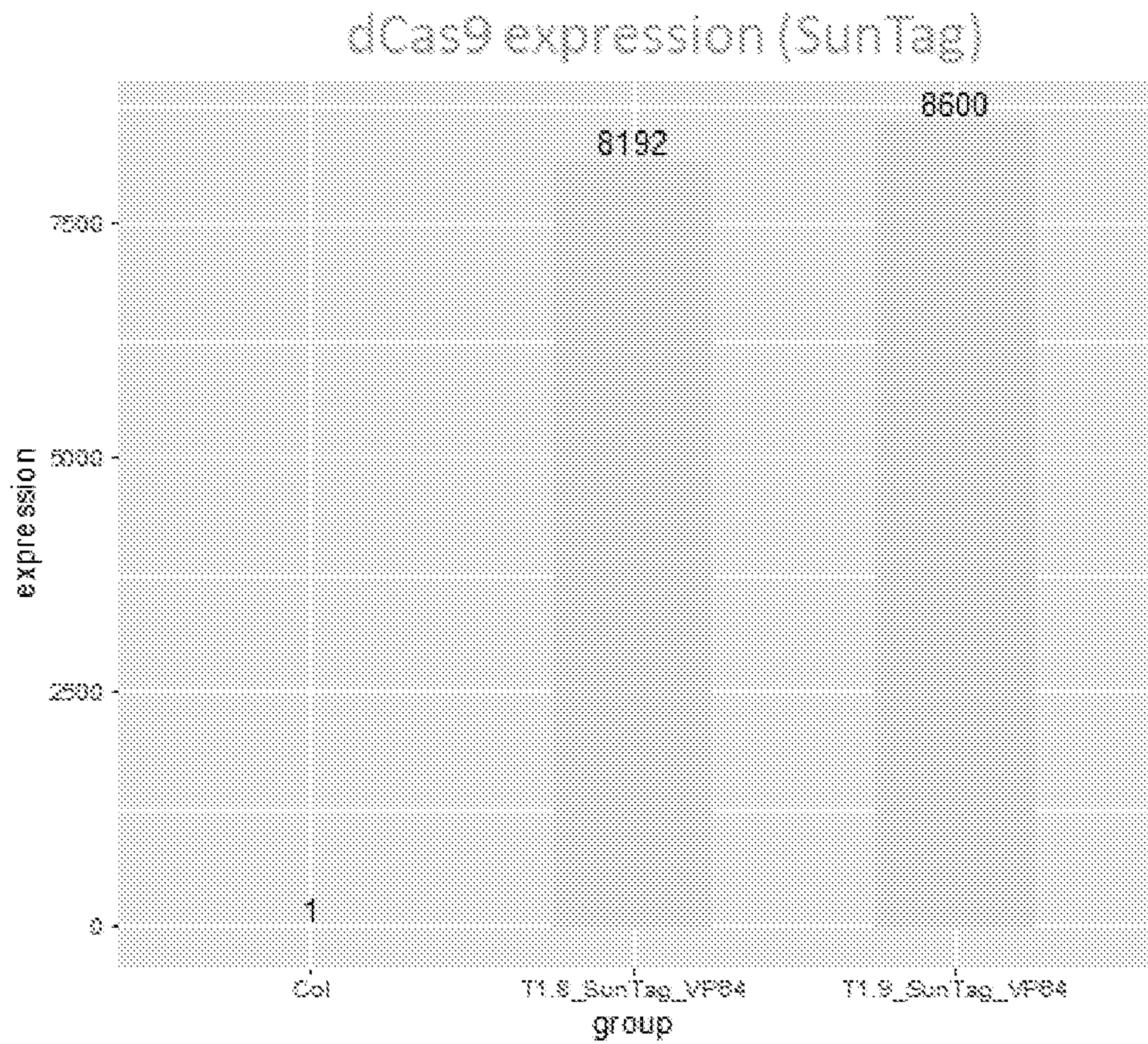


FIG. 2B



**Background
fluorescence
from
chlorophyll**

GFP

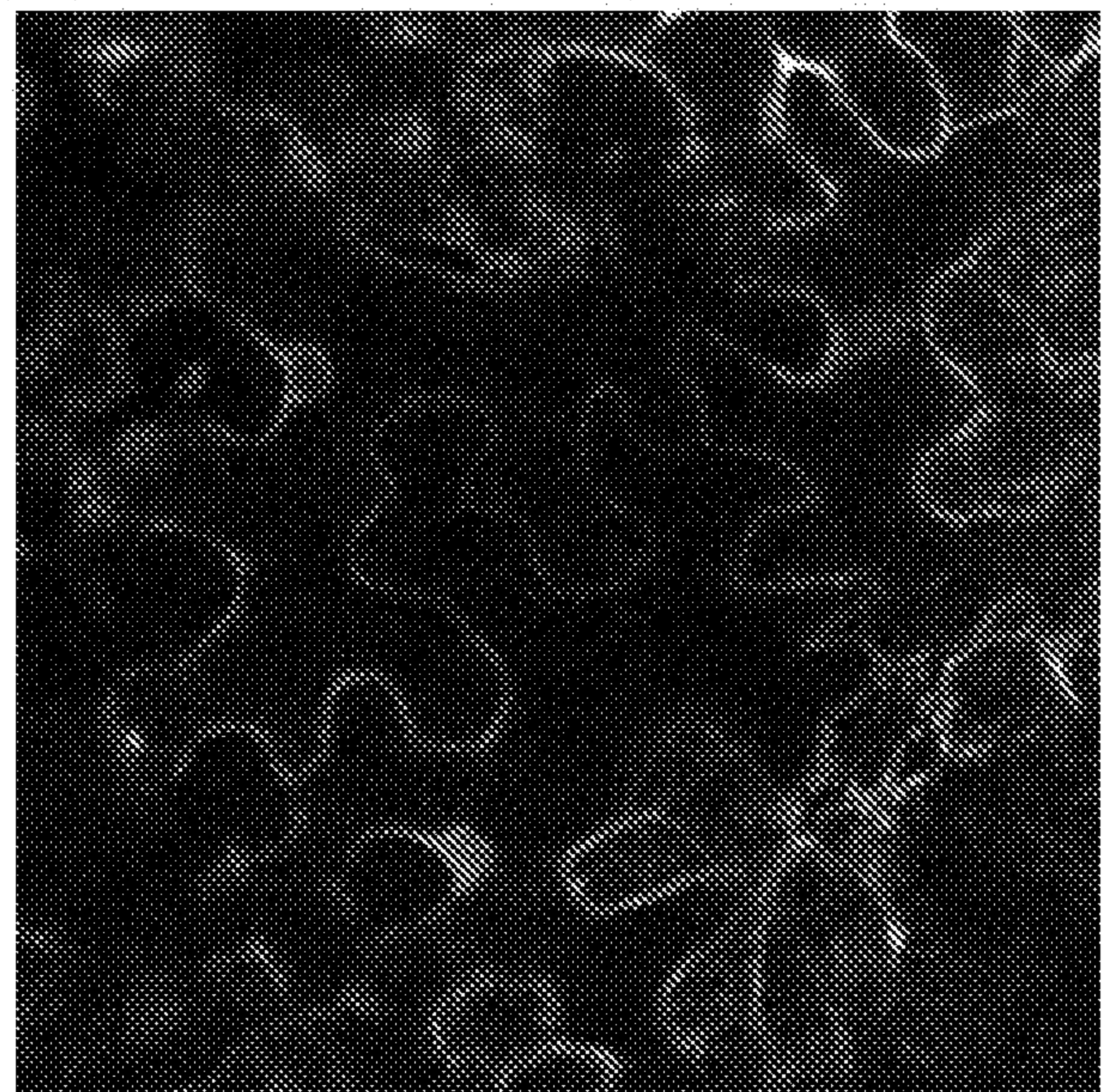
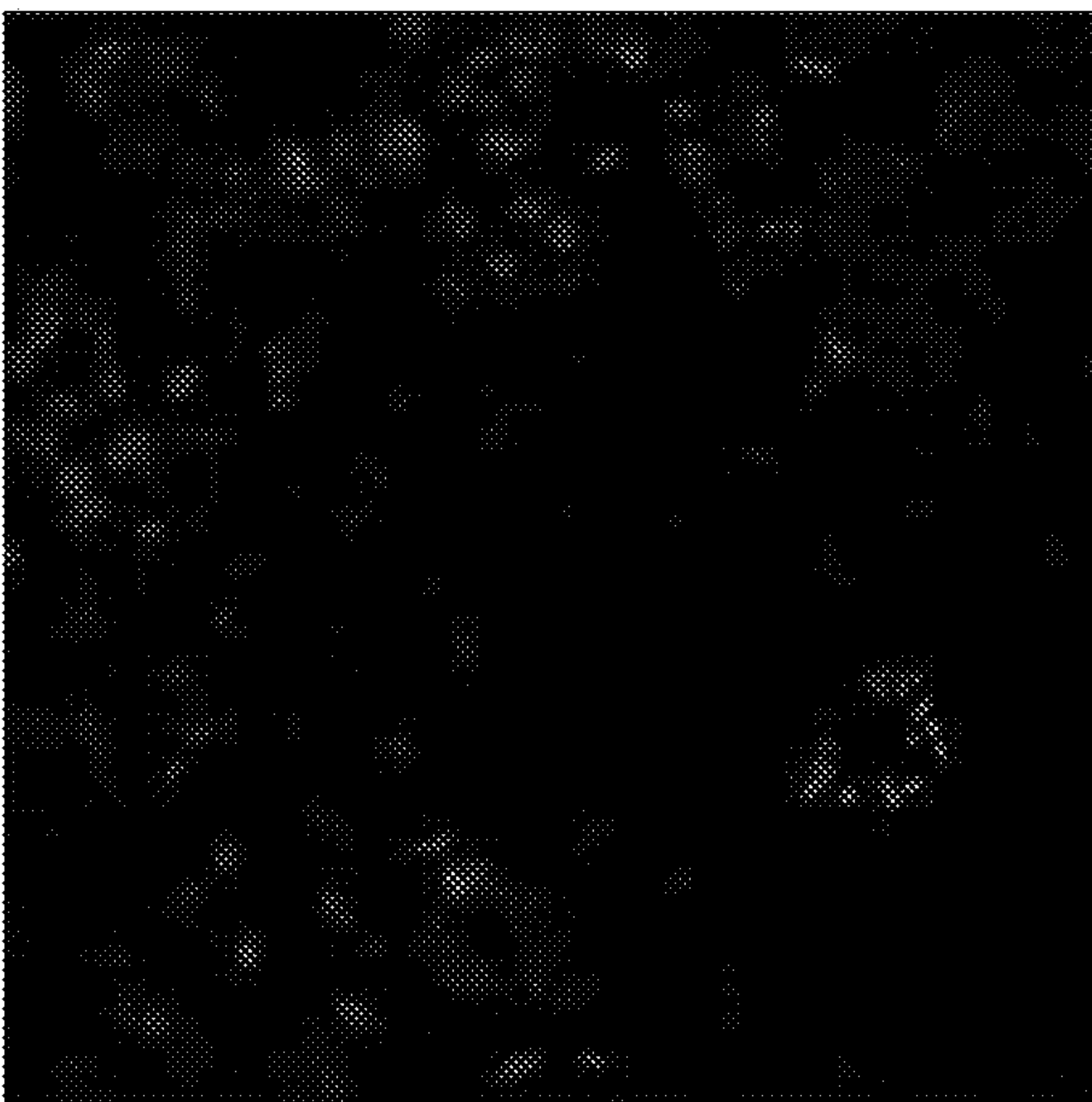
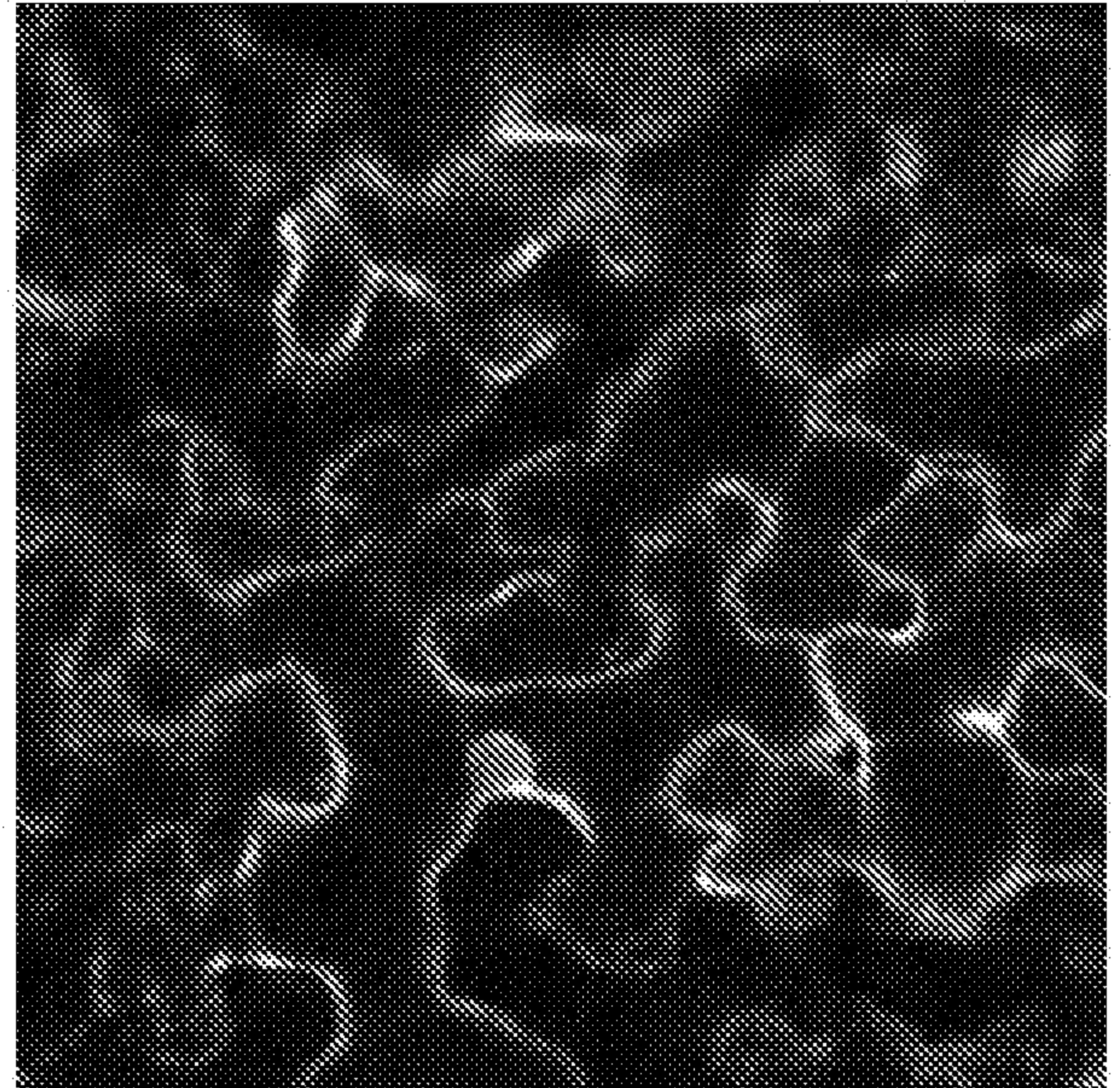
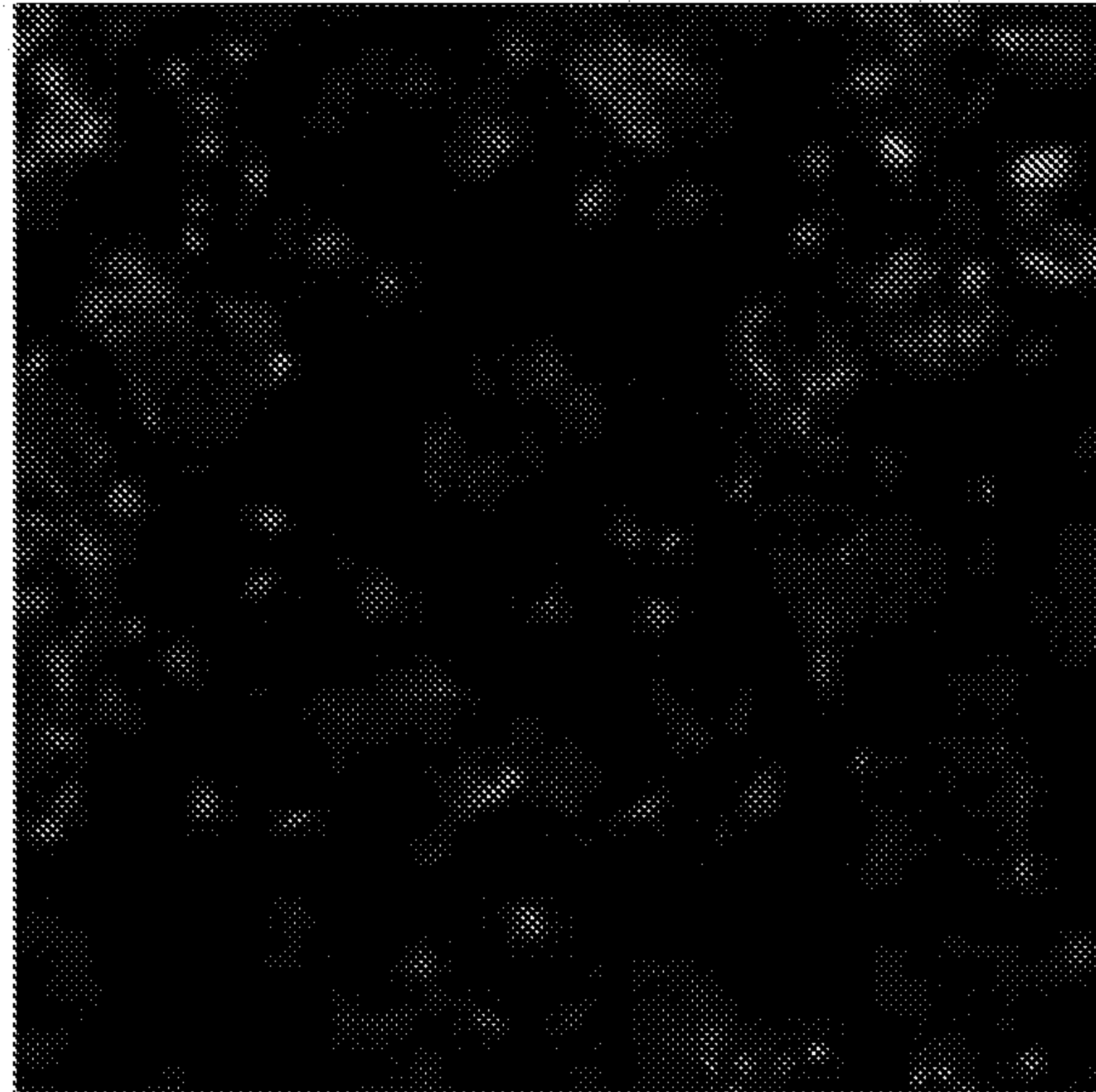
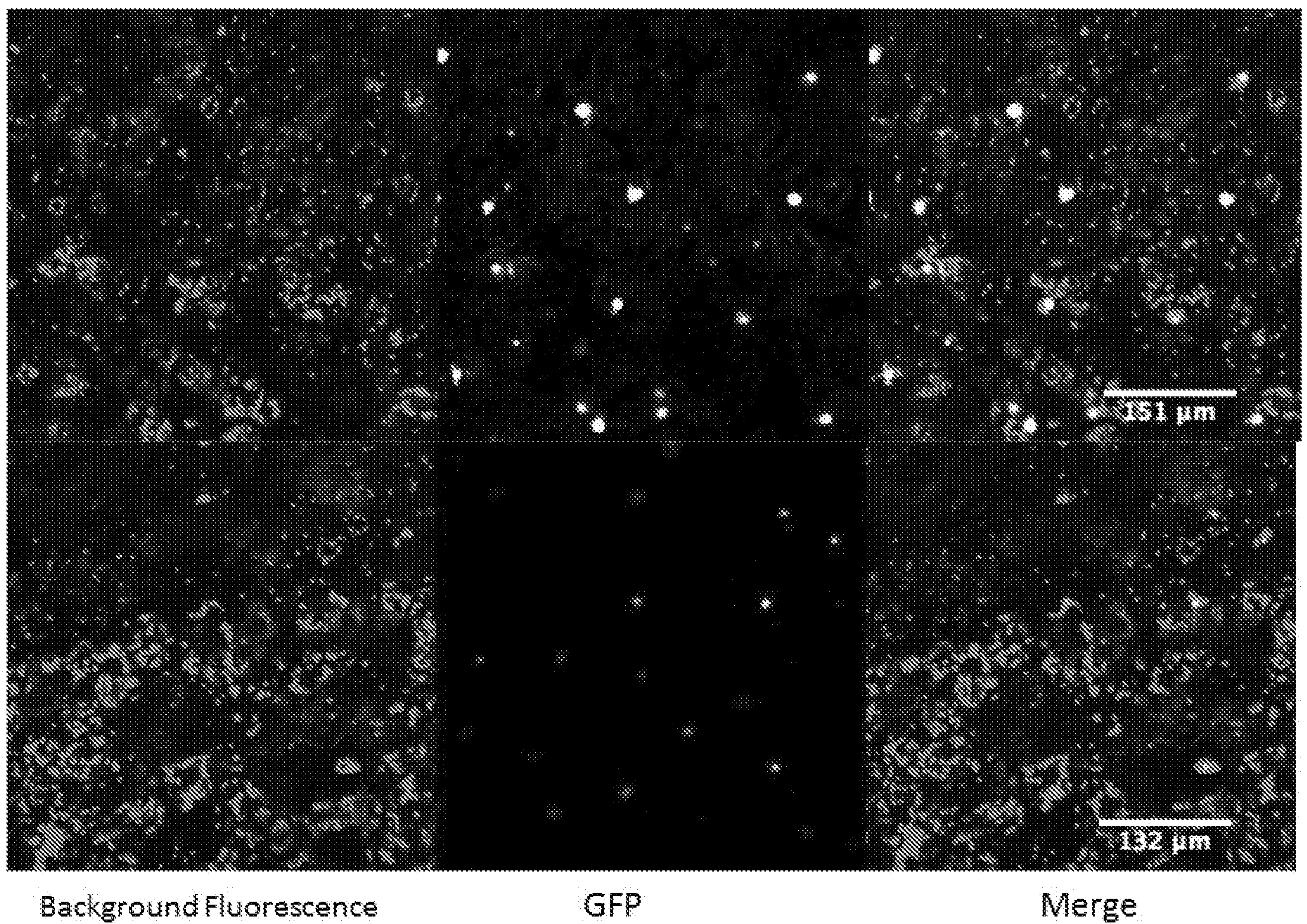
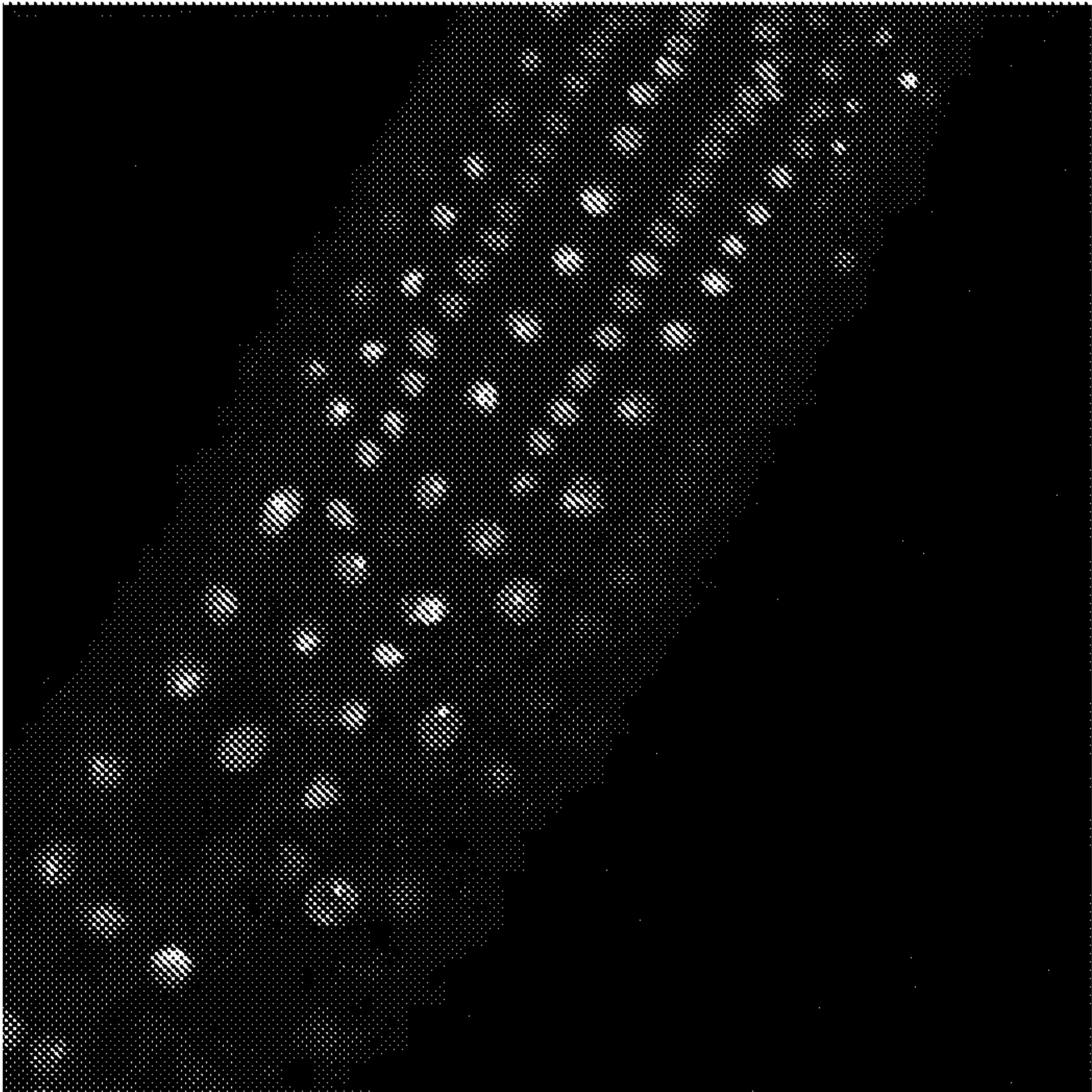


FIG. 3

**FIG. 4**

T2 SunTag VP64 guide4 line 11



T2 SunTag VP64 guide 4 line 11

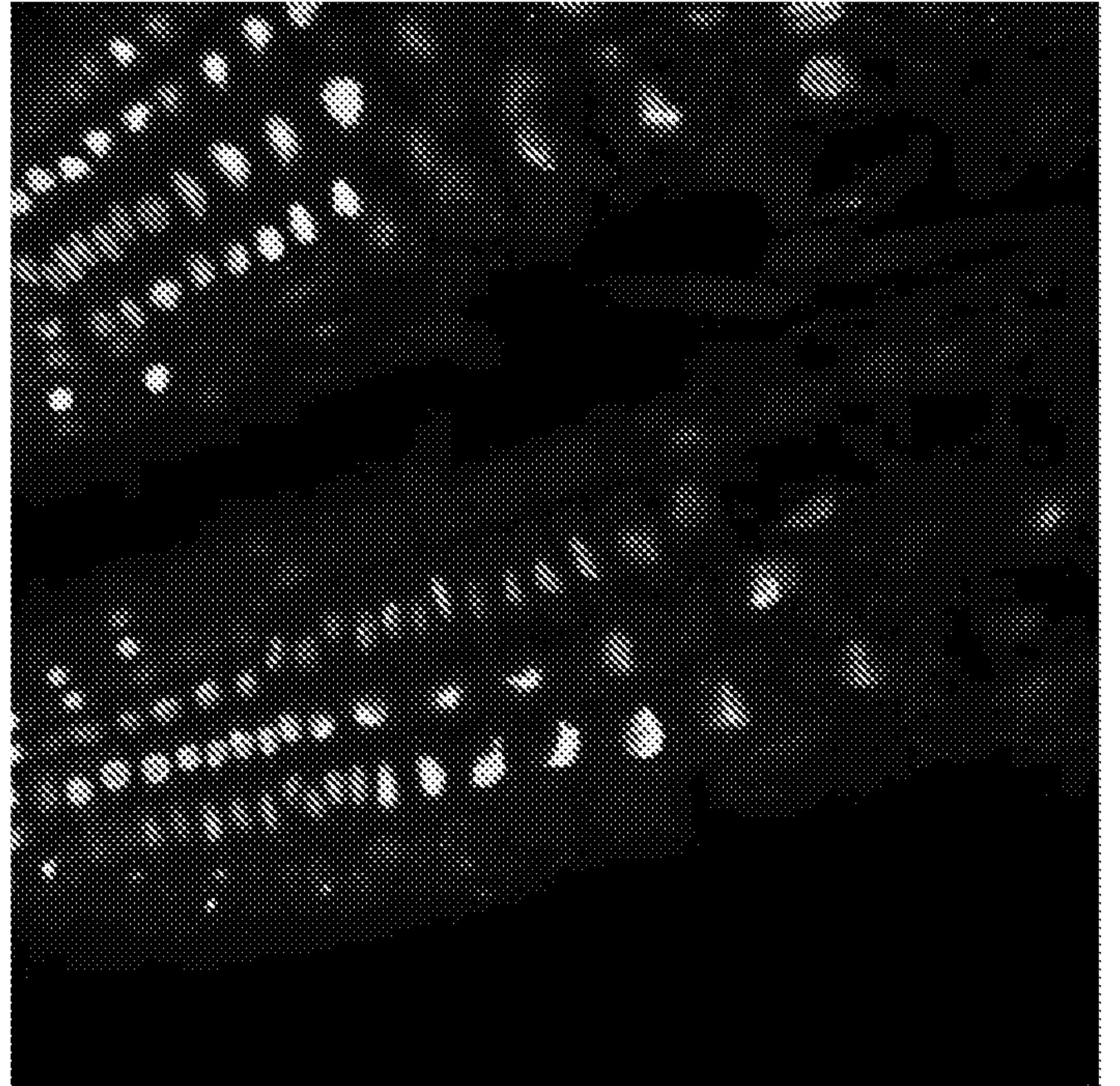


FIG. 5

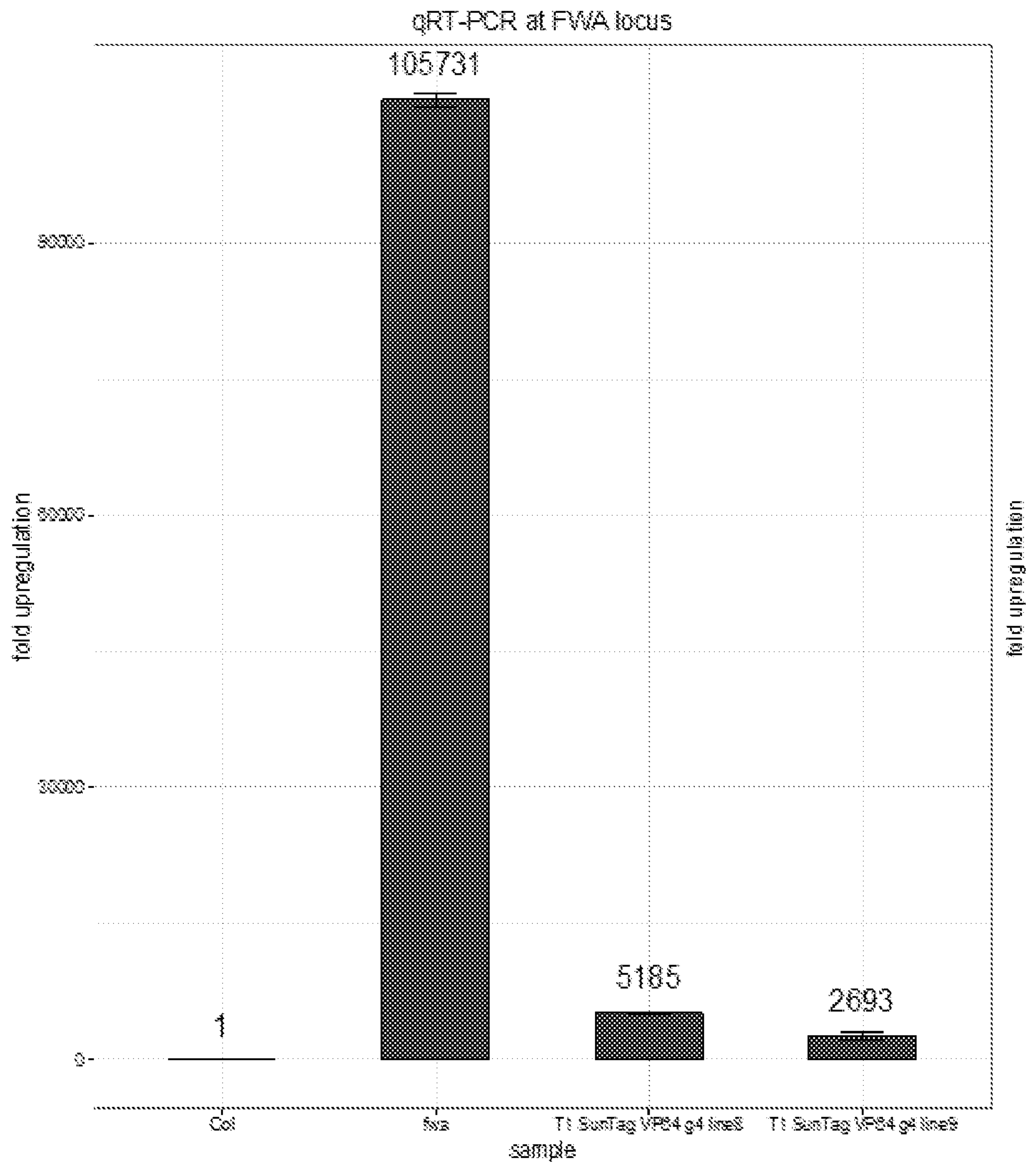


FIG. 6A

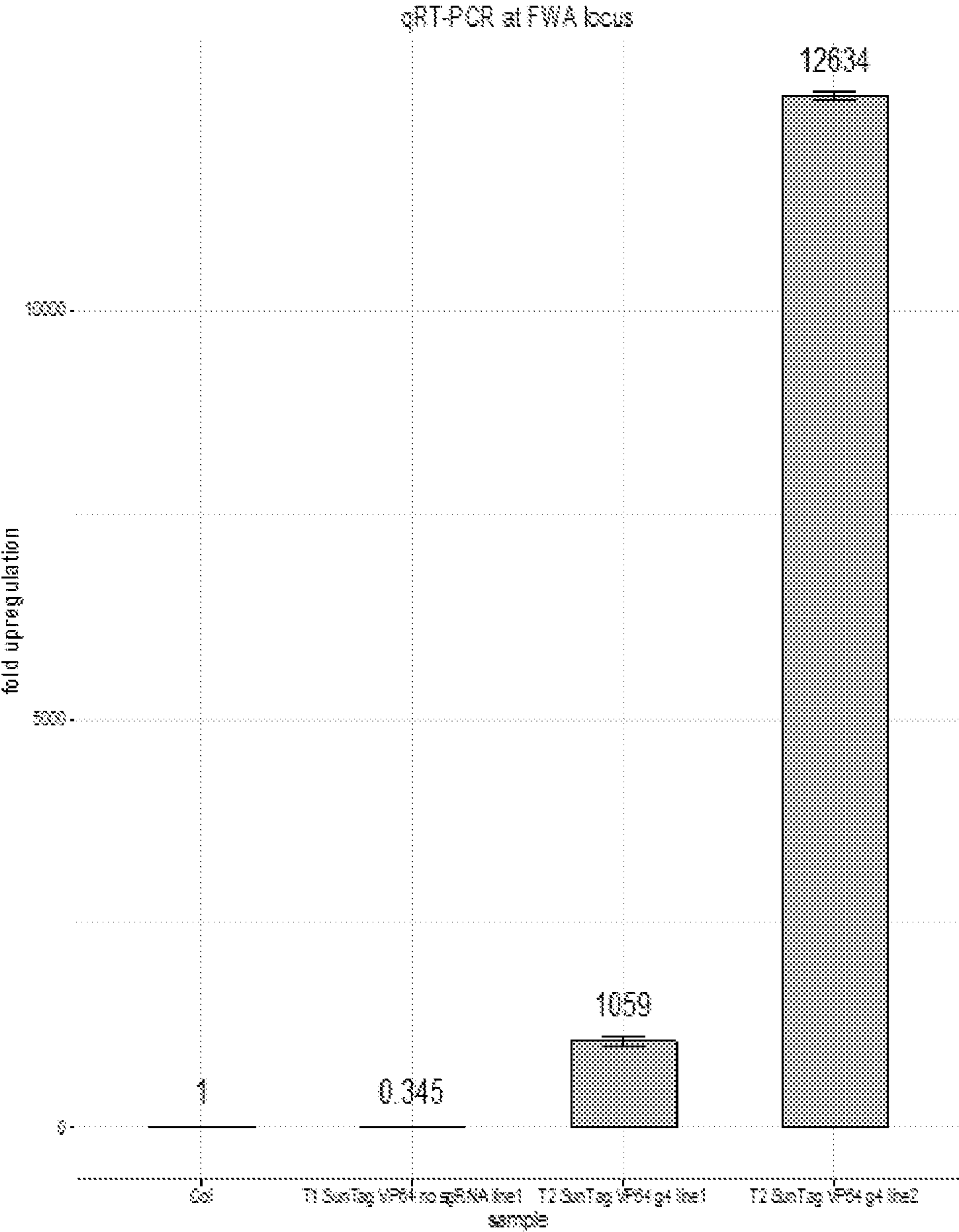


FIG. 6B

8/28

qRT-PCR at FWA locus

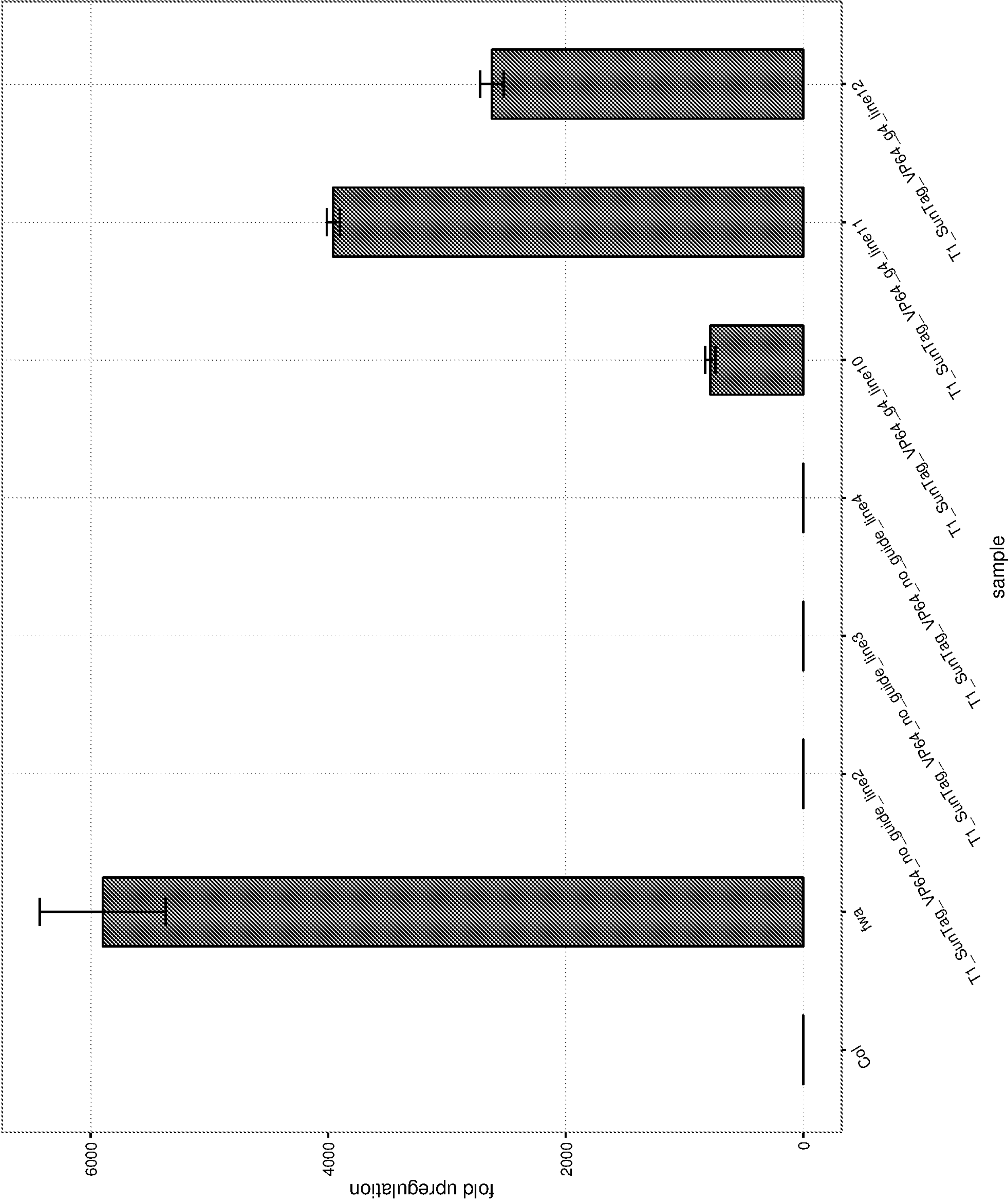
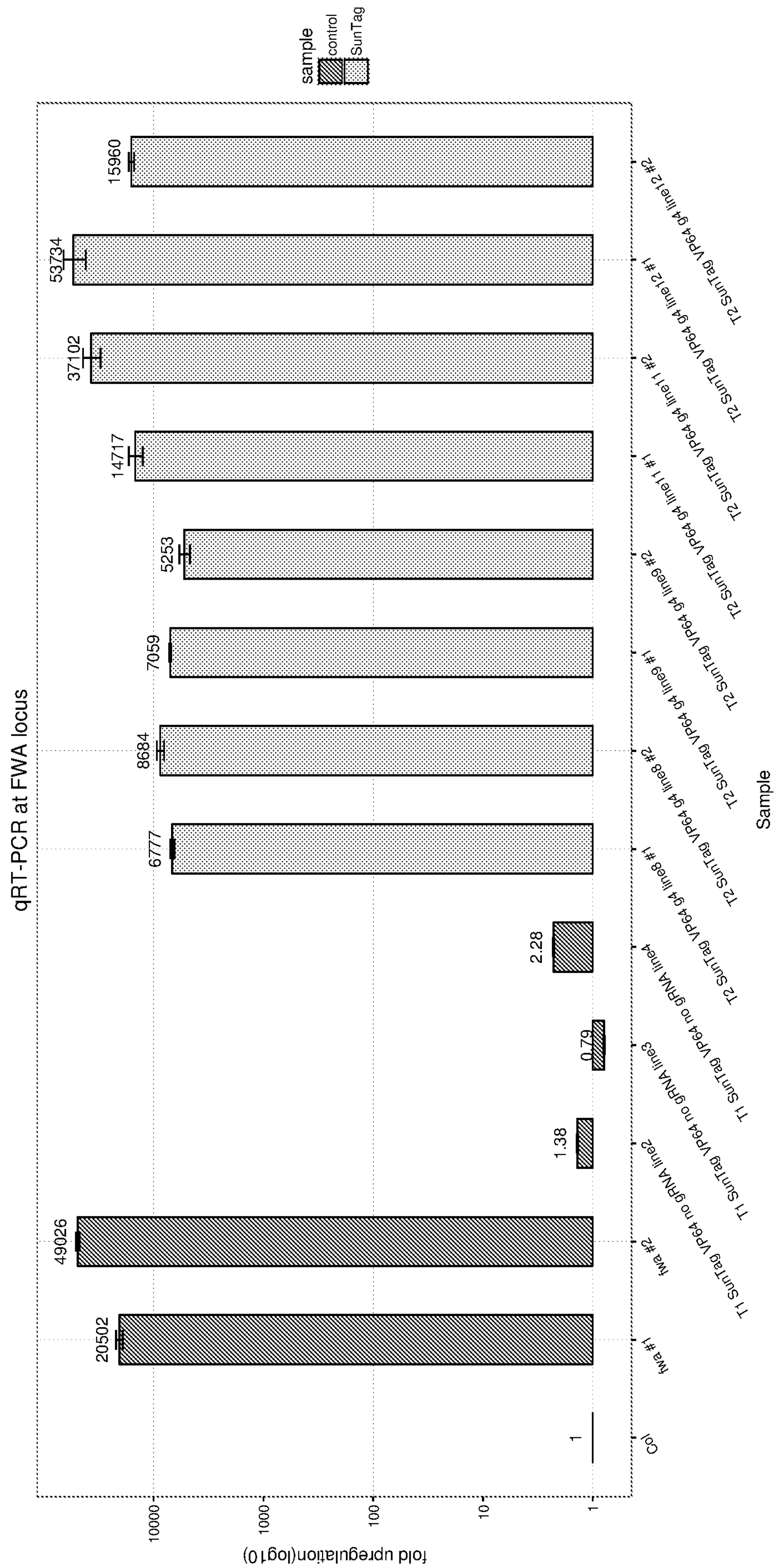


FIG. 7

9/28

FIG. 8



10/28

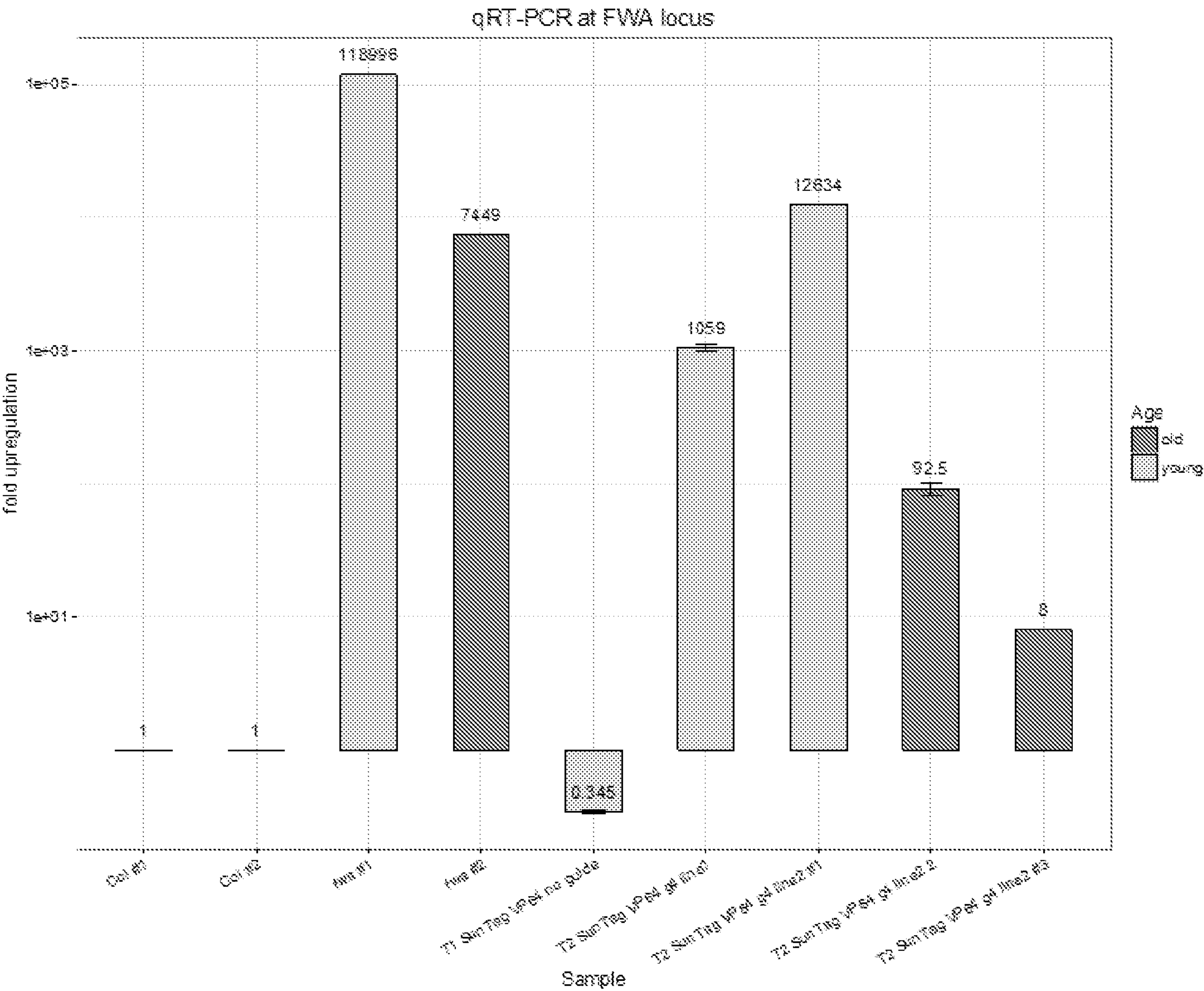


FIG. 9

11/28



No guide

gRNA4

FIG. 10

12/28

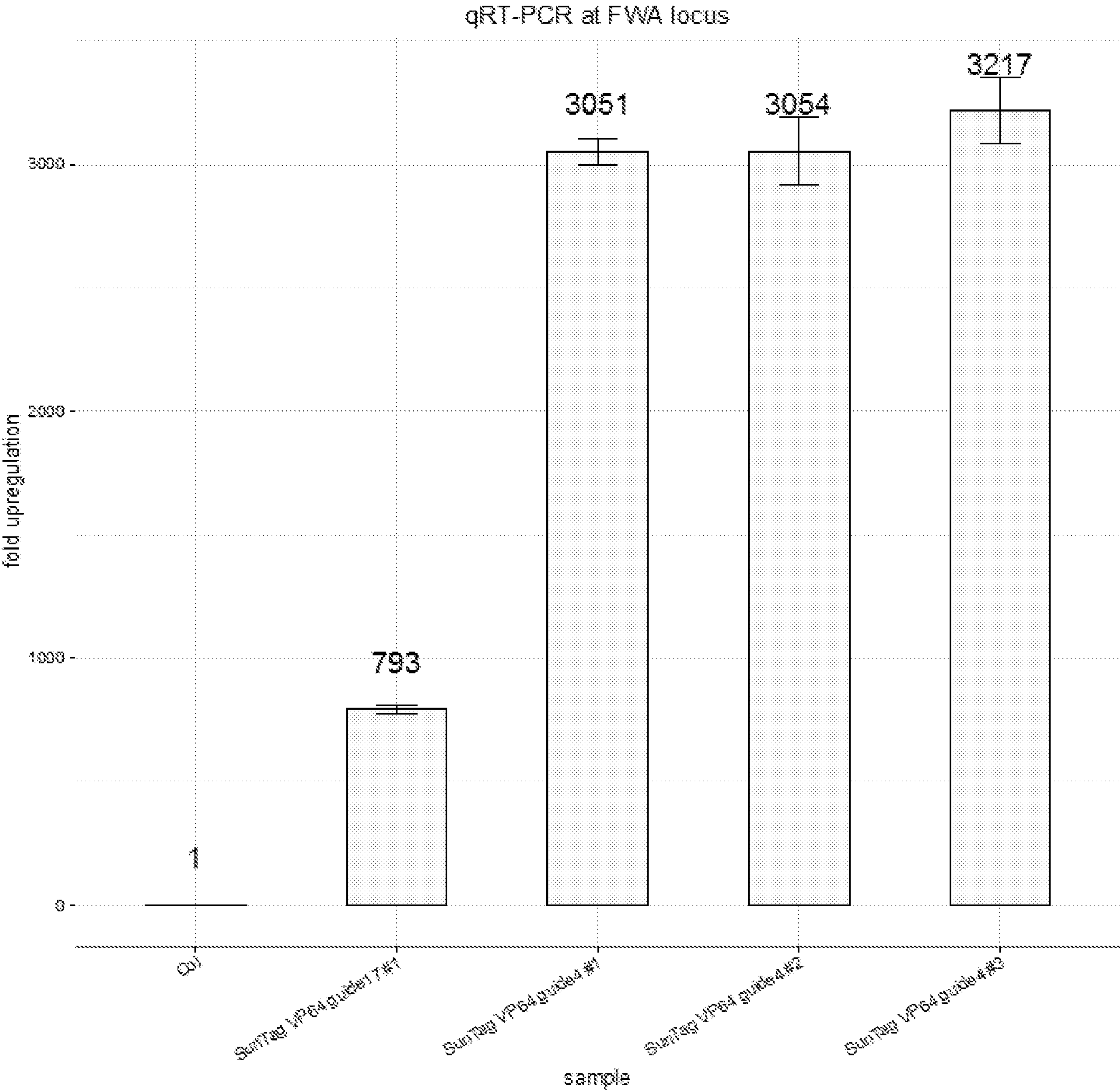


FIG. 11

13/28

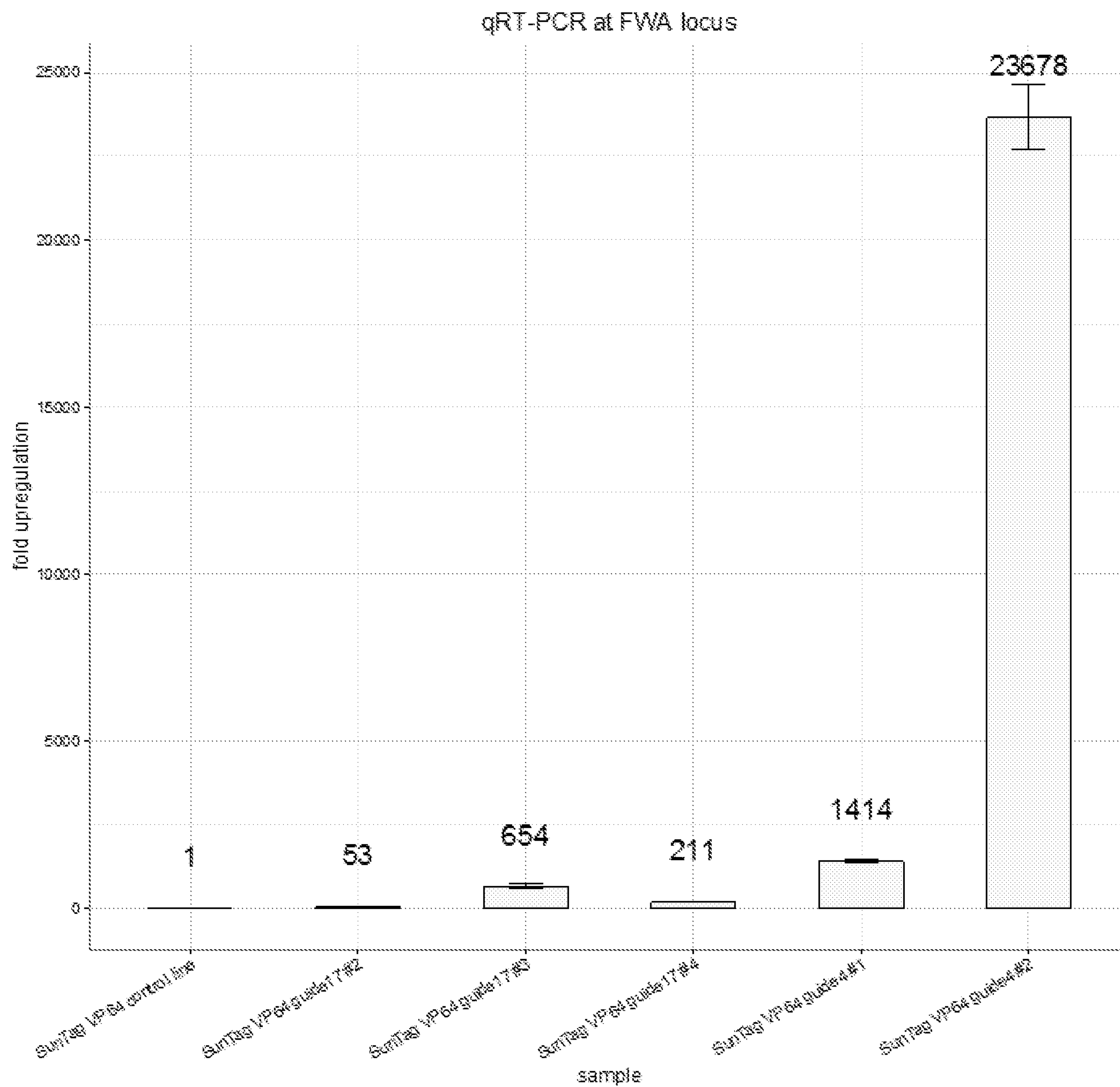
**FIG. 12**

FIG. 13

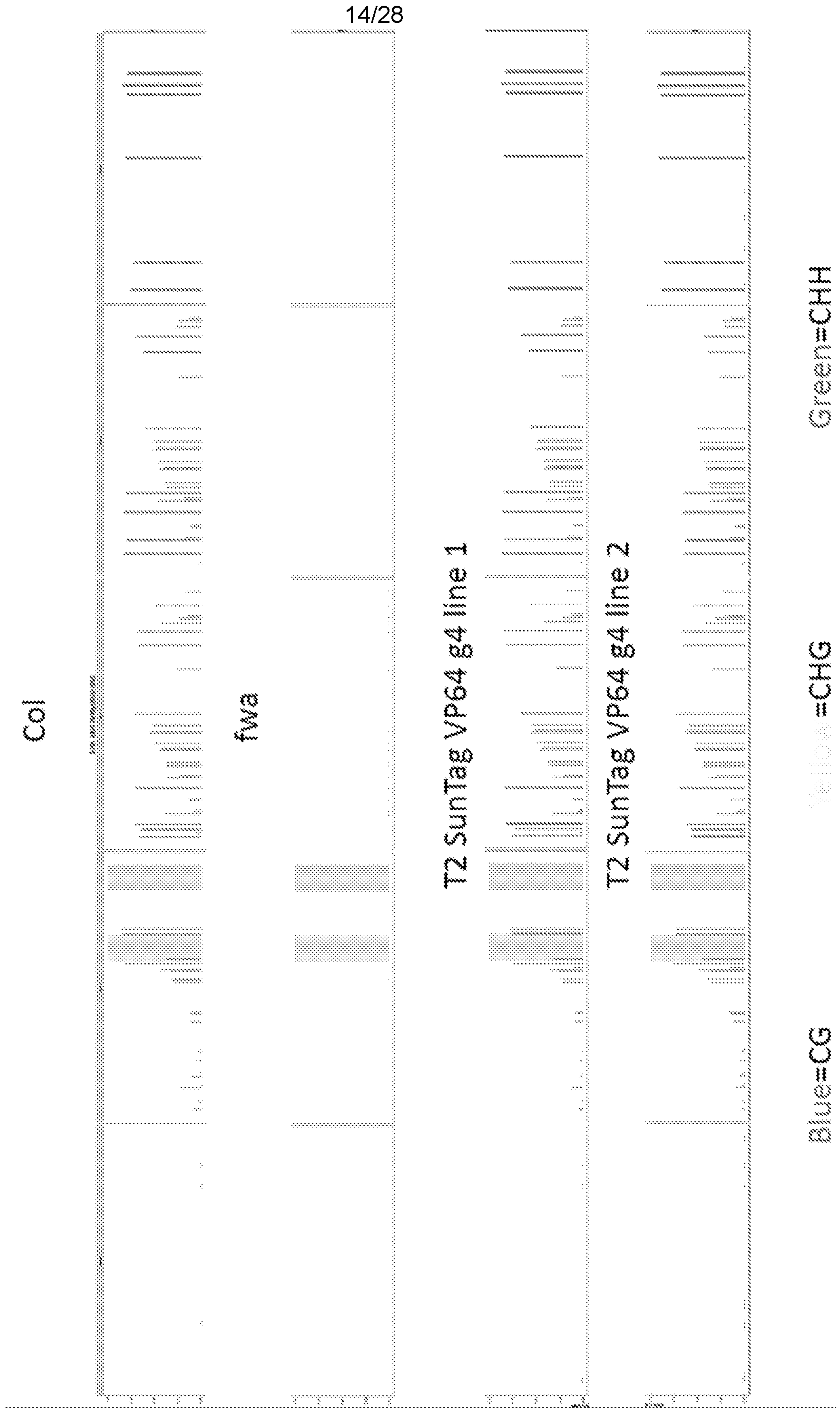
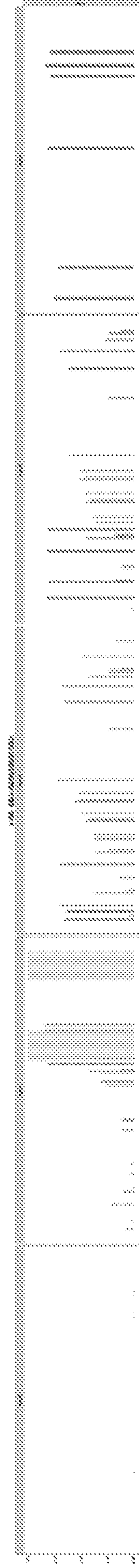


FIG. 14

Col



fwa

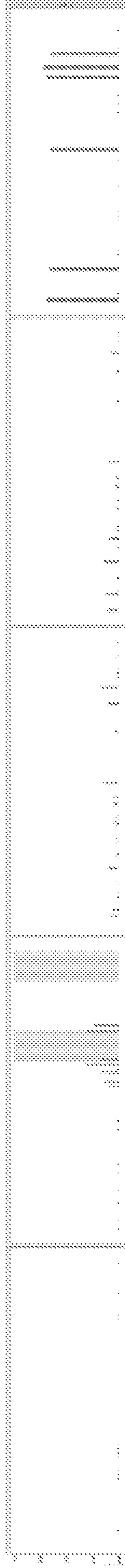


15/28

SunTag VP64 g4 line 11



SunTag VP64 g4 line 12



Blue=CG

Yellow=CHG

Green=CHH

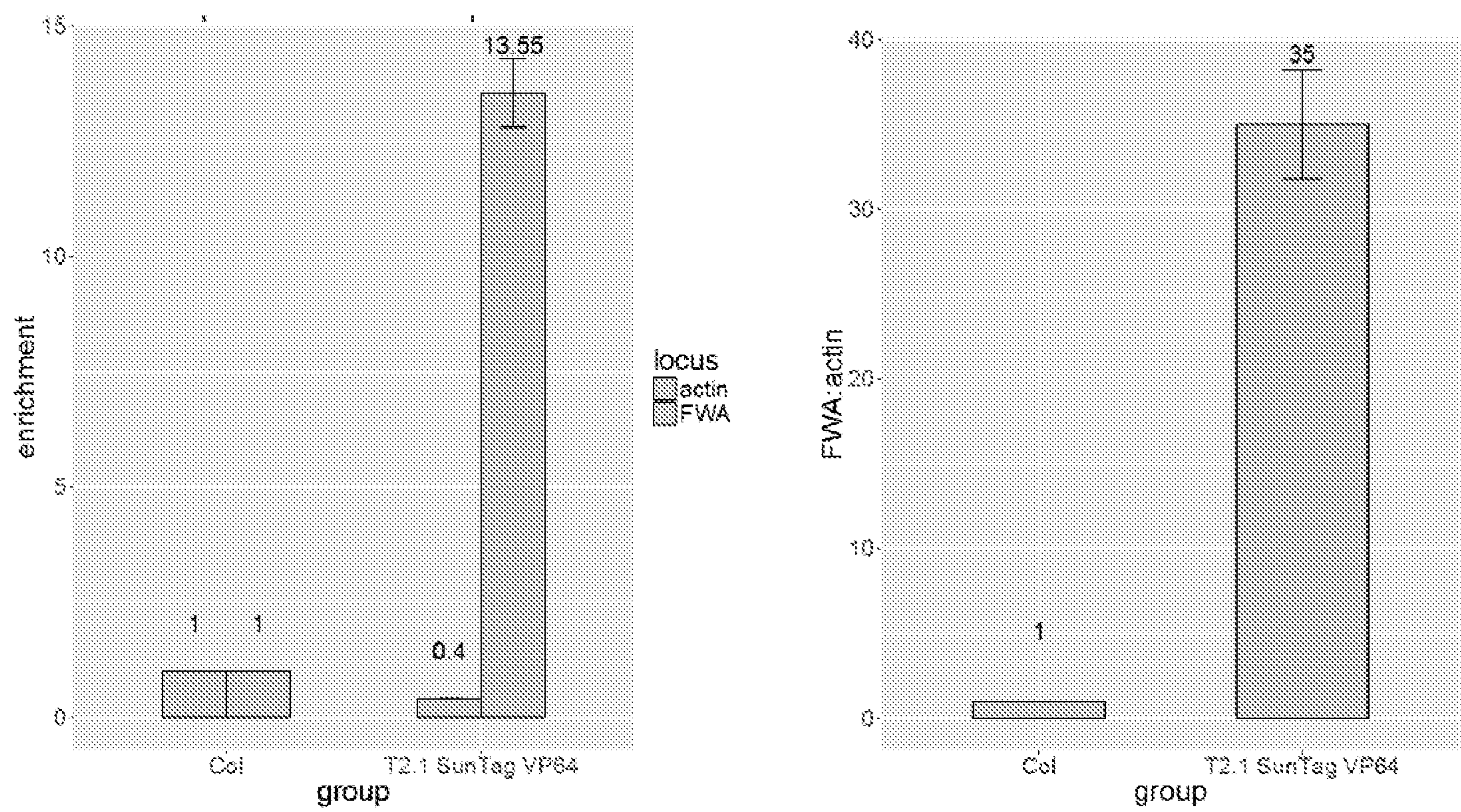


FIG. 15

FIG. 16

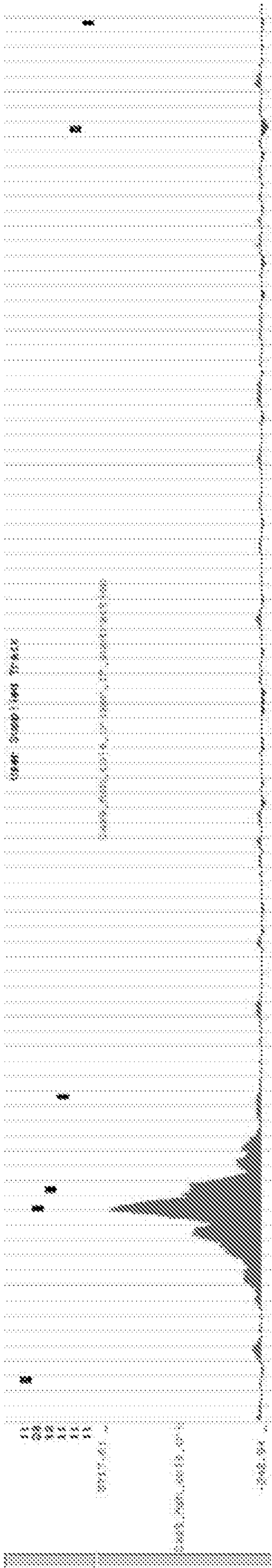
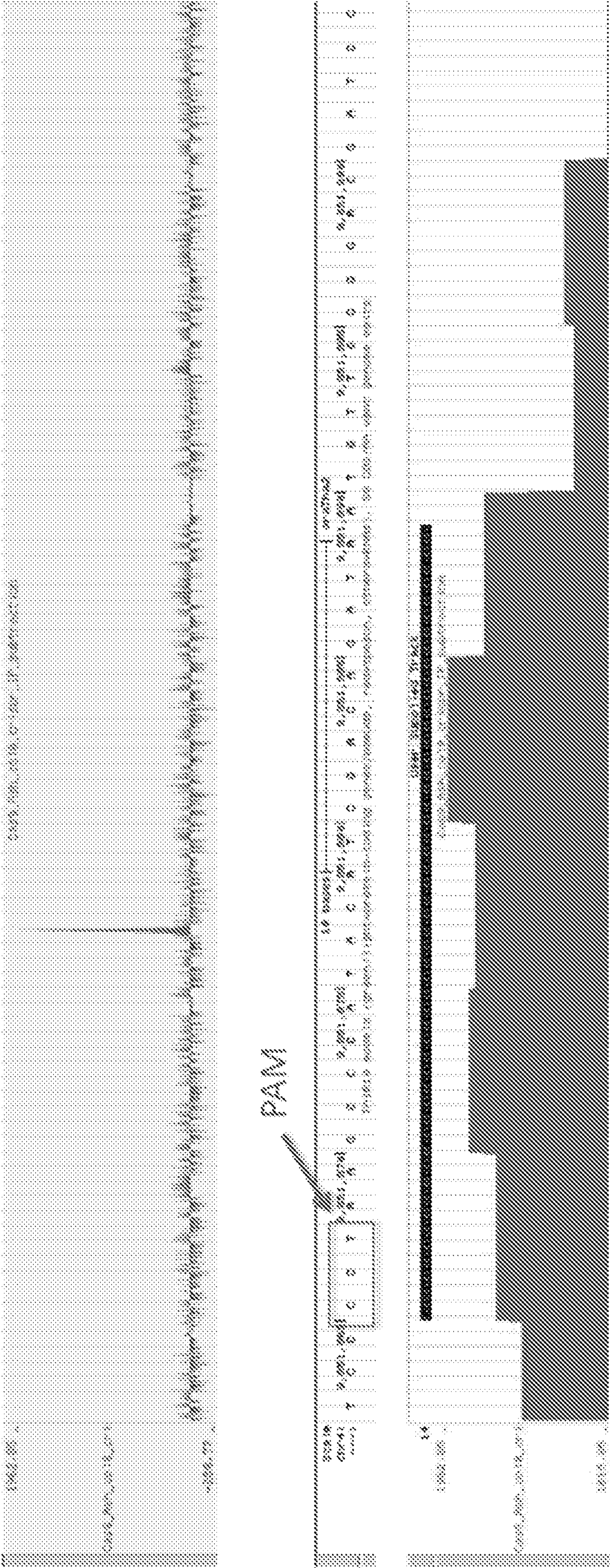


FIG. 17



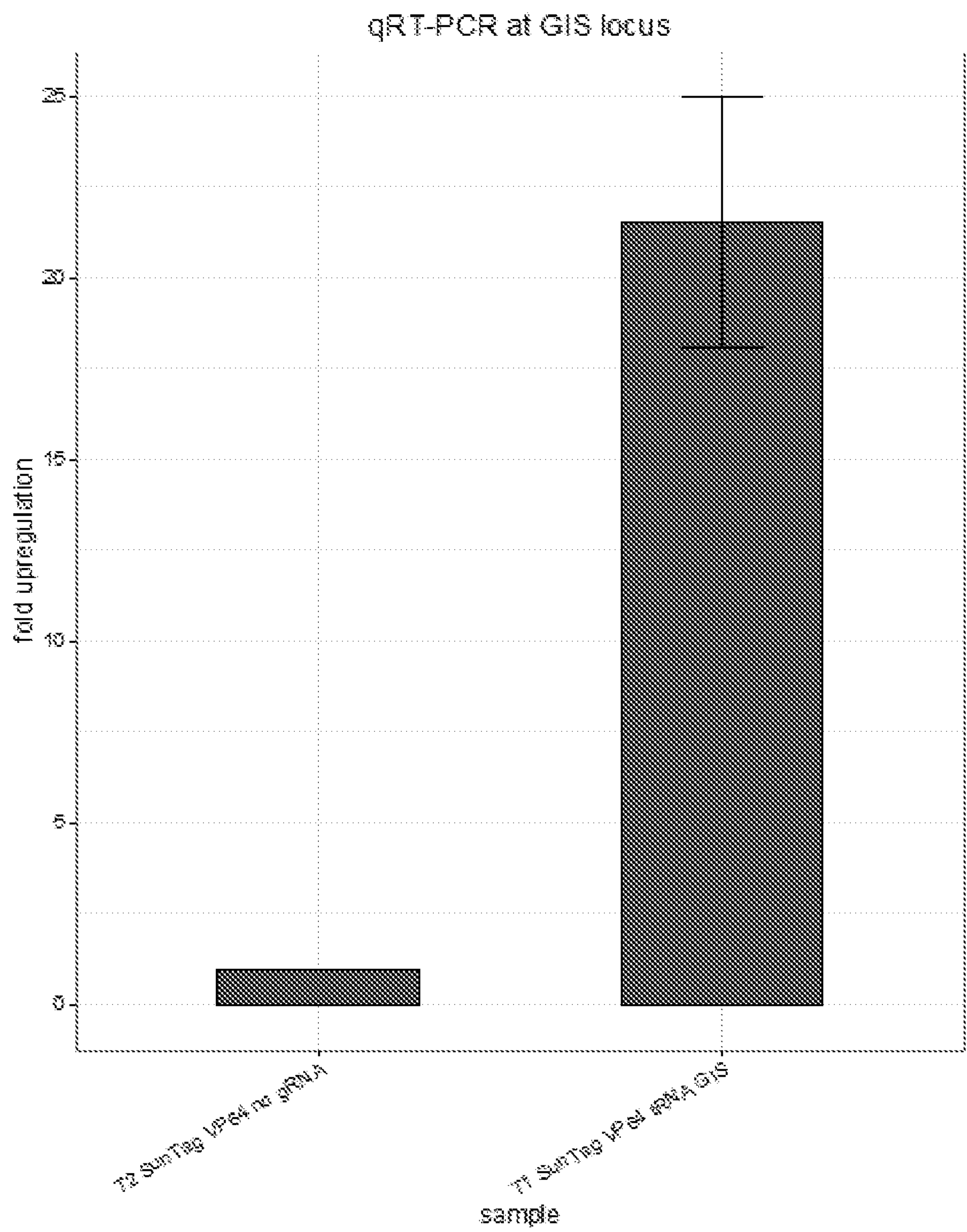
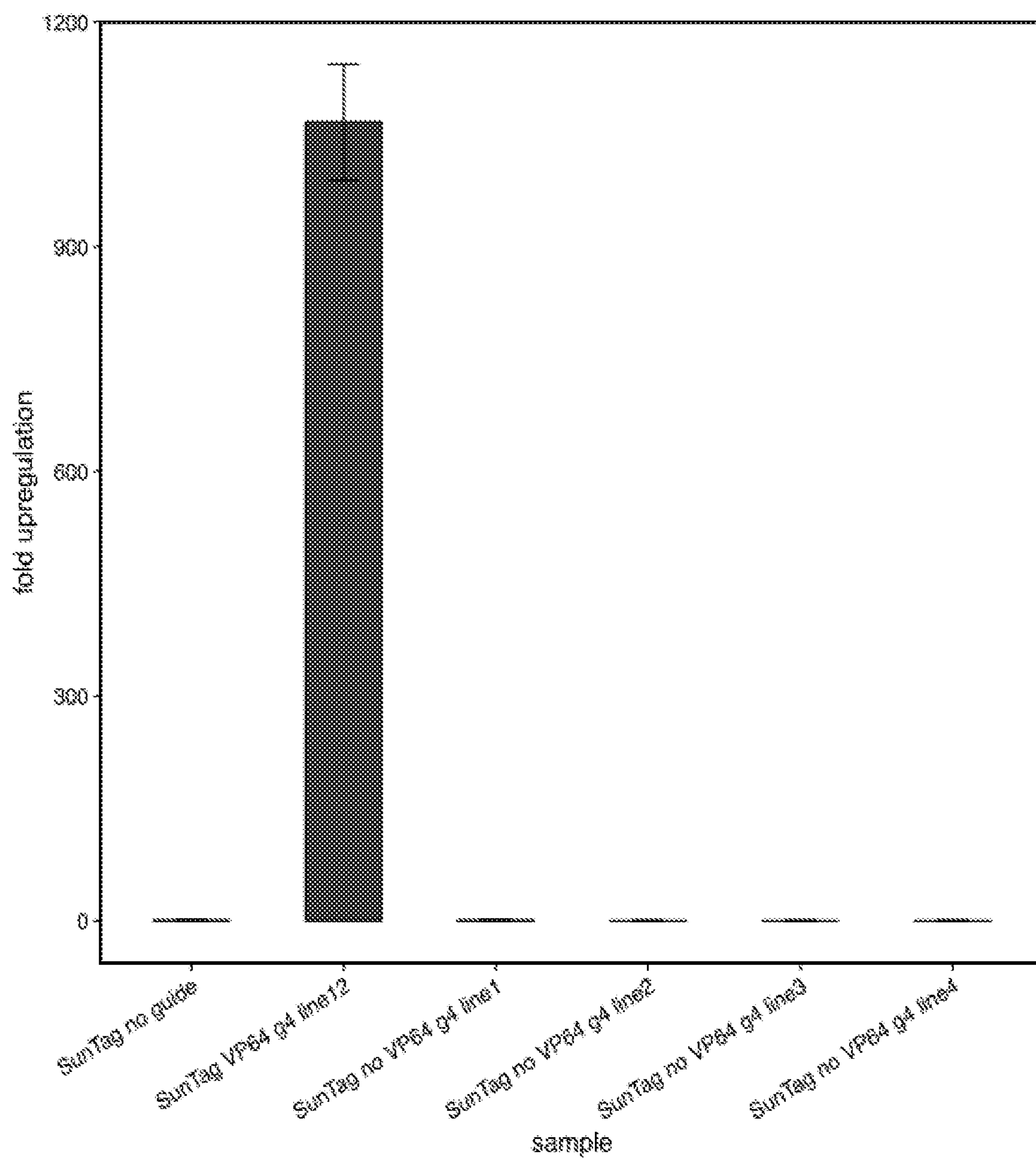
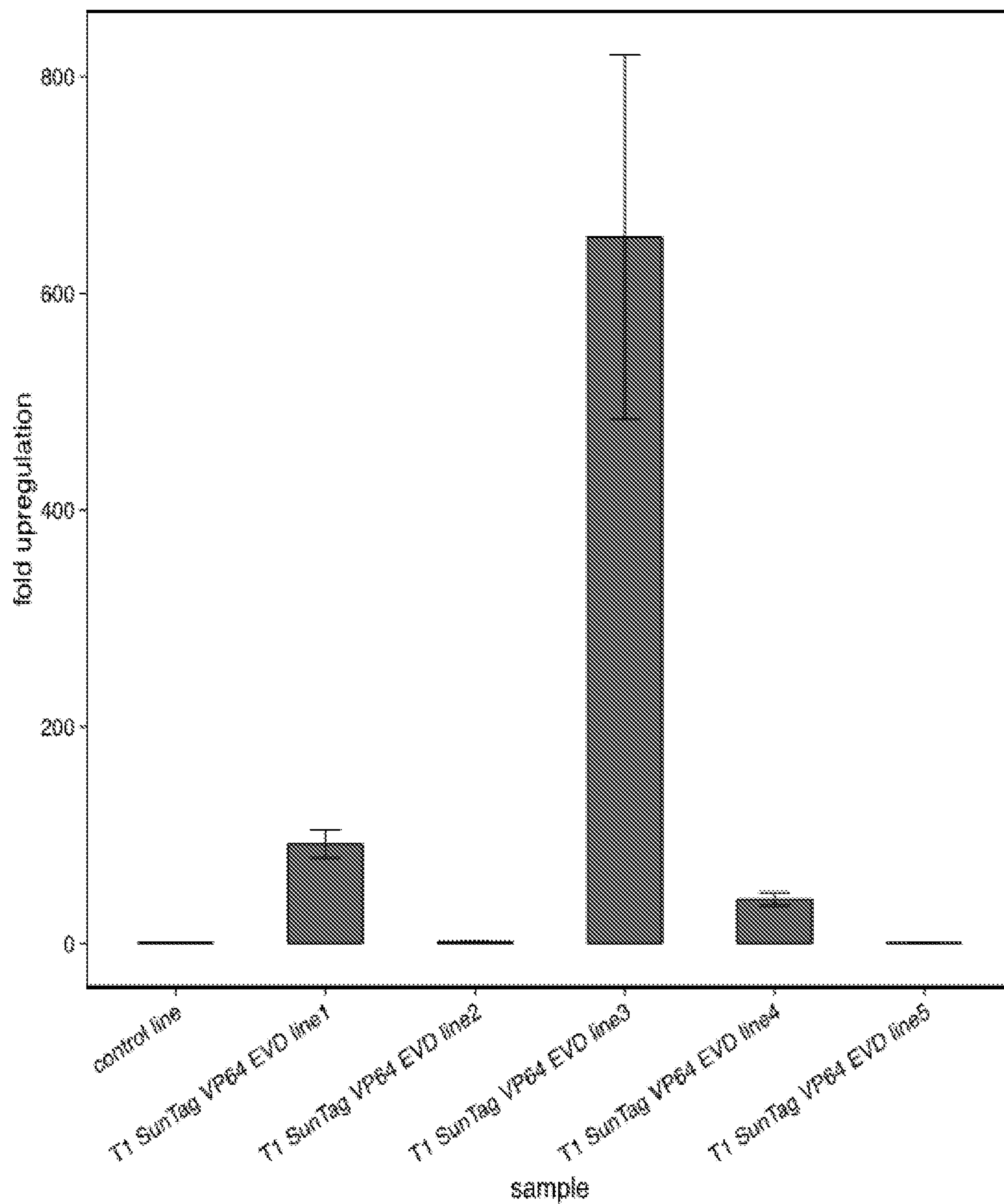


FIG. 18

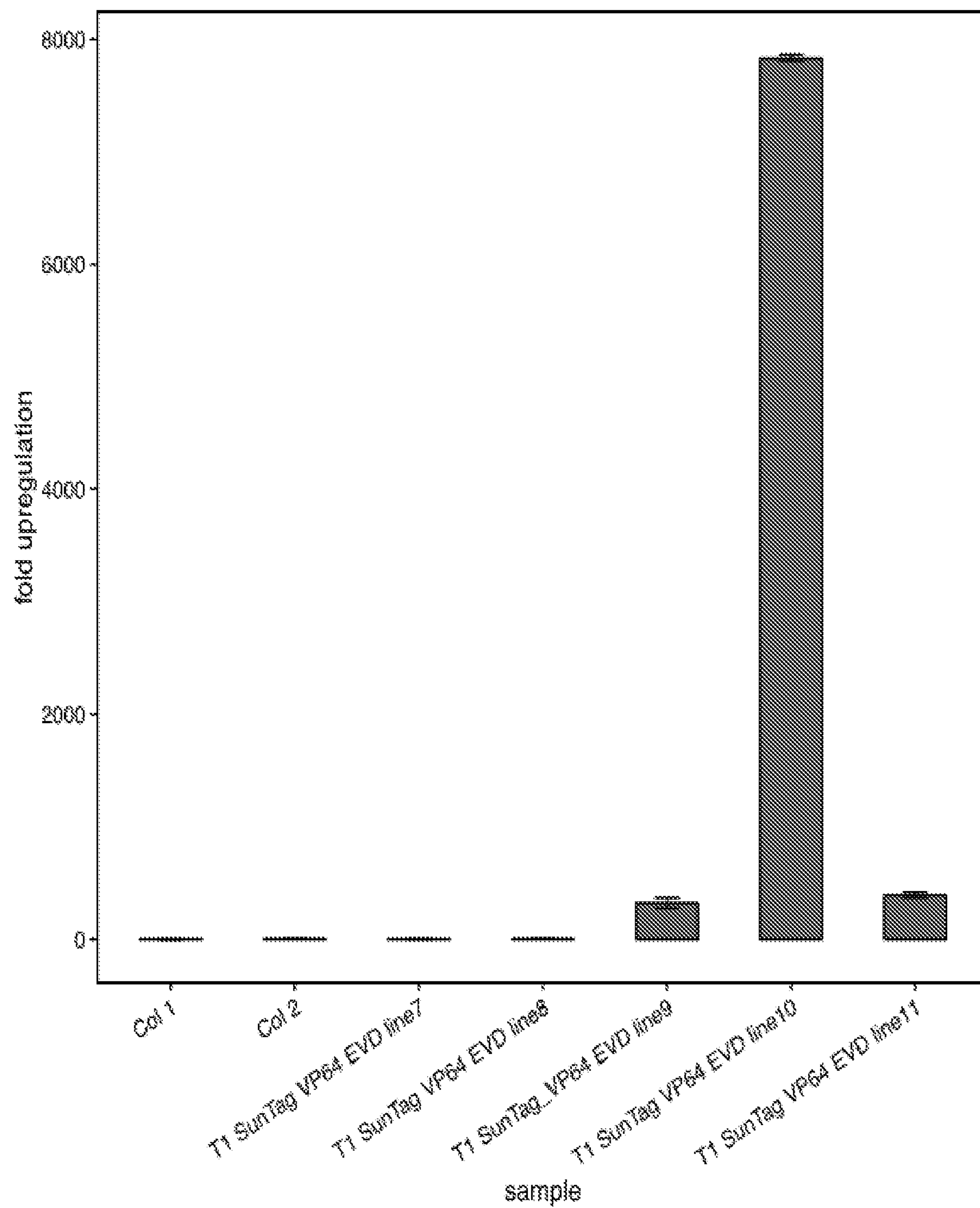
20/28

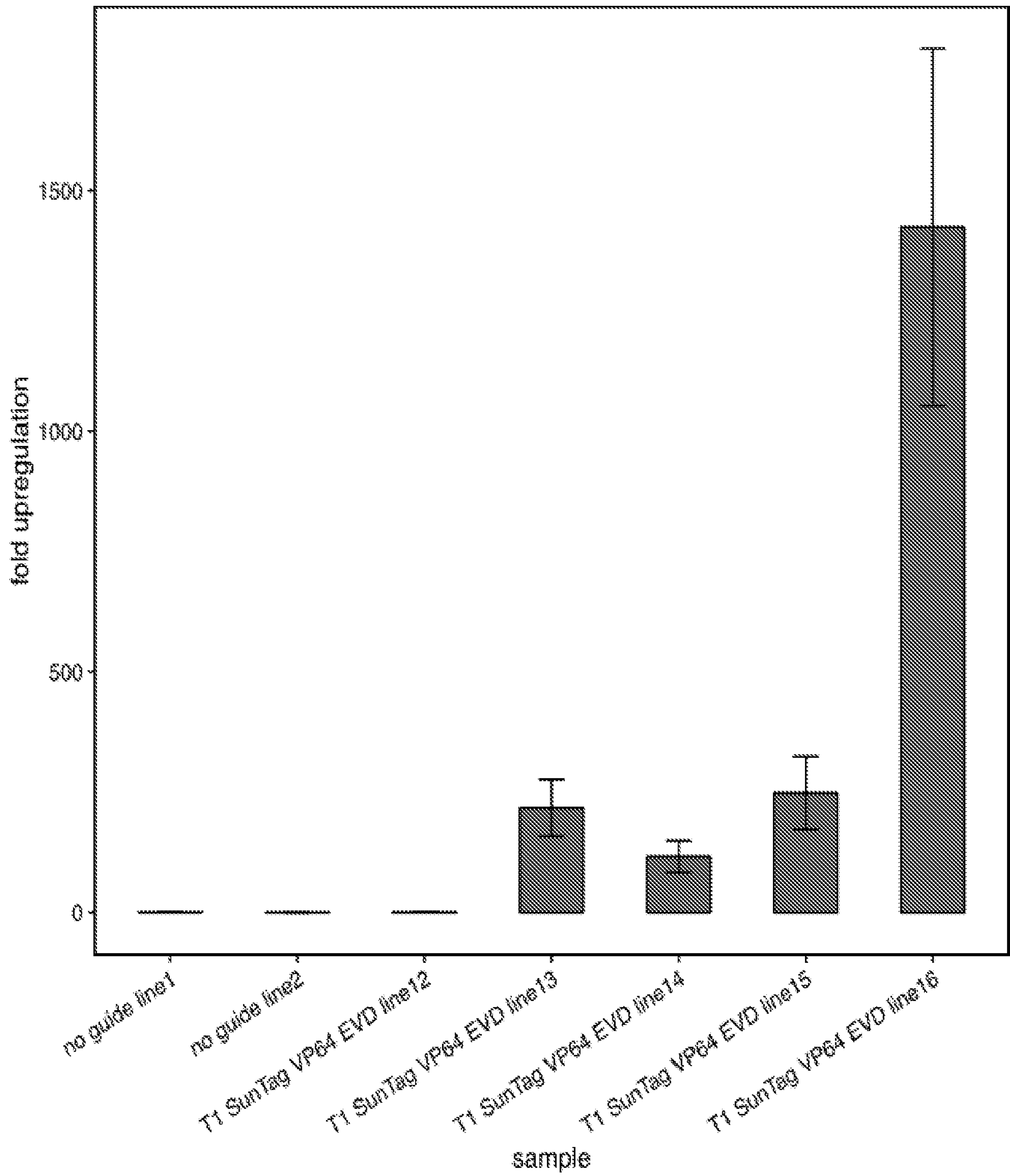
**FIG. 19**

21/28

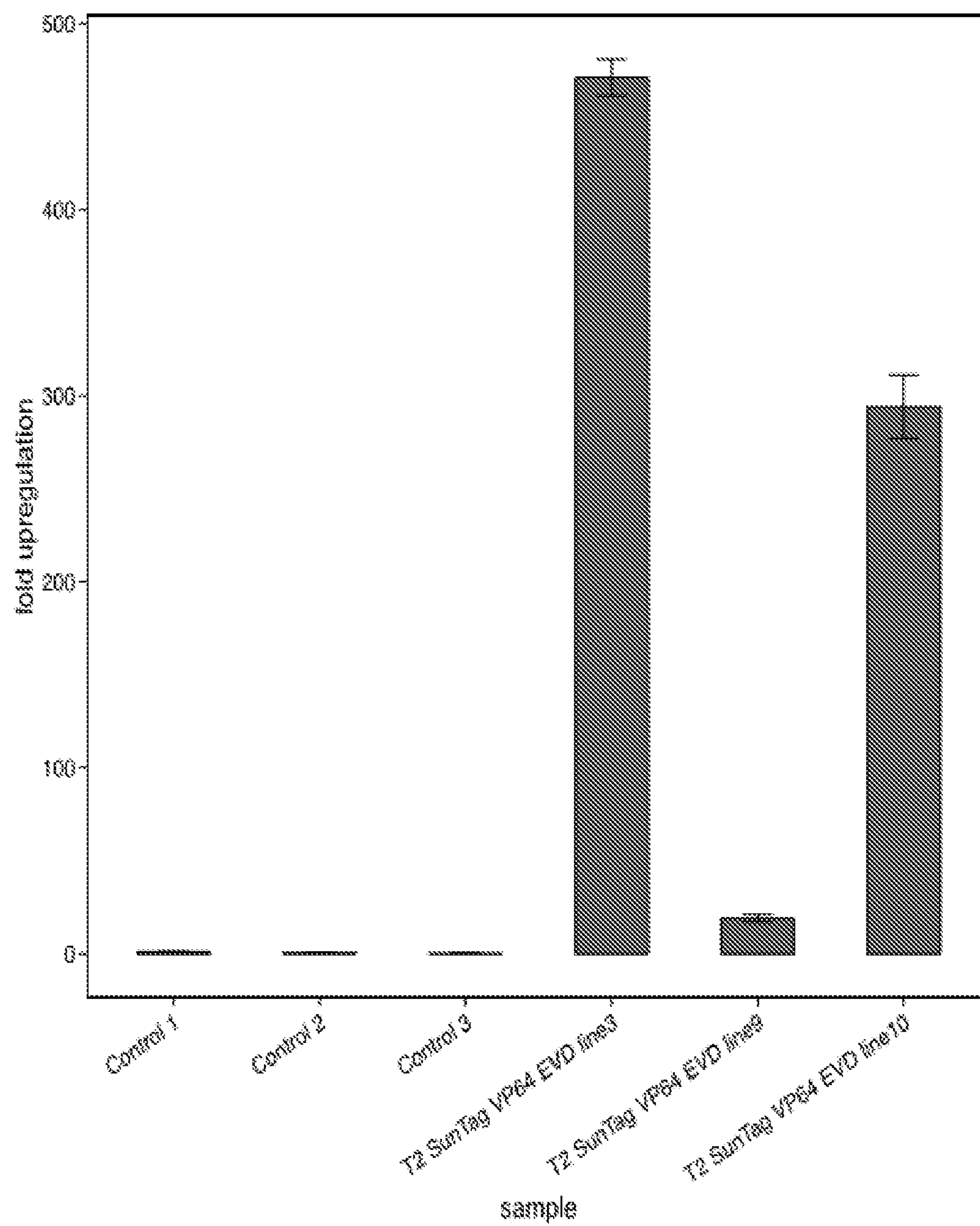
**FIG. 20A**

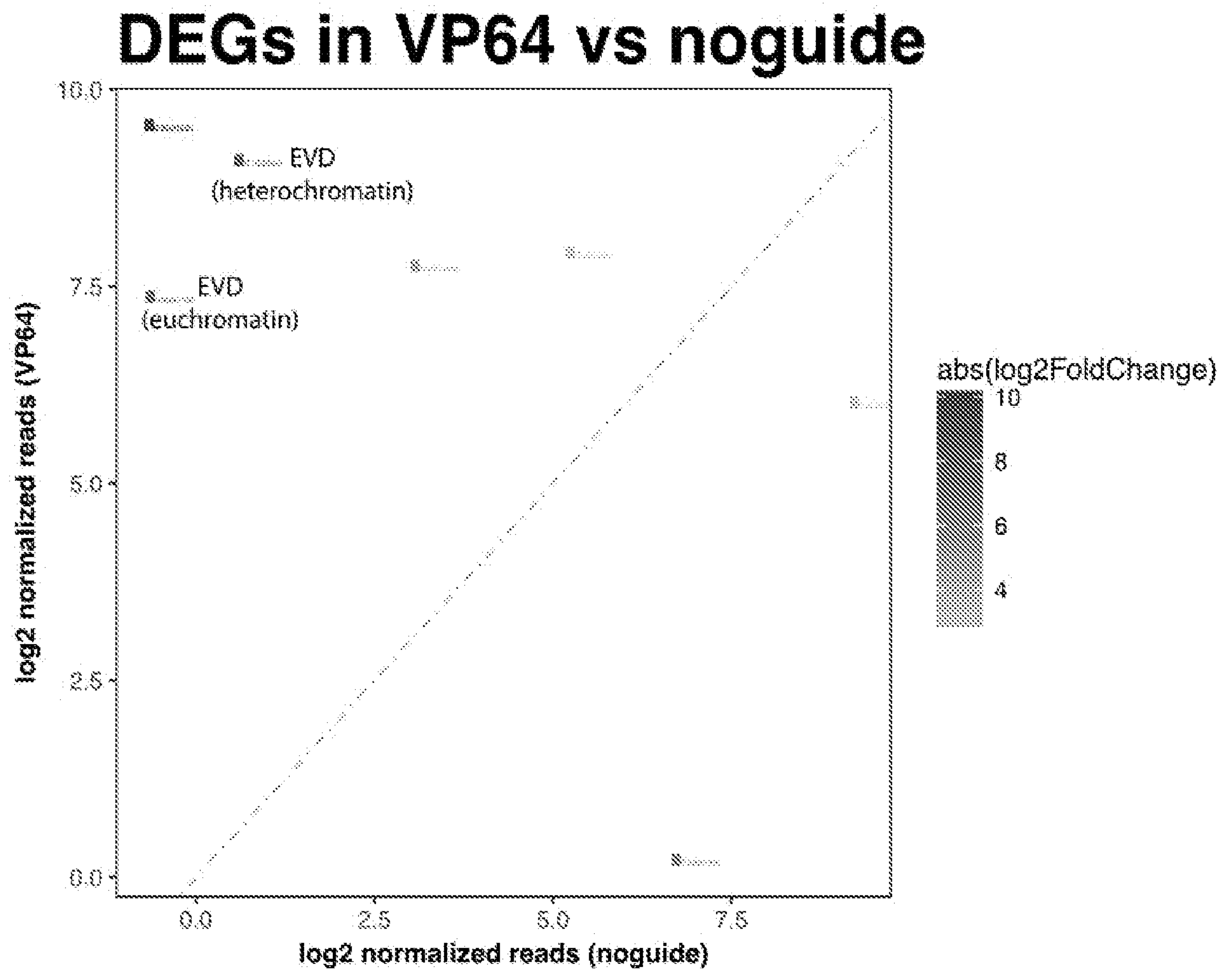
22/28

**FIG. 20B**

**FIG. 20C**

24/28

**FIG. 21**

**FIG. 22A**

DEGs in VP64 vs noguide

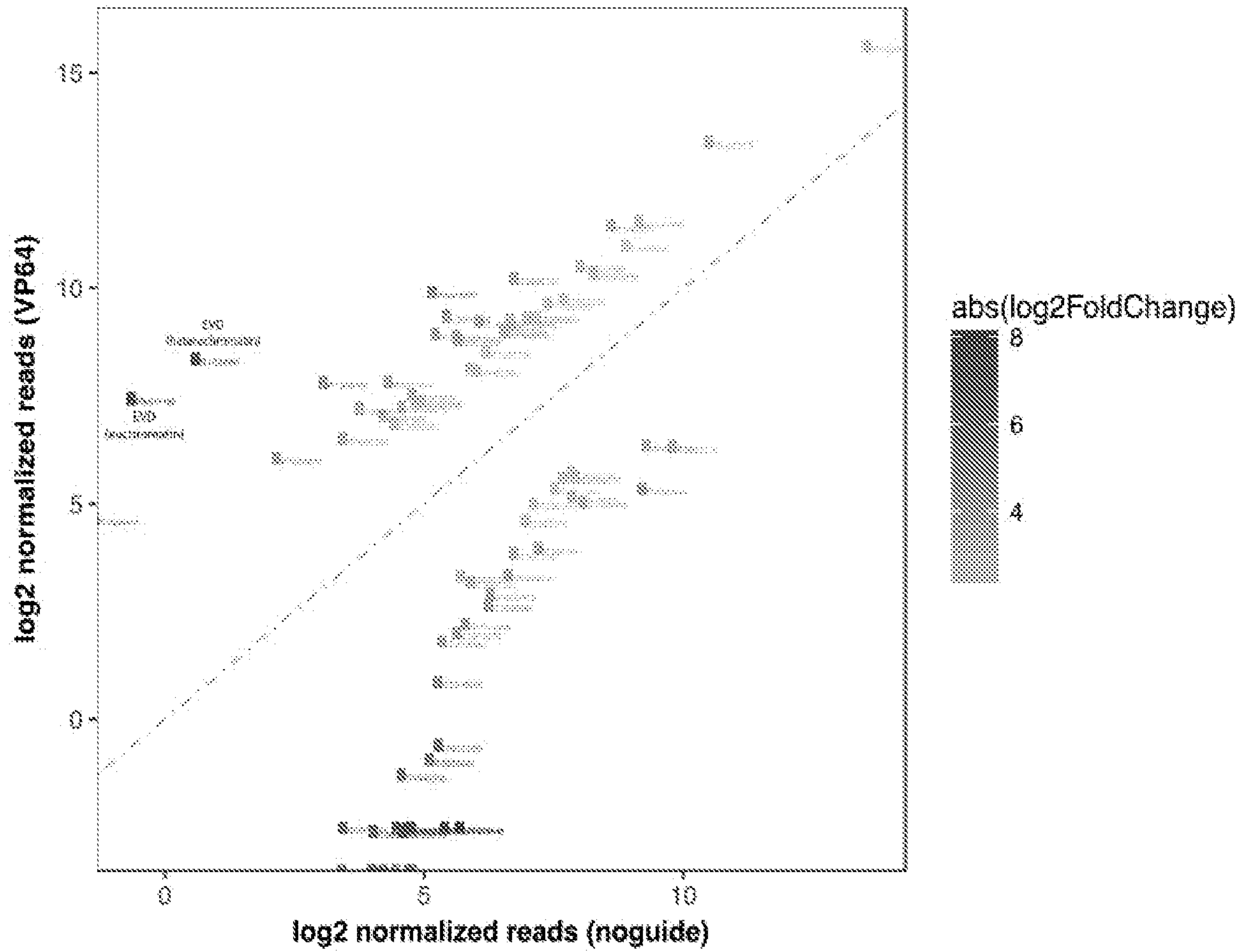
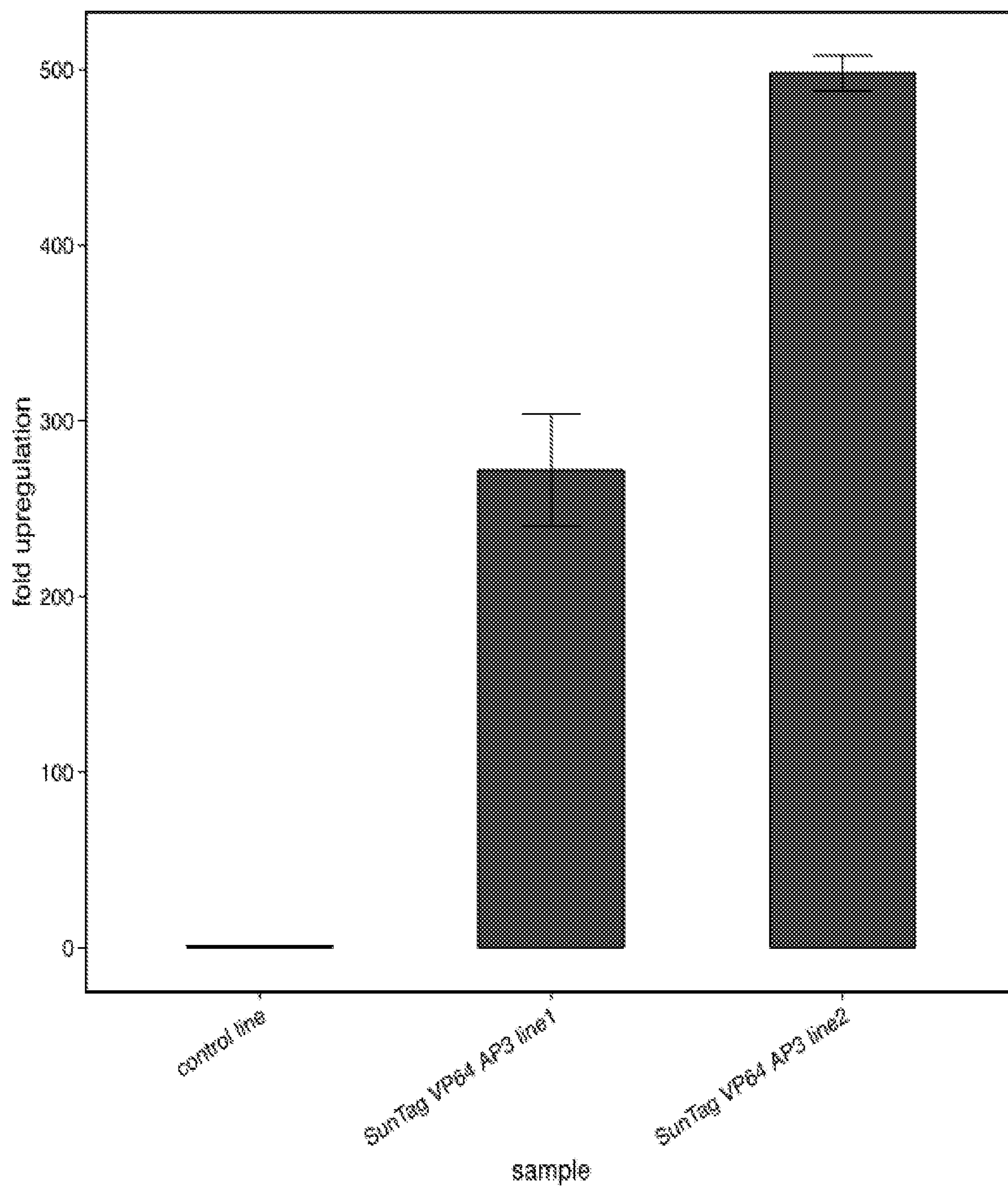
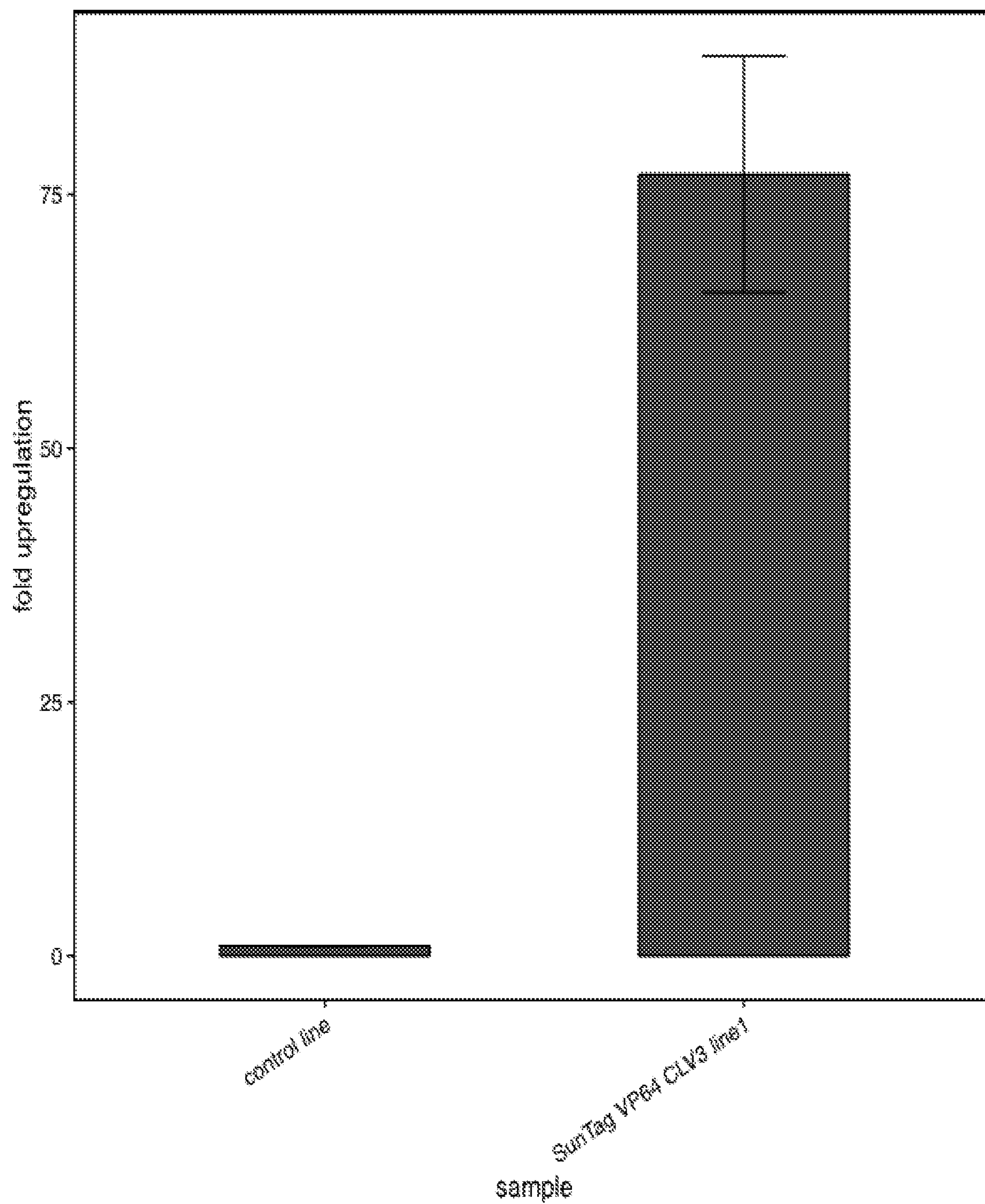


FIG. 22B

27/28

**FIG. 23**

28/28

**FIG. 24**

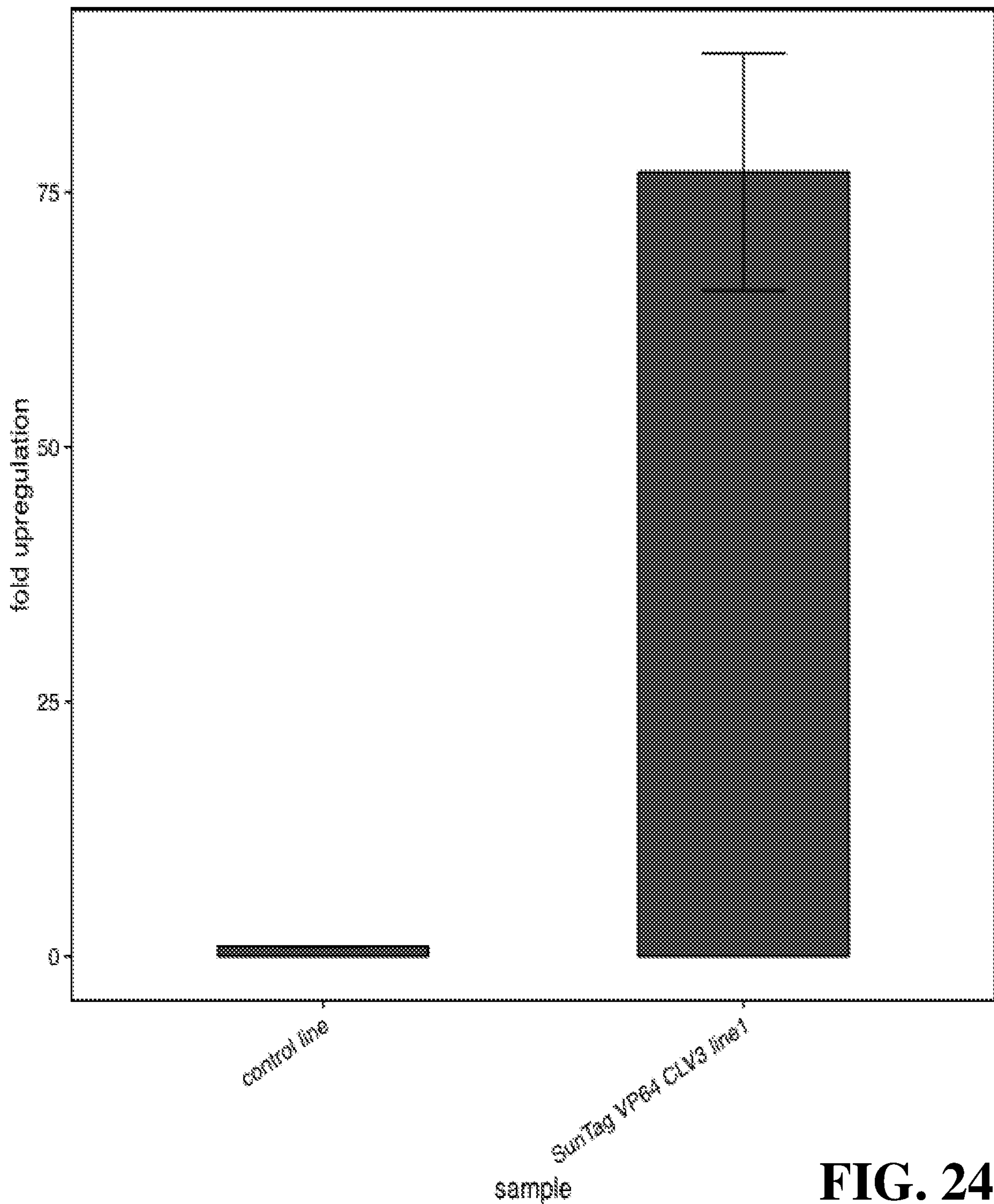


FIG. 24