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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 July 2025 (10.07.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/147656 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12Q 1/68* (2018.01)
C12N 15/11 (2006.01) *C12Q 1/6806* (2018.01)
C12N 15/86 (2006.01)

(21) International Application Number:
PCT/US2025/010297

(22) International Filing Date:
03 January 2025 (03.01.2025)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/617,726 04 January 2024 (04.01.2024) US

(71) Applicant: **UNIVERSITY OF WASHINGTON** [US/US]; 1100 NE Campus Parkway, Suite 200, Seattle, Washington 98195 (US).

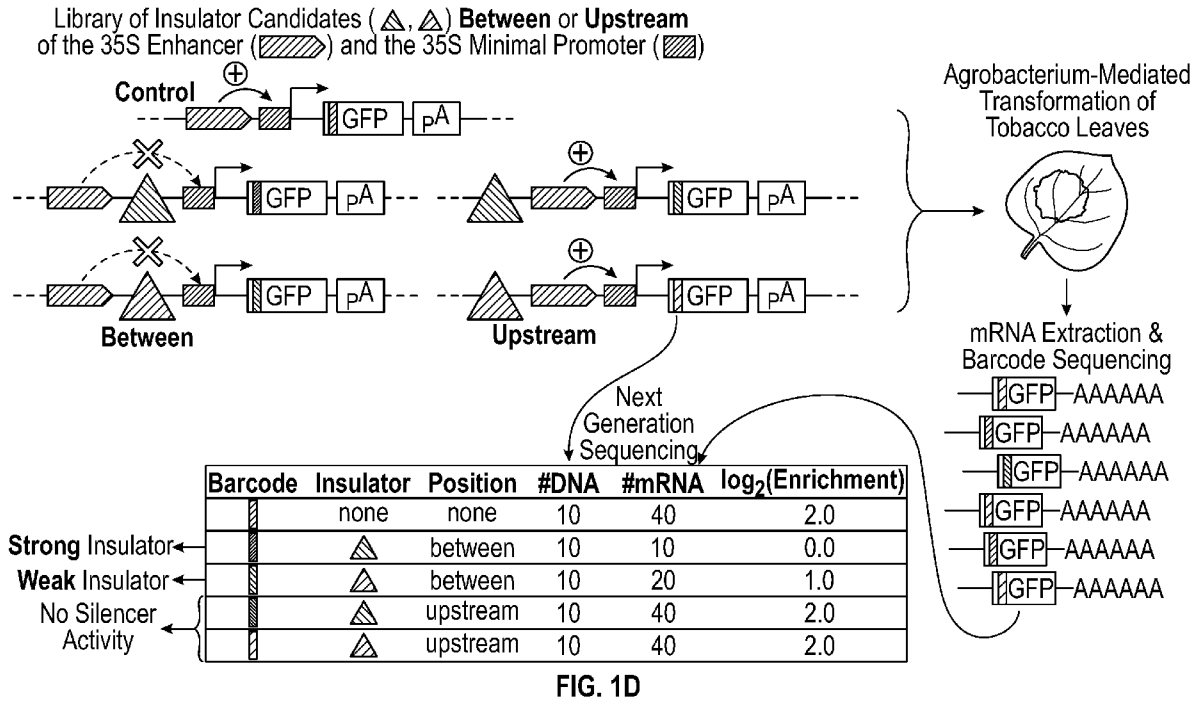
(72) Inventors: **JORES, Tobias**; c/o University of Washington, 1100 NE Campus Parkway, Suite 200, Seattle, Washington 98195 (US). **QUEITSCH, Christine**; c/o University of Washington, 1100 NE Campus Parkway, Suite 200, Seattle, Washington 98195 (US). **CUPERUS, Josh**; c/o University of Washington, 1100 NE Campus Parkway, Suite 200, Seattle, Washington 98195 (US).

(74) Agent: **CAMPBELL, Christopher T.** et al.; Christensen O'Connor Johnson Kindness PLLC, 1201 Third Avenue Suite 3600, Seattle, Washington 98101-3029 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

(54) Title: METHODS AND INSULATORS FOR REGULATION OF TRANSGENE EXPRESSION IN PLANT GENETIC ENVIRONMENTS



Barcode	Insulator	Position	#DNA	#mRNA	log_2(Enrichment)
	none	none	10	40	2.0
		between	10	10	0.0
		between	10	20	1.0
		upstream	10	40	2.0
		upstream	10	40	2.0

FIG. 1D

(57) Abstract: High-throughput methods for screening candidate DNA sequences for functionality as genetic insulators in plant genetic environments, as well as genetic insulators identified by the methods and plants that include the genetic insulators as transgenic elements. Genetic insulators can be cloned with transgenes, and introduced to plant cells, tissues, and organisms, for tighter control of transgene expression.

[Continued on next page]

WO 2025/147656 A1

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, ME, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

METHODS AND INSULATORS FOR REGULATION OF TRANSGENE EXPRESSION IN PLANT GENETIC ENVIRONMENTS

CROSS-REFERENCE(S) TO RELATED APPLICATION(S)

[0001] This PCT application claims the benefit of, and priority to, U.S. Provisional Application No. 63/617,726, filed January 04, 2024; the content of which is incorporated herein by reference in its entirety for all purposes.

STATEMENT REGARDING SEQUENCE LISTING

[0002] The Sequence Listing XML associated with this application is provided in XML format and is hereby incorporated by reference into the specification. The name of the XML file containing the sequence listing is 3915-P1334WO_UW_SL.xml. The XML file is 26,995,762 bytes; was created on December 17, 2024; and is being submitted electronically via Patent Center with the filing of the specification.

STATEMENT OF GOVERNMENT LICENSE RIGHTS

[0003] This invention was made with government support under Grant No. 1748843, awarded by the National Science Foundation. The government has certain rights in the invention.

BACKGROUND

[0004] Faced with accelerating climate change and rapid population growth, there is a significant need for crops with higher yields and greater resilience to ensure food security. Biotechnological approaches to improve agronomic traits in crops commonly rely on transgenes. However, transgenic plants often suffer from transgene mis-expression, silencing, and plant-to-plant variability within plant genetic environments. Regulatory crosstalk between the transgene promoter and cis-regulatory elements from adjacent genes in plant genetic environments can be a source for inconsistent transgene expression in plants.

[0005] In animals, genetic elements, termed enhancer-blocking insulators, have been identified that can interrupt interactions between surrounding cis-regulatory elements. Such enhancer-blocking insulators could help ensure the consistent performance of crop transgenes by preventing their mis-regulation (FIGs 1A-1B). However, to date, only a

handful of insulators that function in plants have been identified, and their large size prevents their effective use in transgenic crops. Furthermore, elements that are functional for enhancer-blocking activity in an animal genetic environment do not necessarily correlate with elements that are functional as insulators in plants.

[0006] Accordingly, there is a need for methods for identification and development of short, but highly effective plant-specific enhancer-blocking insulators, methods for the use of such insulators for improved regulation of transgenes in plants and crops, and transgenic plants and crops comprising such insulators. The present disclosure addresses these and other long-felt and unmet needs in the art.

SUMMARY

[0007] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0008] In an aspect, the disclosure provides a method for determining whether a candidate sequence is an insulator in a plant genetic environment, the method comprising: cloning the candidate sequence between an enhancer and a promoter to produce a test sequence; and attempting to initiate transcription from the test sequence, wherein reduced transcription relative to a control construct without the candidate sequence indicates the candidate sequence or a portion thereof is the insulator.

[0009] In embodiments, the method further comprises: cloning the candidate sequence upstream of the enhancer and the promoter to produce a second test sequence and attempting to initiate transcription from the second test sequence, wherein transcript levels similar to that of the control construct indicate the candidate sequence or the portion thereof is the insulator, and wherein reduced transcription relative to the control construct indicates the candidate sequence or the portion thereof is a silencer.

[0010] In embodiments, the method further comprises: cloning a plurality of candidate sequences between a plurality of enhancers and promoters to produce a plurality of first test sequences that comprise barcodes, transforming the first plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the first plurality of test sequences; cloning at least a portion of candidate sequences upstream of enhancers and promoters to produce a plurality of second test sequences that comprise

barcodes, transforming the second plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the second plurality of test sequences; measuring levels of barcode messenger RNA (mRNA) transcripts of the first plurality of test sequences and the second plurality of test sequences and measuring levels of barcode DNA of the first plurality of test sequences and the second plurality of test sequences, and calculating relative transcription levels as ratios of levels of barcode mRNA transcripts to levels of barcode DNA, wherein reduced relative transcription levels of first test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators or silencers, wherein similar relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators, and wherein reduced relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are silencers.

[0011] In embodiments, transcription levels are measured with a sequencing method.

[0012] In embodiments, the sequencing method comprises next-generation sequencing (NGS).

[0013] In embodiments, the sequencing method comprises a reverse-transcription (RT) step.

[0014] In embodiments, the plant genetic environment is a plant genome, or a portion thereof, of a plant cell, a plant tissue, or a plant organism.

[0015] In embodiments, the plant genetic environment corresponds to an agricultural plant organism or an agricultural crop.

[0016] In embodiments, the plant genetic environment comprises a tobacco genetic environment, a maize genetic environment, a rice genetic environment, a sorghum genetic environment, a tomato genetic environment, or an *Arabidopsis thaliana* genetic environment.

[0017] In embodiments, the enhancer comprises a 35S enhancer and the promoter comprises a 35S minimal promoter.

[0018] In embodiments, one or more candidate sequences are cloned in a forward orientation with respect to the plant genetic context, a reverse orientation with respect to the plant genetic context, or both.

[0019] In embodiments, one or more candidate sequences determined to be insulators comprise a higher frequency of guanine and cytosine (GC) nucleobases relative to a random sequence.

[0020] In an aspect, the disclosure provides an insulator comprising a polynucleotide sequence configured for regulation of transcription within one or more plant genetic environments.

[0021] In embodiments, the insulator does not comprise a naturally-occurring plant sequence.

[0022] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1439 (Synthetic insulator 1 (sIns1)).

[0023] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1440 (Synthetic insulator 2 (sIns2)).

[0024] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence selected from the group consisting of: SEQ ID NOs:1-1,438, SEQ ID NOs:1,441-1,446, SEQ ID NOs:1,447-2,361, SEQ ID NOs:2,362-19,190, SEQ ID NOs:19,191-20,082, SEQ ID NOs:20,083-20,523, SEQ ID NOs:20,524-20,725, and any combination thereof.

[0025] In an aspect, the disclosure provides a plant genetic construct comprising an insulator.

[0026] In an aspect, the disclosure provides a plant cell, plant tissue, or plant comprising a plant genetic construct comprising an insulator, wherein the plant cell, plant tissue, or plant is produced by a genetic engineering method.

[0027] In embodiments, the insulator is stably integrated into a genome of the plant cell, plant tissue, or plant, and the insulator regulates expression of a transgene of the genome by at least partially insulating the transgene from regulatory cross-talk.

DESCRIPTION OF THE DRAWINGS

[0028] The foregoing aspects and many of the attendant advantages of this disclosure will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings.

[0029] FIGs 1A-1J show enhancer-blocking insulators can prevent transgene mis-regulation, according to aspects of the disclosure. FIG. 1A shows enhancers (enh) and promoters (pro) can interact with each other, even over long distances. Such interactions can lead to the mis-regulation of neighboring genes. FIG. 1B shows enhancer-blocking insulators (ins) can prevent regulatory crosstalk when placed between enhancers and promoters. FIG. 1C shows an approach for distinguishing between insulator and silencer activity, according to aspects of the disclosure. The activity of an enhancer-blocking insulator can only be detected when it is situated between an enhancer and a promoter. In contrast, a silencer is also active when it resides upstream of both the enhancer and promoter. FIG. 1D shows an example high-throughput method to identify enhancer-blocking insulators active in plants, according to aspects of the disclosure. FIGs 1E-1J show synthetic insulators block the 35S enhancer in plants, according to aspects of the disclosure. FIGs 1E and 1H show the synthetic insulators sIns1 and sIns2 were cloned in the forward (fwd) or reverse (rev) orientation between the 35S enhancer and 35S minimal promoter or upstream of both elements. The enrichment of the constructs was determined in transiently transformed tobacco leaves (FIG. 1E) or maize protoplasts (FIG. 1H), normalized to a control construct without an enhancer and insulator (0, solid line) and compared to a control construct without an insulator (dashed line). FIGs 1F and 1I show a construct with a Luciferase (Luc) gene driven by the constitutive UBQ10 promoter and a NanoLuciferase (NanoLuc) gene controlled by the 35S minimal promoter and 35S enhancer was integrated into the genome of *Arabidopsis thaliana*. Insulators were cloned between the 35S enhancer and minimal promoter. Leaf discs from transgenic T2 plants were harvested and their proteins extracted. The luminescence of NanoLuc and Luc was measured in these extracts (FIG. 1F). The relative expression level of NanoLuc was determined by normalizing its

activity with that of the internal control Luc [log2(NanoLuc/Luc)]. The relative expression of NanoLuc for constructs with the synthetic insulators is shown alongside that of controls without an enhancer and insulator (noEnh) or with the 35S enhancer but without an insulator (noIns) (FIG. 1I). FIGs 1G and 1J show 170 bp fragments of the synthetic insulator were cloned in the forward (right-pointing arrows) or reverse (left-pointing arrows) orientation between the 35S enhancer and 35S minimal promoter or upstream of both elements and the activity of the resulting constructs was determined. The start and end points of the arrows indicate the start and end coordinates of the fragments in the full-length insulator. The enrichment of the constructs was determined in transiently transformed tobacco leaves (FIG. 1G) or maize protoplasts (FIG. 1J), normalized to a control construct without an enhancer and insulator (0, solid line) and compared to a control construct without an insulator (dashed line).

[0030] FIGs 2A-2F show short fragments exhibit enhancer-blocking insulator activity, according to aspects of the disclosure. FIG. 2A shows known insulators were split into partially overlapping 170-bp fragments. The insulator fragments were cloned in the forward or reverse orientation between a 35S, AB80, or Cab-1 enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. Constructs without an enhancer (none) but with insulator fragments were also created. FIG. 2B shows all insulator fragment constructs were pooled and subjected to Plant STARR-seq in tobacco leaves (tobacco) and maize protoplasts (maize). Reporter mRNA enrichment was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0). The enrichment of a control construct without an insulator is indicated as a black dot. Violin plots represent the kernel density distribution and the box plots inside represent the median (center line), upper and lower quartiles, and 1.5× interquartile range (whiskers) for all corresponding constructs. Numbers at the bottom of each violin indicate the number of samples in each group. FIG. 2C shows correlation between the enrichment of insulator fragments in constructs with the 35S enhancer in tobacco leaves and maize protoplasts. FIG. 2D shows enrichment of constructs with insulator fragments cloned between the 35S enhancer and minimal promoter. The position along the full-length insulator and the orientation (arrow pointing right, fwd; arrow pointing left, rev) of the fragments is indicated by arrows. FIG. 2E shows correlation between insulator fragment enrichment and GC content for constructs with the 35S enhancer. FIG. 2F shows correlation between insulator fragment enrichment in tobacco leaves in constructs with the indicated enhancers. The dashed line represents a

y = x line fitted through the point corresponding to a control construct without an insulator (black dot). Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated in FIGs 2C, 2E, and 2F.

[0031] FIGs 3A-3K show insulators are active in stable transgenic lines in *Arabidopsis*, rice, and maize, according to aspects of the disclosure. FIG. 3A shows transgenic *Arabidopsis* and rice lines were generated with T-DNAs harboring a constitutively expressed luciferase (Luc) gene and a NanoLuciferase (NanoLuc) gene under control of a 35S minimal promoter coupled to the 35S or AB80 enhancer (as indicated above the plots) with insulator candidates inserted between the enhancer and promoter. NanoLuciferase activity was measured in at least 4 plants from these lines and normalized to the activity of luciferase. The NanoLuc/Luc ratio was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0). FIGs 3B and 3C show the activity of full-length insulators was measured in *Arabidopsis* lines (FIG. 3B) and compared to the corresponding results from Plant STARR-seq in tobacco leaves (FIG. 3C). FIGs 3D and 3E show the activity of synthetic full-length insulators was measured in rice lines (FIG. 3D) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (FIG. 3E). FIGs 3F and 3G show the activity of insulator fragments was measured in *Arabidopsis* lines (FIG. 3F) and compared to the corresponding results from Plant STARRseq in tobacco leaves (FIG. 3G). FIG. 3H shows for transgenic maize lines, a reporter gene driven by a moderate-strength constitutive promoter (ZmPro) and an upstream 35S enhancer was created and insulator fragments were inserted between the enhancer and promoter. The reporter gene cassette was inserted in the maize genome by site-directed integration and the expression of the reporter gene was measured in various tissues/developmental stages by ELISA. FIGs 3I and 3J show the activity of insulator fragments was measured in R1 leaves of transgenic maize lines (FIG. 3I) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (FIG. 3J). FIG. 3K shows correlation (Pearson's R^2) between the expression of all tested constructs across different tissues and developmental stages. The correlation with Plant STARR-seq results from maize protoplasts is also shown. Box plots in FIGs 3B, 3D, 3F, and 3I represent the median (center line), upper and lower quartiles (box limits), 1.5× interquartile range (whiskers), and outliers (points) for all corresponding samples from two to three independent replicates. Numbers at the bottom of each box plot indicate the number of samples in each group. In FIGs 3C, 3E, 3G, and 3J, the dashed line represents a linear

regression line and error bars represent the 95% confidence interval. Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated.

[0032] FIGs 4A-4G show insulator fragments can be stacked to create very strong enhancer-blocking insulators, according to aspects of the disclosure. FIG. 4A shows one, two, or three 170-bp fragments of known insulators were cloned between a 35S enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. FIG. 4B shows all insulator constructs were pooled and subjected to Plant STARR-seq in tobacco leaves (tobacco) and maize protoplasts (maize). Reporter mRNA enrichment was normalized to a control construct without an enhancer or insulator (log2 set to 0). Violin plots are as defined in FIG. 2B. The enrichment of a control construct without an insulator is indicated as dotted line. FIG. 4C shows a linear model was trained to predict the enrichment of stacked insulator constructs based on the activity of individual insulator fragments and their position within the construct. The correlation between the model's prediction (prediction) and experimentally determined enrichment values (measurement) is shown as a hexbin plot (shade represents the count of points in each hexagon). Pearson's R^2, Spearman's ρ, and number (n) of fragments are indicated. FIG. 4D shows coefficients assigned by the linear model to insulator fragments in the indicated positions of the stacked constructs. FIGs 4E and 4F show the activity of insulator fragment combinations in constructs as in FIG. 2H was measured in R1 leaves of transgenic maize lines (FIG. 4E) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (FIG. 4F). Box plots are as defined in FIGs 2A-2F. The enrichment of a control construct without an insulator (noIns) is indicated as a dotted line. In FIG. 4F, the dashed line represents a linear regression line and error bars represent the 95% confidence interval. Pearson's R^2 , Spearman's ρ, and number (n) of constructs are indicated. FIG. 4G shows correlation (Pearson's R^2) between the expression of all tested constructs across different tissues and developmental stages. The correlation with Plant STARR-seq results from maize protoplasts is also shown.

[0033] FIGs 5A-5F show insulators exhibit silencer activity in at least some contexts, according to aspects of the disclosure. FIG. 5A shows insulator fragments were cloned upstream of an AB80 or Cab-1 enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. Half of the constructs also harbored a 35S enhancer upstream of the insulator fragments (with 35S) while the other half lacked an upstream enhancer (without 35S). FIG. 5B shows all constructs were pooled and subjected

to Plant STARR-seq in tobacco leaves. Reporter mRNA enrichment was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0). The enrichment of a control construct without an insulator is indicated as a black dot. FIG. 5C shows correlation between insulator fragment activity in constructs with or without the upstream 35S enhancer. The dashed line represents a y = x line fitted through the point corresponding to a control construct without an insulator (black dot). FIG. 5D shows insulator fragments were cloned in between (insulator construct) or upstream of (silencer construct) a 35S enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. FIG. 5E shows all constructs were pooled and subjected to Plant STARR-seq in tobacco leaves (tobacco) or maize protoplasts (maize). Reporter mRNA enrichment was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0).The enrichment of a control construct without an insulator is indicated as a black dot. FIG. 5F shows comparison of the enrichment of insulator fragments in insulator or silencer constructs. A linear regression line is shown as a solid line and its slope and goodness-of-fit (R^2) is indicated. Violin plots in FIGs 5B and 5E are as defined in FIG. 2B.

[0034] FIGs 6A-6D show silencer activity depends on enhancer strength, according to aspects of the disclosure. FIG. 6A shows selected insulators and insulator fragments were cloned in between (insulator construct) or upstream of (silencer construct) an enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. Eight different enhancers were used to build these constructs. All constructs were pooled and subjected to Plant STARR-seq in tobacco leaves (tobacco) or maize protoplasts (maize). FIG. 6B shows strength of the eight enhancers in constructs without an insulator. Reporter mRNA enrichment was normalized to a control construct without an enhancer (none; $\log_2$ set to 0). Box plots represent the median (center line), upper and lower quartiles, and 1.5× interquartile range (whiskers) for all corresponding barcodes from two independent replicates. Numbers at the bottom of the plot indicate the number of samples in each group. FIG. 6C shows comparison of the enrichment of insulators and insulator fragments in insulator or silencer constructs. A linear regression line is shown as a solid line and its slope and goodness-of-fit (R^2) is indicated. FIG. 6D shows correlation between the slope of the regression lines from FIG. 6C and the strength of the corresponding enhancer from FIG. 6B. Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated. A linear regression line is shown as a dashed line.

[0035] FIGs 7A-7B show Plant STARR-seq detects activity of enhancer-blocking insulators, according to aspects of the disclosure. FIG. 7A shows full-length insulators were cloned in the forward (fwd) or reverse (rev) orientation between a 35S enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. FIG. 7B shows all insulator constructs were pooled and subjected to Plant STARR-seq in tobacco leaves (tobacco) and maize protoplasts (maize). Reporter mRNA enrichment was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0). Box plots represent the median (center line), upper and lower quartiles, and 1.5× interquartile range (whiskers) for all corresponding barcodes from two independent replicates. Numbers at the bottom of the plot indicate the number of samples in each group. The enrichment of a control construct without an insulator (noIns) is indicated as a dotted line.

[0036] FIGs 8A-8F show Plant STARR-seq yields highly reproducible results, according to aspects of the disclosure. FIGs 8A-8G show correlation between biological replicates of Plant STARR-seq for the full-length insulator library used in FIGs 7A-7B (FIG. 8A), the insulator fragment library used in FIGs 2A-2F (FIG. 8B), the insulator fragment combination library used in FIGs 4A-4G (FIG. 8C), the downstream enhancer library (FIG. 8D) and the insulator/silencer library used in FIGs 5A-5F (FIG. 8E), and the enhancer-insulator combination library used in FIGs 6A-6C (FIG. 8F). Experiments were performed in tobacco leaves (tobacco) or maize protoplasts (maize) as indicated. Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated. The shade in the hexbin plots in FIG. 8C represents the count of points in each hexagon.

[0037] FIGs 9A-9B show activity of insulator fragments in different maize tissues, according to aspects of the disclosure. FIGs 9A and 9B show transgenic maize lines were created using constructs as in FIG. 3H. The activity of insulator fragments was measured in the indicated tissues (FIG. 9A) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (FIG. 9B). Box plots in FIG. 9A are as defined in FIGs 3A-3K.

[0038] FIGs 10A-10B show activity of insulator fragment combinations in different maize tissues, according to aspects of the disclosure. FIGs 10A and 10B show transgenic maize lines were created using insulator fragment combinations in constructs as in FIG. 3H. The activity of insulator fragments was measured in the indicated tissues (FIG. 10A) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (FIG. 10B). Box plots in (FIG. 10A) are as defined in FIG. 3.

[0039] FIG. 11 shows enhancers downstream of insulator fragments slightly reduce their activity, according to aspects of the disclosure. Correlation between the activity of insulator fragments cloned between a 35S enhancer and a 35S minimal promoter with or without an additional AB80 or Cab-1 enhancer inserted between the insulator fragment and 35S minimal promoter. The dashed line represents a y = x line fitted through the point corresponding to a control construct without an insulator (black dot). Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated.

[0040] FIGs 12A-12D show enhancer-dependent silencer activity in stable transgenic plants, according to aspects of the disclosure. FIG. 12A shows transgenic *Arabidopsis* lines were generated with T-DNAs harboring a constitutively expressed luciferase (Luc) gene and a NanoLuciferase (NanoLuc) gene under control of a 35S minimal promoter coupled to the 35S or AB80 enhancer (as indicated above the plots) with insulator candidates inserted upstream of the enhancer. NanoLuciferase activity was measured in at least 4 plants from these lines and normalized to the activity of luciferase. The NanoLuc/Luc ratio was normalized to a control construct without an enhancer or insulator (noEnh; $\log_2$ set to 0). FIGs 12B and 12C show the activity of full-length insulators was measured in *Arabidopsis* lines (FIG. 12B) and compared to the corresponding results from Plant STARR-seq in tobacco leaves (FIG. 12C). Box plots in FIG. 12B are as defined in FIG. 3. In FIG. 12C, the dashed line represents a linear regression line and error bars represent the 95% confidence interval. Pearson's R^2, Spearman's ρ, and number (n) of constructs are indicated. FIG. 12D shows comparison of the mean NanoLuc/Luc ratio of full-length insulators in insulator (FIG. 3B) or silencer constructs (FIG. 3B). A linear regression line is shown as a solid line and its slope and goodness-of-fit (R^2) is indicated.

DETAILED DESCRIPTION

[0041] The present disclosure provides high-throughput methods for characterizing candidate polynucleotide sequences as insulators or silencers on a large scale, as well as methods for positively identifying sequences as insulators for use in methods of plant genetic engineering. Screening methods can be performed in whole or in part, in any order of steps, by a person having ordinary skill in the art assisted by one or more computational devices or systems. Novel insight and the resultant development of a screening platform, as disclosed herein, leverages differences in regulatory crosstalk and gene expression

observed as a result of the placement of candidate sequences relative to other plant genetic elements. These approaches can be used to identify polynucleotide sequences with insulator and/or silencer activity (FIGs 1A-1D). Insulators can be cloned in combination with one or more transgenes for delivery into and expression by a plant host, for reduced regulatory cross-talk and spurious expression of the one or more transgenes by the plant host. The insulators can comprise one or more plant DNA sequences, one or more non-plant DNA sequences, or both. In at least some embodiments, insulators of the disclosure are synthetic and contain no plant DNA sequences.

[0042] The present disclosure also provides plant insulators identified with such screening methods, as well as plant genetic constructs, plant cells, plant tissues, and plant organisms that stably incorporate one or more plant insulators for tighter regulation of transgenes in transgenic plants and reduced regulatory crosstalk with other genetic elements within the plant genomic environment. Plants that are genetically engineered with one or more insulators of the disclosure exhibit reduced spurious or undesired expression of transgenes. In various aspects, the disclosure provides fundamental and significant improvements in approaches for plant gene regulation that collectively enable further advancement in plant genetic engineering techniques for improved crop stability, resilience, yield, and other desirable or necessary agricultural characteristics for reliable crop production in the face of climate change.

[0043] Accordingly, in an aspect, the disclosure provides a method for determining whether a candidate sequence is an insulator in a plant genetic environment. The method comprises cloning the candidate sequence between an enhancer and a promoter to produce a test sequence and attempting to initiate transcription from the test sequence. The attempt to initiate transcription can include inducing a condition that would be expected to initiate transcription in the absence of the candidate sequence or with a random or control sequence in place of the candidate sequence. Once transcription initiation is attempted, transcription can be measured with any suitable technique, including but not limited to reverse-transcription PCR (RT-PCR), RNA sequencing, a next-generation sequencing (NGS) technique, and the like. Reduced transcription relative to a control construct (without the candidate sequence) indicates the candidate sequence, or a portion thereof, is functional as an insulator. In this manner, using a high-throughput approach, a large number of candidate sequences can be tested for insulator activity in plant genetic environments.

[0044] In embodiments, the method further comprises cloning the candidate sequence upstream of the enhancer and the promoter to produce a second test sequence and attempting to initiate transcription from the second test sequence. Transcript levels observed to be at least about the same or similar to that of the control construct indicate the candidate sequence, or the portion thereof, is the insulator, while reduced transcription relative to the control construct indicates the candidate sequence, or the portion thereof, is functional as a silencer.

[0045] In various embodiments, the method comprises testing a large number of candidate sequences as part of a high-throughput screening method. In embodiments, the method comprises cloning a plurality of candidate sequences between a plurality of enhancers and promoters to produce a plurality of first test sequences that comprise barcodes, transforming the first plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the first plurality of test sequences. In embodiments, the method further comprises cloning at least a portion of candidate sequences upstream of enhancers and promoters to produce a plurality of second test sequences that comprise barcodes, transforming the second plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the second plurality of test sequences. Once transcription initiation is attempted, the method proceeds to a step of measuring levels of barcode messenger RNA (mRNA) transcripts of the first plurality of test sequences and the second plurality of test sequences and measuring levels of barcode DNA of the first plurality of test sequences and the second plurality of test sequences, and calculating relative transcription levels as ratios of levels of barcode mRNA transcripts to levels of barcode DNA. With reference to the transcript level results, reduced relative transcription levels of first test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators or silencers; relative transcription levels of second test sequences that are observed to be at least about the same or similar to those of the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators; and reduced relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are silencers.

[0046] In embodiments, transcription levels are measured with a sequencing method, such as a next-generation sequencing (NGS) method; however, alternate methods of transcription level measurement are envisioned without departing from the scope and

spirit of the disclosure, including but not limited to quantitative reverse transcription polymerase chain reaction (RT-PCR).

[0047] Any of various plant genetic environments can be used for developing, screening, and validating candidate sequences as insulators, including any suitable plant species, cell type, tissue type, and the like. In embodiments, a plant genetic environment is a plant genome, or a portion thereof, of a plant cell, a plant tissue, or a plant organism. The plant genetic environment can correspond to a commercially valuable species, such as an agricultural plant organism or an agricultural crop. In example embodiments, the plant genetic environment comprises a tobacco genetic environment, a maize genetic environment, a rice genetic environment, a sorghum genetic environment, a tomato genetic environment, or an *Arabidopsis thaliana* genetic environment, or any combination thereof. In addition, while any of various genetic elements can be used in methods of the disclosure, in example embodiments, an enhancer comprises a 35S enhancer and a promoter comprises a 35S minimal promoter.

[0048] In embodiments, one or more candidate sequences are cloned in a forward orientation with respect to the plant genetic context, a reverse orientation with respect to the plant genetic context, or both. Cloning candidate sequences in either or both orientations can enable a better understanding of the relationship between positioning of sequence elements of candidate sequences and positioning of enhancer and promoter sequences to be regulated by an insulator functionality. In addition, while any of various biophysical characteristics of candidate sequences can contribute to the ability of the candidate sequences to function as insulators, in at least some embodiments, one or more candidate sequences determined to be insulators comprise a higher frequency of guanine and cytosine (GC) nucleobases relative to a random sequence. In embodiments, the higher frequency of GC nucleobases can be expressed relative to a reference sequence (*e.g.*, a reference sequence having about 30%, about 35%, about 40%, about 45%, or about 50% GC content). The higher frequence of GC nucleobases can comprise about *n*% GC nucleobases, wherein *n* is an integer selected from the group consisting of 51-99 and the percentage is expressed as percent nucleobases that are G or C out of the total number of nucleobases (*i.e.*, G, C, adenine (A), and thymine (T) nucleobases).

[0049] In an aspect, the disclosure provides an insulator comprising a polynucleotide sequence configured for regulation of transcription within one or more plant genetic environments, including implementations wherein the insulator is positioned

between distinct gene regulatory elements that are desired to be regulated with little or no regulatory cross-talk. In embodiments, the insulator does not comprise a naturally-occurring plant sequence and can be generated at random or as a result of another non-plant polynucleotide sequence, including but not necessarily limited to those described herein.

[0050] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1439 (Synthetic insulator 1 (sIns1)).

[0051] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1440 (Synthetic insulator 2 (sIns2)).

[0052] In embodiments, the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence selected from the group consisting of: SEQ ID NOs:1-1,438, SEQ ID NOs:1,441-1,446, SEQ ID NOs:1,447-2,361, SEQ ID NOs:2,362-19,190, SEQ ID NOs:19,191-20,082, SEQ ID NOs:20,083-20,523, SEQ ID NOs:20,524-20,725, and any combination thereof.

[0053] In an aspect, the disclosure provides a plant genetic construct comprising an insulator. In another aspect, the plant genetic construct can be an element of a kit and can be provided in a format for facilitation of use of the plant genetic construct in a method of generating a plant cell, plant tissue, or plant comprising the plant genetic construct. In a further aspect, the disclosure provides a plant cell, plant tissue, or plant comprising a plant genetic construct comprising an insulator, wherein the plant cell, plant tissue, or plant is produced by a genetic engineering method. In various embodiments, the insulator is stably integrated into a genome of the plant cell, plant tissue, or plant, and the insulator regulates expression of a transgene of the genome by at least partially insulating the transgene from regulatory cross-talk.

Terminology

[0054] The complete disclosure of all patents, patent applications, and publications, and electronically available material cited herein are incorporated by reference in their entirety. Supplementary materials referenced in publications (such as supplementary tables, supplementary figures, supplementary materials and methods, and/or supplementary experimental data) are likewise incorporated by reference in their entirety. In the event that any inconsistency exists between the disclosure of the present application and the disclosure(s) of any document incorporated herein by reference, the disclosure of the present application shall govern. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The disclosure is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the disclosure defined by the claims.

[0055] The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While the specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure.

[0056] Specific elements of any foregoing embodiments can be combined or substituted for elements in other embodiments. Moreover, the inclusion of specific elements in at least some of these embodiments can be optional, wherein further embodiments can include one or more embodiments that specifically exclude one or more of these specific elements. Furthermore, while advantages associated with certain embodiments of the disclosure have been described in the context of these embodiments, other embodiments can also exhibit such advantages, and not all embodiments need necessarily exhibit such advantages to fall within the scope of the disclosure.

[0057] Unless otherwise specified herein, terms of the disclosure have the same meaning imbued into them by the knowledge of the person having ordinary skill in the art to which the disclosure most closely belongs.

[0058] As used herein, a "candidate sequence" refers to a polynucleotide or DNA sequence that is at least potentially suitable for use as an insulator in one or more types of plants. Candidate sequences can be verified as insulators with screening methods of the disclosure, and if verified, used as insulators in the creation of plant genetic constructs and delivery vectors, transgenic plant cells, transgenic plants and crops, and the like.

[0059] As used herein, an "insulator" in a plant refers to a *cis* regulatory element that enables plant gene or transgene regulation and/or chromosomal organization within a plant genetic environment (*e.g.*, a plant genome, a plant chromosome, or the like), including transgene regulation that results in reduced regulatory cross-talk and/or reduced spurious or uncontrolled expression of the transgene relative to a suitable control.

[0060] As used herein, "percent identity" (*e.g.*, "80% identity") refers to the number of nucleobases that are the same between polynucleotide sequences that are optimally aligned, expressed as a percentage, with the total number of nucleobases as the denominator. For example, the polynucleotide sequence **AT**GGCTGGTT has 80% identity with the polynucleotide sequence **GC**GGCTGGTT (**bolded** nucleobases differ).

[0061] As used herein, a "kit" can comprise an "instructional material" which includes a publication, a recording, a diagram, or any other medium of expression which can be used to communicate the usefulness of one or more elements of the kit for carrying out a testing method, a screening method, a treatment method, or other method of the disclosure, including methods for regulation of expression of a plant transgene as described herein or as known in the art. Optionally, or alternately, the instructional material can describe one or more methods of alleviating diseases or disorders in a cell or a tissue of a plant. The instructional material of the kit can, for example, be affixed to a container which contains an identified compound or can be shipped together with a container which contains the identified compound. Alternatively, the instructional material can be shipped separately from the container with the intention that the instructional material and the compound be used cooperatively.

[0062] As used herein, the term "non-transitory machine-readable storage medium," refers to a computer readable medium or processor readable medium that can store information as data, as well as instructions for performance of logic operations by one or more processors for manipulation of the data. The non-transitory machine-readable storage medium can include any type of memory, such as volatile memory like random access memory (RAM), dynamic random-access memory (DRAM), static random-access memory (SRAM), or non-volatile memory like read-only memory (ROM), flash memory, magnetic or optical disks, or compact-disc read-only memory (CD-ROM), among other devices used to store data or programs on a temporary or permanent basis. The non-transitory machine-readable storage medium can be configured to store instructions. The instructions are executable by the one or more processors to cause the computing device to

perform any of the functions or methods described herein. The non-transitory machine-readable storage medium can also be configured to store a computational model. The computational model can take the form of a convolutional neural network or any other type of artificial neural network. The computational model can take other forms as well.

[0063] As used herein, the term "computational device" refers to a device, *e.g.*, an electronic device, capable of performing logic operations. Example computational devices include computers, laptops, and tablets, as are known in the art. A computing device includes one or more processors, a non-transitory computer readable medium, a communication interface, a display, and a user interface. Components of the computing device are linked together by a system bus, network, or other connection mechanism. The one or more processors can be any type of processor(s), such as a microprocessor, a digital signal processor, a multicore processor, *etc.*, coupled to the non-transitory computer readable medium. The communication interface can include hardware to enable communication within the computational device and/or between the computational device and one or more other devices. The hardware can include transmitters, receivers, and antennas, for example. The communication interface can be configured to facilitate communication with one or more other devices, in accordance with one or more wired or wireless communication protocols. For example, the communication interface can be configured to facilitate wireless data communication for the computational device according to one or more wireless communication standards, such as one or more Institute of Electrical and Electronics Engineers (IEEE) 801.11 standards, ZigBee standards, Bluetooth standards, *etc.* As another example, the communication interface can be configured to facilitate wired data communication with one or more other devices. The communication interface can also include analog-to-digital converters (ADCs) or digital-to-analog converters (DACs) that the computational device can use to control various components. The display can be any type of display component configured to display data. As one example, the display can include a touchscreen display. As another example, the display can include a flat-panel display, such as a liquid-crystal display (LCD) or a light-emitting diode (LED) display. A user interface can be included as part of the computational device and can include one or more pieces of hardware used to provide data and control signals to the computing device. For instance, the user interface can include a mouse or a pointing device, a keyboard or a keypad, a microphone, a touchpad, or a touchscreen, among other possible types of user input devices. Generally, the user interface can enable

an operator to interact with a graphical user interface (GUI) provided by the computing device (*e.g.*, displayed by the display)

[0064] As used herein, the term "system" refers to one or more computational devices, or one or more elements thereof, configured to perform one or more tasks, methods, or processes. A system of the disclosure can comprise a computing device. The computing device includes one or more processors, a non-transitory computer readable medium, a communication interface, a display, and a user interface. Components of the computing device are linked together by a system bus, network, or other connection mechanism.

[0065] As used herein and unless otherwise indicated, the terms "a" and "an" are taken to mean "one," "at least one" or "one or more." Unless otherwise required by context, singular terms used herein shall include pluralities and plural terms shall include the singular.

[0066] Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise," "comprising," and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to".

[0067] Unless the context clearly requires otherwise, the phrase "consisting essentially of" limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristic(s) of the claim.

[0068] Unless the context clearly requires otherwise, the phrase "consisting of" excludes any element, step, or ingredient not specified.

[0069] If an element is described or claimed herein such that it "comprises" a feature, that description or claim also includes embodiments wherein the element "consists essentially of" and embodiments wherein the element "consists of" the feature, unless something else is specifically stated to the contrary.

[0070] Words using the singular or plural number also include the plural and singular number, respectively. Additionally, the words "herein," "above," and "below" and words of similar import, when used in this application, shall refer to this application as a whole and not to any particular portions of the application.

[0071] Unless otherwise indicated, all numbers expressing quantities of components, molecular weights, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless

otherwise indicated to the contrary, the numerical parameters set forth in the specification and claims are approximations that can vary depending upon the desired properties sought to be obtained by the present disclosure. At the very least, and not as an attempt to limit the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques.

[0072] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. All numerical values, however, inherently contain a range necessarily resulting from the standard deviation found in their respective testing measurements.

[0073] All headings are for the convenience of the reader and should not be used to limit the meaning of the text that follows the heading, unless so specified.

[0074] All of the references cited herein are incorporated by reference. Aspects of the disclosure can be modified, if necessary, to employ the systems, functions, and concepts of the above references and application to provide yet further embodiments of the disclosure. These and other changes can be made to the disclosure in light of the detailed description.

[0075] It will be appreciated that, although specific embodiments of the disclosure have been described herein for purposes of illustration, various modifications can be made without deviating from the spirit and scope of the disclosure. Accordingly, the disclosure is not limited except as stated by the claims.

Summary of Tables

Table	Description
1	Insulator fragments used in fragment combination library
9	Insulator fragment combinations tested in stable transgenic maize lines
10	Insulators and insulator fragments used in the enhancer-insulator combination
11	Full-length insulator sequences

[0076] Table 1. Insulator fragments used in fragment combination library.

Insulator	**Start**	**Stop**	**Orientation**	**Insulator activity in tobacco**
Fragments for positions 1 and 2				
β-phaseolin	230	399	fwd and rev	bottom 25% (both orientations)
β-phaseolin	383	552	fwd and rev	bottom 25% (both orientations)
β-phaseolin	1148	1317	fwd and rev	top 25% (rev orientation)
β-phaseolin	1317	1486	fwd and rev	top 25% (rev orientation)
β-phaseolin	1395	1564	fwd and rev	top 25% (both orientations)
β-phaseolin	1633	1802	fwd and rev	top 25% (fwd orientation)
β-phaseolin	1712	1881	fwd and rev	top 25% (both orientations)
β-phaseolin	1791	1960	fwd and rev	top 25% (both orientations)
β-phaseolin	2266	2435	fwd and rev	top 25% (both orientations)

Insulator	**Start**	**Stop**	**Orientation**	**Insulator activity in tobacco**
β-phaseolin	2345	2514	fwd and rev	top 25% (fwd orientation)
β-phaseolin	2741	2910	fwd and rev	top 25% (fwd orientation)
β-phaseolin	3058	3227	fwd and rev	top 25% (both orientations)
β-phaseolin	3454	3623	fwd and rev	bottom 25% (both orientations)
TBS	252	421	fwd and rev	top 25% (both orientations)
TBS	588	757	fwd and rev	top 25% (both orientations)
TBS	756	925	fwd and rev	bottom 25% (both orientations)
TBS	1681	1850	fwd and rev	top 25% (both orientations)
TBS	1765	1934	fwd and rev	top 25% (rev orientation)
λ-EXOB	1	170	fwd and rev	top 25% (both orientations)

Insulator	**Start**	**Stop**	**Orientation**	**Insulator activity in tobacco**
λ-EXOB	83	252	fwd and rev	top 25% (both orientations)
λ-EXOB	166	335	fwd and rev	top 25% (both orientations)
λ-EXOB	249	418	fwd and rev	top 25% (both orientations)
λ-EXOB	332	501	fwd and rev	top 25% (both orientations)
λ-EXOB	415	584	fwd and rev	top 25% (both orientations)
λ-EXOB	663	832	fwd and rev	top 25% (rev orientation)
λ-EXOB	829	998	fwd and rev	top 25% (fwd orientation)
BEAD-1C	246	415	fwd and rev	top 25% (both orientations)
UASrpg	157	326	fwd and rev	top 25% (fwd orientation)
sIns1	54	223	fwd and rev	top 25% (rev orientation)

Insulator	Start	Stop	Orientation	Insulator activity in tobacco
sIns2	335	504	fwd and rev	top 25% (both orientations)
gypsy	1	170	fwd and rev	bottom 25% (both orientations)
gypsy	54	223	fwd and rev	bottom 25% (both orientations)
Fragments for position 3				
β-phaseolin	1633	1802	fwd	top 5%
β-phaseolin	1712	1881	rev	top 5%
λ-EXOB	1	170	fwd	top 5%
λ-EXOB	332	501	fwd	top 5%
λ-EXOB	415	584	rev	top 5%

[0077] Table 2. Insulator fragment combinations tested in stable transgenic maize lines.

Name	Number of fragments	Fragment 3	Fragment 2	Fragment 1	Insulator activity in maize protoplasts
D2	2		β-phaseolin 1148-1317, fwd	sIns2 335-504, fwd	strong

Name	Number of fragments	Fragment 3	Fragment 2	Fragment 1	Insulator activity in maize protoplasts
T30	3	β-phaseolin 1633-1802, fwd	λ-EXOB 663-832, fwd	β-phaseolin 1564-1395, rev	intermediate
T21	3	β-phaseolin 1881-1712, rev	λ-EXOB 663-832, fwd	λ-EXOB 170-1, rev	weak
T27	3	λ-EXOB 1-170, fwd	λ-EXOB 584-415, rev	TBS 925-756, rev	strong
T32	3	λ-EXOB 584-415, rev	λ-EXOB 832-663, rev	β-phaseolin 1960-1791, rev	strong
T24	3	λ-EXOB 584-415, rev	β-phaseolin 3227-3058, rev	λ-EXOB 170-1, rev	strong
T25	3	λ-EXOB 1-170, fwd	β-phaseolin 1802-1633, rev	β-phaseolin 1317-1148, rev	strong
T19	3	λ-EXOB 1-170, fwd	β-phaseolin 1148-1317, fwd	sIns2 335-504, fwd	strong
T9	3	λ-EXOB 584-415, rev	TBS 1765-1934, fwd	sIns2 335-504, fwd	strong

[0078] Table 3. Insulators and insulator fragments used in the enhancer-insulator combination library.

Type	Insulator	Start	Stop	Orientation
full-length insulator	λ-EXOB	1	998	fwd
full-length insulator	BEAD-1C	1	538	fwd
full-length insulator	UASrpg	1	378	fwd
full-length insulator	sIns1	1	386	fwd
full-length insulator	sIns2	1	504	fwd
full-length insulator	gypsy	1	386	fwd
insulator fragment	β-phaseolin	1395	1564	fwd
insulator fragment	TBS	756	925	fwd
insulator fragment	λ-EXOB	1	170	fwd
insulator fragment	BEAD-1C	246	415	fwd
insulator fragment	UASrpg	1	170	fwd

Type	Insulator	Start	Stop	Orientation
insulator fragment	gypsy	54	223	fwd

Examples

Example 1: High-throughput identification of functional plant insulators and characterization of synthetic insulators.

A high-throughput method to identify plant enhancer-blocking insulators

[0079] To detect the activity of enhancer-blocking insulators, they can be cloned between an enhancer and a minimal promoter. Since the insulator will prevent the enhancer from activating the minimal promoter in such a configuration, the expression level of the corresponding reporter gene will be lower than that of a control construct without an insulator. However, a silencer could lead to a similar decrease in expression (FIG. 1C, left). Therefore, to distinguish between true insulators and silencers, a second construct can be implemented. Here, the insulator candidates are cloned upstream of both the enhancer and minimal promoter. In this context, only a silencer can cause decreased expression, while an insulator will not affect the expression level (FIG. 1C, right).

[0080] Based on these considerations, a method for the high-throughput identification of enhancer-blocking insulators active in plants is disclosed. The method is configured to 1) work in plants and 2) enable the detection of insulator and silencer activity.

[0081] Insulator candidates are cloned between the 35S enhancer and 35S minimal promoter, or upstream of both elements to measure insulator or silencer activity, respectively (FIG. 1D). The candidate sequences are linked to a barcode located in a reporter gene via next generation sequencing. The pooled library of reporter constructs is then used for transient transformation of tobacco leaves or maize protoplasts. After 1-2 days incubation, the transformed plant tissue is harvested, and the reporter mRNA is extracted. Finally, the abundance of the barcode sequences in the input DNA and the extracted mRNA is determined by next generation sequencing. For each insulator candidate, the enrichment of all linked barcodes in the RNA relative to the DNA is calculated. The difference between the enrichment of an insulator candidate and a control construct without an insulator is a measure of the insulator or silencer activity of the candidate sequence (FIG. 1D).

Characterization of synthetic insulators

[0082] Using the method described above, two novel, synthetic insulators (named sIns1 and sIns2) are disclosed and characterized. Both insulators substantially decrease the effect of the very strong 35S enhancer on downstream genes in transiently transformed tobacco and maize cells, as well as in stable *Arabidopsis* lines. Smaller (170bp) fragments of the synthetic insulators possess similar activity. With the disclosed aspects, characterization of the synthetic insulators in stable rice lines can be performed.

Origin and sequence of the synthetic insulators

[0083] The synthetic insulator sIns1 is derived from plasmid pZS*11-yfp0 (Addgene #53241) and covers part of the β-lactamase (AmpR) coding sequence, the AmpR promoter, and the neighboring plasmid backbone sequence. The synthetic insulator sIns2 is derived from plasmid pEvolvR-enCas9-PolI3M-TBD (Addgene #113077) and covers part of the enCas9 coding sequence.

[0084] In various aspects, high-throughput insulator screening methods of the disclosure can be used for identification of plant insulators. Using the disclosed methods, plant DNA sequences can be identified as having insulator activity.

[0085] For example, one or more of SEQ ID NOs:1-1440, or one or more portions thereof, can have insulator activity. One or more of SEQ ID NOs:1-1440, or one or more portions thereof, can be used as part of a genetic engineering method for reducing regulatory cross-talk and improving expression of transgenes in plants. In particular embodiments, sIns1 (SEQ ID NO:1439), sIns2 (SEQ ID NO:1440), or both, can have insulator activity and/or can be used as part of a genetic engineering method for reducing regulatory cross-talk and improving expression of transgenes in plants.

[0086] sIns1:

AGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC
TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA
TCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT
TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAA
AATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCCAACAT
ATGGCACCGGAGGCTTTCGTCTTCAC (SEQ ID NO:1439)

[0087] sIns2:

GCTGAACGCAAAGCTGATCACTCAGCGGAAGTTCGACAATCTCACTAAGGCTGAGAGGGG

CGGACTGAGCGAACTGGACAAAGCAGGATTCATTAAACGGCAACTTGTGGAGACTCGGCA
GATTACTAAACATGTCGCCCAAATCCTTGACTCACGCATGAATACCAAGTACGACGAAAA
CGACAAACTTATCCGCGAGGTGAAGGTGATTACCCTGAAGTCCAAGCTGGTCAGCGATTT
CAGAAAGGACTTTCAATTCTACAAAGTGCGGGAGATCAATAACTATCATCATGCTCATGA
CGCATATCTGAATGCCGTGGTGGGAACCGCCCTGATCAAGAAGTACCCAGCACTGGAAAG
CGAGTTCGTGTACGGAGACTACAAGGTCTACGACGTGCGCAAGATGATTGCCAAATCTGA
GCAGGAGATCGGAAAGGCCACCGCAAAGTACTTCTTCTACAGCAACATCATGAATTTCTT
CAAGACCGAAATCACCCTTGCAAA (SEQ ID NO:1440)

Example 2: Small DNA elements can act as both insulators and silencers in plants (for additional information, see: Jores, T. *et al.* Small DNA elements can act as both insulators and silencers in plants. bioRxiv pre-print, September 19, 2024. doi: https://doi.org/10.1101/2024.09.13.612883., the contents of which are incorporated by reference herein in their entirety for all purposes.)

Abstract

[0088] Plant insulators can act as silencers. Insulators include cis-regulatory elements that separate transcriptional units, whereas silencers include elements that repress transcription regardless of their position. In plants, these elements remain largely uncharacterized. In this example, the massively parallel reporter assay Plant self-transcribing active regulatory region sequencing (STARR-seq) is used with short fragments of eight large insulators to newly characterize more than 100 fragments that unexpectedly block enhancer activity. The short fragments can be combined to generate more powerful insulators that abolish the capacity of the strong viral 35S enhancer to activate the 35S minimal promoter. Unexpectedly, when tested upstream of weak enhancers, these fragments function as silencers and repress transcription. Thus, these elements are capable of both insulating as well as repressing transcription, dependent upon regulatory context. The findings are validated in stable transgenic *Arabidopsis*, maize, and rice plants. The short elements identified herein are useful building blocks for plant biotechnology efforts.

Introduction

[0089] Precise control of gene expression is crucial for plants to grow and develop in a changing environment, however, genomic approaches to study plant gene regulation have focused mainly on promoters and enhancers. In contrast, repressive elements such as silencers and insulators have received far less attention.

[0090] Insulators compartmentalize genomes into discrete transcriptional units. Insulators have one or both of two principal functions: blocking enhancers from interacting with core promoters (enhancer-blocking insulators) and forming barriers against the spread of repressive heterochromatin (barrier insulators). Enhancer-blocking insulators are defined by their ability to act when situated between an enhancer and promoter, but not when the order is reversed such that the enhancer is closer to the promoter than the insulator. Insulators can be thought to prevent ectopic gene expression, maintain chromatin accessibility, and enable differentially regulated genes to reside in close proximity to one another.

[0091] To date, most research on insulators has been performed in animal models. In contrast, only a handful of plant sequences have been shown to act as insulators in transient or stable transgenic plant reporter assays. For example, the Transformation Booster Sequence (TBS) from *Petunia hybrida*, the β-phaseolin gene from *Phaseolis vulgaris*, and a gypsy-like sequence from *Arabidopsis thaliana* function as enhancer blocking insulators in transgenic plants. In addition, a few heterologous sequences show enhancer blocking insulator activity in plants, including λ-EXOB from phage λ, BEAD-1C from humans, and UASrpg from yeast. In prior studies, insulator activity was inferred from β-glucuronidase (GUS) staining or fluorescence of a reporter gene, both measures with limitations of dynamic range, quantification accuracy, and throughput. Because the enhancer-blocking activity of insulators is detected as reduced transcription in prior commonly used reporter assays, care must be taken to distinguish between insulators and silencers, which could also cause reduced transcription in these assays. Silencers recruit repressive transcription factors and, like enhancers, can act in a position-independent manner (*i.e.*, upstream or downstream of an enhancer). This position-independency is thought to be a key difference between silencers and insulators that differentiates between the two element types.

[0092] To date, no general principles are known that typify insulator or silencer function in plants, nor are there high-throughput methods to identify these elements. Short and strong insulators will facilitate synthetic biology applications to ensure predictable expression of transgenes, blocking inappropriate enhancer-promoter interactions and alleviating chromatin position effects. Similarly, silencers can enable fine-tuning of transgene expression and minimize expression noise. Furthermore, understanding the

sequence features of functional plant insulators and silencers can enable targeting similar elements in plant genomes to engineer gene expression.

[0093] In this example, Plant STARR-seq, a massively parallel reporter assay, was applied to test the insulator and silencer activity of over 100 short (170 bp) fragments derived from either previously described enhancer-blocking insulators or two novel synthetic insulator sequences (*i.e.*, sIns1, sIns2). The disclosed assay distinguishes enhancer-blocking activity from transcriptional repression and reveals that the insulator-derived elements harbor both insulator-like and silencer-like activities. Promising elements were tested and verified in stable transgenic *Arabidopsis*, rice, and maize plants.

Results

Plant STARR-seq detects the activity of enhancer-blocking insulators

[0094] Plant STARR-seq can identify and characterize cis-regulatory elements. To test whether Plant STARR-seq can identify enhancer-blocking insulators, a reporter construct that includes a barcoded green fluorescent protein (GFP) gene under the control of a 35S minimal promoter coupled to a 35S enhancer was created; insulator candidates were placed between this enhancer and promoter (FIG. 7A). Insulator sequences included four heterologous sequences that show insulator activity in plants (λ-EXOB (SEQ ID NO:1443), BEAD-1C (SEQ ID NO:1444), UASrpg (SEQ ID NO:1445), and a *Drosophila* gypsy element (SEQ ID NO:1446)), and two synthetic sequences (sIns1 (SEQ ID NO:1439) and sIns2 (SEQ ID NO:1440)) for which preliminary data suggested they can act as insulators. The synthetic sequences sIns1 and sIns2 were derived from a plasmid backbone and a human codon-optimized coding sequence of Cas9, respectively, and engineered for insulator activity. The insulator candidate sequences were cloned in the forward or reverse orientation, and their insulator activity was determined by Plant STARR-seq in tobacco (*Nicotiana benthamiana*) leaves and maize (*Zea mays*) protoplasts. Constructs without the 35S enhancer and without an insulator (noEnh) or with the 35S enhancer and without an insulator (noIns) were included as controls (FIG. 7A).

[0095] Insulator activity was measured as reduced enrichment compared to the enrichment of the no insulator (noIns) control. Except for the gypsy element, the other five tested insulator candidates resulted in reduced enrichment, indicating that they function as enhancer-blocking insulators in this assay (FIG. 7B). The gypsy element shows enhancer-blocking and barrier insulator activities in *Drosophila*; however, it lacks enhancer-blocking activity in plants, consistent with these results. For some of the insulators, orientation-

dependent activity was observed (FIG. 7B). Taken together, the results herein demonstrate that Plant STARR-seq reproducibly (FIGs 8A-8F) measures insulator activity.

[0096] The large size of prior known enhancer-blocking insulators precludes their application in plant biotechnology. To identify and develop short sequences with insulator activity, overlapping 170-bp fragments of each of the six insulators in addition to two plant sequences with insulator activity (β-phaseolin and TBS) were array-synthesized, and the enhancer-blocking activity of these fragments measured (FIG. 2A). Many fragments retained partial insulator activity in tobacco and maize (FIG. 2B and SEQ ID NOs:1,447-2,361), but their activity varied between the two assay systems, possibly pointing to species-specific differences (FIG. 2C). Overall, seven of the eight insulators, excluding only the gypsy element, harbored clusters of fragments that partially blocked the 35S enhancer (FIG. 2D). Without wishing to be bound by any particular theory, this clustering of active fragments may be driven by local nucleotide composition, because GC content strongly correlated with a fragment's insulator activity (FIG. 2E). However, GC content does not fully explain insulator activity: Many insulator-derived fragments showed orientation-dependent activity (FIG. 2D). Furthermore, the insulator-derived fragments were tested with the AB80 enhancer from *Pisum sativum* and the Cab-1 enhancer from *Triticum aestivum*, which drive the expression of chlorophyl a-b binding proteins, and it was found that the activity of these fragments was largely enhancer independent (FIGs 2B and 2F).

[0097] To validate the findings, insulator activity was measured in stable transgenic plants (FIGs 3A-3K). Full-length insulators and fragments thereof showed enhancer-blocking insulator activity in *Arabidopsis*, rice (*Oryza sativa*), and maize, well correlated with the Plant STARR-seq results (FIG. 3 and FIG. 9). In maize, insulator activity was measured in four tissues (leaf, stalk, silk, and husk) and in two developmental stages (V6 and R1), and similar results were obtained, indicating that these insulators do not act in a tissue-specific manner (FIG. 3K).

Active fragments can be assembled into strong insulators

[0098] It was asked whether insulator activity can be increased by combining up to three fragments. Twenty-six (26) fragments with high insulator activity (top 25% of all fragments) in tobacco and six (6) fragments with low insulator activity (bottom 25% of all fragments) in tobacco (Table 1) were evaluated. These fragments were used in the forward and reverse orientation to build constructs with both the individual fragments and with the

over 2,900 randomly generated two-fragment combinations. Additionally, over 13,000 three-fragment combinations were built that added one of five fragments with very high insulator activity (top 5% of all fragments; Table 1) upstream of the randomly generated two-fragment combinations. Fragments (SEQ ID NOs:1,447-2,361) and fragment combinations (SEQ ID NOs:2,362-19,190) were cloned between the 35S enhancer and 35S minimal promoter (FIG. 4A). As a trend, increasing the number of insulator fragments increased insulator activity. In tobacco, most constructs with three insulator fragments completely blocked the 35S enhancer (FIG. 4B and SEQ ID NOs:1-1446). In addition, combinations of fragments derived from different full-length insulators showed a similar activity distribution to combinations of fragments derived from the same full-length insulator. Similarly, the activity distribution of combinations with two copies of the same fragment was largely indistinguishable from that of combinations with two non-identical fragments.

[0099] A linear model was trained based on the insulator activity of the individual fragments and their position in the construct to predict the insulator activity of two-fragment and three-fragment combinations in tobacco and maize (FIG. 4C). Model accuracy was similar for the two-fragment and three-fragment combinations in tobacco (R^2 of 0.67 and 0.62, respectively). In maize, prediction accuracy was higher for the two-fragment combinations than for the three-fragment combinations (R^2 of 0.60 and 0.48, respectively). The model coefficients showed that the fragment closest to the minimal promoter contributes the most to the combined insulator activity, while the fragment closest to the enhancer contributes the least (FIG. 4D). Taken together, the insulator activity of the individual fragments appears to be a determinant for the activity of the fragment combinations.

[0100] Next, the activity of one two-fragment combination and nine three-fragment combinations was tested (Table 2) in stable maize plants. Most of these fragment combinations showed insulator activity in the transgenic maize plants (FIG. 4E and FIG. 10A). However, their activity was weaker than observed in the Plant STARR-seq experiments; without wishing to be bound by any particular theory, this may have been because a moderate-strength promoter from maize was used for the transgenic maize reporter constructs instead of the minimal 35S promoter used in Plant STARR-seq (FIG. 4F and FIG. 10B). To further increase insulator activity, the two-fragment combination D2 was cloned downstream of the three-fragment combinations T9, T32, and T27 (Table 2) to

yield three constructs of five fragments (T9+D2, T32+D2, and T27+D2). These five-fragment combinations showed similar insulator activity as the corresponding two- or three-fragment combinations (FIG. 4E and FIG. 10A), indicating diminishing returns from stacking increasing numbers of fragments. Because most insulator combinations reached the detection limit in the Plant STARR-seq assay but not in the stable maize plants, the correlation between the ELISA and Plant STARR seq data was low (FIGs 3F and 3G, and FIG. 10B). However, a strong correlation was observed between ELISA results for samples obtained from different plant tissues (leaf, stalk, and root) and developmental stages (V6 and R1). This observation is consistent with the results for single insulator fragments and indicates that insulator activity is not strongly affected by tissue identity or developmental stage.

Insulator-derived fragments also exhibit silencer activity

[0101] Without wishing to be bound by any particular theory, the comparison of the Plant STARR-seq and stable maize data can suggest that insulator activity may be promoter-dependent. To investigate this hypothesis, constructs with hybrid promoters were built by inserting the AB80 or Cab-1 enhancer between the 35S minimal promoter and the insulator fragments and it was tested whether an additional downstream enhancer affected the ability of the insulator-derived fragments to block an upstream 35S enhancer (FIG. 5A, top). Many fragments showed insulator activity with both downstream enhancers (FIG. 5B, left and SEQ ID NOs:19,191-20,082) and this activity was only slightly weaker than in constructs without a downstream enhancer (FIGs 12A-12D). Without wishing to be bound by any particular theory, this finding can suggest that the insulator-derived fragments remain active at a greater distance and work with more complex promoters than the short 35S minimal promoter.

[0102] A set of control constructs without the upstream 35S enhancer (FIG. 5A, bottom) was also tested and it was found that many insulator fragments resulted in lower enrichment than a control construct without an insulator fragment (FIG. 5B, right and SEQ ID NOs:19,191-20,082), indicating transcriptional repression. The enrichment of reporter constructs with and without the upstream 35S enhancer was well correlated (FIG. 5C). Without wishing to be bound by any particular theory, these results demonstrate that fragments derived from characterized insulators with enhancer blocking activity in Plant-STARR-seq can also function as transcriptional silencers. To rigorously assess whether the insulator-derived fragments had silencer activity, a new library with two different construct

layouts was built: (i) in the 'insulator' construct, the fragments were inserted between the 35S enhancer and 35S minimal promoter; and (ii) in the 'silencer' construct, the fragments were inserted upstream of the 35S enhancer (FIG. 5D). As with other results in this disclosure, many fragments led to a reduced enrichment of the reporter gene when inserted between the enhancer and promoter (*i.e.*, in the insulator construct). The insulator-derived fragments showed little to no activity in the silencer construct in tobacco; however, some silencer activity was observed in maize (FIG. 5E and SEQ ID NOs:20,083-20,523).

[0103] It was reasoned that the activity of fragments in the insulator construct might be a combination of enhancer-blocking activity and silencer activity. To quantify what fraction of the apparent insulator activity could be explained by transcriptional repression rather than insulation, the activities of all fragments in the insulator construct were plotted against their activities in the silencer construct (FIG. 5F). The slope of the regression line in these plots is a proxy for the maximal contribution of transcriptional repression to the apparent insulator activity. Up to 6% and 43% of the observed activity in the insulator construct could be explained by silencer activity in tobacco and maize, respectively (FIG. 5F).

Silencer activity depends on enhancer strength

[0104] Because evidence of silencer activity was found in tobacco leaves in constructs containing the AB80 or Cab-1 enhancer (FIGs 5B and 5C), but not in those with the strong 35S enhancer (FIGs 5E and 5F), insulator and silencer constructs with eight different enhancers were built (FIGs 6A and 6B). These enhancers showed a wide range of strength in tobacco but were all, apart from the 35S enhancer, weak in maize (FIG. 6B).

[0105] These enhancers were tested with six full-length insulators and six insulator-derived fragments (Table 3). Insulators and insulator fragments showed little activity as silencers with strong enhancers (like the 35S, At-9661, and Sl-12881 enhancers in tobacco and the 35S enhancer in maize) but much more activity as silencers with weak enhancers (FIG. 6C and SEQ ID NOs:20,524-20,725). This result conclusively demonstrates that these previously identified insulators and their fragments can function as enhancer-blocking insulators or as silencers depending on regulatory context.

[0106] As before, the enrichment of fragments in insulator constructs was plotted against their enrichment in silencer constructs. The slope of a linear regression line was used as a proxy to determine how much of the apparent insulator activity could be explained by silencer activity. For constructs with strong enhancers, between 6% and 27% of the

apparent insulator activity could be explained by silencer activity. This proportion increased with weak enhancers, such that silencer activity could explain up to 94% of the observed activity in the insulator construct. Overall, the slopes negatively correlated with the strength of the corresponding enhancer (FIG. 6D). To test whether the insulators showed silencer activity when integrated into the genome, dual-luciferase reporter constructs were used with the insulator residing upstream of the 35S or AB80 enhancer to generate stable transgenic *Arabidopsis* plants (FIG. 12A). As in the transient Plant STARR-seq experiments, the insulators showed no silencer activity with the strong 35S enhancer and partial silencer activity with the somewhat weaker AB80 enhancer in transgenic *Arabidopsis* plants (FIGs 12B-12D). Taken together, these results are consistent with the observation that previously identified insulators show silencer activity that is inversely correlated with the strength of the enhancer with which they are paired.

Discussion

[0107] Using the high throughput Plant STARR-seq assay on fragments of insulators known to be functional in plants, more than 100 170-bp fragments with enhancer-blocking activity were identified and characterized. These short fragments can be combined to generate stronger insulators, at least some of which are capable of completely blocking the activity of the viral 35S enhancer. The fragments were active as insulators with different enhancers and promoters and across diverse plant tissues. Surprisingly, these insulators and their fragments showed silencer activity when coupled with weak enhancers.

[0108] This unexpected finding showcases the complexity of regulatory grammar, wherein cis-regulatory elements can have multiple activities that may be observed only in specific conditions or contexts. For example, mesoderm-specific *Drosophila* silencers often function as enhancers in other cell types. Thus, regulatory elements must be tested systematically in different contexts – *e.g.*, as insulators, silencers, or enhancers, and across species and tissues – to understand the mechanistic underpinnings of their potentially complex functions. The elements developed herein behaved like classical enhancer-blocking insulators in combination with strong enhancers, as they reduced reporter expression only when inserted between the enhancer and promoter. In contrast, with weak enhancers, the same elements behaved like typical silencers that repressed transcription in a position independent manner. With intermediate strength enhancers, a continuum between these two extremes was observed: The elements reduced transcription when placed either upstream or downstream of the enhancer, but the effect was stronger in the

downstream context. The observed activity can be a combination of distinct insulator and silencer functions or another novel, unknown regulatory mechanism. Interestingly, independent of enhancer strength, the activity of fragments in insulator and silencer constructs is well correlated; without wishing to be bound by any particular theory, this may suggest that the latter might be the case.

[0109] To date, the molecular mechanisms underlying plant insulator function are unknown. In animals, several DNA-binding proteins, including su(Hw), BEAF-32, and Zw5 in *Drosophila* and CTCF in humans, play a role in insulator function. However, homologs of these proteins have not been identified in plants. The number of fragments with insulator activity tested here may be too small to derive putative protein-binding motifs with confidence. Moreover, there is no evidence that insulation in plants requires protein binding.

[0110] In contrast to enhancer activity, it was found herein that insulator activity was orientation-dependent. In some prior instances, orientation-dependence was a consequence of composite elements with both insulator and enhancer activities. An alternative explanation for orientation dependence is that structural properties of the insulator DNA contribute to insulator function, a hypothesis that is also consistent with the finding that GC content is a contributor to insulator activity.

[0111] Short insulator elements, as disclosed herein, are useful for plant biotechnology to minimize the size of transgene cassettes to ensure efficient transformation. Prior efforts to produce transgene cassettes, including those composed of multiple genes, had unsatisfactory and unpredictable expression patterns, even when regulation of the individual genes was well-characterized. The insulators identified herein are proposed to be building blocks to make expression more predictable and thus plant engineering more economically feasible.

[0112] The insulator activity as disclosed herein exhibited some specificity to the tobacco or maize system, which can suggest that insulators can be designed for either dicots or monocots, as well as both dicots and monocots. Although the present disclosure shows that the use of insulators in transgene cassettes can account for both their silencer and insulator activities, existing plant biotechnology efforts tend to use strong constitutive promoters, such that silencer activity is negligible. Moreover, when used with tissue- or condition-specific enhancers, insulators with enhancer-dependent silencer activity can be beneficial. Such insulator-enhancer combinations as enabled by the present disclosure can

repress leaky expression in tissues or conditions in which the enhancer is inactive and insulate expression when the enhancer becomes fully active. Similarly, the dual-function elements disclosed herein can be used to fine-tune transgene expression by repressing overly-active transcription while simultaneously isolating the transgene from other surrounding regulatory elements.

Materials and methods

Library design and construction

[0113] The full-length λ-EXOB, BEAD-1C, UASrpg, and gypsy insulators were ordered as synthesized DNA fragments. The synthetic insulators sIns1 (SEQ ID NO:1439) and sIns2 (SEQ ID NO:1440) were PCR amplified from pZS*11_4enh (www.addgene.org/149423/) and pEvolvR-enCas9-PolI3M-TBD (Addgene no. 113077; https://www.addgene.org/113077/), respectively. Insulator fragments were ordered as an oligonucleotide array from Twist Bioscience® with 15-bp flanking sequences for amplification. The 35S, AB80, and Cab-1 enhancers were PCR amplified from pZS*11_4enh. The At-9661, Sl-12881, Sb-11289, Zm-23177, and Sl-774 enhancers were ordered as synthesized DNA fragments. The sequences of the full-length insulators and the oligonucleotides used in this example are listed in SEQ ID NOs:1,439-1,442 and SEQ ID NOs:20,726-20,871, respectively. All libraries used in this example were constructed using pPSm, a shortened version of pPSup (Addgene no. 149416; https://www.addgene.org/149416/) lacking the BlpR cassette, as the base plasmid. The plasmid's T-DNA region harbors a GFP reporter construct terminated by the poly(A) site of the *Arabidopsis* (*Arabidopsis thaliana*) ribulose bisphosphate carboxylase small chain 1A gene. Two versions of pPSm were created to receive insulators in the forward (pPSmF) or reverse (pPSmR) orientation by changing the BsaI scars to ACTC and CTGT or ACAG and GAGT, respectively. The plasmids were deposited at Addgene® (Addgene no. 226912 and 226913; https://www.addgene.org/226912/; https://www.addgene.org/226913/). Gibson assembly was used to insert enhancers into pPSm plasmids. The 35S minimal promoter followed by the 5′ UTR from a maize histone H3 gene (Zm00001d041672), an ATG start codon and a 18-bp random barcode (VNNVNNVNNVNNVNNVNN; wherein V = A, C, or G) was cloned in front of the second codon of GFP by Golden Gate cloning using BbsI-HF (New England Biolabs®; NEB®). To distinguish between sub-libraries, positions 1, 4, 7, 10, 13, and 16 of the barcodes were set to fixed bases. Insulators and insulator fragments were inserted into the pPSm plasmids by Golden Gate cloning using

BsaI-HFv2 (NEB®). The resulting libraries were bottlenecked to yield about 20–50 barcodes per enhancer. The base plasmid for dual-luciferase constructs was derived from pDL (Addgene no. 208978; https://www.addgene.org/208978/;) by changing the BsaI scars to ACTC and CTGT. The 35S or AB80 enhancer was inserted into this plasmid upstream or downstream of the BsaI Golden Gate cassette via Gibson assembly. Full-length insulators and insulator fragments were inserted by Golden Gate cloning using BsaI-HFv2 (NEB®). For rice dual-luciferase constructs, the BlpR cassette was replaced by a hygromycin resistance gene under control of the switchgrass polyubiquitin 2 promoter and the 35S terminator derived from plasmid JD633 (Addgene no. 160393; https://www.addgene.org/160393/). The expression cassettes for the *Agrobacterium*-based transformation vectors to generate transgenic corn plants included a reporter gene driven by a moderate strength constitutive promoter coupled to a heterologous intron with either the CaMV 35S enhancer upstream of the promoter (negative control) or no enhancer (positive control). The same terminator was used in cassettes to terminate transcription. The insulators were tested using the expression cassette with the 35S enhancer by inserting them between the 35S enhancer and the promoter.

Tobacco cultivation and transformation

[0114] Tobacco (*Nicotiana benthamiana*) was grown in soil (Sunshine Mix no. 4) at 25°C in a long-day photoperiod (16 h light and 8 h dark; cool-white fluorescent lights [Philips TL-D 58 W/840]; intensity 300 µmol m^{-2} s^{-1}). Plants were transformed approximately 3 weeks after germination. For transient transformation of tobacco leaves, Plant STARR-seq libraries were introduced into *Agrobacterium tumefaciens* strain GV3101 (harboring the virulence plasmid pMP90 and the helper plasmid pMisoG) by electroporation. An overnight culture of the transformed *A. tumefaciens* was diluted into 100 ml YEP medium (1% [w/v] yeast extract and 2% [w/v] peptone) and grown at 28°C for 8 h. A 5-ml input sample of the cells was collected, and plasmids were isolated from it using the QIAprep® Spin Miniprep® Kit (QIAGEN®) according to the manufacturer's instructions. The remaining cells were harvested and resuspended in 100 ml induction medium (M9 medium [3 g/L KH_2PO_4, 0.5 g/L NaCl, 6.8 g/L Na_2HPO_4, and 1 g/L NH_4Cl] supplemented with 1% [w/v] glucose, 10 mM MES, pH 5.2, 100 µM $CaCl_2$, 2 mM $MgSO_4$, and 100 µM acetosyringone). After overnight growth, the *Agrobacteria* were harvested, resuspended in infiltration solution (10 mM MES, pH 5.2, 10 mM $MgCl_2$, 150 µM acetosyringone, and 5 µM lipoic acid) to an optical density (OD) of 1 and infiltrated into

leaves 3 and 4 of two (full-length insulator library) or four (all other libraries) tobacco plants. The plants were further grown for 48 h under normal conditions (16 h light and 8 h dark) or in the dark before mRNA extraction.

Maize cultivation and transformation

[0115] For Plant STARR-seq in maize (Zea mays L. cultivar B73), a PEG transformation method was used. Maize seeds were germinated in soil at 25°C in a long-day photoperiod (16 h light and 8 h dark; cool-white fluorescent lights [Philips® TL-D 58 W/840]; intensity 300 µmol m^{-2} s^{-1}). After 3 days, the seedlings were moved to complete darkness at 25°C and grown for 10–11 days. From each seedling, 10 cm sections from the second and third leaf were cut into thin 0.5 mm strips perpendicular to veins and immediately submerged in 10 ml of protoplasting enzyme solution (0.6 M mannitol, 10 mM MES pH 5.7, 15 mg/ml cellulase R10, 3 mg/ml macerozyme, 1 mM $CaCl_2$, 0.1% [w/v] BSA, and 5 mM beta-mercaptoethanol). The mixture was covered in foil to keep out light, vacuum infiltrated for 3 min, and incubated on a shaker at 40 rpm for 2.5 hours. Protoplasts were released by incubating an extra 10 min at 80 rpm. To quench the reaction, 10 mL ice-cold MMG (0.6 M Mannitol, 4 mM MES pH 5.7, 15 mM $MgCl_2$) was added to the enzyme solution and the whole solution was filtered through a 40 µM cell strainer. To pellet protoplasts, the filtrate was split into equal volumes of no more than 10 mL in chilled round-bottom glass centrifuge vials and centrifuged at 100 x g for 4 min at room temperature (RT). Pellets were resuspended in 1 mL cold MMG each and combined into a single round-bottom vial. To wash, MMG was added to make a total volume of 5 mL and the solution was centrifuged at 100 x g for 3 min at RT. This wash step was repeated two more times. The final pellet was resuspended in 1–2 mL of MMG. A sample of the resuspended protoplasts was diluted 1:20 in MMG and used to count the number of viable cells using Fluorescein Diacetate as a dye. For each replicate, one to ten million protoplasts were mixed with 15–150 µg of the Plant STARR-seq plasmid library in a fresh tube, topped with MMG to a volume of 114.4 µL per million protoplasts, and incubated on ice for 30 min. For PEG transformation, 105.6 µL per million protoplasts of PEG solution (0.6 M Mannitol, 0.1 M $CaCl_2$, 25% [w/v] poly-ethylene glycol MW 4000) was added to reach a final concentration of 12% (w/v) PEG. The mixture was incubated for 10 min in the dark at RT. After incubation, the transformation solution was diluted with five volumes incubation solution (0.6 M Mannitol, 4 mM MES pH 5.7, 4 mM KCl), and centrifuged at 100 x g for 4 min at RT. The protoplast pellet was washed with 5 mL of incubation solution,

centrifuged at 100 x g for 3 min at RT, and resuspended in incubation solution to a concentration of 500 cells/μL. Protoplasts were incubated overnight in the dark at RT to allow for transcription of the plasmid library and then pelleted (4 min, 100 x g, RT). The pellet was washed with 1–5 mL incubation solution and centrifuged (3 min, 100 x g, RT). The pellet was finally resuspended in 1–5 mL incubation solution. An aliquot of the solution was used to check transformation efficiency under a microscope. Cells were pelleted (4 min, 100 x g, RT) and resuspended in 1–2 mL Trizol for subsequent mRNA extraction. An aliquot of the plasmid library used for PEG transformation was used as the input sample for Plant STARR-seq. Stable transgenic maize plants were then generated using an existing method.

***Arabidopsis* cultivation and transformation**

[0116] *Arabidopsis thaliana* Col-0 was grown in soil (Sunshine Mix no. 4) at 20°C in a long-day photoperiod (16 h light and 8 h dark; cool-white fluorescent lights [Sylvania® FO32/841/ECO 32W]; intensity 100 μmol m^{-2} s^{-1}). For transformation, dual-luciferase plasmids were introduced into *Agrobacterium tumefaciens* strain GV3101 (harboring the virulence plasmid pMP90 and the helper plasmid pMisoG) by electroporation. Transgenic *Arabidopsis* plants were generated by floral dipping and selected for by spraying with a 0.01% (w/v) Glufosinate solution.

Rice cultivation and transformation

[0117] The rice (*Oryza sativa L. ssp. japonica*) cultivar Kitaake was used for genetic transformation following an existing protocol with slight modifications. The mature seeds were sterilized with a 7.5% (w/v) sodium hypochlorite solution for 20 minutes, followed by three sterile water rinses. The seeds were placed on callus induction medium (4.4 g/L MS salts with vitamins, 30 g/L sucrose, 2 mg/L 2,4-dichlorophenoxyacetic acid, 8 g/L agar, pH 5.8) to induce callus cells from scutellum for 10 days. The calli were co-cultivated on callus induction medium supplemented with 200 μM of acetosyringone for 3 days with the *Agrobacterium* strain EHA101 (OD = 0.5) carrying individual insulator constructs. The callus cells were transferred to callus induction medium supplemented with 300 mg/L timentin and 50 mg/L hygromycin for two rounds of selection. The hygromycin resistant callus cells of individual lines were transferred to regeneration medium (4.4 g/L MS salts with vitamins, 30 g/L sucrose, 3 mg/L 6-benzylaminopurine, 0.5 mg/L 1-naphthaleneacetic acid, 8 g/L agar, 25 mg/L hygromycin, 150 mg/L timentin, pH 5.8) for about two rounds to regenerate shoots. The shoots were transferred to rooting medium (4.4

g/L MS salts with vitamins, 30 g/L sucrose, 25 mg/L hygromycin, 8 g/L agar, pH 5.8) and were grown till healthy roots were produced before transferring to soil. The plantlets were transferred to a plastic box containing topsoil from the research farm at the University of Missouri flooded with water. The plantlets were grown in a greenhouse with a short-day photoperiod (12 h light and 12 h dark) at 28°C and 24°C during the day and night, respectively.

Plant STARR-seq

[0118] For all tobacco Plant STARR-seq experiments, two independent biological replicates were performed. Different plants and fresh *Agrobacterium* cultures were used for each biological replicate. Tobacco leaves were harvested 2 days after infiltration and partitioned into batches of 4 leaves. The leaf batches were frozen in liquid nitrogen, finely ground with mortar and pestle, and immediately resuspended in 10 mL QIAzol (Qiagen®). The suspensions were cleared by centrifugation (5 min, 4,000 x g, 4°C). The supernatant was transferred to a 15 mL MaXtract High Density tube (Qiagen®) and mixed with 2.5 mL chloroform. After centrifugation (10 min, 1,000 x g, 4°C), the supernatant (approximately 7 mL) was poured into a new tube, and mixed by inversion with 3.5 mL high salt buffer (0.8 M sodium citrate, 1.2 M NaCl) and 3.5 mL isopropanol. The solution was incubated for 15 min at RT to precipitate the RNA and centrifuged (30 min, 4,000 x g, 4°C). The pellet was washed with 10 mL ice-cold 70% ethanol, centrifuged (5 min, 4000 x g, 4°C), and air-dried. The pellet was resuspended in 625 µL of warm (65°C) nuclease-free water and transferred to a new tube. The solution was supplemented with 70 µL 20X DNase I buffer (1 mM $CaCl_2$, 100 mM Tris pH 7.4), 70 µL 200 mM $MnCl_2$, 5 µL DNase I (ThermoFisher Scientific®), and 1 µL RNaseOUT (ThermoFisher Scientific®). After 1 h incubation at 37°C, the reaction was stopped with 50 µL 500 mM EDTA. To precipitate the RNA, 375 µL high salt buffer and 375 µL isopropanol were added. After incubation for 15 min at room temperature (RT), the RNA was pelleted by centrifugation (20 min, 20,000 x g, 4°C). The pellet was washed with 1 mL ice-cold 70% ethanol, centrifuged (5 min, 20,000 x g, 4°C), air-dried, and resuspended in 50 µL nuclease-free water. All batches of the same sample were pooled, and the solution was supplemented with 0.5 µL RNaseOUT™. For cDNA synthesis, two to four reactions with 11 µL RNA solution, 1 µL 10 µM GFP496 specific reverse transcription primer, and 1 µL 10 mM dNTPs were incubated at 65°C for 5 min then immediately placed on ice. The reactions were supplemented with 4 µL 5X SuperScript IV buffer, 1 µL 100 mM DTT, 1 µL

RNaseOUT™, and 1 µL SuperScript IV reverse transcriptase (ThermoFisher Scientific®). To ensure that the samples were largely free of DNA contamination, four reactions were used as controls, where the reverse transcriptase and RNaseOUT™ were replaced with water. Reactions were incubated for 10 min at 55°C, followed by 10 min at 80°C. Sets of 4 reactions each were pooled. The cNDA was purified with the Clean&Concentrate-5 kit (Zymo Research®) and eluted in 20 µL 10 mM Tris. The barcode was amplified with 10-20 cycles of polymerase chain reaction (PCR) and read out by next generation sequencing.

[0119] For Plant STARR-seq in maize protoplasts, the protoplast-containing Trizol solution from PEG transformation was transferred to 2 mL Phasemaker tubes (1 mL per tube; ThermoFisher Scientific®), mixed thoroughly with 300 µL chloroform, and centrifuged (5 min, 15,000 x g, 4°C). RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN®). The supernatant was transferred to a QIAshredder column (QIAGEN®) and centrifuged (2 min, 20,000 x g, RT). The flowthrough was transferred to a new 1.5 mL tube and mixed with 300 µL 100% ethanol. Up to 500 µL of the solution was loaded on an RNeasy mini spin column (QIAGEN®). After centrifugation (10 seconds, 16,100 x g, RT) the flowthrough was discarded. This was repeated until the whole solution had been added to the column. The column was washed with 350 µL RW1 buffer followed by centrifugation (30 sec, 16,100 x g, RT). An on-column DNase I digestion was performed with 70 µL RDD buffer and 10 µL DNase I (QIAGEN®) for 15 min at RT. The column was washed once with 350 µL RW1 buffer and twice with 500 µL RPE buffer. After each wash step, the column was centrifuged (30 sec, 16,100 x g, RT) and the flowthrough was discarded. The column was dried with an extra centrifugation step (30 sec, 16,100 x g, RT) and transferred to a 1.5 mL collection tube. For elution, 50 µL of RNase-free water was added, and the column was incubated for 1 min, and centrifuged (1 min, 16,100 x g, RT). This elution step was repeated with an additional 40 µL of RNase-free water. The eluate was treated with DNase I (5 µL of 20x DNaseI buffer, 5 µL 200 mM $MnCl_2$, 1 µL RNaseOUT™, and 2 µL DNase I) for 1 h at 37°C. The solution was supplemented with 20 µL 500 mM EDTA, 1 µL 20 mg/mL glycogen, 12 µL ice-cold 8M LiCl, and 300 µL ice-cold 100% ethanol. The solution was incubated 15 min at −80°C, centrifuged (20 min, 20,000 x g, 4°C). The pellet was washed with 500 µL ice-cold 70% ethanol, and centrifuged (3 min, 20,000 x g, 4°C). The pellet was air-dried and resuspended in 100 µL RNase-free water. Reverse transcription, purification, PCR amplification and sequencing were performed as for the tobacco samples.

Subassembly and barcode sequencing

[0120] Paired-end sequencing on an Illumina® NextSeq™ 550 platform or Illumina® NextSeq™ 2000 platform was used to link insulator fragments to their respective barcodes. The insulator region was sequenced using paired reads (100–150 bp), and two 18-bp indexing reads were used to sequence the barcodes. The paired insulator fragment and barcode reads were assembled using PANDAseq (version 2.11). Insulator fragment-barcode pairs with less than 5 reads and insulator fragments with a mutation or truncation were discarded. For each Plant STARR-seq experiment, barcodes were sequenced using paired-end reads on an Illumina® NextSeq™ 550 platform or Illumina® NextSeq™ 2000 platform. The paired barcode reads were assembled using PANDAseq.

Non-Limiting Embodiments

[0121] While general features of the disclosure are described and shown and particular features of the disclosure are set forth in the claims, the following non-limiting embodiments relate to features, and combinations of features, that are explicitly envisioned as being part of the disclosure. The following non-limiting Embodiments contain elements that are modular and can be combined with each other in any number, order, or combination to form a new non-limiting Embodiment, which can itself be further combined with other non-limiting Embodiments.

[0122] Embodiment 1. A method for determining whether a candidate sequence is an insulator in a plant genetic environment, the method comprising: cloning the candidate sequence between an enhancer and a promoter to produce a test sequence; and attempting to initiate transcription from the test sequence, wherein reduced transcription relative to a control construct without the candidate sequence indicates the candidate sequence or a portion thereof is the insulator.

[0123] Embodiment 2. The method of Embodiment 1 or any other Embodiment, further comprising: cloning the candidate sequence upstream of the enhancer and the promoter to produce a second test sequence and attempting to initiate transcription from the second test sequence, wherein transcript levels similar to that of the control construct indicate the candidate sequence or the portion thereof is the insulator, and wherein reduced transcription relative to the control construct indicates the candidate sequence or the portion thereof is a silencer.

[0124] Embodiment 3. The method of any one of Embodiments 1-2 or any other Embodiment, further comprising: cloning a plurality of candidate sequences between a plurality of enhancers and promoters to produce a plurality of first test sequences that comprise barcodes, transforming the first plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the first plurality of test sequences; cloning at least a portion of candidate sequences upstream of enhancers and promoters to produce a plurality of second test sequences that comprise barcodes, transforming the second plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the second plurality of test sequences; measuring levels of barcode messenger RNA (mRNA) transcripts of the first plurality of test sequences and the second plurality of test sequences and measuring levels of barcode DNA of the first plurality of test sequences and the second plurality of test sequences, and calculating relative transcription levels as ratios of levels of barcode mRNA transcripts to levels of barcode DNA, wherein reduced relative transcription levels of first test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators or silencers, wherein similar relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators, and wherein reduced relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are silencers.

[0125] Embodiment 4. The method of any one of Embodiments 1-3 or any other Embodiment, wherein transcription levels are measured with a sequencing method.

[0126] Embodiment 5. The method of any one of Embodiments 1-4 or any other Embodiment, wherein the sequencing method comprises next-generation sequencing (NGS).

[0127] Embodiment 6. The method of any one of Embodiments 1-5 or any other Embodiment, wherein the sequencing method comprises a reverse-transcription (RT) step.

[0128] Embodiment 7. The method of any one of Embodiments 1-6 or any other Embodiment, wherein the plant genetic environment is a plant genome, or a portion thereof, of a plant cell, a plant tissue, or a plant organism.

[0129] Embodiment 8. The method of any one of Embodiments 1-7 or any other Embodiment, wherein the plant genetic environment corresponds to an agricultural plant organism or an agricultural crop.

[0130] Embodiment 9. The method of any one of Embodiments 1-8 or any other Embodiment, wherein the plant genetic environment comprises a tobacco genetic environment, a maize genetic environment, a rice genetic environment, a sorghum genetic environment, a tomato genetic environment, or an *Arabidopsis thaliana* genetic environment.

[0131] Embodiment 10. The method of any one of Embodiments 1-9 or any other Embodiment, wherein the enhancer comprises a 35S enhancer and the promoter comprises a 35S minimal promoter.

[0132] Embodiment 11. The method of any one of Embodiments 1-10 or any other Embodiment, wherein one or more candidate sequences are cloned in a forward orientation with respect to the plant genetic context, a reverse orientation with respect to the plant genetic context, or both.

[0133] Embodiment 12. The method of any one of Embodiments 1-11 or any other Embodiment, wherein one or more candidate sequences determined to be insulators comprise a higher frequency of guanine and cytosine (GC) nucleobases relative to a random sequence.

[0134] Embodiment 13. An insulator comprising a polynucleotide sequence configured for regulation of transcription within one or more plant genetic environments.

[0135] Embodiment 14. The insulator of Embodiment 13 or any other Embodiment, wherein the insulator does not comprise a naturally-occurring plant sequence.

[0136] Embodiment 15. The insulator of any one of Embodiments 13-14 or any other Embodiment, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1439 (Synthetic insulator 1 (sIns1)).

[0137] Embodiment 16. The insulator of any one of Embodiments 13-14 or any other Embodiment, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at

least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1440 (Synthetic insulator 2 (sIns2)).

[0138] Embodiment 17. The insulator of Embodiment 13 or any other Embodiment, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence selected from the group consisting of: SEQ ID NOs:1-1,438, SEQ ID NOs:1,441-1,446, SEQ ID NOs:1,447-2,361, SEQ ID NOs:2,362-19,190, SEQ ID NOs:19,191-20,082, SEQ ID NOs:20,083-20,523, SEQ ID NOs:20,524-20,725, and any combination thereof.

[0139] Embodiment 18. A plant genetic construct comprising the insulator of any one of Embodiments 13-17 or any other Embodiment.

[0140] Embodiment 19. A plant cell, plant tissue, or plant comprising the plant genetic construct of Embodiment 18 or any other Embodiment, wherein the plant cell, plant tissue, or plant is produced by a genetic engineering method.

[0141] Embodiment 20. The plant cell, plant tissue, or plant of Embodiment 19 or any other Embodiment, wherein the insulator is stably integrated into a genome of the plant cell, plant tissue, or plant, and wherein the insulator regulates expression of a transgene of the genome by at least partially insulating the transgene from regulatory cross-talk.

[0142] While illustrative embodiments have been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the disclosure.

CLAIMS

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A method for determining whether a candidate sequence is an insulator in a plant genetic environment, the method comprising:

cloning the candidate sequence between an enhancer and a promoter to produce a test sequence; and

attempting to initiate transcription from the test sequence, wherein reduced transcription relative to a control construct without the candidate sequence indicates the candidate sequence or a portion thereof is the insulator.

2. The method of claim 1, further comprising:

cloning the candidate sequence upstream of the enhancer and the promoter to produce a second test sequence and attempting to initiate transcription from the second test sequence, wherein transcript levels similar to that of the control construct indicate the candidate sequence or the portion thereof is the insulator, and wherein reduced transcription relative to the control construct indicates the candidate sequence or the portion thereof is a silencer.

3. The method of claim 1, further comprising:

cloning a plurality of candidate sequences between a plurality of enhancers and promoters to produce a plurality of first test sequences that comprise barcodes, transforming the first plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the first plurality of test sequences;

cloning at least a portion of candidate sequences upstream of enhancers and promoters to produce a plurality of second test sequences that comprise barcodes, transforming the second plurality of test sequences into plant cells or tissues, and attempting to initiate transcription from the second plurality of test sequences;

measuring levels of barcode messenger RNA (mRNA) transcripts of the first plurality of test sequences and the second plurality of test sequences and measuring levels of barcode DNA of the first plurality of test sequences and the second plurality of test sequences, and calculating relative transcription levels as ratios of levels of barcode mRNA transcripts to levels of barcode DNA,

wherein reduced relative transcription levels of first test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators or silencers,

wherein similar relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are insulators, and

wherein reduced relative transcription levels of second test sequences relative to the control construct indicates one or more candidate sequences thereof, or portions thereof, are silencers.

4. The method of claim 1, wherein transcription levels are measured with a sequencing method.

5. The method of claim 4, wherein the sequencing method comprises next-generation sequencing (NGS).

6. The method of claim 4, wherein the sequencing method comprises a reverse-transcription (RT) step.

7. The method of claim 1, wherein the plant genetic environment is a plant genome, or a portion thereof, of a plant cell, a plant tissue, or a plant organism.

8. The method of claim 1, wherein the plant genetic environment corresponds to an agricultural plant organism or an agricultural crop.

9. The method of claim 8, wherein the plant genetic environment comprises a tobacco genetic environment, a maize genetic environment, a rice genetic environment, a sorghum genetic environment, a tomato genetic environment, or an *Arabidopsis thaliana* genetic environment.

10. The method of claim 1, wherein the enhancer comprises a 35S enhancer and the promoter comprises a 35S minimal promoter.

11. The method of claim 1, wherein one or more candidate sequences are cloned in a forward orientation with respect to the plant genetic context, a reverse orientation with respect to the plant genetic context, or both.

12. The method of claim 1, wherein one or more candidate sequences determined to be insulators comprise a higher frequency of guanine and cytosine (GC) nucleobases relative to a random sequence.

13. An insulator comprising a polynucleotide sequence configured for regulation of transcription within one or more plant genetic environments.

14. The insulator of claim 13, wherein the insulator does not comprise a naturally-occurring plant sequence.

15. The insulator of claim 14, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1439 (Synthetic insulator 1 (sIns1)).

16. The insulator of claim 14, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence of SEQ ID NO:1440 (Synthetic insulator 2 (sIns2)).

17. The insulator of claim 13, wherein the insulator comprises a polynucleotide sequence having at least 80% identity, at least 85% identity, at least 90% identity, at least 91% identity, at least 92% identity, at least 93% identity, at least 94% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, or 100% identity to a polynucleotide sequence selected from the group consisting of: SEQ ID NOs:1-1,438, SEQ ID NOs:1,441-1,446, SEQ ID NOs:1,447-2,361, SEQ ID NOs:2,362-19,190, SEQ ID NOs:19,191-20,082, SEQ ID NOs:20,083-20,523, SEQ ID NOs:20,524-20,725, and any combination thereof.

18. A plant genetic construct comprising the insulator of claim 13.

19. A plant cell, plant tissue, or plant comprising the plant genetic construct of claim 18, wherein the plant cell, plant tissue, or plant is produced by a genetic engineering method.

20. The plant cell, plant tissue, or plant of claim 19, wherein the insulator is stably integrated into a genome of the plant cell, plant tissue, or plant, and wherein the insulator regulates expression of a transgene of the genome by at least partially insulating the transgene from regulatory cross-talk.

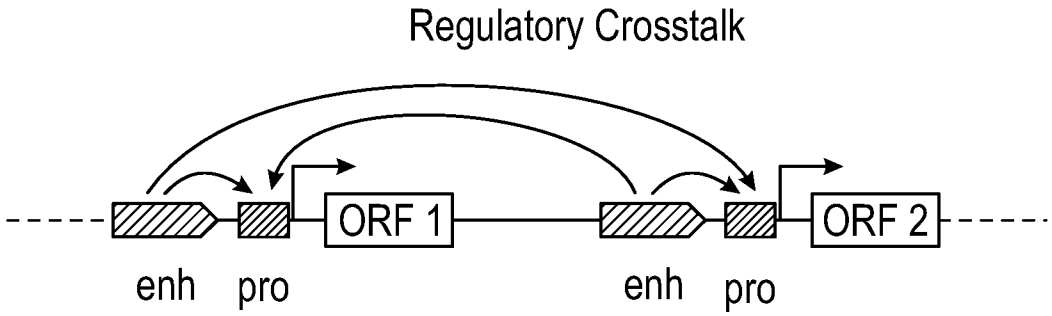

Regulatory Crosstalk
ORF 1
ORF 2
enh
pro
enh
pro

FIG. 1A

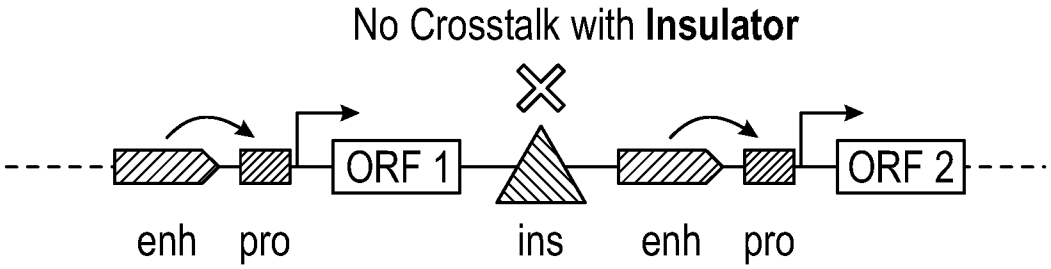

No Crosstalk with **Insulator**
ORF 1
ORF 2
enh
pro
ins
enh
pro

FIG. 1B

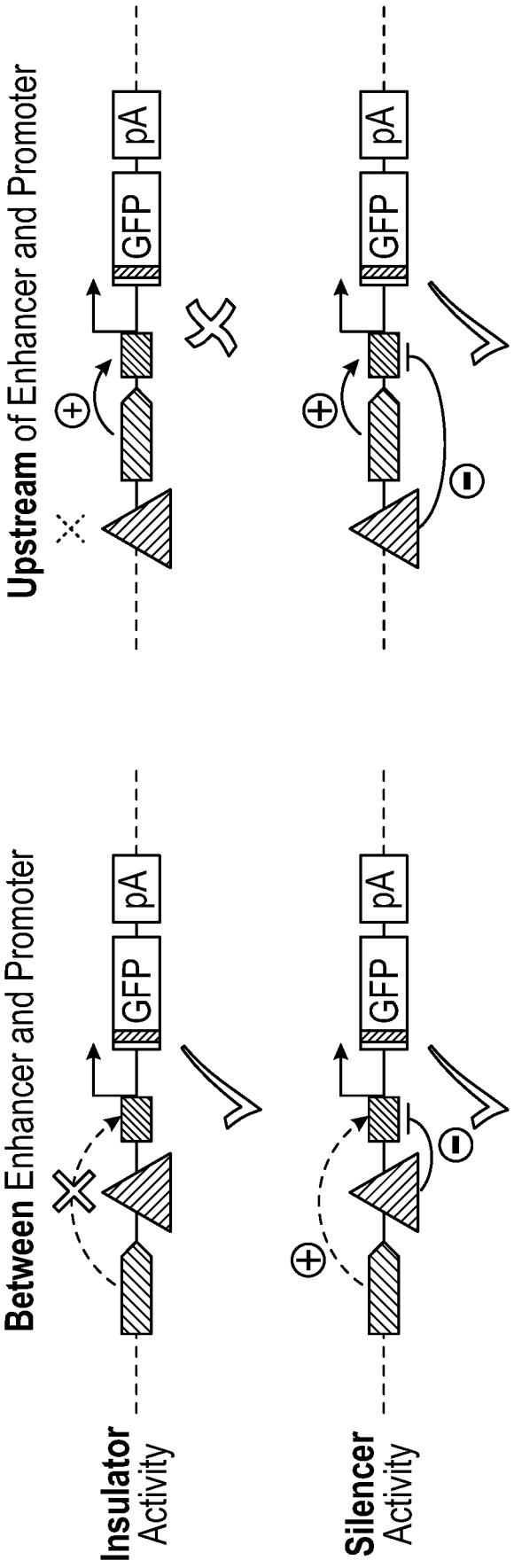

Upstream of Enhancer and Promoter
Between Enhancer and Promoter
GFP
pA
Insulator Activity
Silencer Activity

FIG. 1C

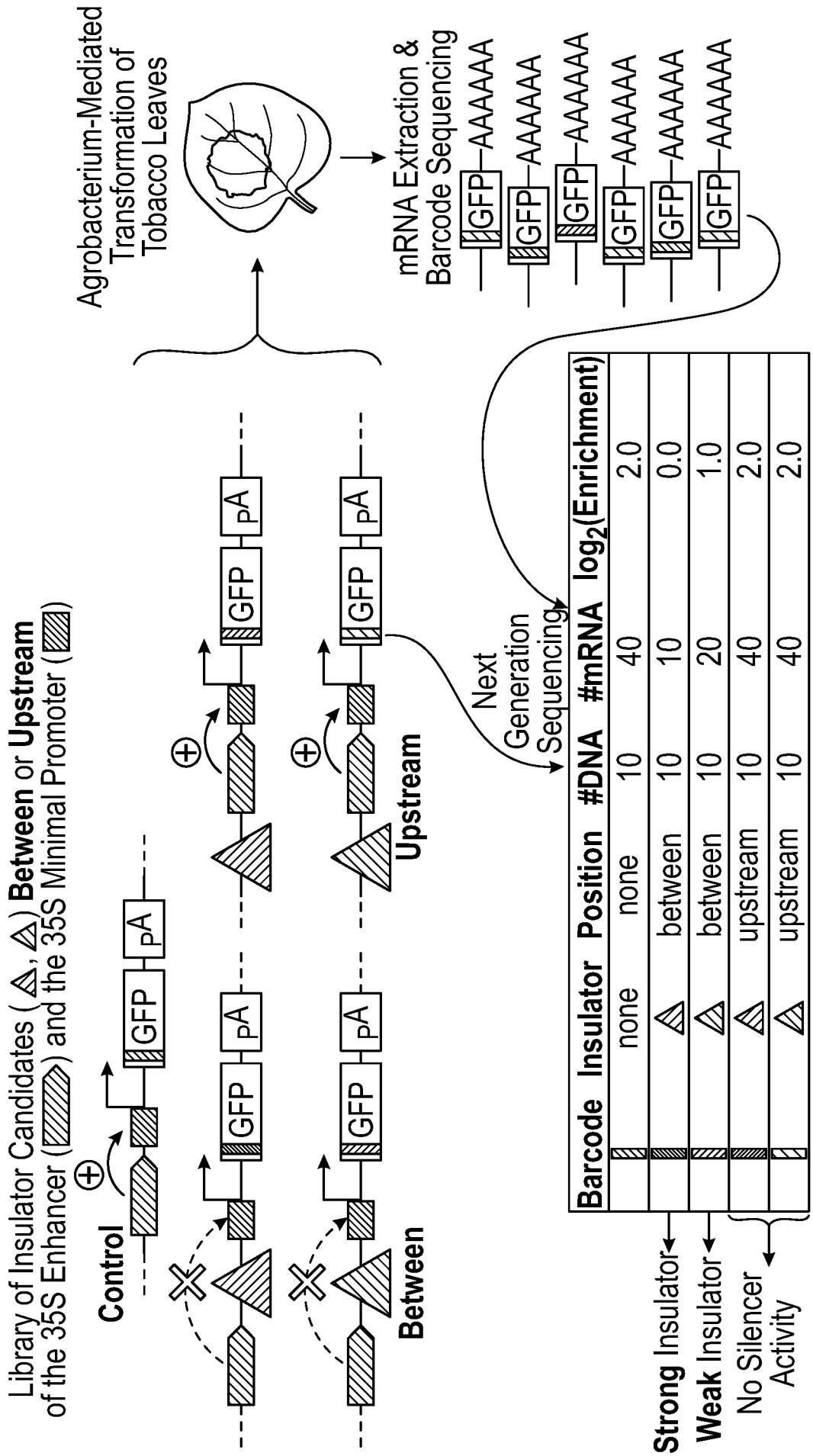

Library of Insulator Candidates (,) Between or Upstream of the 35S Enhancer ( ) and the 35S Minimal Promoter ( )
Control
Between
Upstream
GFP
pA
Agrobacterium-Mediated Transformation of Tobacco Leaves
mRNA Extraction & Barcode Sequencing
AAAAAA
Next Generation Sequencing
Barcode Insulator Position #DNA #mRNA log2(Enrichment)
none none 10 40 2.0
between 10 10 0.0
between 10 20 1.0
upstream 10 40 2.0
upstream 10 40 2.0
Strong Insulator
Weak Insulator
No Silencer Activity

FIG. 1D

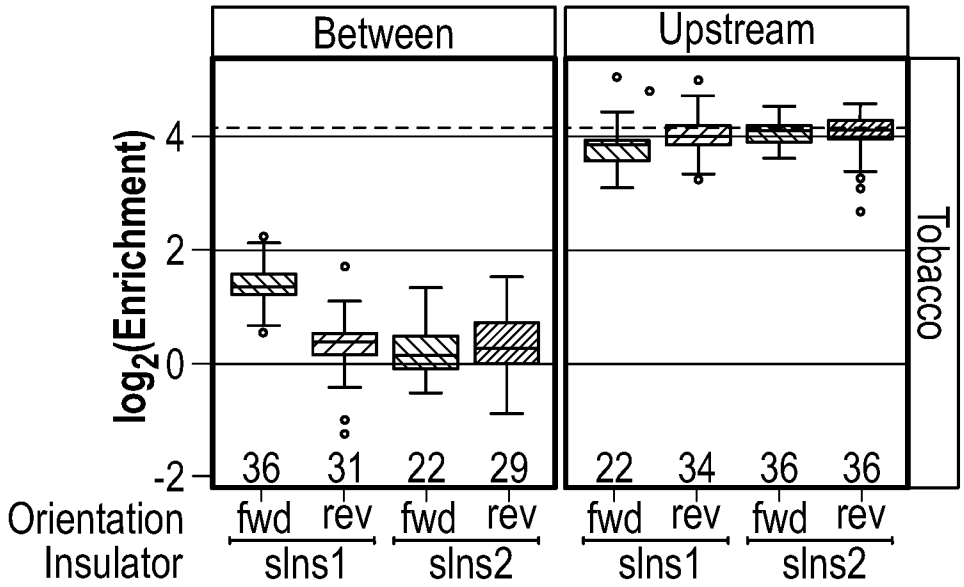

Between
Upstream
Tobacco
log2(Enrichment)
4
2
0
-2
36 31 22 29
22 34 36 36
Orientation
fwd rev fwd rev
fwd rev fwd rev
Insulator
slns1 slns2
slns1 slns2

FIG. 1E

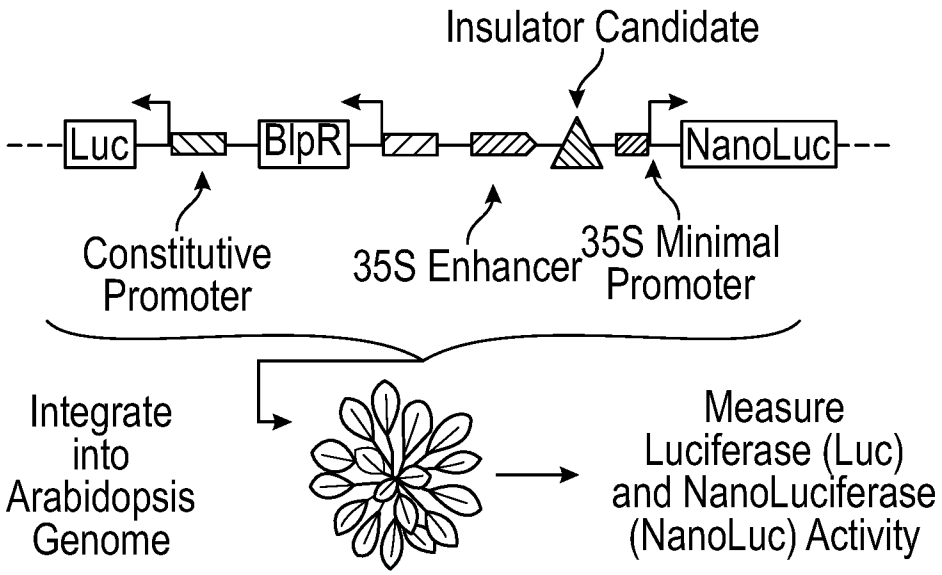

Insulator Candidate
Luc
BlpR
NanoLuc
Constitutive Promoter
35S Enhancer
35S Minimal Promoter
Integrate into Arabidopsis Genome
Measure Luciferase (Luc) and NanoLuciferase (NanoLuc) Activity

FIG. 1F

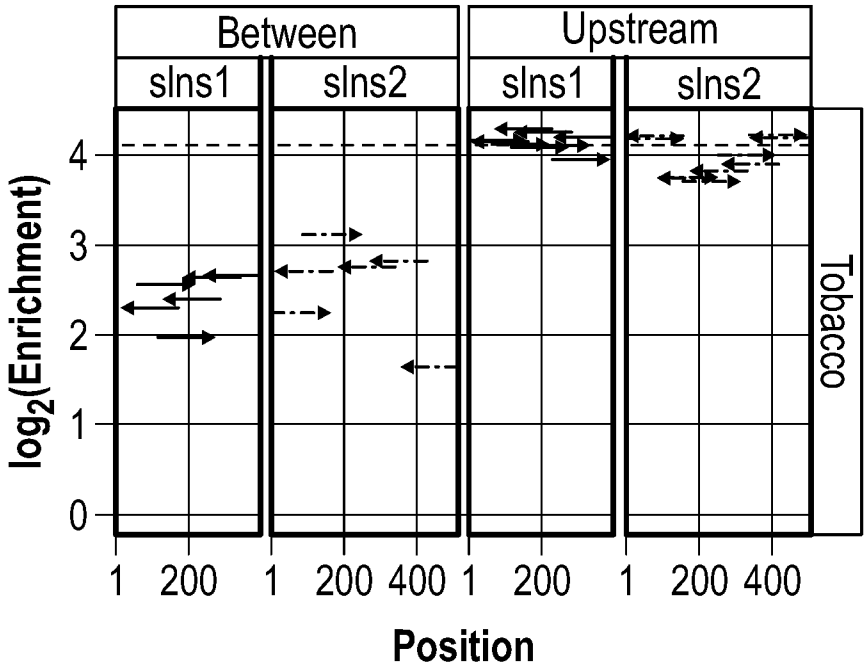

Between
Upstream
slns1
slns2
slns1
slns2
Tobacco
log2(Enrichment)
0
1
2
3
4
1
200
1
200
400
1
200
1
200
400
Position

FIG. 1G

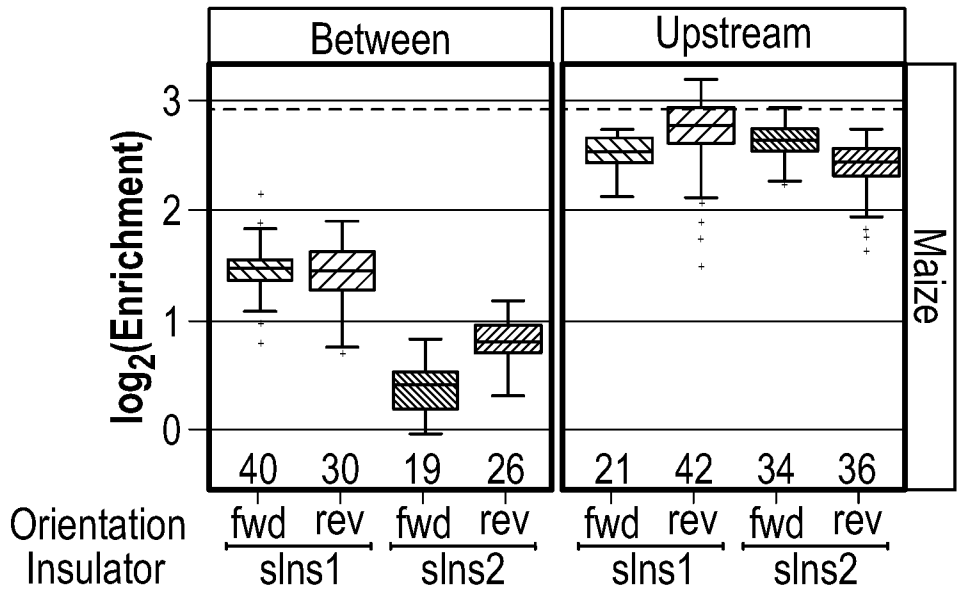

Between
Upstream
Maize
log2(Enrichment)
0
1
2
3
40
30
19
26
21
42
34
36
Orientation
fwd
rev
fwd
rev
fwd
rev
fwd
rev
Insulator
slns1
slns2
slns1
slns2

FIG. 1H

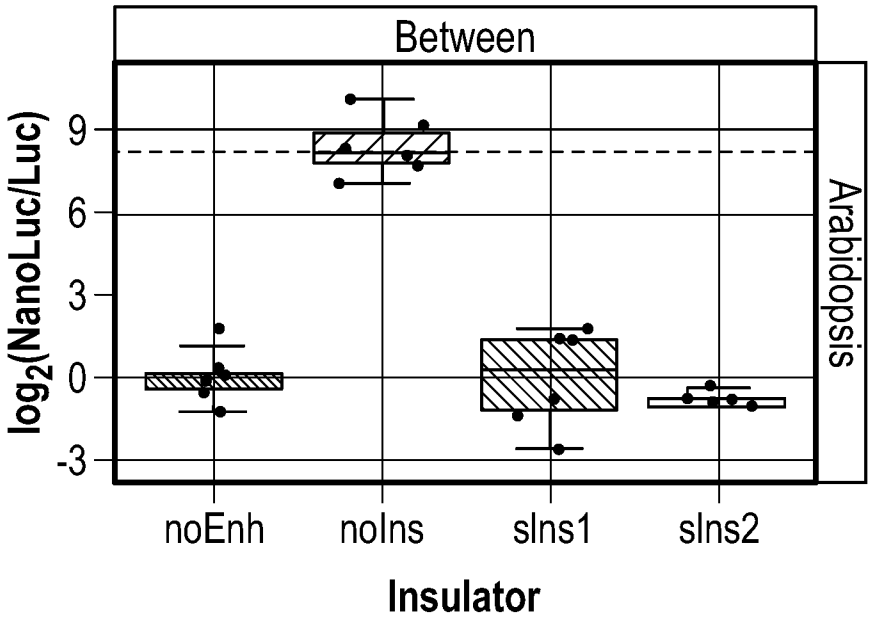

Between
Arabidopsis
log2(NanoLuc/Luc)
9
6
3
0
-3
noEnh
noIns
sIns1
sIns2
Insulator

FIG. 1I

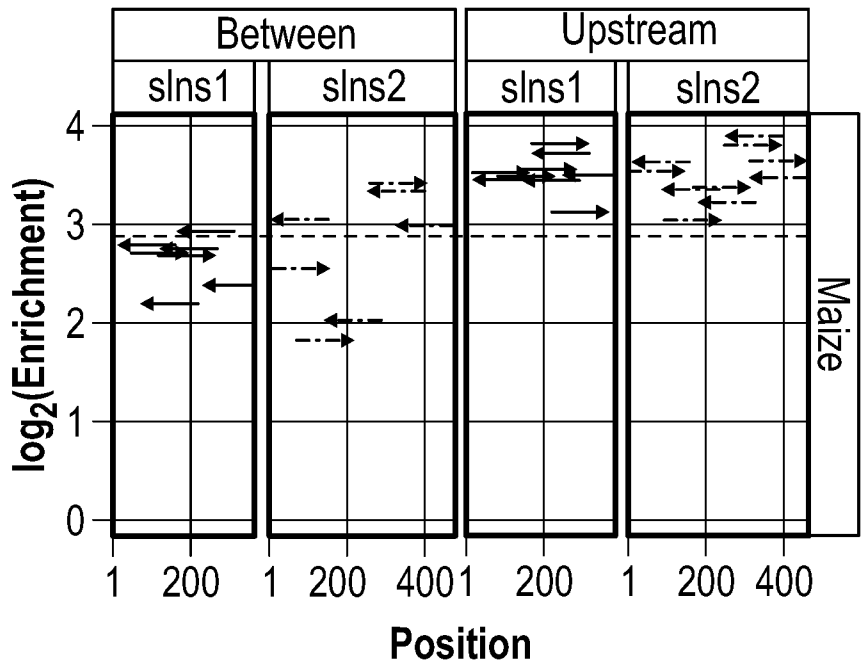

Between
Upstream
sIns1
sIns2
sIns1
sIns2
Maize
log2(Enrichment)
4
3
2
1
0
1
200
1
200
400
1
200
1
200
400
Position

FIG. 1J

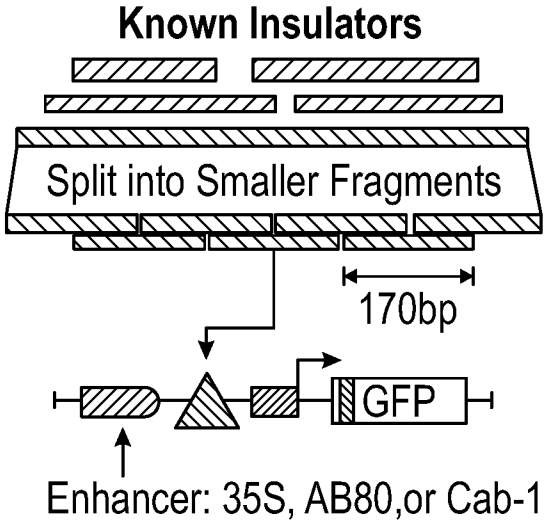


FIG. 2A

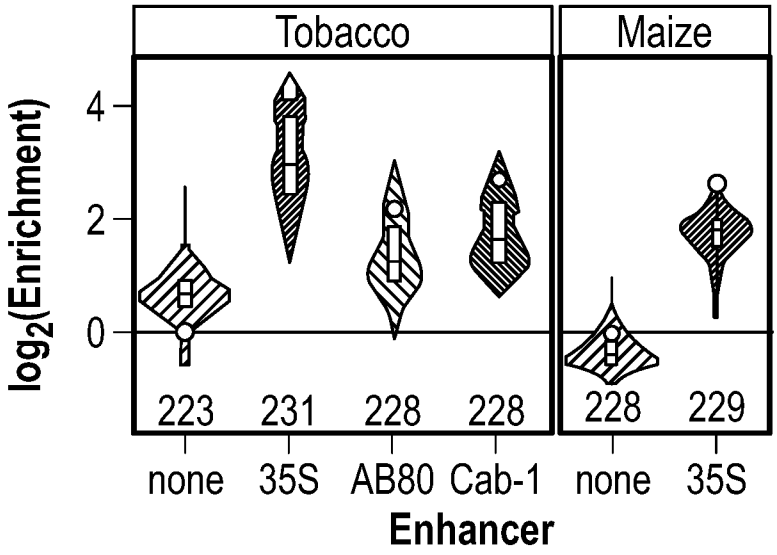


FIG. 2B

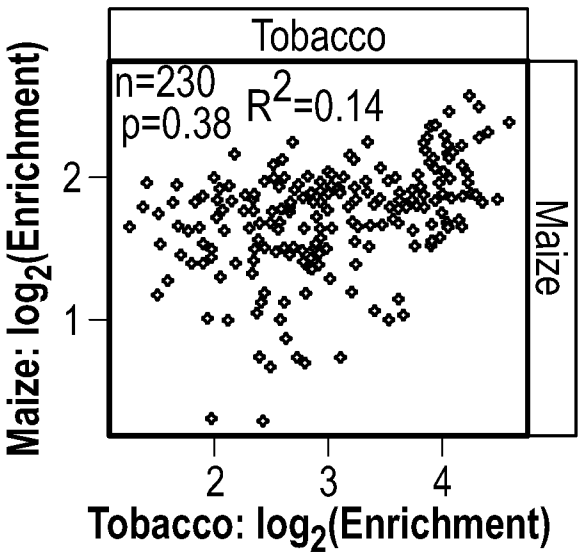


FIG. 2C

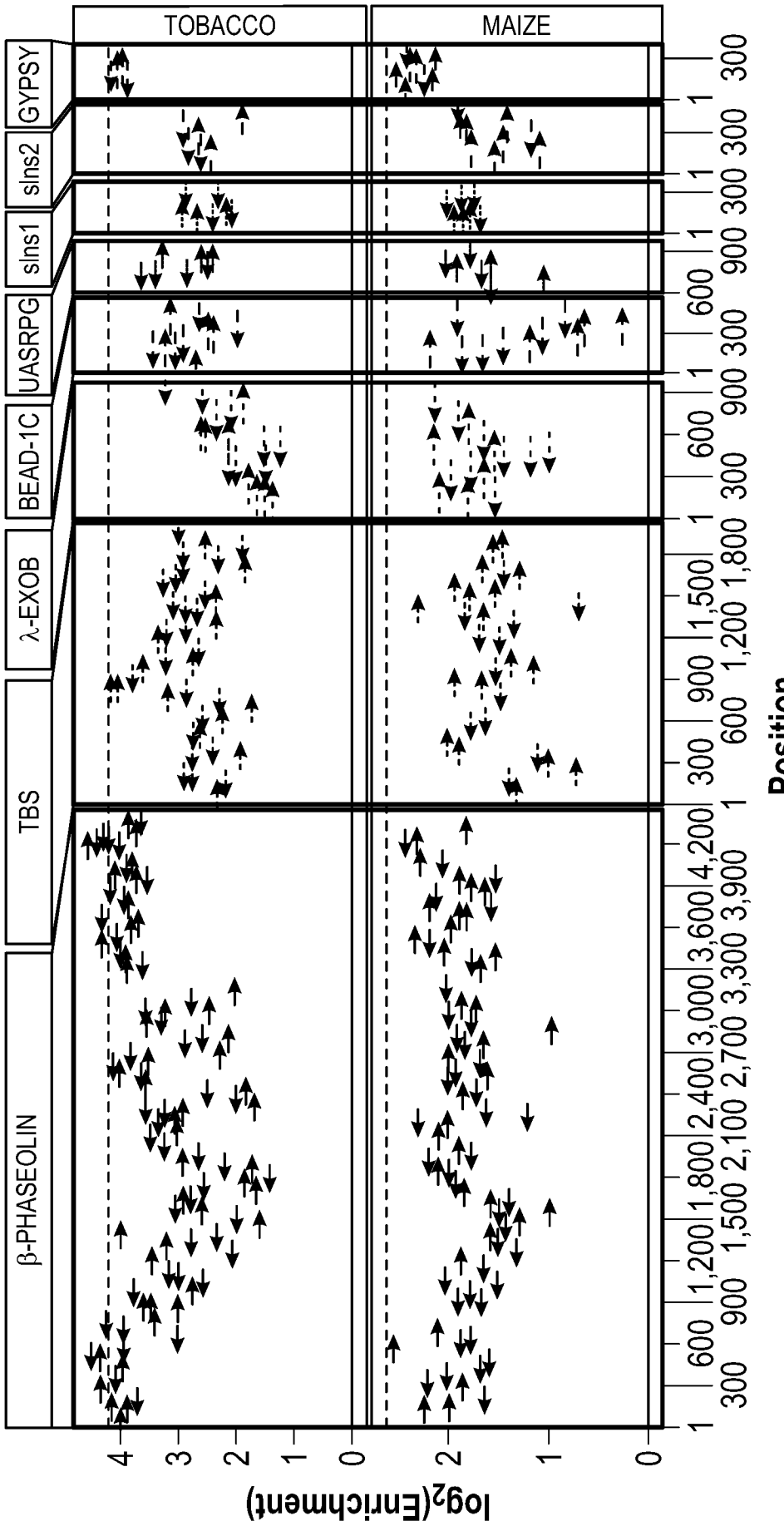

TOBACCO
MAIZE
β-PHASEOLIN
TBS
λ-EXOB
BEAD-1C
UASRPG
sIns1
sIns2
GYPSY
log2(Enrichment)
Position

FIG. 2D

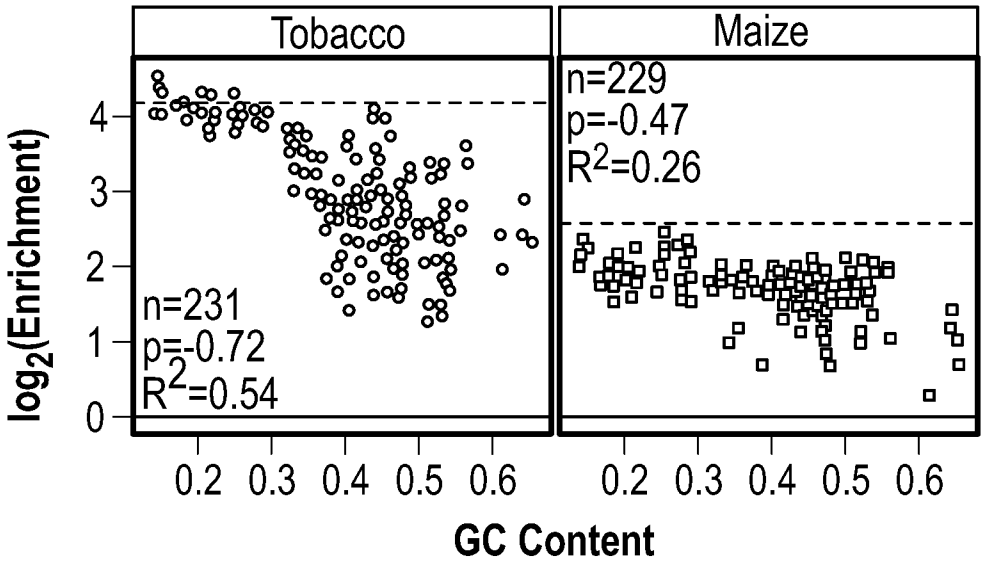

Tobacco
Maize
n=231
p=-0.72
R2=0.54
n=229
p=-0.47
R2=0.26
log2(Enrichment)
GC Content
0.2 0.3 0.4 0.5 0.6
0 1 2 3 4

FIG. 2E

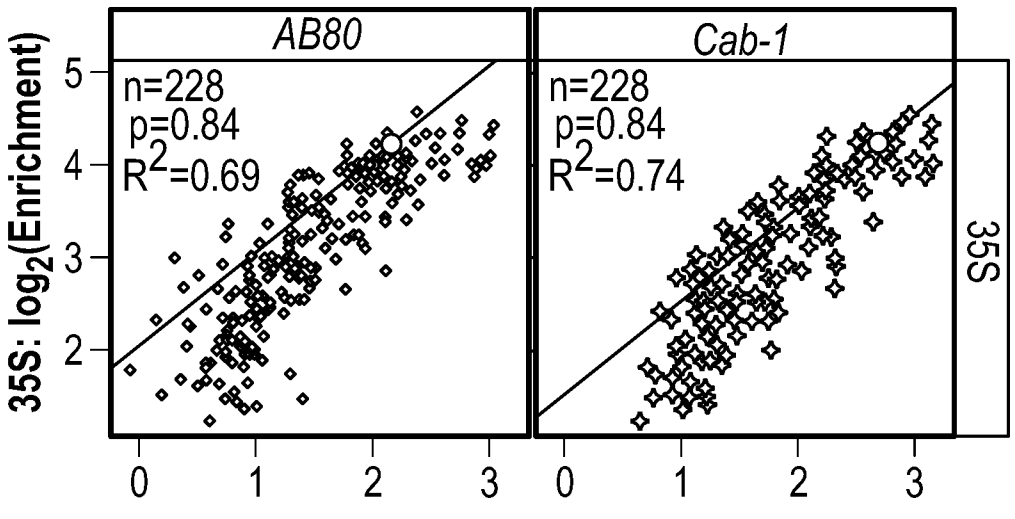

AB80
Cab-1
n=228
p=0.84
R2=0.69
n=228
p=0.84
R2=0.74
35S: log2(Enrichment)
AB80 or Cab-1: log2(Enrichment)
35S
0 1 2 3
2 3 4 5

FIG. 2F

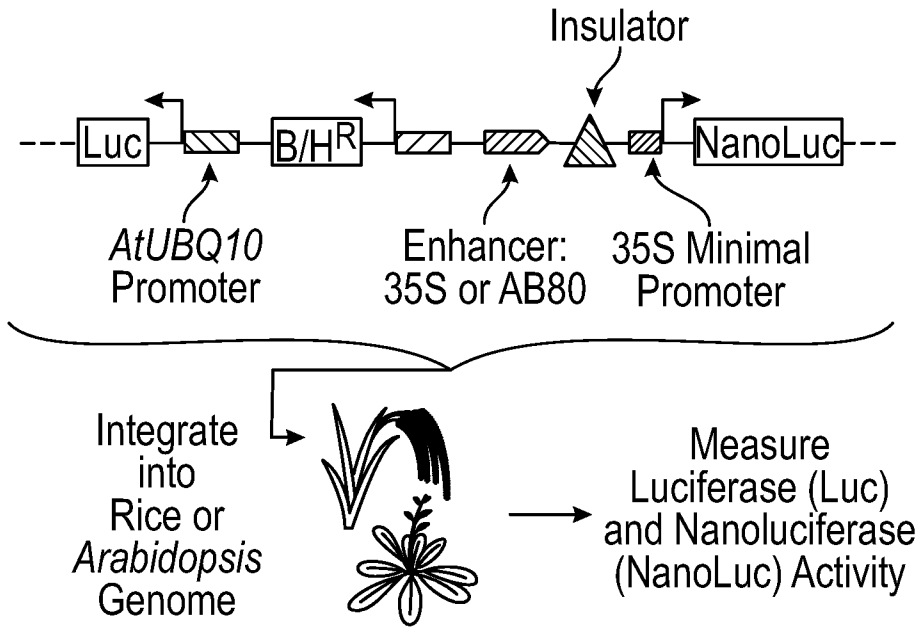

Insulator
Luc
B/H^R
NanoLuc
AtUBQ10 Promoter
Enhancer: 35S or AB80
35S Minimal Promoter
Integrate into Rice or Arabidopsis Genome
Measure Luciferase (Luc) and Nanoluciferase (NanoLuc) Activity

FIG. 3A

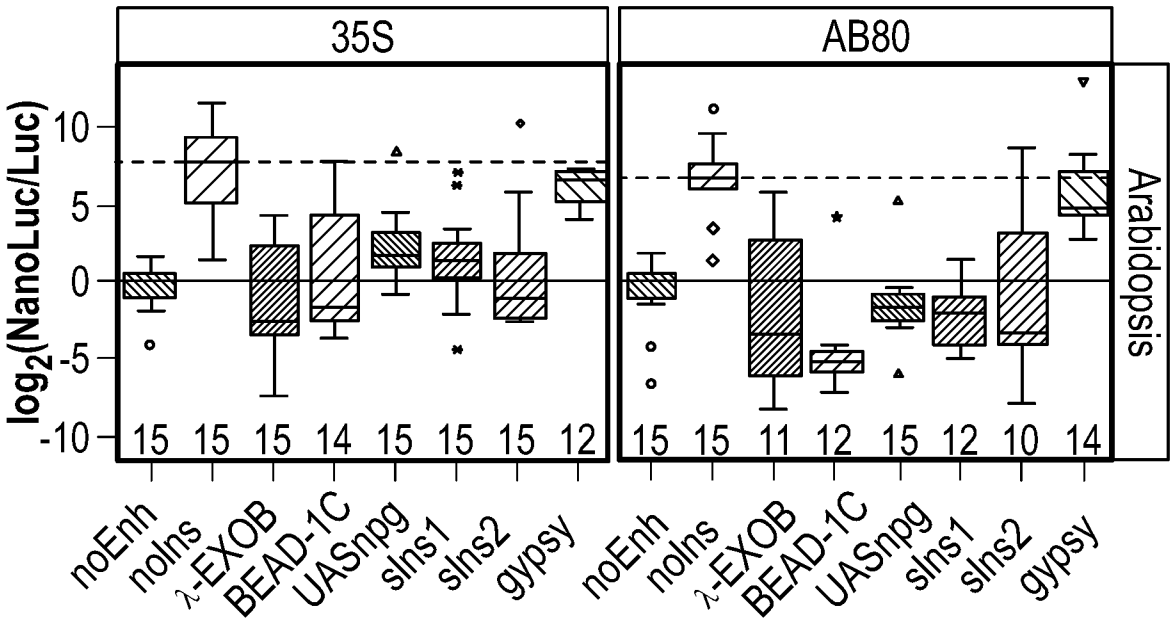

35S
AB80
Arabidopsis
log2(NanoLuc/Luc)
10
5
0
-5
-10
15 15 15 14 15 15 15 12
15 15 11 12 15 12 10 14
noEnh
noIns
λ-EXOB
BEAD-1C
UASnpg
sIns1
sIns2
gypsy

FIG. 3B

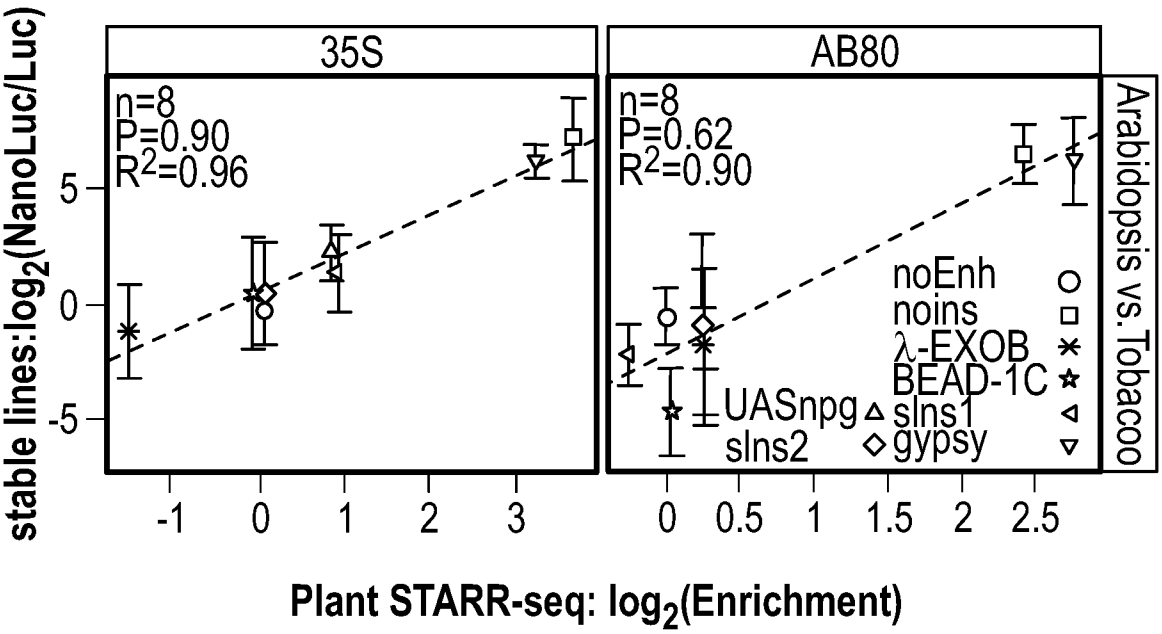

35S
AB80
n=8
P=0.90
R2=0.96
n=8
P=0.62
R2=0.90
Arabidopsis vs. Tobacco
noEnh
noins
λ-EXOB
BEAD-1C
UASnpg
slns1
slns2
gypsy
stable lines:log2(NanoLuc/Luc)
Plant STARR-seq: log2(Enrichment)

FIG. 3C

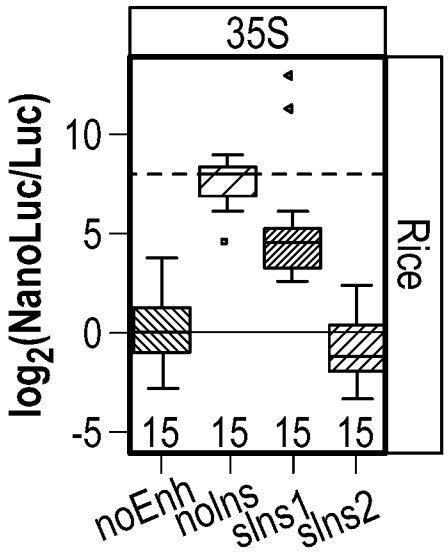

35S
Rice
log2(NanoLuc/Luc)
15 15 15 15
noEnh
noIns
slns1
slns2

FIG. 3D

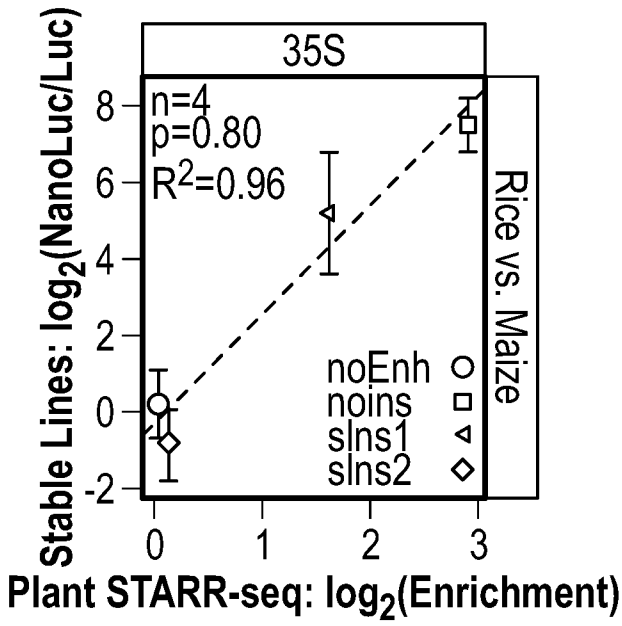

35S
n=4
p=0.80
R2=0.96
Rice vs. Maize
noEnh
noins
sIns1
sIns2
Stable Lines: log2(NanoLuc/Luc)
Plant STARR-seq: log2(Enrichment)

FIG. 3E

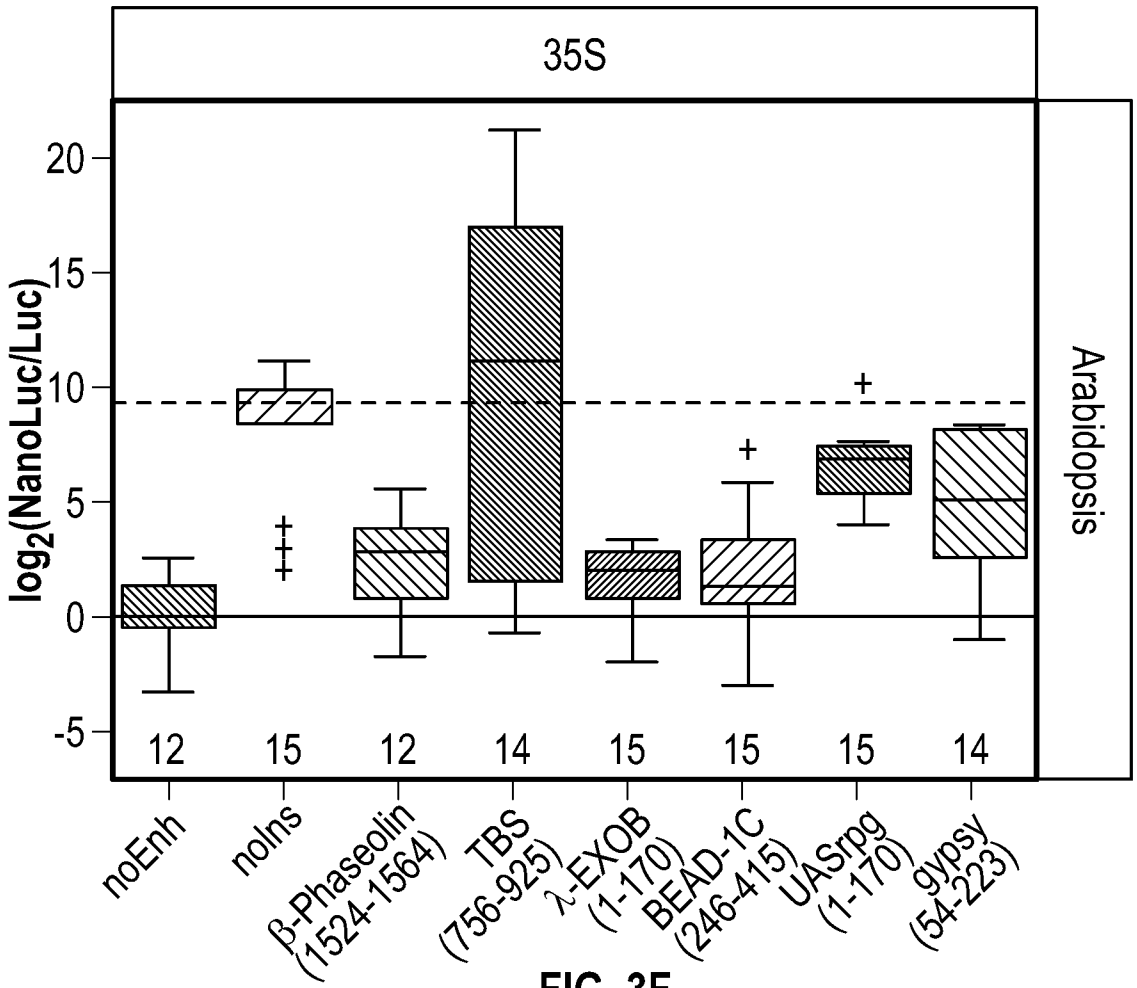

35S
Arabidopsis
log2(NanoLuc/Luc)
12
15
12
14
15
15
15
14
noEnh
noIns
β-Phaseolin (1524-1564)
TBS (756-925)
λ-EXOB (1-170)
BEAD-1C (246-415)
UASrpg (1-170)
gypsy (54-223)

FIG. 3F

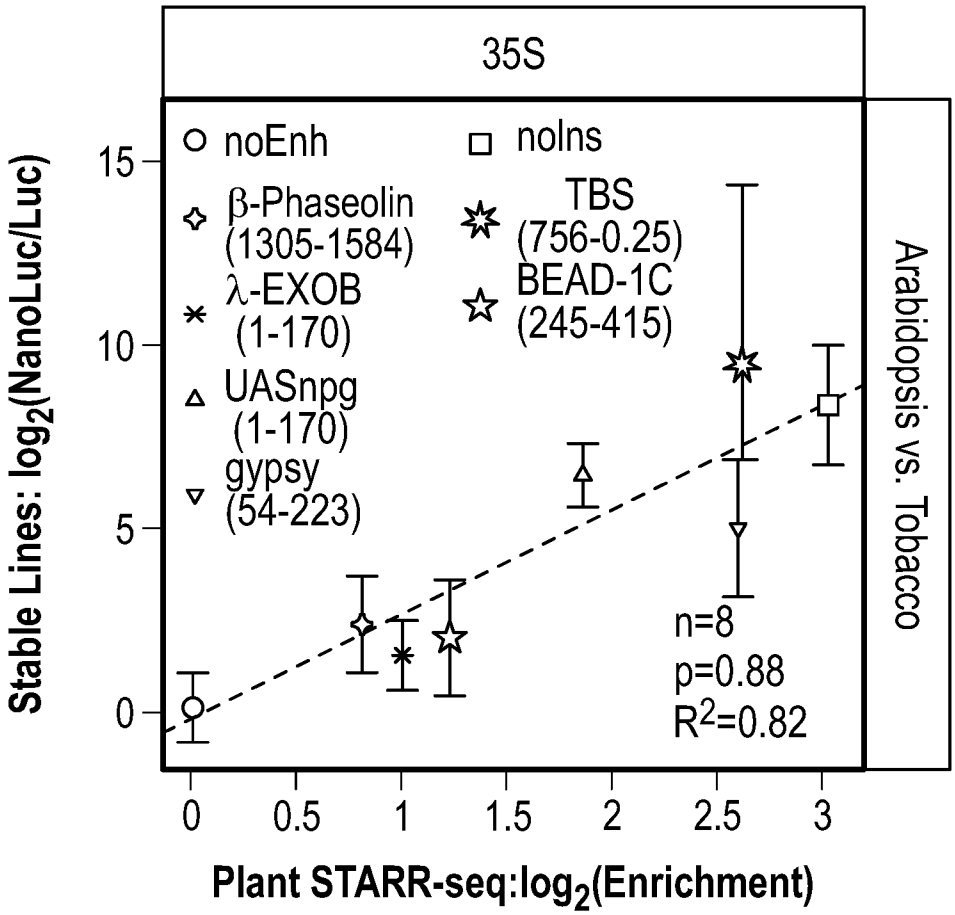

35S
noEnh
noIns
β-Phaseolin (1305-1584)
TBS (756-0.25)
λ-EXOB (1-170)
BEAD-1C (245-415)
UASnpg (1-170)
gypsy (54-223)
n=8
p=0.88
R2=0.82
Arabidopsis vs. Tobacco
Stable Lines: log2(NanoLuc/Luc)
Plant STARR-seq:log2(Enrichment)
0
5
10
15
0
0.5
1
1.5
2
2.5
3

FIG. 3G

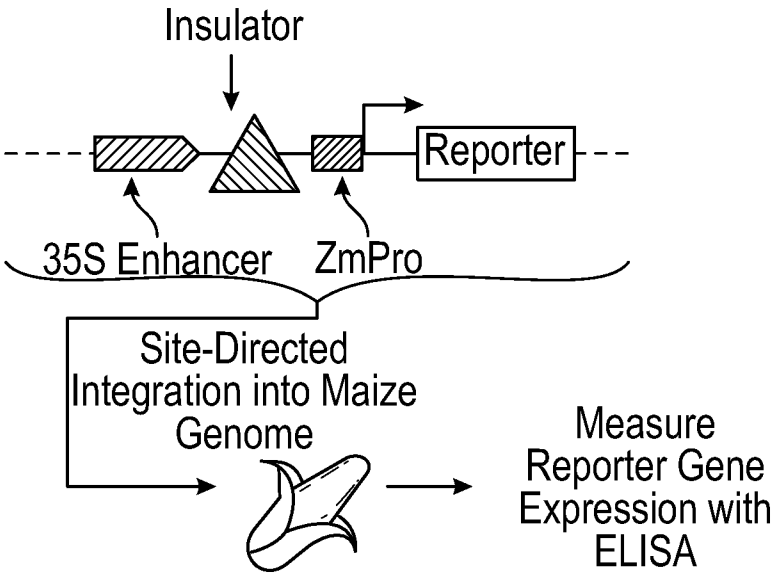

Insulator
Reporter
35S Enhancer
ZmPro
Site-Directed Integration into Maize Genome
Measure Reporter Gene Expression with ELISA

FIG. 3H

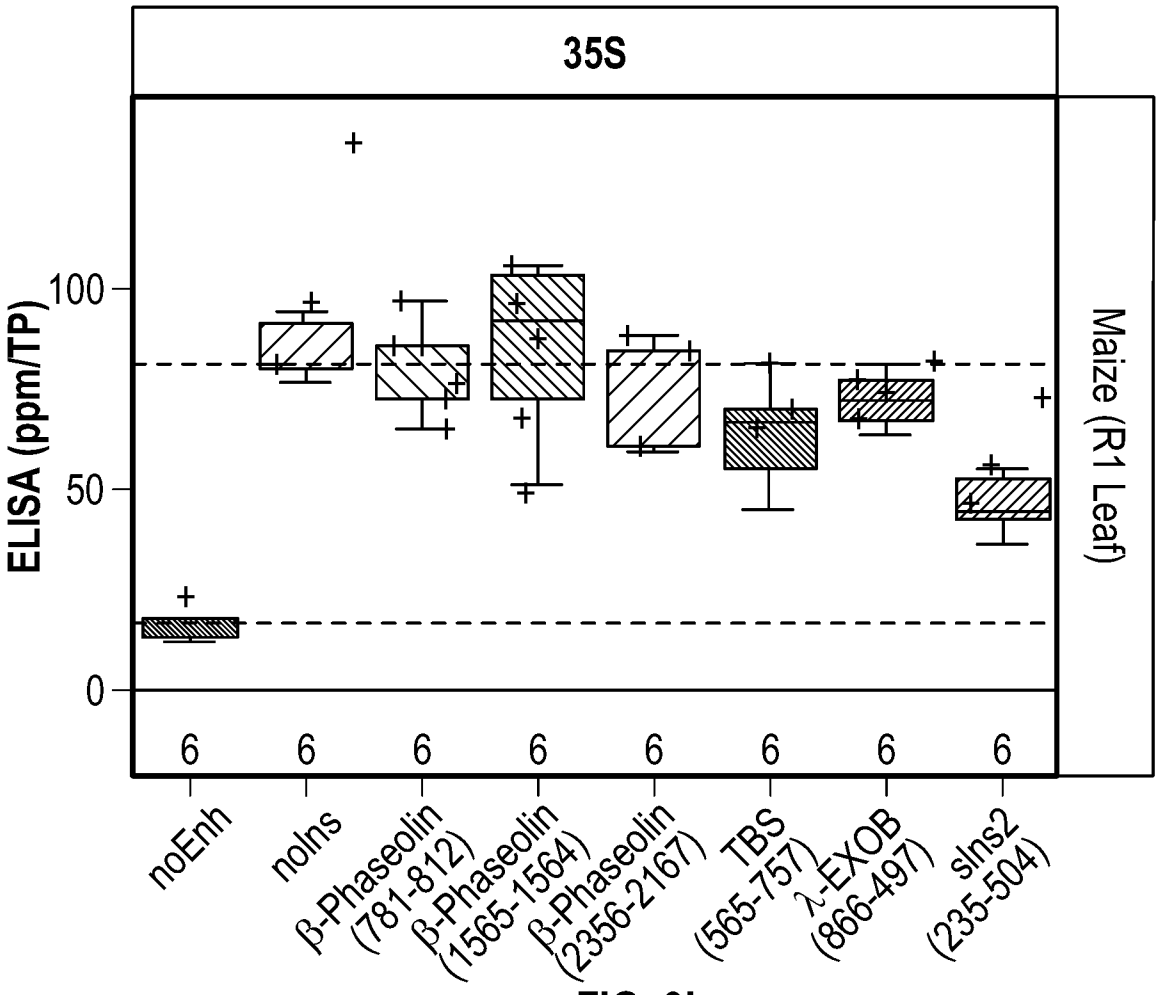

35S
ELISA (ppm/TP)
100
50
0
6
6
6
6
6
6
6
6
Maize (R1 Leaf)
noEnh
noIns
β-Phaseolin (781-812)
β-Phaseolin (1565-1564)
β-Phaseolin (2356-2167)
TBS (565-757)
λ-EXOB (866-497)
sIns2 (235-504)

FIG. 3I

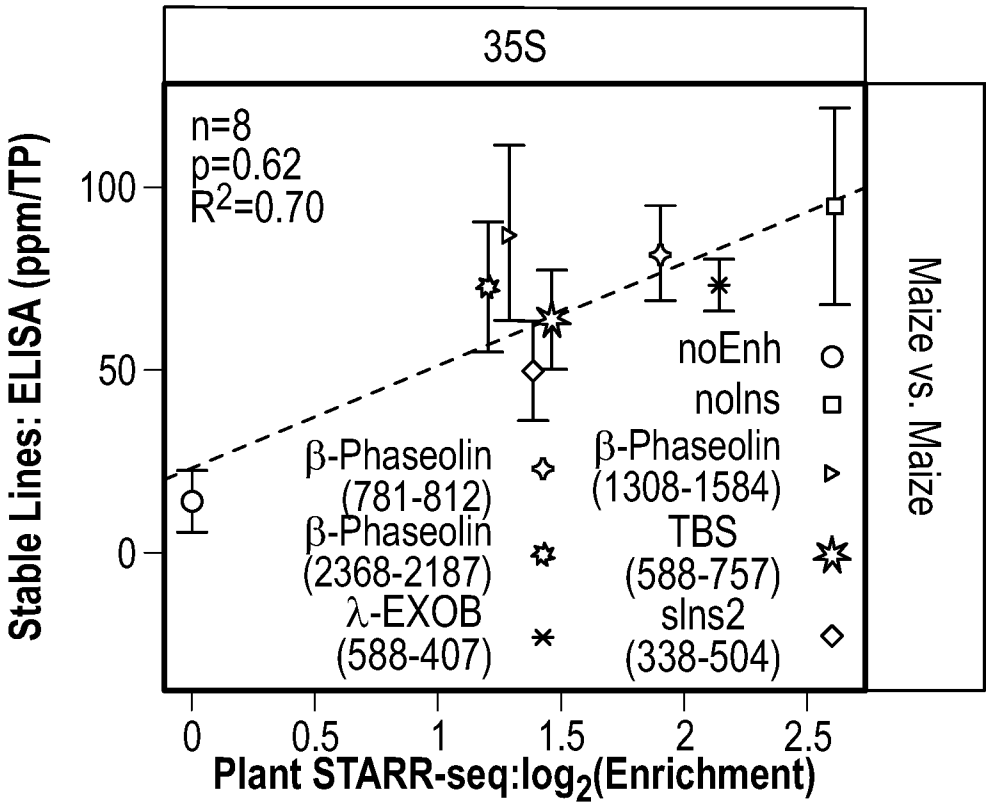

35S
n=8
p=0.62
R2=0.70
Stable Lines: ELISA (ppm/TP)
100
50
0
noEnh
noIns
β-Phaseolin (781-812)
β-Phaseolin (1308-1584)
β-Phaseolin (2368-2187)
TBS (588-757)
λ-EXOB (588-407)
sIns2 (338-504)
Maize vs. Maize
0
0.5
1
1.5
2
2.5
Plant STARR-seq:log2(Enrichment)

FIG. 3J

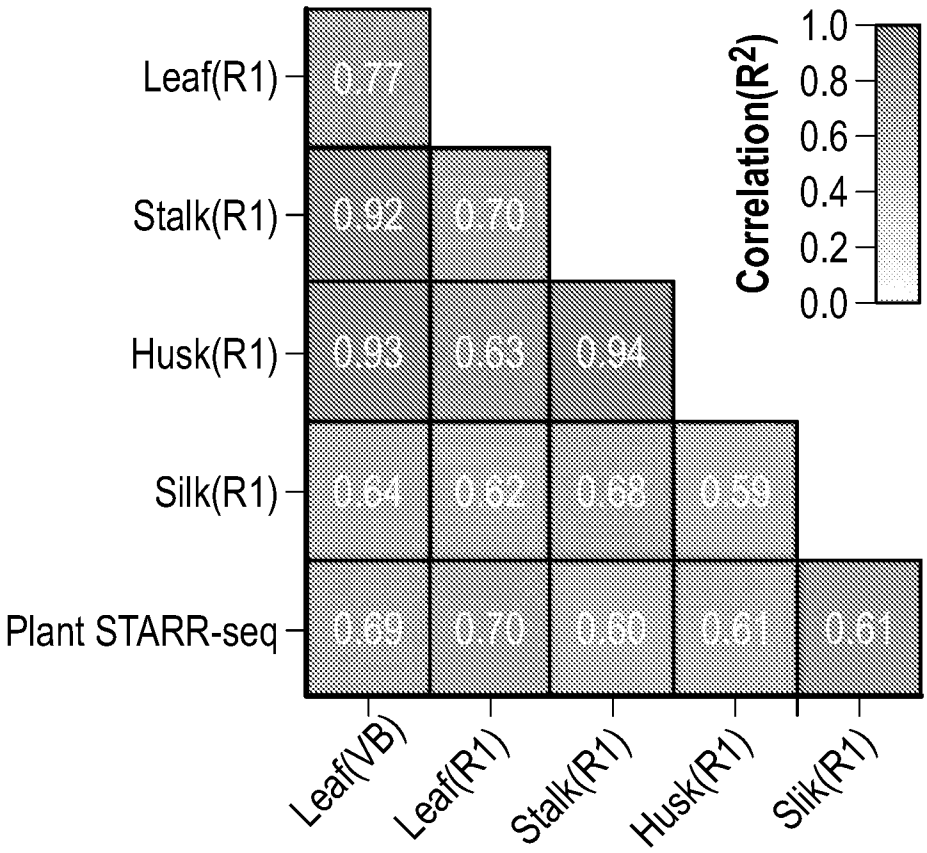

Leaf(R1)
Stalk(R1)
Husk(R1)
Silk(R1)
Plant STARR-seq
0.77
0.92
0.70
0.93
0.63
0.94
0.64
0.62
0.68
0.59
0.69
0.70
0.60
0.61
0.61
Leaf(VB)
Leaf(R1)
Stalk(R1)
Husk(R1)
Slik(R1)
Correlation(R2)
1.0
0.8
0.6
0.4
0.2
0.0

FIG. 3K

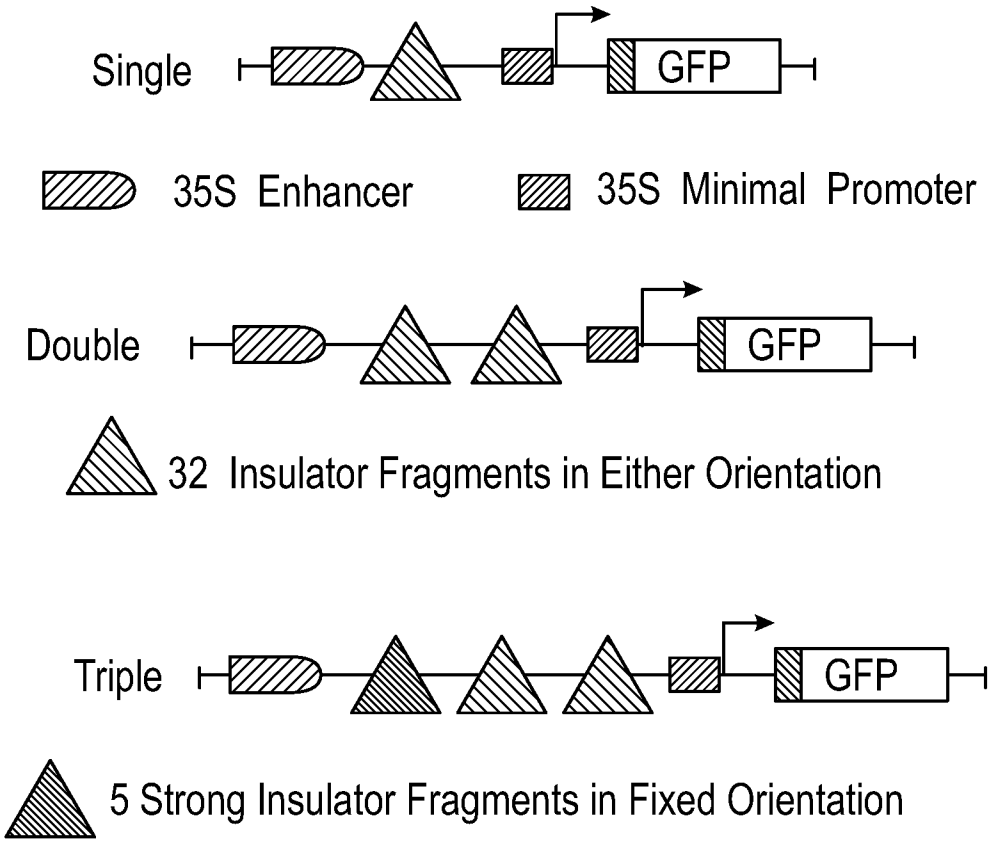

Single
GFP
35S Enhancer
35S Minimal Promoter
Double
GFP
32 Insulator Fragments in Either Orientation
Triple
GFP
5 Strong Insulator Fragments in Fixed Orientation

FIG. 4A

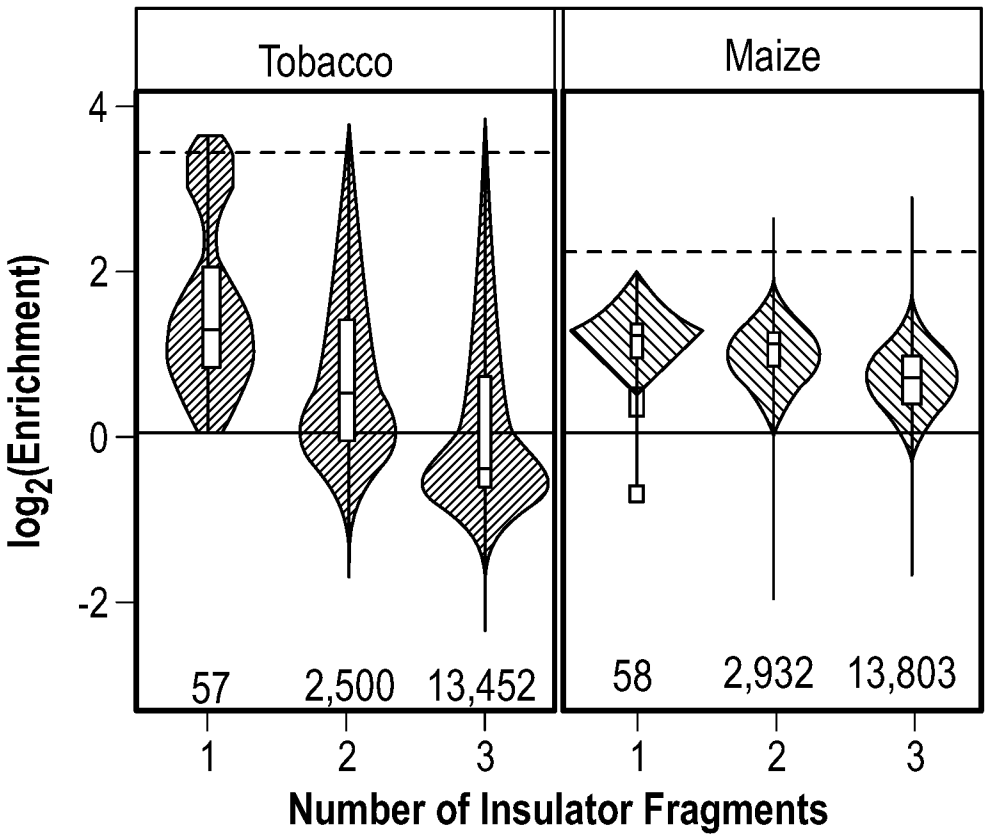

Tobacco
Maize
4
2
0
-2
log2(Enrichment)
57
2,500
13,452
58
2,932
13,803
1
2
3
1
2
3
Number of Insulator Fragments

FIG. 4B

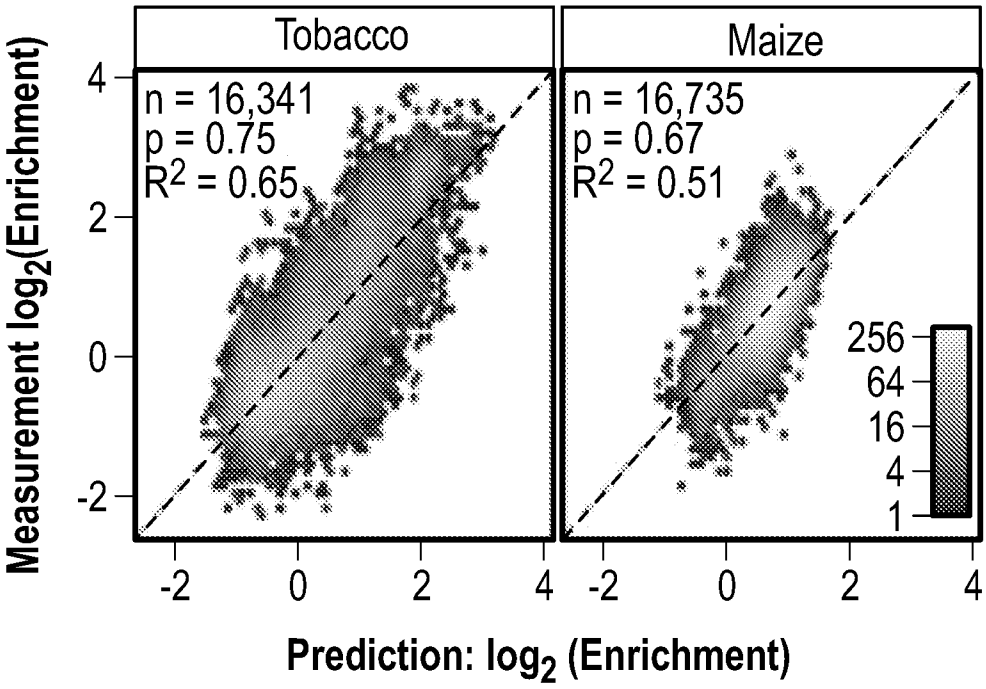

Tobacco
Maize
n = 16,341
p = 0.75
R2 = 0.65
n = 16,735
p = 0.67
R2 = 0.51
256
64
16
4
1
Measurement log2(Enrichment)
Prediction: log2 (Enrichment)

FIG. 4C

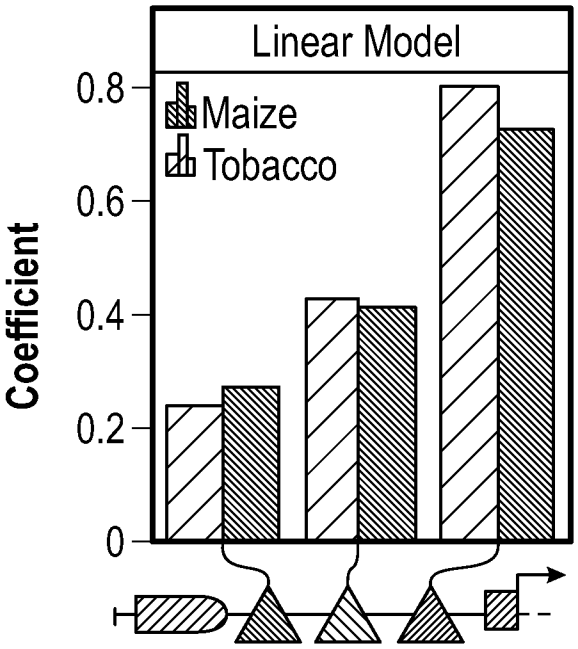

Linear Model
Maize
Tobacco
Coefficient

FIG. 4D

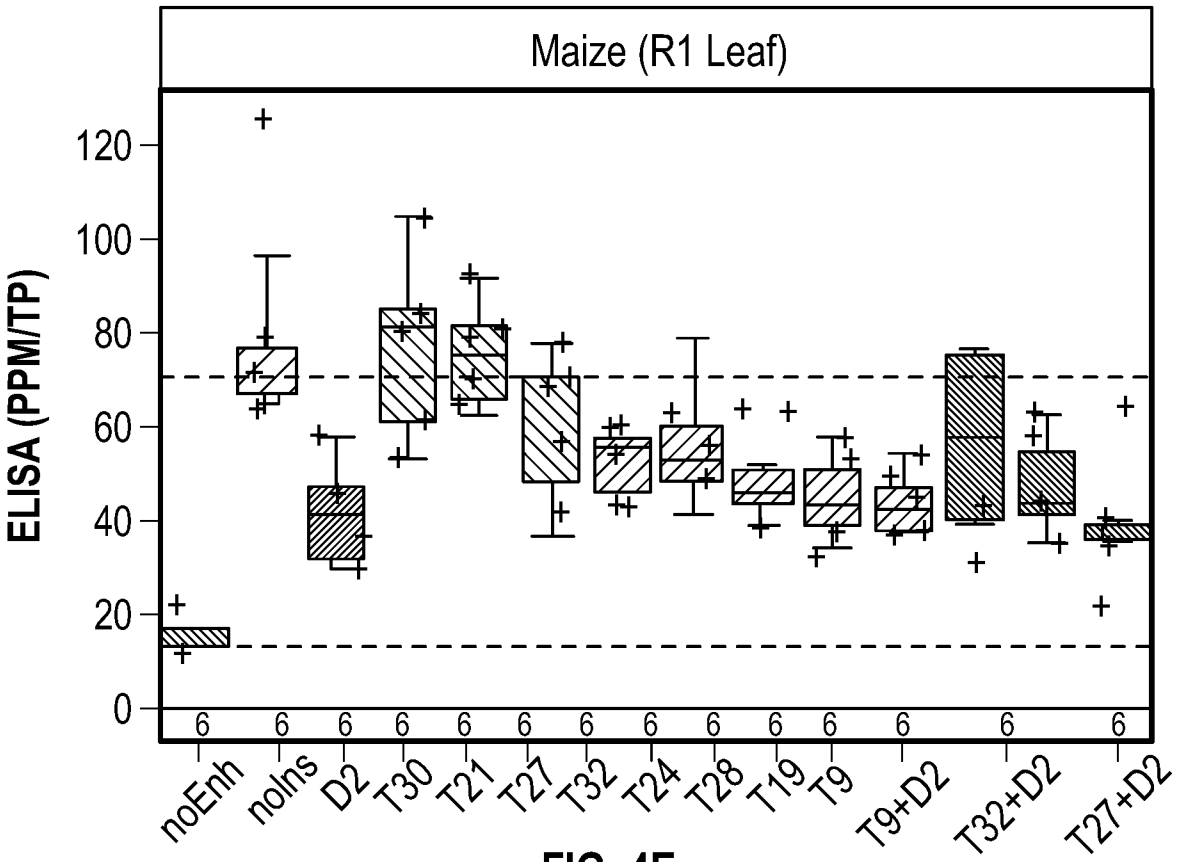

Maize (R1 Leaf)
ELISA (PPM/TP)
120
100
80
60
40
20
0
6
noEnh
noIns
D2
T30
T21
T27
T32
T24
T28
T19
T9
T9+D2
T32+D2
T27+D2

FIG. 4E

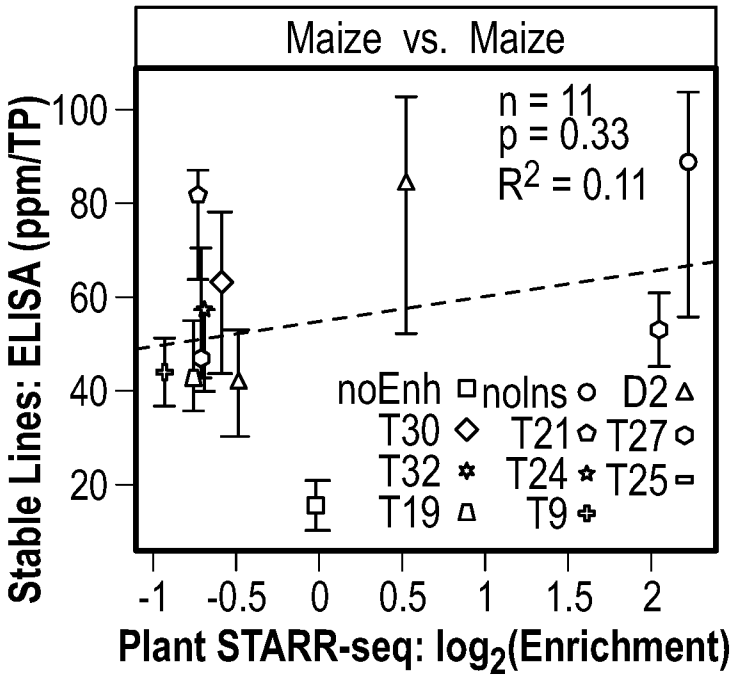

Maize vs. Maize
n = 11
p = 0.33
R2 = 0.11
Stable Lines: ELISA (ppm/TP)
100
80
60
40
20
noEnh
noIns
D2
T30
T21
T27
T32
T24
T25
T19
T9
-1
-0.5
0
0.5
1
1.5
2
Plant STARR-seq: log2(Enrichment)

FIG. 4F

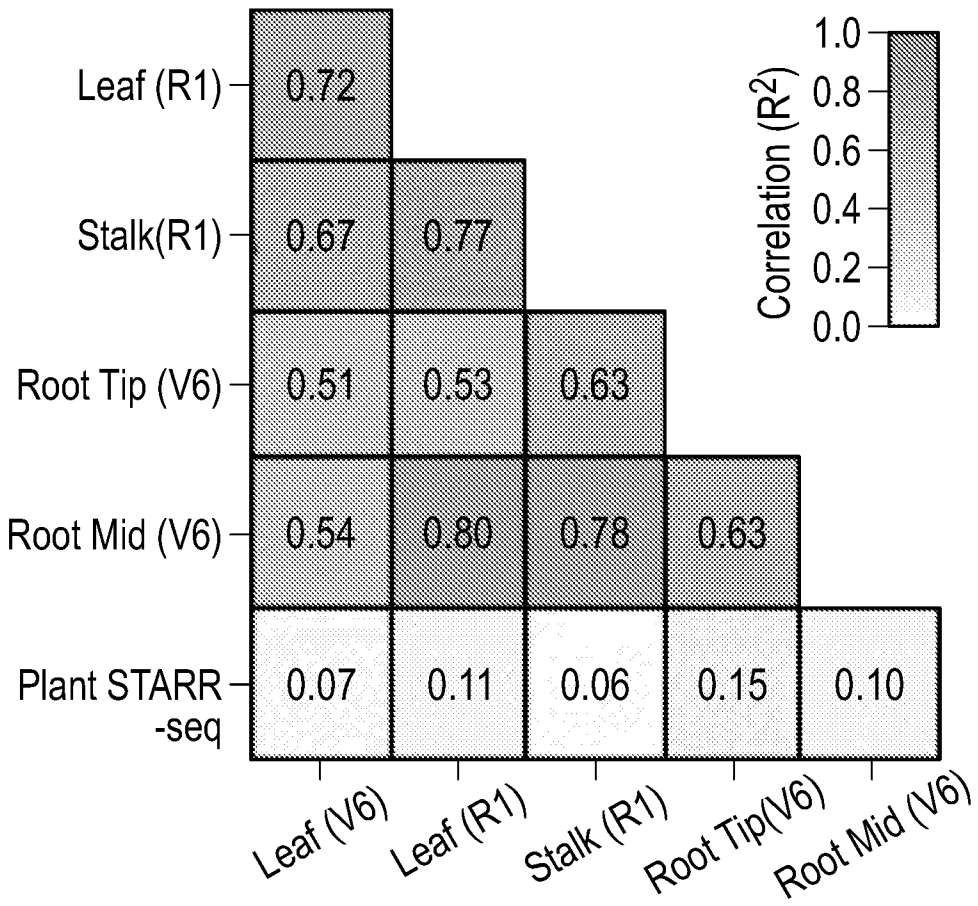

Leaf (R1)
Stalk(R1)
Root Tip (V6)
Root Mid (V6)
Plant STARR -seq
0.72
0.67
0.77
0.51
0.53
0.63
0.54
0.80
0.78
0.63
0.07
0.11
0.06
0.15
0.10
Leaf (V6)
Leaf (R1)
Stalk (R1)
Root Tip(V6)
Root Mid (V6)
Correlation (R^2)
1.0
0.8
0.6
0.4
0.2
0.0

FIG. 4G

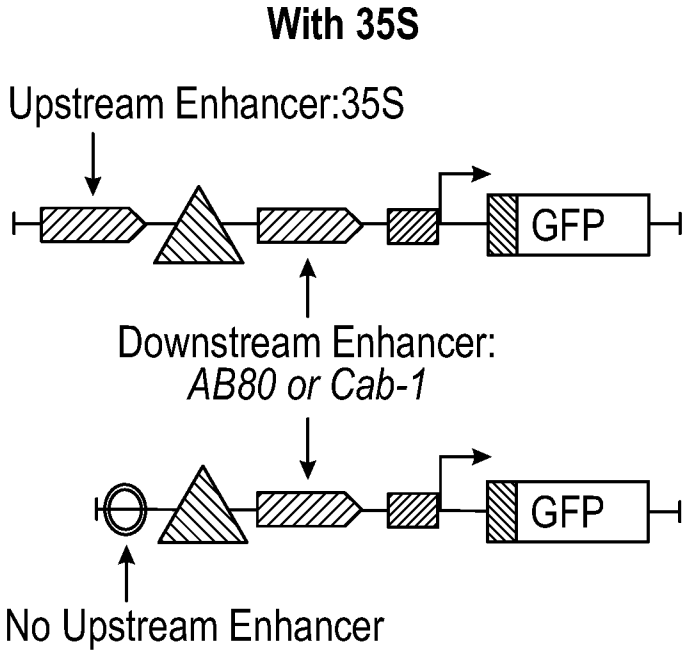

With 35S
Upstream Enhancer:35S
GFP
Downstream Enhancer:
AB80 or Cab-1
GFP
No Upstream Enhancer
Without 35S

FIG. 5A

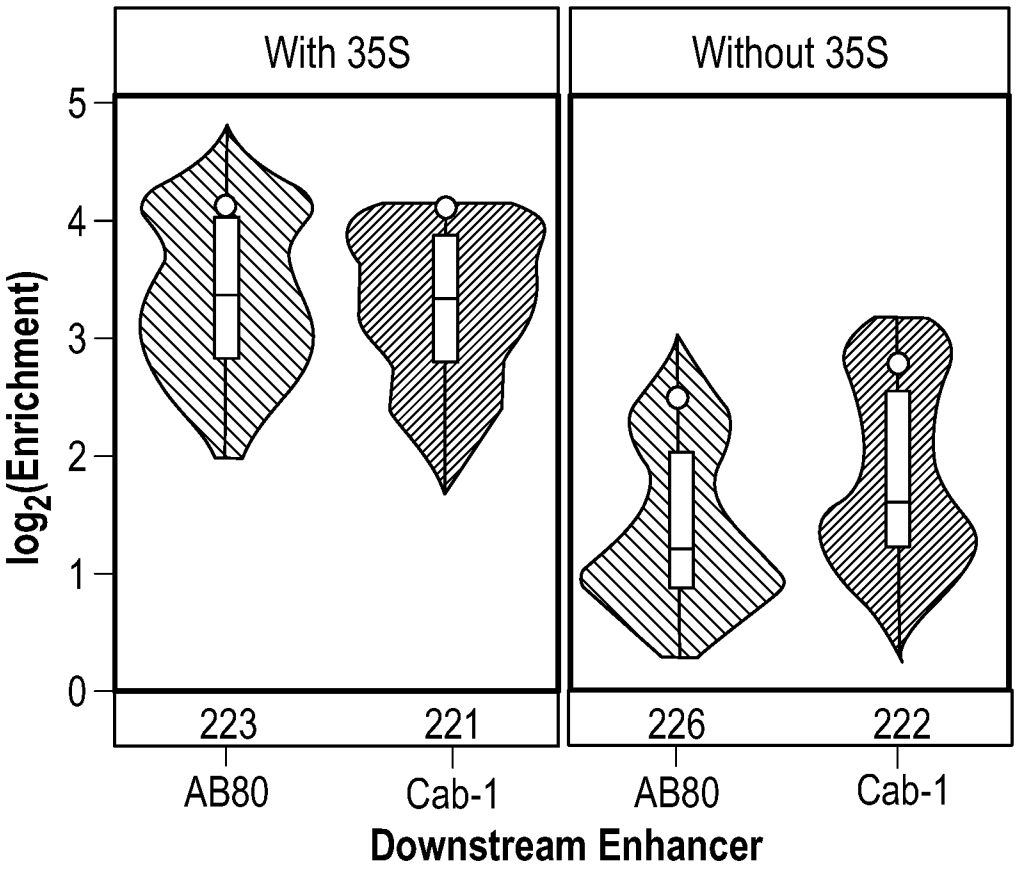

With 35S
Without 35S
5
4
3
2
1
0
log2(Enrichment)
223
221
226
222
AB80
Cab-1
AB80
Cab-1
Downstream Enhancer

FIG. 5B

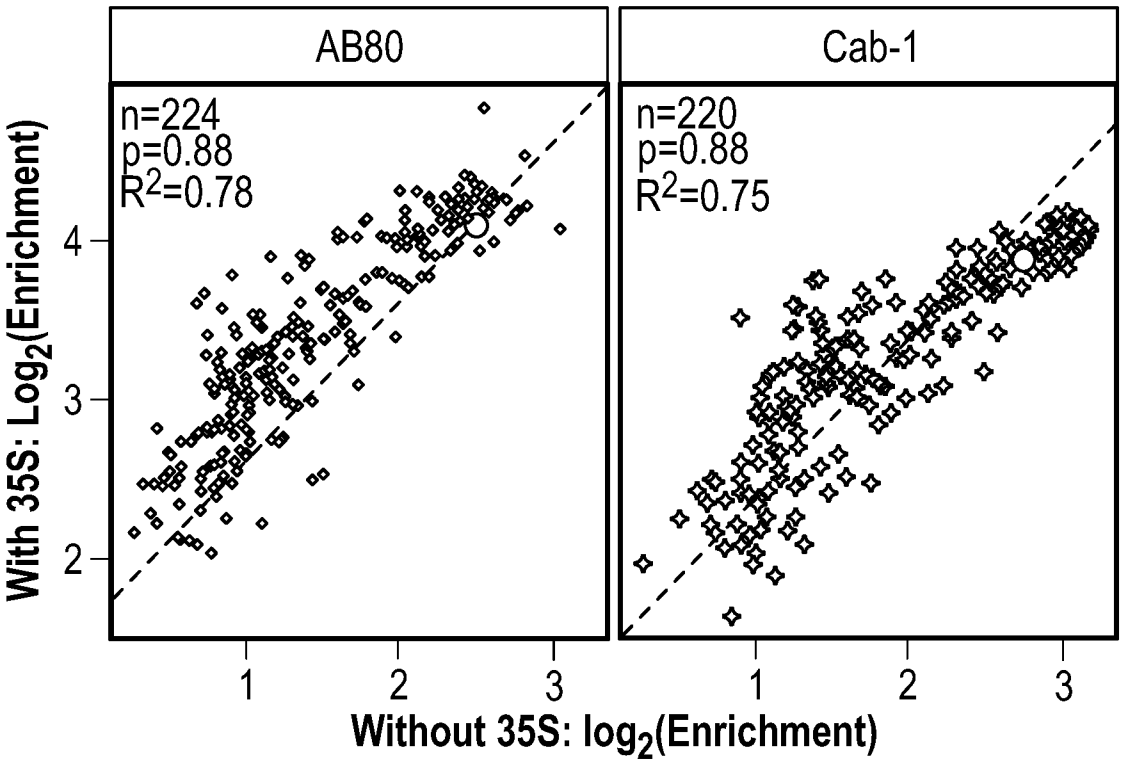

AB80
Cab-1
n=224
p=0.88
R2=0.78
n=220
p=0.88
R2=0.75
With 35S: Log2(Enrichment)
Without 35S: log2(Enrichment)
1
2
3
4

FIG. 5C

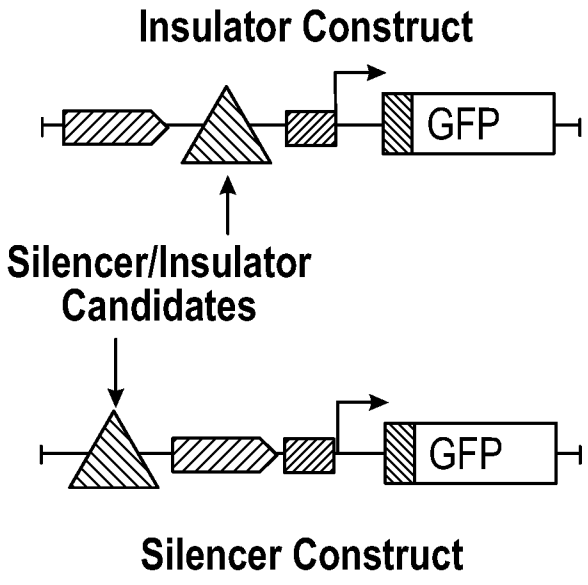

Insulator Construct
GFP
Silencer/Insulator
Candidates
GFP
Silencer Construct

FIG. 5D

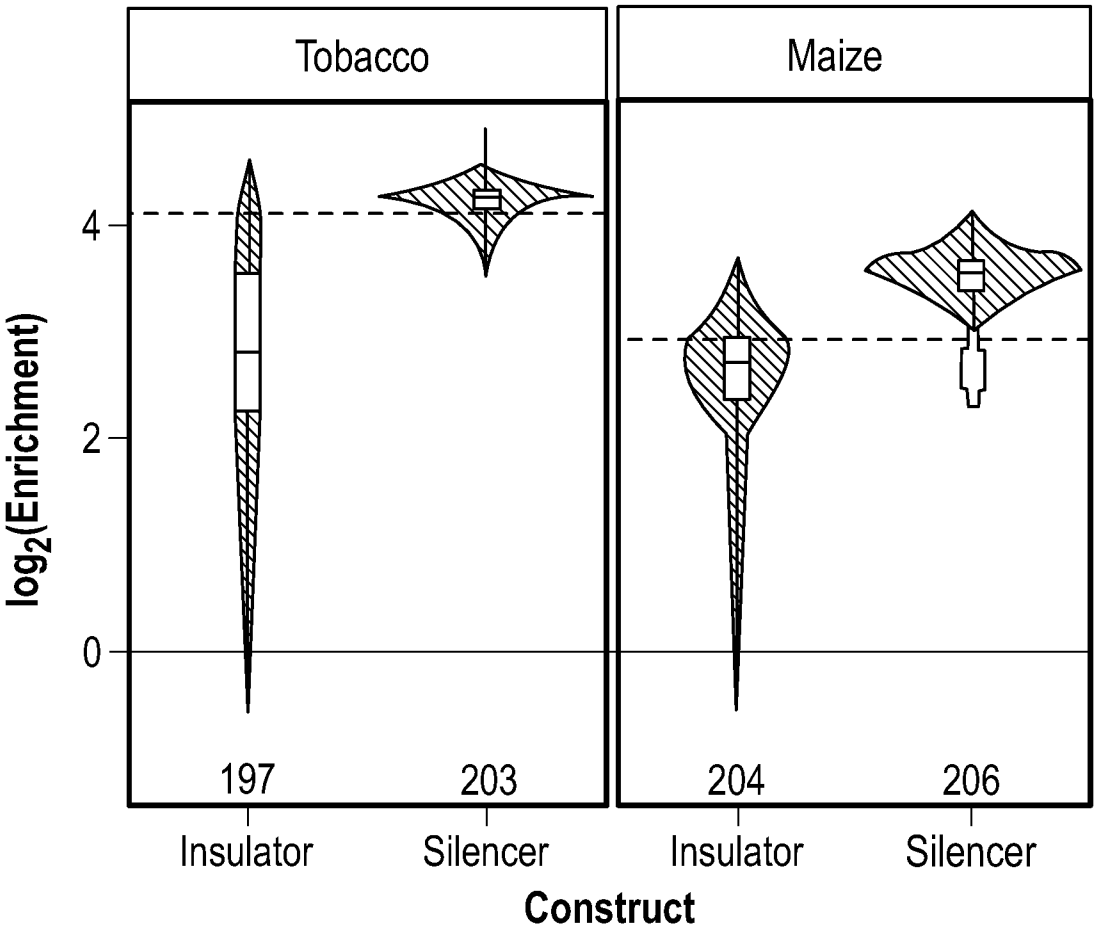

Tobacco
Maize
log2(Enrichment)
4
2
0
197
203
204
206
Insulator
Silencer
Insulator
Silencer
Construct

FIG. 5E

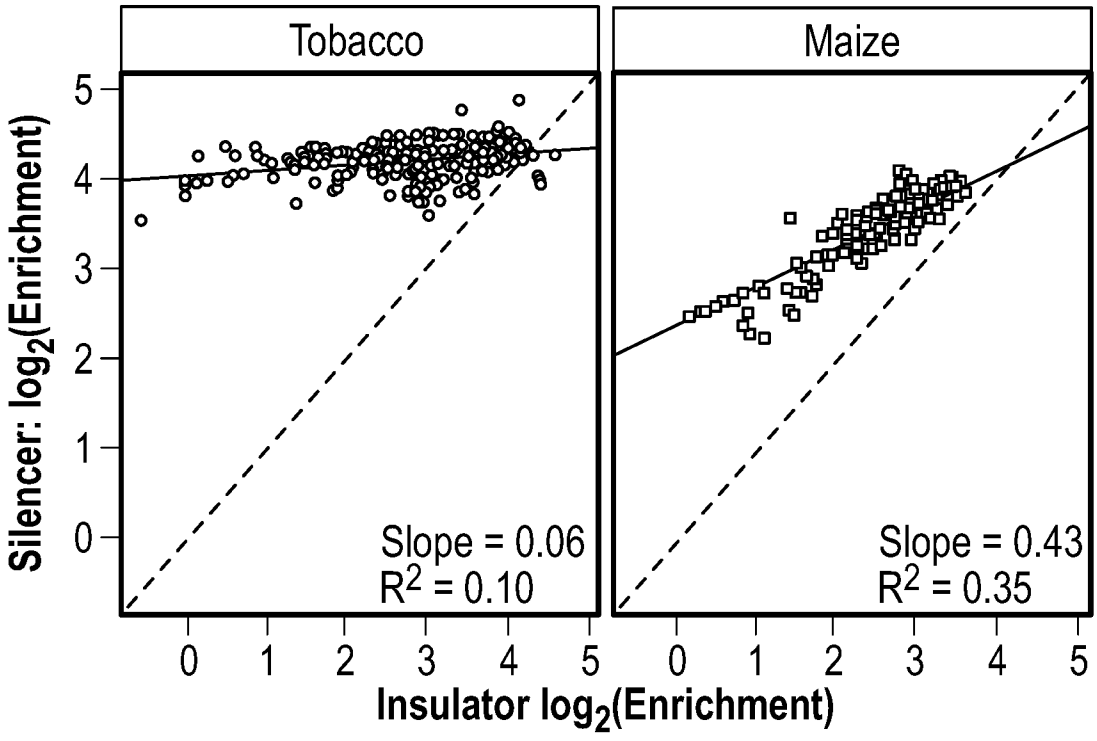

Tobacco
Maize
Silencer: log2(Enrichment)
5
4
3
2
1
0
Slope = 0.06
R2 = 0.10
Slope = 0.43
R2 = 0.35
0
1
2
3
4
5
Insulator log2(Enrichment)

FIG. 5F

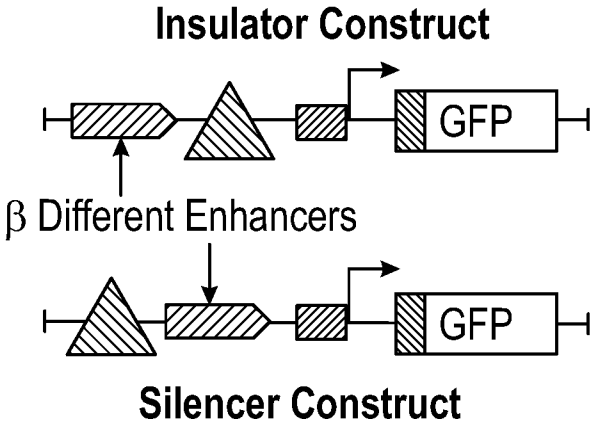

Insulator Construct
GFP
β Different Enhancers
GFP
Silencer Construct

FIG. 6A

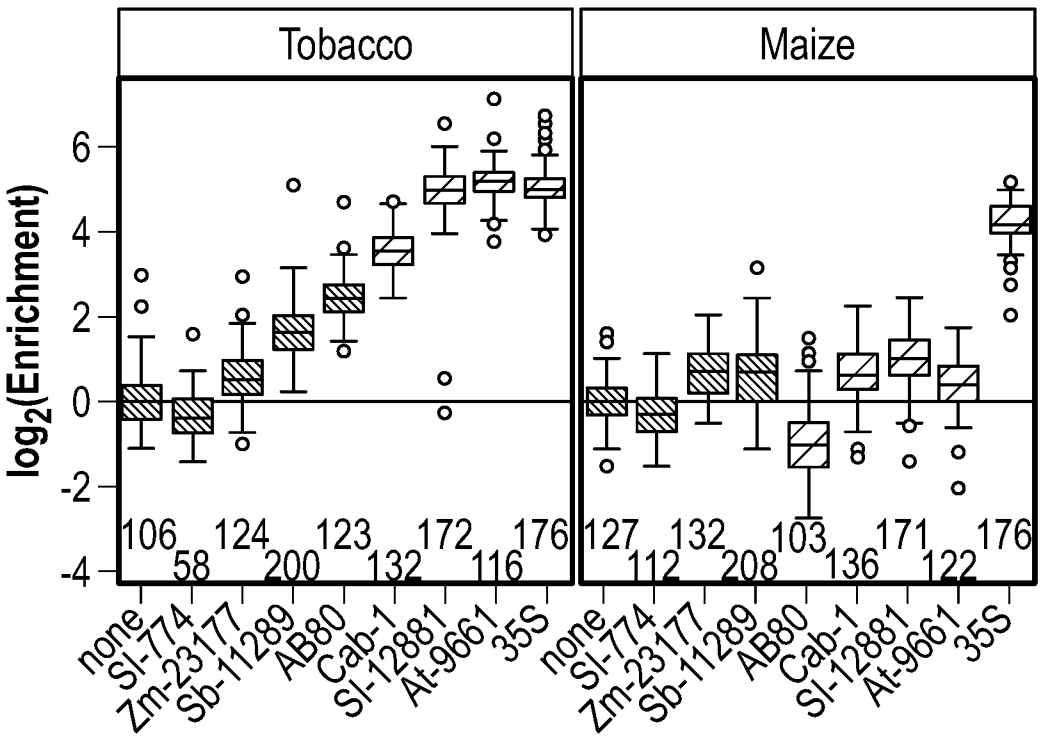

Tobacco
Maize
log2(Enrichment)
6
4
2
0
-2
-4
106
58
124
200
123
132
172
116
176
127
112
132
208
103
136
171
122
176
none
Sl-774
Zm-23177
Sb-11289
AB80
Cab-1
Sl-12881
At-9661
35S

FIG. 6B

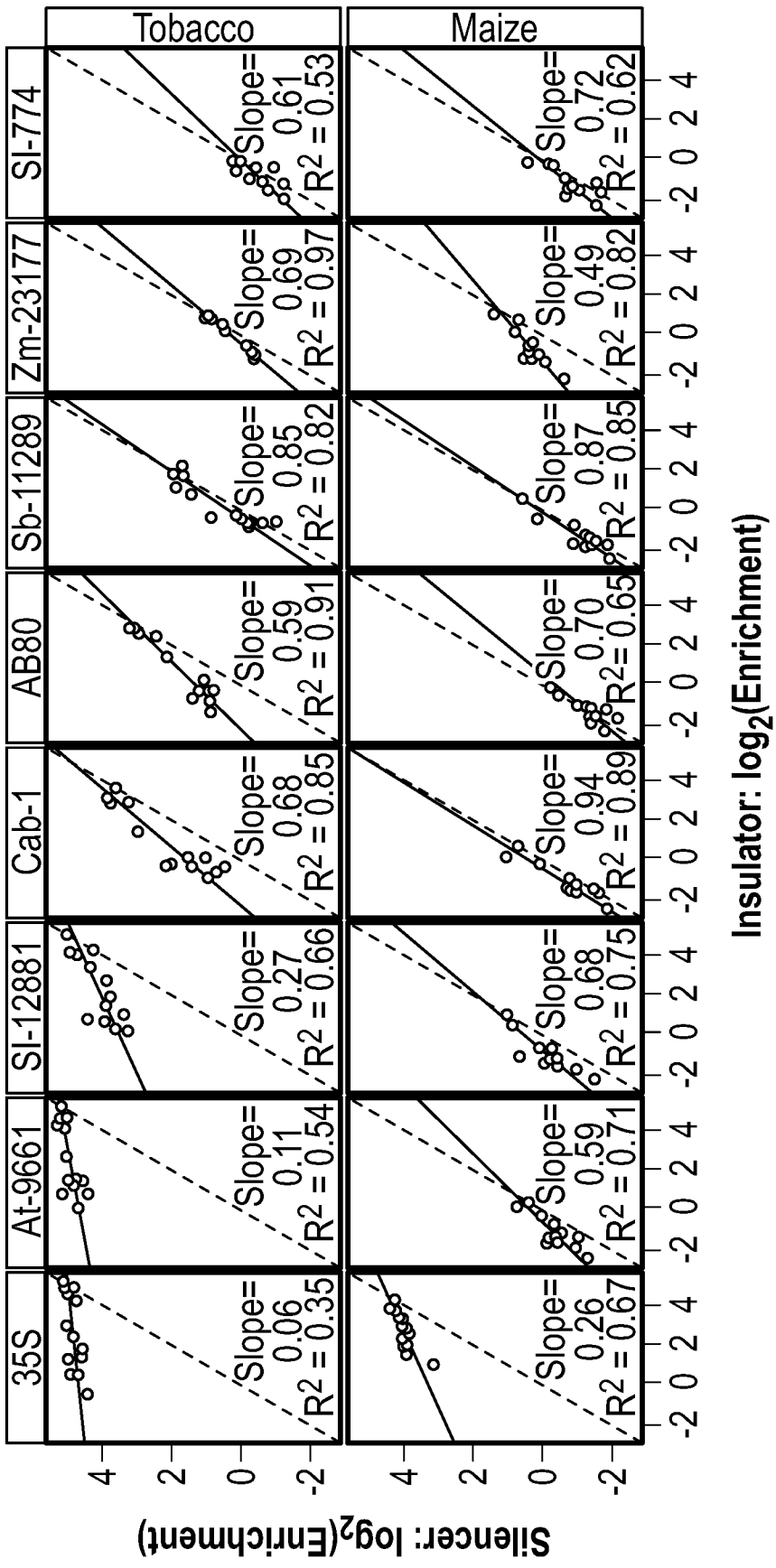

Tobacco
Maize
35S
At-9661
SI-12881
Cab-1
AB80
Sb-11289
Zm-23177
SI-774
Slope= 0.06
R2 = 0.35
Slope= 0.11
R2 = 0.54
Slope= 0.27
R2 = 0.66
Slope= 0.68
R2 = 0.85
Slope= 0.59
R2 = 0.91
Slope= 0.85
R2 = 0.82
Slope= 0.69
R2 = 0.97
Slope= 0.61
R2 = 0.53
Slope= 0.26
R2 = 0.67
Slope= 0.59
R2 = 0.71
Slope= 0.68
R2 = 0.75
Slope= 0.94
R2 = 0.89
Slope= 0.70
R2 = 0.65
Slope= 0.87
R2 = 0.85
Slope= 0.49
R2 = 0.82
Slope= 0.72
R2 = 0.62
-2 0 2 4
4
2
0
-2
Silencer: log2(Enrichment)
Insulator: log2(Enrichment)

FIG. 6C

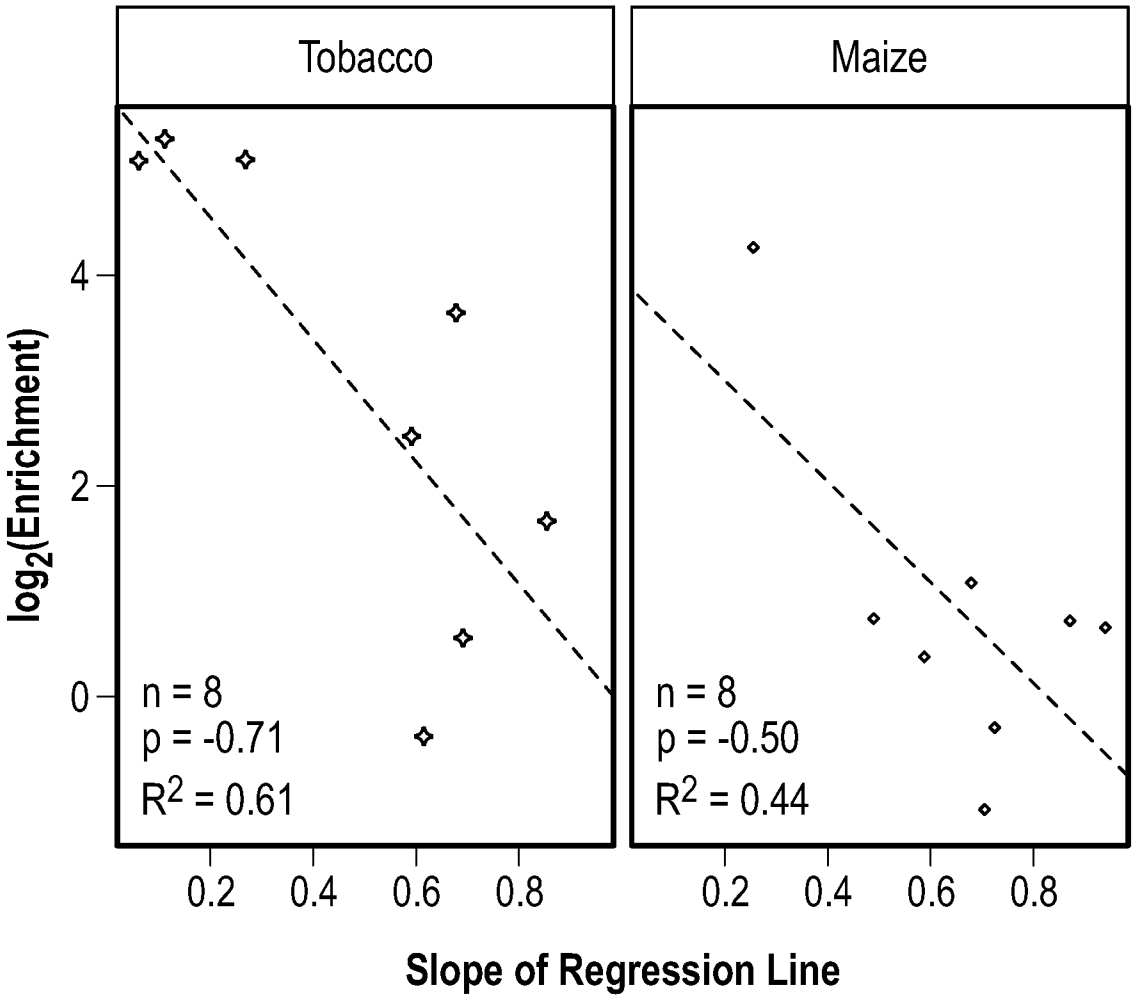

Tobacco
Maize
log2(Enrichment)
4
2
0
n = 8
p = -0.71
R2 = 0.61
n = 8
p = -0.50
R2 = 0.44
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
Slope of Regression Line

FIG. 6D

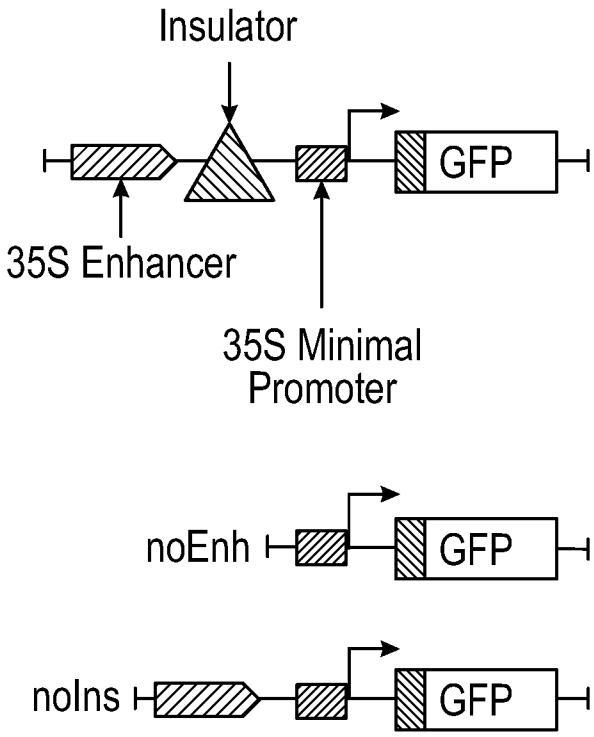

Insulator
GFP
35S Enhancer
35S Minimal Promoter
noEnh
GFP
noIns
GFP

FIG. 7A

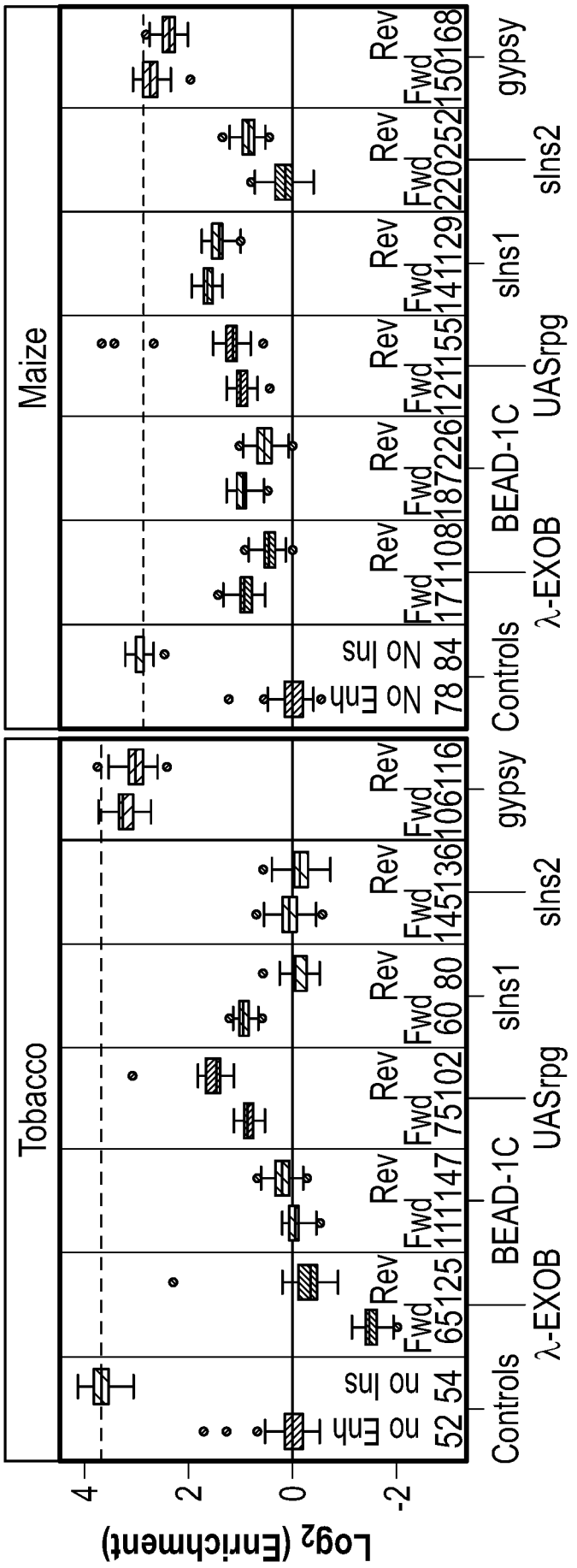

Log2 (Enrichment)
4
2
0
-2
Tobacco
Maize
52 54 65 125 111 147 75 102 60 80 145 136 106 116 78 84 171 108 187 226 121 155 141 129 220 252 150 168
no Enh
no Ins
Rev
Fwd
Controls
λ-EXOB
BEAD-1C
UASrpg
slns1
slns2
gypsy
Insulator

FIG. 7B

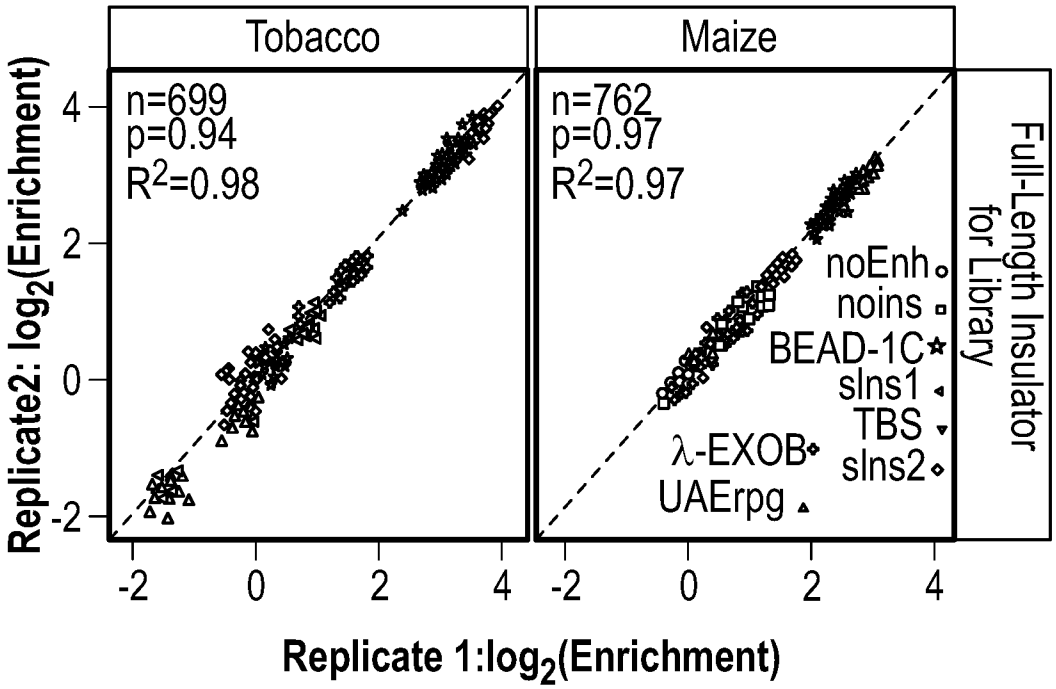

Tobacco
Maize
n=699
p=0.94
R2=0.98
n=762
p=0.97
R2=0.97
noEnh
noins
BEAD-1C
sIns1
TBS
sIns2
λ-EXOB
UAErpg
Full-Length Insulator for Library
Replicate2: log2(Enrichment)
Replicate 1:log2(Enrichment)

FIG. 8A

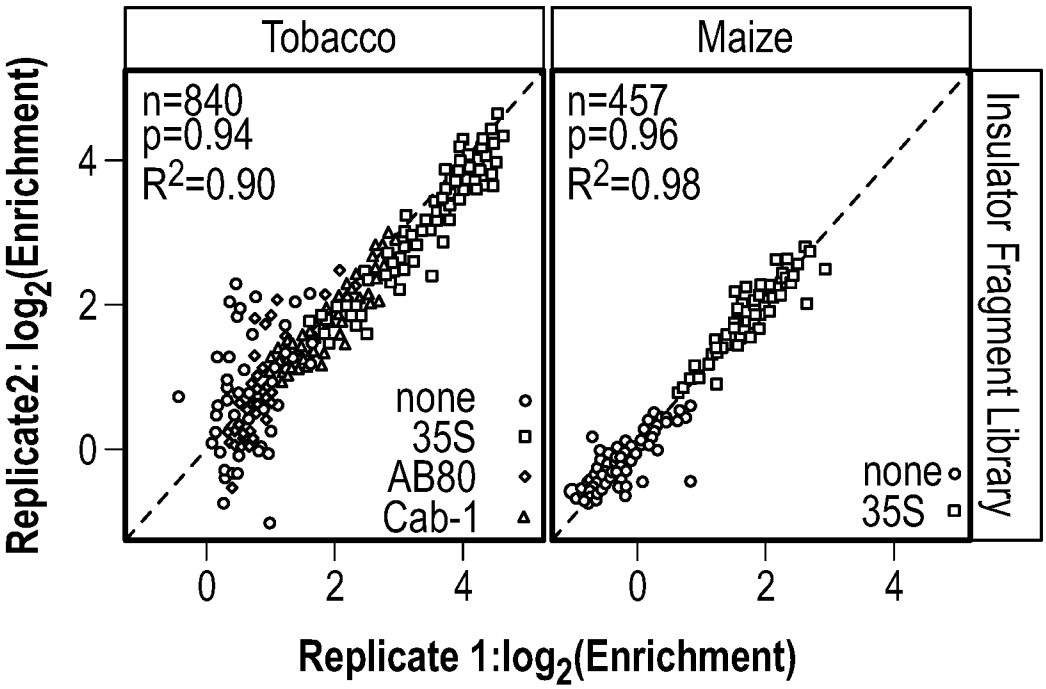

Tobacco
Maize
n=840
p=0.94
R2=0.90
n=457
p=0.96
R2=0.98
none
35S
AB80
Cab-1
Insulator Fragment Library
Replicate2: log2(Enrichment)
Replicate 1:log2(Enrichment)

FIG. 8B

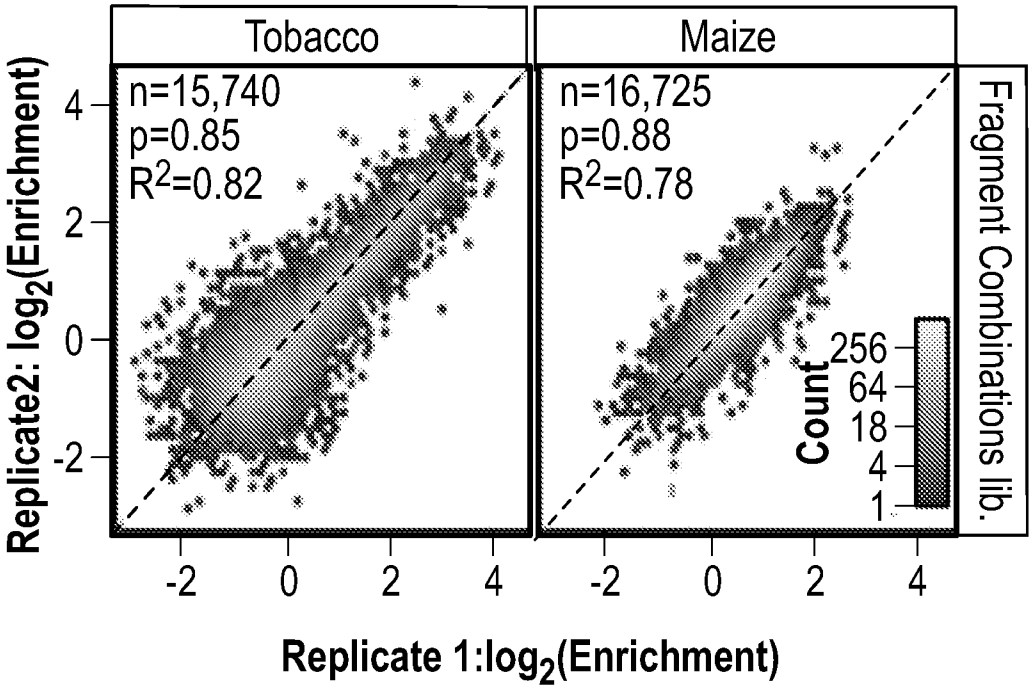

Tobacco
Maize
n=15,740
p=0.85
R2=0.82
n=16,725
p=0.88
R2=0.78
Fragment Combinations lib.
Count
256
64
18
4
1
Replicate2: log2(Enrichment)
Replicate 1:log2(Enrichment)

FIG. 8C

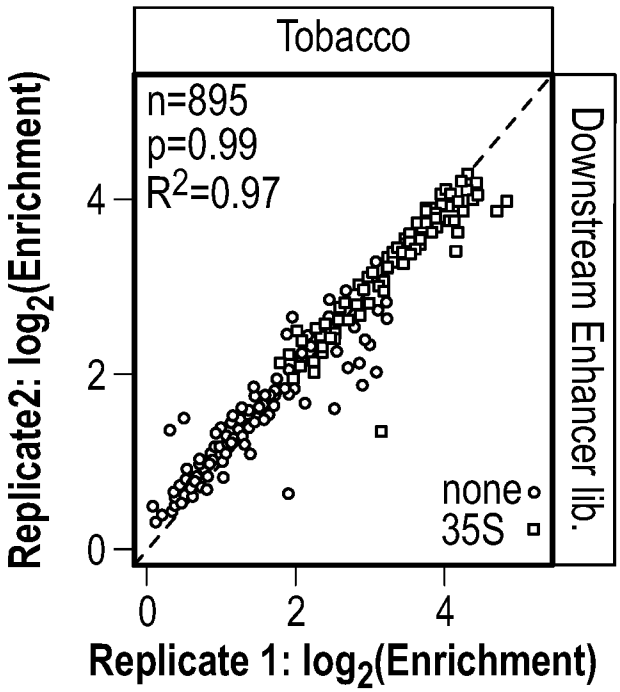

Tobacco
n=895
p=0.99
R2=0.97
Downstream Enhancer lib.
none
35S
Replicate2: log2(Enrichment)
Replicate 1: log2(Enrichment)

FIG. 8D

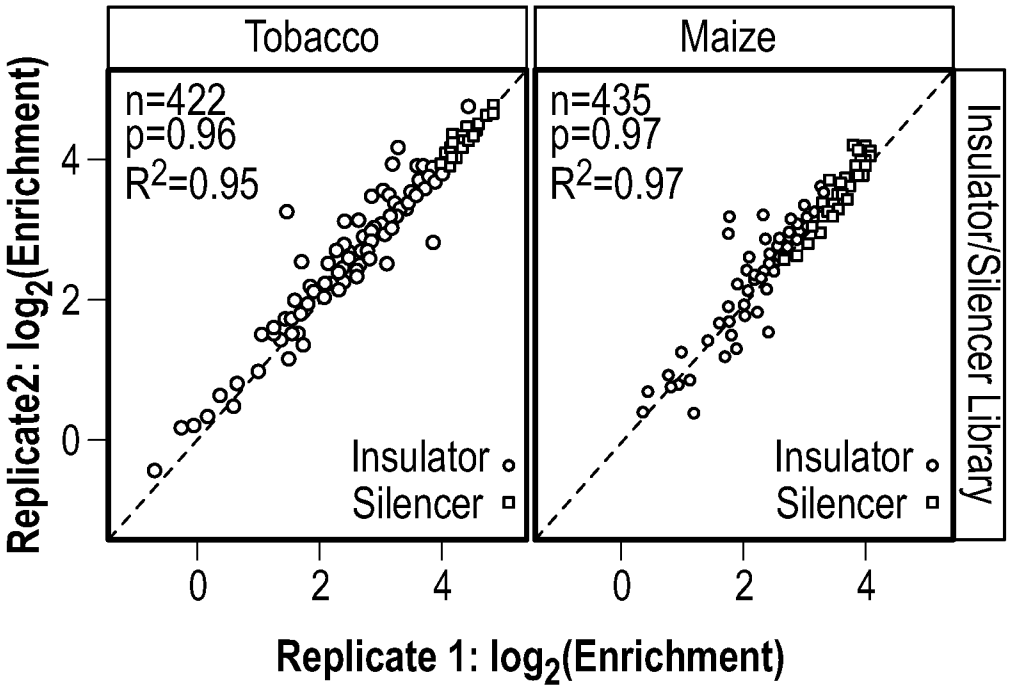

Tobacco
Maize
n=422
p=0.96
R2=0.95
n=435
p=0.97
R2=0.97
Insulator
Silencer
Insulator/Silencer Library
Replicate2: log2(Enrichment)
Replicate 1: log2(Enrichment)

FIG. 8E

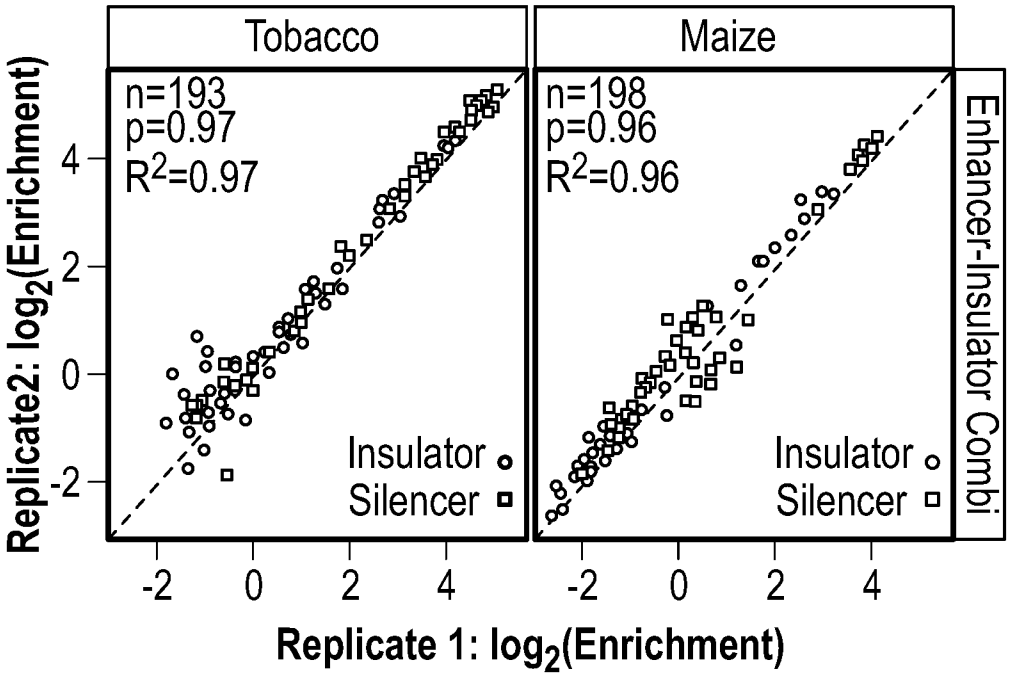

Tobacco
Maize
n=193
p=0.97
R2=0.97
n=198
p=0.96
R2=0.96
Insulator
Silencer
Enhancer-Insulator Combi
Replicate2: log2(Enrichment)
Replicate 1: log2(Enrichment)

FIG. 8F

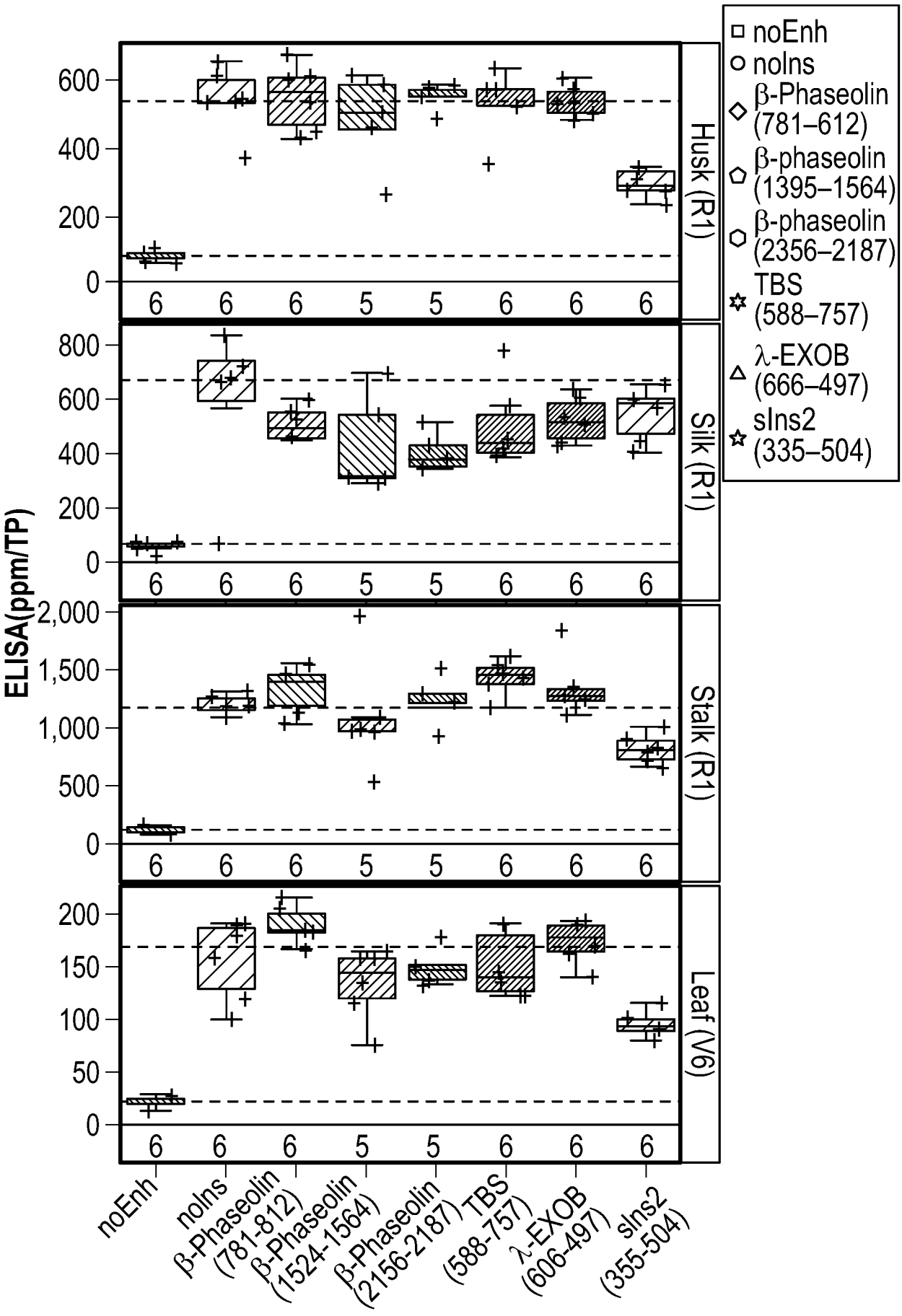

ELISA(ppm/TP)
Husk (R1)
Silk (R1)
Stalk (R1)
Leaf (V6)
noEnh
noIns
β-Phaseolin (781–612)
β-phaseolin (1395–1564)
β-phaseolin (2356–2187)
TBS (588–757)
λ-EXOB (666–497)
sIns2 (335–504)
noEnh
noIns
β-Phaseolin (781-812)
β-Phaseolin (1524-1564)
β-Phaseolin (2156-2187)
TBS (588-757)
λ-EXOB (606-497)
sIns2 (355-504)

FIG. 9A

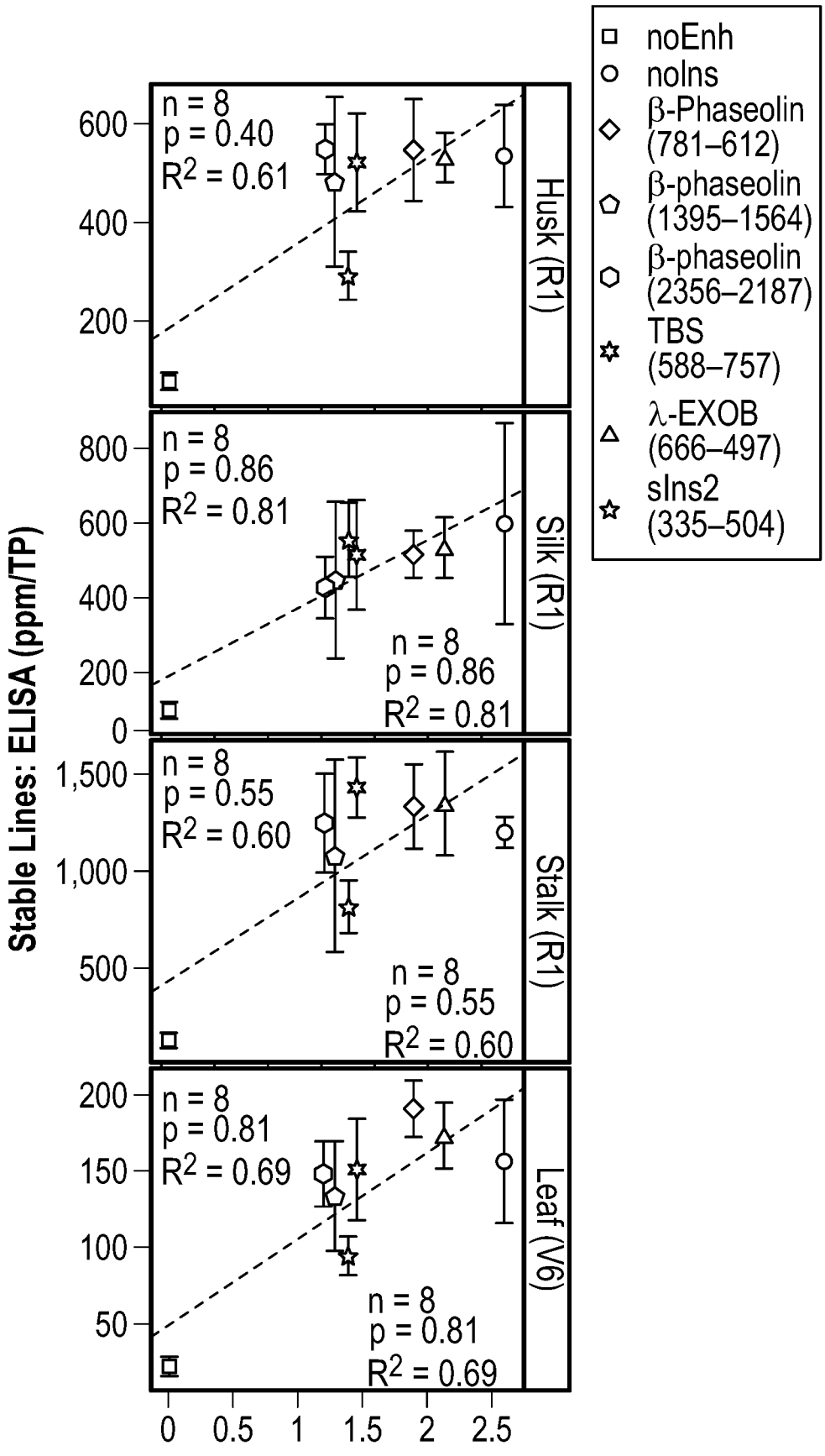

noEnh
noIns
β-Phaseolin (781–612)
β-phaseolin (1395–1564)
β-phaseolin (2356–2187)
TBS (588–757)
λ-EXOB (666–497)
sIns2 (335–504)
Husk (R1)
Silk (R1)
Stalk (R1)
Leaf (V6)
n = 8
p = 0.40
R2 = 0.61
n = 8
p = 0.86
R2 = 0.81
n = 8
p = 0.55
R2 = 0.60
n = 8
p = 0.81
R2 = 0.69
Stable Lines: ELISA (ppm/TP)
Plant STARR-seq: log2(Enrichment)

FIG. 9B

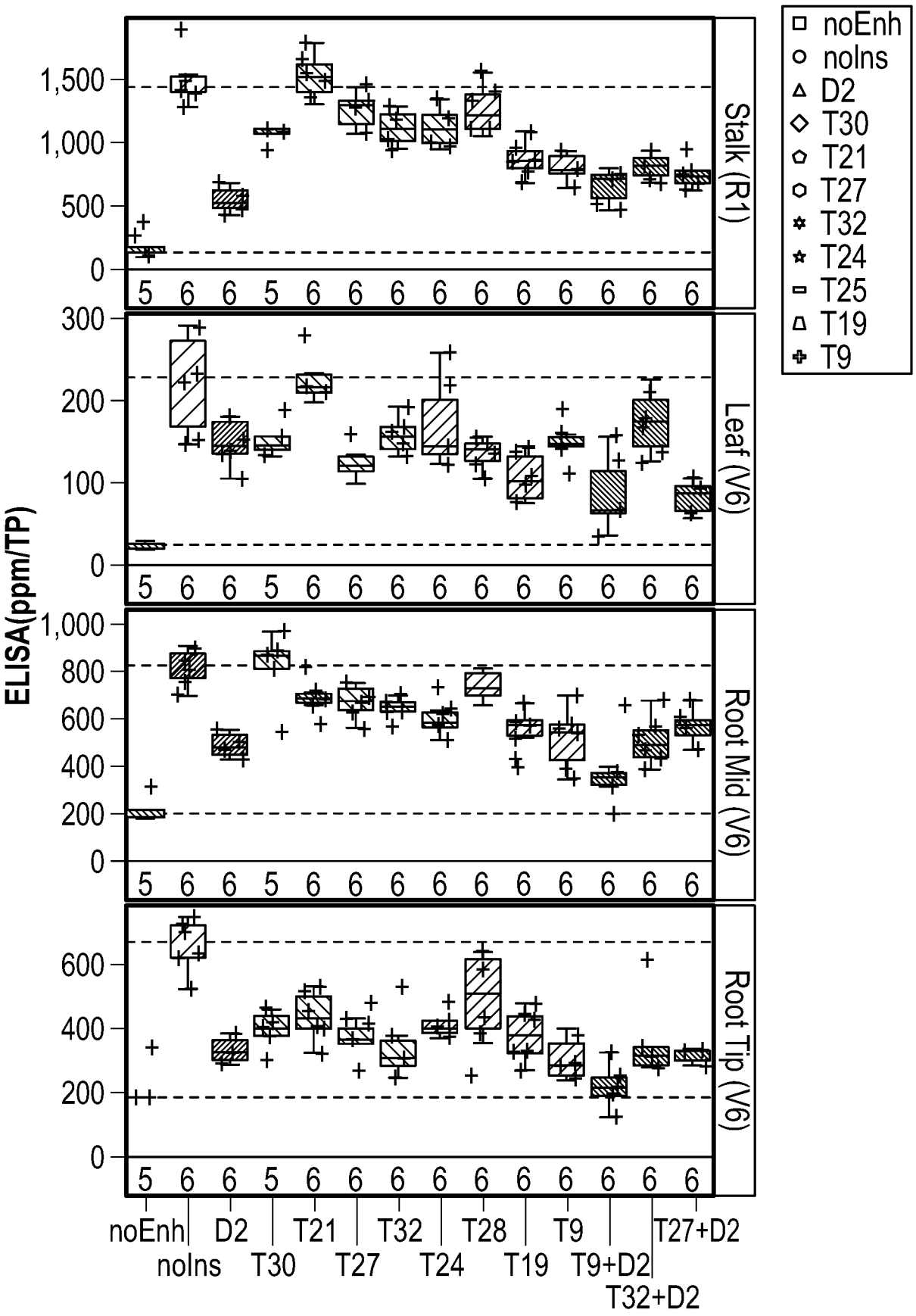

ELISA(ppm/TP)
Stalk (R1)
Leaf (V6)
Root Mid (V6)
Root Tip (V6)
noEnh
noIns
D2
T30
T21
T27
T32
T24
T25
T19
T9
noEnh
noIns
D2
T30
T21
T27
T32
T24
T28
T19
T9
T9+D2
T32+D2
T27+D2

FIG. 10A

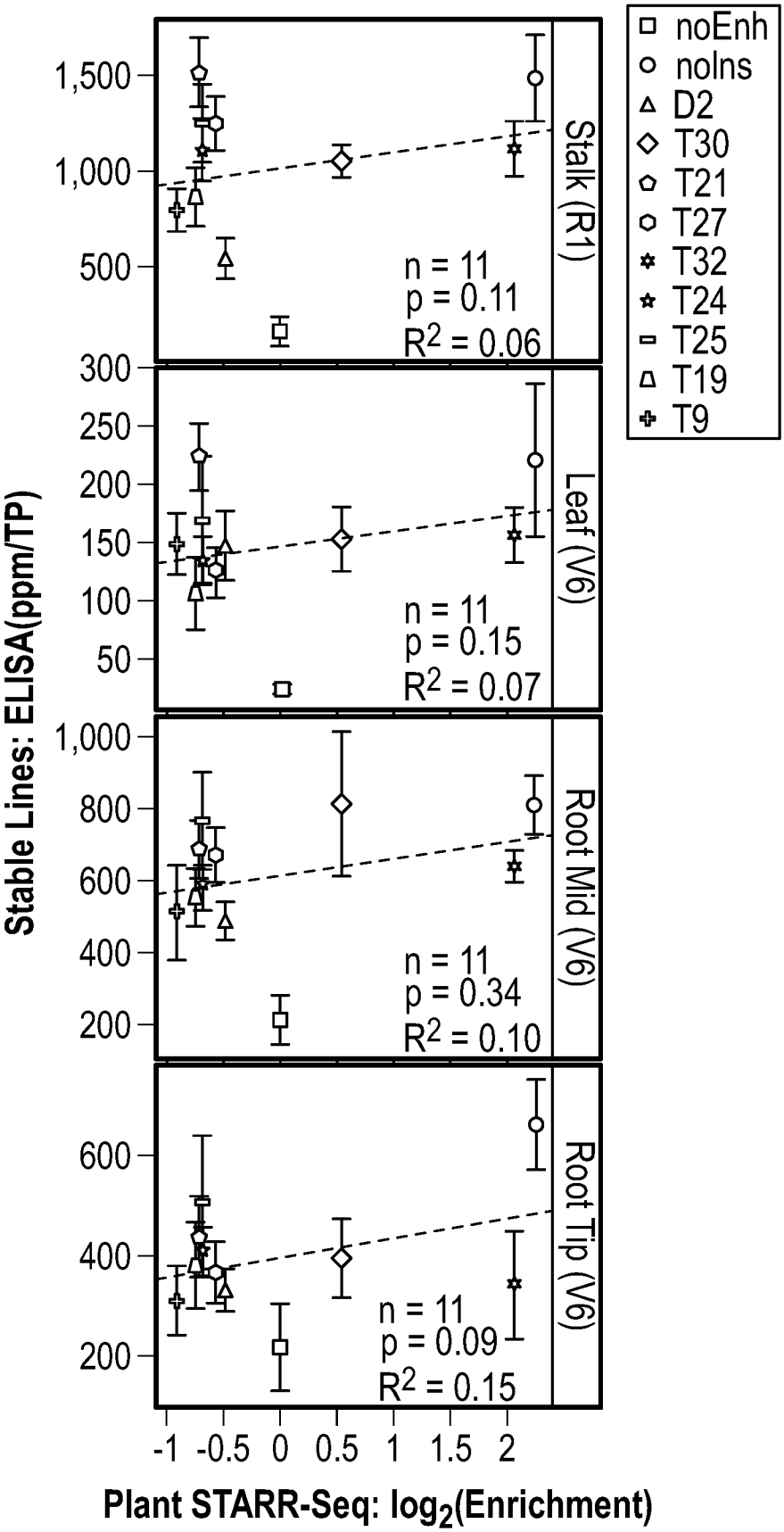

noEnh
noIns
D2
T30
T21
T27
T32
T24
T25
T19
T9
Stalk (R1)
n = 11
p = 0.11
R2 = 0.06
Leaf (V6)
n = 11
p = 0.15
R2 = 0.07
Root Mid (V6)
n = 11
p = 0.34
R2 = 0.10
Root Tip (V6)
n = 11
p = 0.09
R2 = 0.15
Stable Lines: ELISA(ppm/TP)
Plant STARR-Seq: log2(Enrichment)

FIG. 10B

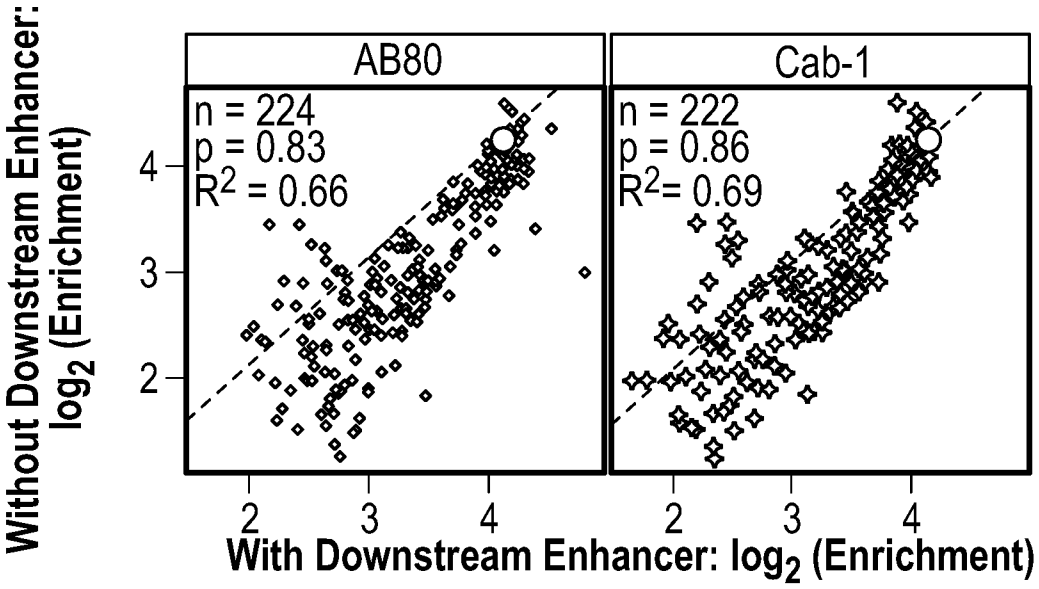

AB80
Cab-1
n = 224
p = 0.83
R2 = 0.66
n = 222
p = 0.86
R2= 0.69
Without Downstream Enhancer: log2 (Enrichment)
With Downstream Enhancer: log2 (Enrichment)
2
3
4

FIG. 11

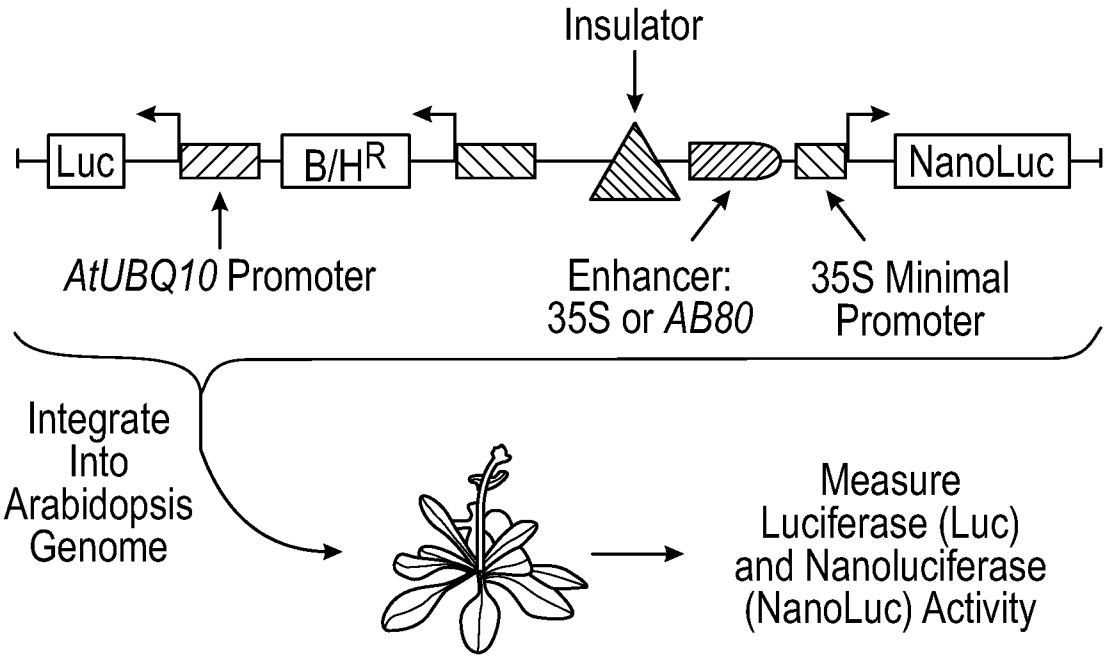


FIG. 12A

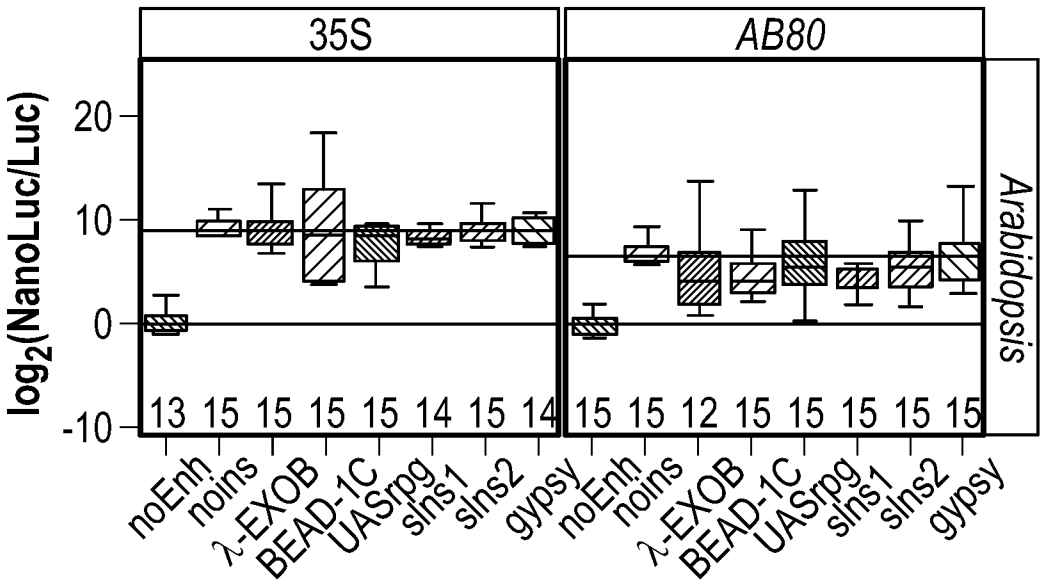


FIG. 12B

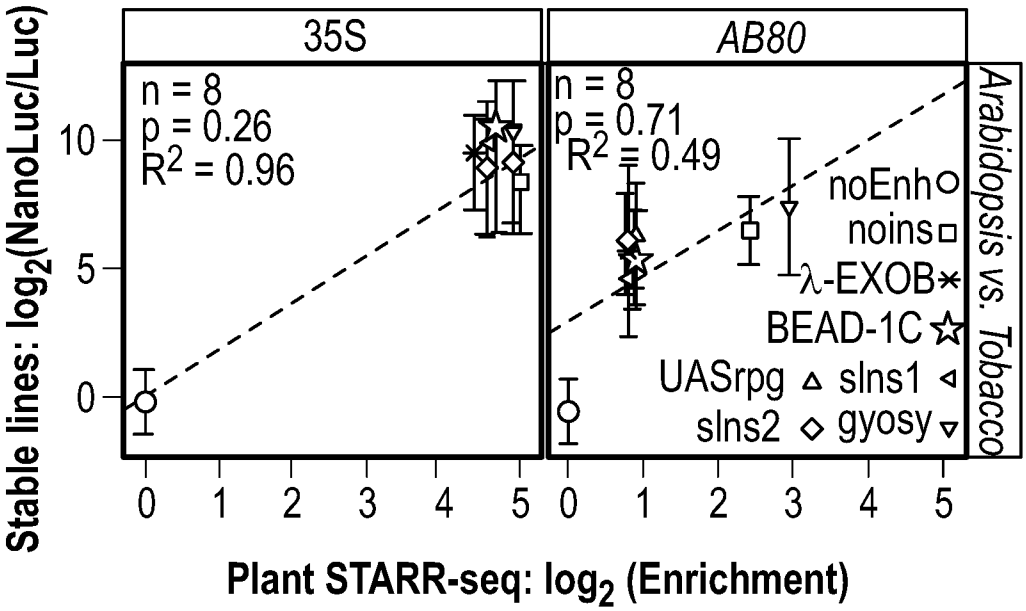


FIG. 12C

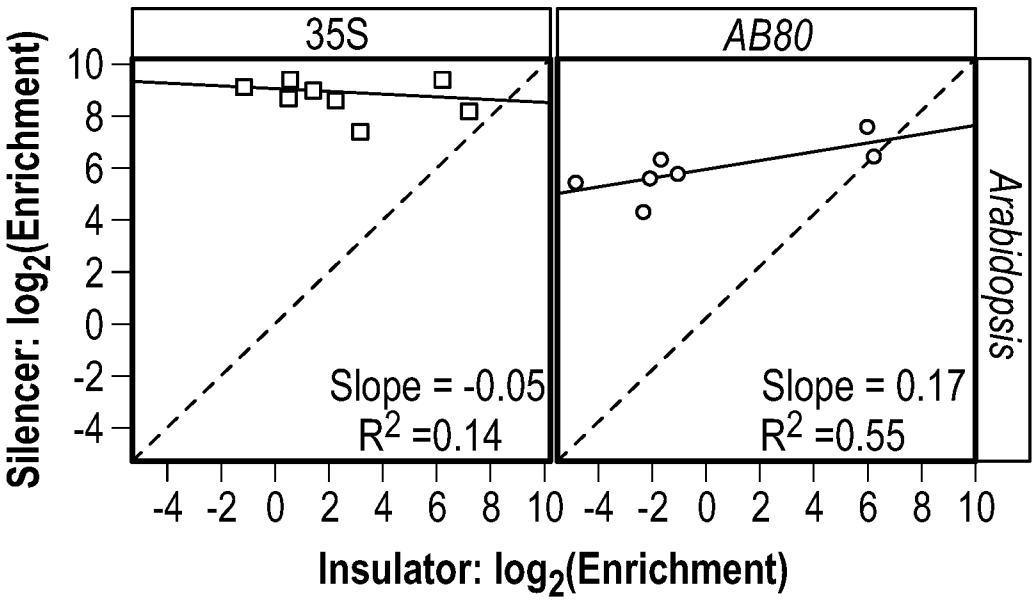


FIG. 12D

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2025/010297

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***C12N 15/10*** (2025.01); ***C12N 15/11*** (2025.01); ***C12N 15/86*** (2025.01); ***C12Q 1/68*** (2025.01); ***C12Q 1/6806*** (2025.01)

CPC: ***C12N 15/1086; C12N 15/111; C12N 15/86; C12Q 1/68; C12Q 1/6806;*** *C12N 2830/008; C12Q 2563/179*

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	US 2010/0083408 A1 (YI et al.) 01 April 2010 (01.04.2010)	
X	abstract; para [0007]; [0039]; [0076]; [0101]-[0102]	1, 7-11
Y	abstract; para [0007]; [0039]; [0076]; [0101]-[0102]	2-6, 12
	US 2018/0291370 A1 (DUKE UNIVERSITY) 11 October 2018 (11.10.2018)	
Y	abstract; para [0120]-[0121]; [0151]; [0179]-[0180]; [0204]; [0219]	4-6
	US 2022/0170910 A1 (ENCODED THERAPEUTICS, INC.) 02 June 2022 (02.06.2022)	
Y	para [0003]; [0010]; [0095]	3
	PETRYKOWSKA et al., Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome research. August 2008; Vol 18, No 8, pp: 1238-46	
Y	abstract; p. 1239, col 1, para 1; Figure 1	2, 3, 12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 April 2025 (30.04.2025)	**12 May 2025 (12.05.2025)**

Name and mailing address of the ISA/US	Authorized officer
COMMISSIONER FOR PATENTS **MAIL STOP PCT, ATTN: ISA/US** **P.O. Box 1450** **Alexandria, VA 22313-1450** **UNITED STATES OF AMERICA**	**KARI RODRIQUEZ**
Facsimile No. **571-273-8300**	Telephone No. **PCT Help Desk: 571-272-4300**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2025/010297

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2025/010297

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

Group I, claims 1-12, directed to a method for determining whether a candidate sequence is an insulator in a plant genetic environment.

Group II, claims 13-20, directed to an insulator comprising a polynucleotide sequence configured for regulation of transcription within one or more plant genetic environments, or a plant genetic construct, plant cell, plant tissue, or plant comprising the insulator.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-12**

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.