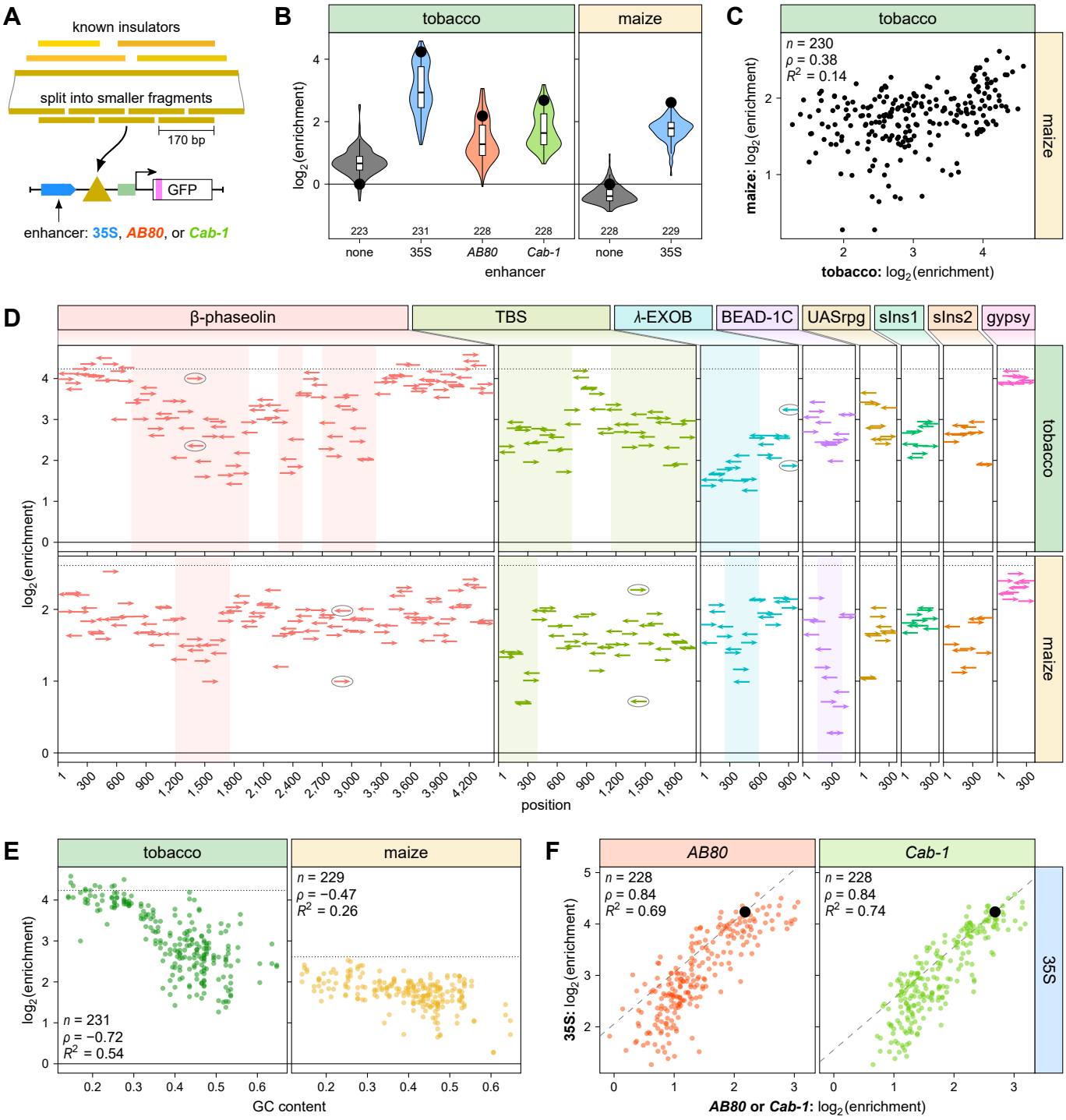

A
known insulators
split into smaller fragments
170 bp
GFP
enhancer: 35S, AB80, or Cab-1
B
tobacco
maize
log2(enrichment)
223
231
228
228
228
229
none
35S
AB80
Cab-1
none
35S
enhancer
C
tobacco
n = 230
ρ = 0.38
R2 = 0.14
maize
maize: log2(enrichment)
tobacco: log2(enrichment)
D
β-phaseolin
TBS
λ-EXOB
BEAD-1C
UASrpg
sIns1
sIns2
gypsy
tobacco
maize
log2(enrichment)
position
E
tobacco
maize
n = 231
ρ = −0.72
R2 = 0.54
n = 229
ρ = −0.47
R2 = 0.26
log2(enrichment)
GC content
F
AB80
Cab-1
n = 228
ρ = 0.84
R2 = 0.69
n = 228
ρ = 0.84
R2 = 0.74
35S
35S: log2(enrichment)
AB80 or Cab-1: log2(enrichment)

**Figure 1.** Short fragments exhibit enhancer-blocking insulator activity. **A)** 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. **B)** 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; log2 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. **C)** Correlation between the enrichment of insulator fragments in constructs with the 35S enhancer in tobacco leaves and maize protoplasts. **D)** 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. Clusters of active fragments are shown as shaded areas. Insulators with highly orientation-dependent activity are circled. **E)** Correlation between insulator fragment enrichment and GC content for constructs with the 35S enhancer. **F)** 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 $\rho$, and number ($n$) of constructs are indicated in (**C**), (**E**), and (**F**). The dotted line in (**D**) and (**E**) represents the enrichment of a control construct without an insulator.

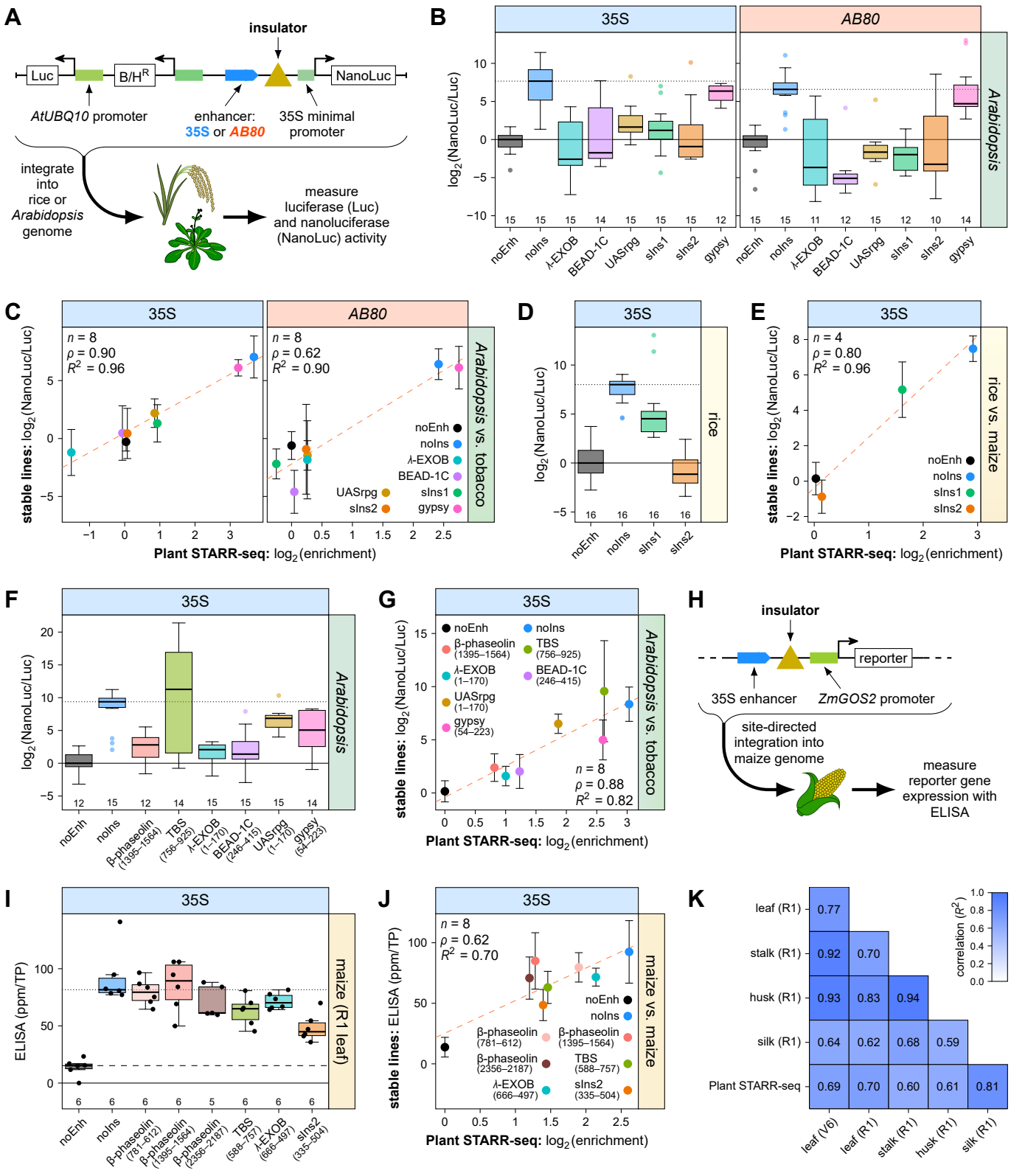

A
insulator
Luc
B/H^R
NanoLuc
AtUBQ10 promoter
enhancer:
35S or AB80
35S minimal
promoter
integrate
into
rice or
Arabidopsis
genome
measure
luciferase (Luc)
and nanoluciferase
(NanoLuc) activity
B
35S
AB80
Arabidopsis
log2(NanoLuc/Luc)
noEnh
noIns
λ-EXOB
BEAD-1C
UASrpg
sIns1
sIns2
gypsy
C
35S
AB80
n = 8
ρ = 0.90
R2 = 0.96
n = 8
ρ = 0.62
R2 = 0.90
Arabidopsis vs. tobacco
stable lines: log2(NanoLuc/Luc)
Plant STARR-seq: log2(enrichment)
D
35S
rice
E
35S
n = 4
ρ = 0.80
R2 = 0.96
rice vs. maize
F
35S
Arabidopsis
β-phaseolin
(1395–1564)
TBS
(756–925)
λ-EXOB
(1–170)
BEAD-1C
(246–415)
UASrpg
(1–170)
gypsy
(54–223)
G
35S
n = 8
ρ = 0.88
R2 = 0.82
Arabidopsis vs. tobacco
H
insulator
reporter
35S enhancer
ZmGOS2 promoter
site-directed
integration into
maize genome
measure
reporter gene
expression with
ELISA
I
35S
maize (R1 leaf)
ELISA (ppm/TP)
β-phaseolin
(781–612)
β-phaseolin
(2356–2187)
sIns2
(335–504)
J
35S
n = 8
ρ = 0.62
R2 = 0.70
maize vs. maize
K
leaf (R1)
stalk (R1)
husk (R1)
silk (R1)
Plant STARR-seq
leaf (V6)
correlation (R2)
0.77
0.92
0.70
0.93
0.83
0.94
0.64
0.62
0.68
0.59
0.69
0.70
0.60
0.61
0.81

**Figure 2.** Insulators are active in stable transgenic lines in *Arabidopsis*, rice, and maize. **A)** 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; log2 set to 0). **B, C)** The activity of full-length insulators was measured in *Arabidopsis* lines (**B**) and compared to the corresponding results from Plant STARR-seq in tobacco leaves (**C**). **D, E)** The activity of synthetic full-length insulators was measured in rice lines (**D**) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (**E**). **F, G)** The activity of insulator fragments was measured in *Arabidopsis* lines (**F**) and compared to the corresponding results from Plant STARR-seq in tobacco leaves (**G**). **H)** For transgenic maize lines, a reporter gene driven by the constitutive, moderate-strength *ZmGOS2* promoter 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. **I, J)** The activity of insulator fragments was measured in R1 leaves of transgenic maize lines (**I**) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (**J**). **K)** 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 (**B**), (**D**), (**F**), and (**I**) 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. For groups with less than 10 samples, individual data points are shown as black dots. In (**C**), (**E**), (**G**), and (**J**), the dashed line represents a linear regression line and error bars represent the 95% confidence interval. Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of constructs are indicated. The dotted line in (**B**), (**D**), (**F**) and (**I**) represents the median enrichment of a control construct without an insulator, and the dashed line in (**I**) represents the median enrichment of a control construct without an insulator and without the 35S enhancer.

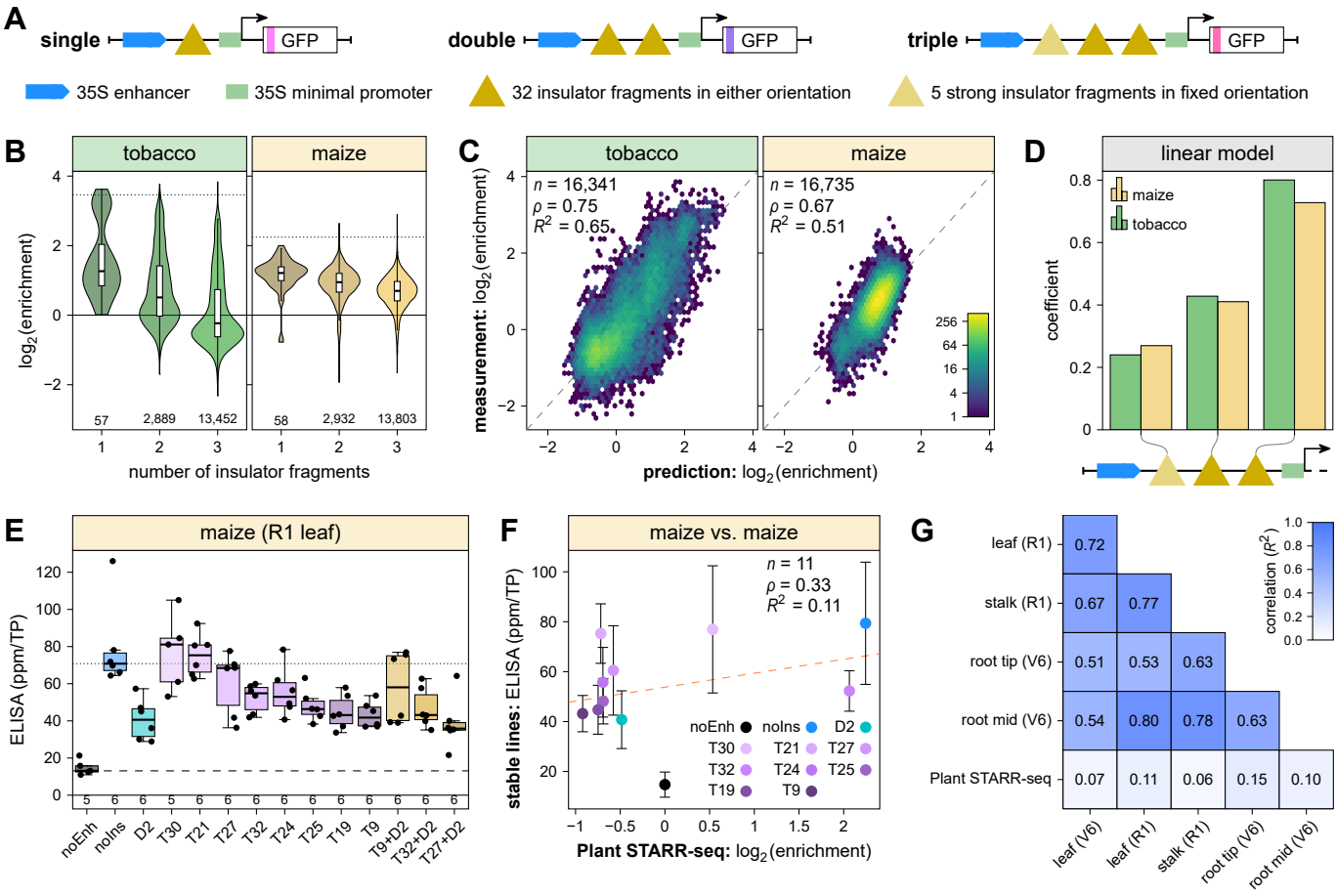



**Figure 3.** Insulator fragments can be stacked to create very strong enhancer-blocking insulators. **A)** One, two, or three 170-bp fragments of known insulators were cloned between a 35S enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. **B)** All insulator 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 Figure 1B. **C)** 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 (color represents the count of points in each hexagon). Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of fragments are indicated. **D)** Coefficients assigned by the linear model to insulator fragments in the indicated positions of the stacked constructs. **E, F)** The activity of insulator fragment combinations in constructs as in Figure 2H was measured in R1 leaves of transgenic maize lines (**E**) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (**F**). Box plots are as defined in Figure 2. The enrichment of a control construct without an insulator (noIns) is indicated as a dotted line. In (**F**), the dashed line represents a linear regression line and error bars represent the 95% confidence interval. Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of constructs are indicated. **G)** 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. The dotted line in (**B**) and (**E**) represents the enrichment of a control construct without an insulator, and the dashed line in (**E**) represents the enrichment of a control construct without an insulator and without the 35S enhancer.

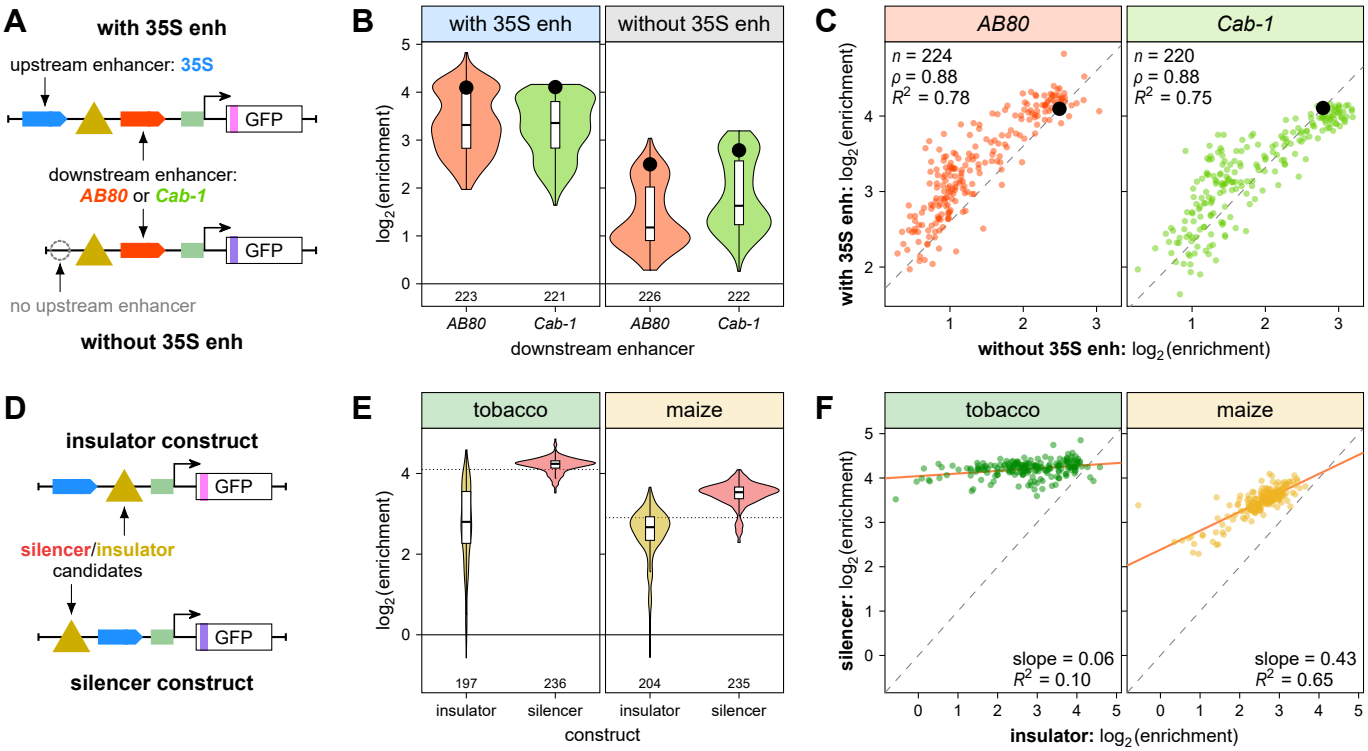



**Figure 4.** Insulators exhibit silencer activity in some contexts. **A)** Insulator fragments were cloned upstream of a *AB80* or *Cab-1* enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. Half of the constructs also harbored a 35S enhancer upstream of the insulator fragments (with 35S enh) while the other half lacked an upstream enhancer (without 35S enh). **B)** 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; log2 set to 0). The enrichment of a control construct without an insulator is indicated as a black dot. **C)** 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). **D)** 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. **E)** 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; log2 set to 0). The enrichment of a control construct without an insulator is indicated as a dotted line. **F)** 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 (**B**) and (**E**) are as defined in Figure 1B.

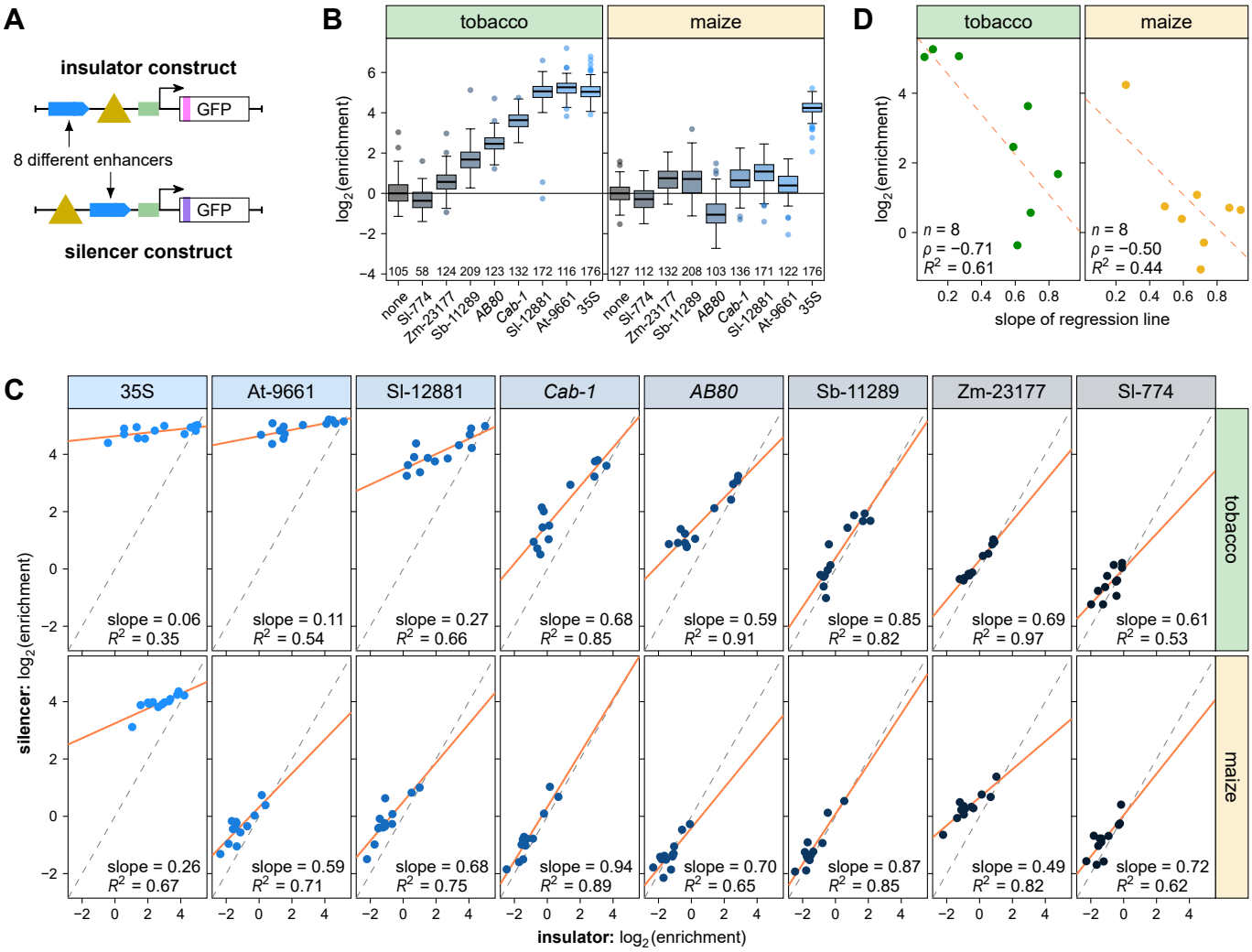
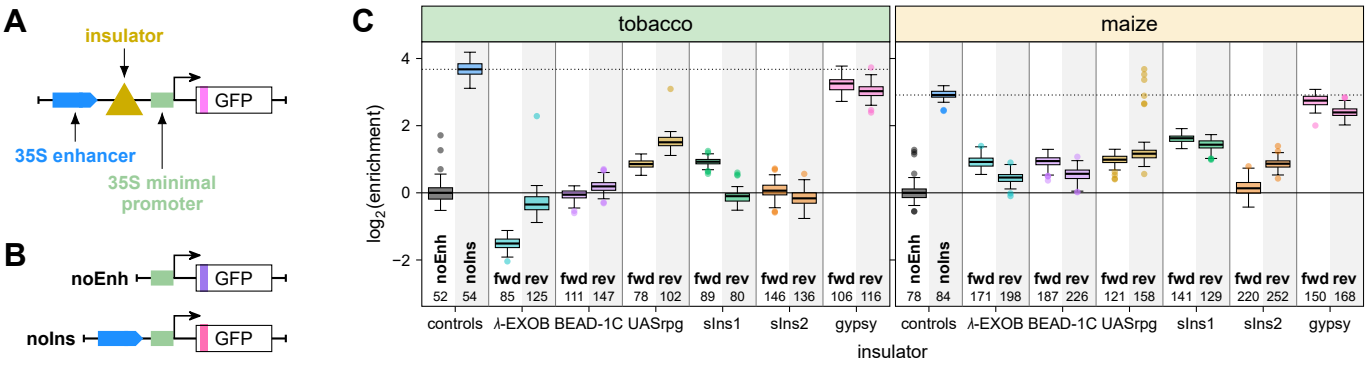
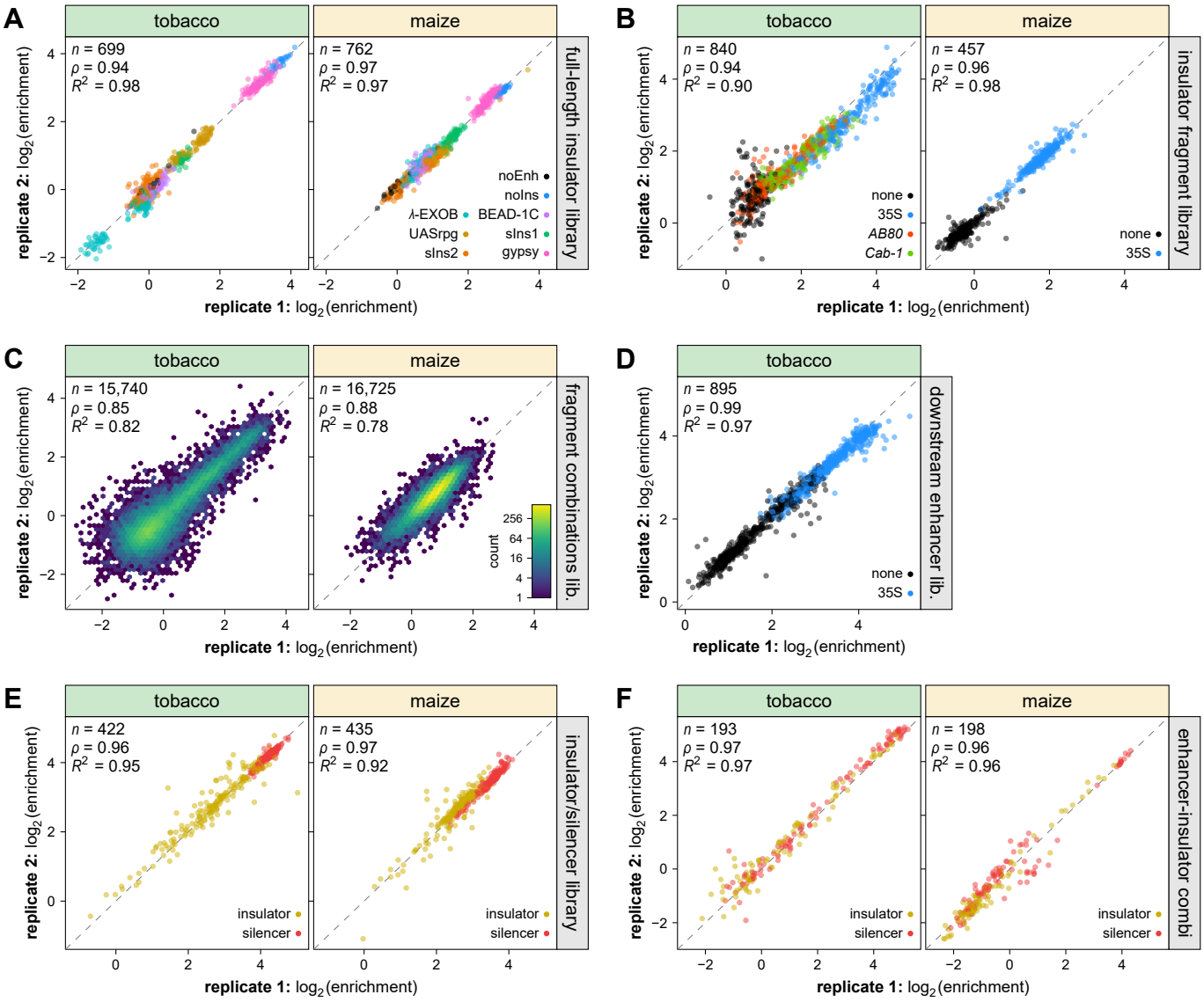


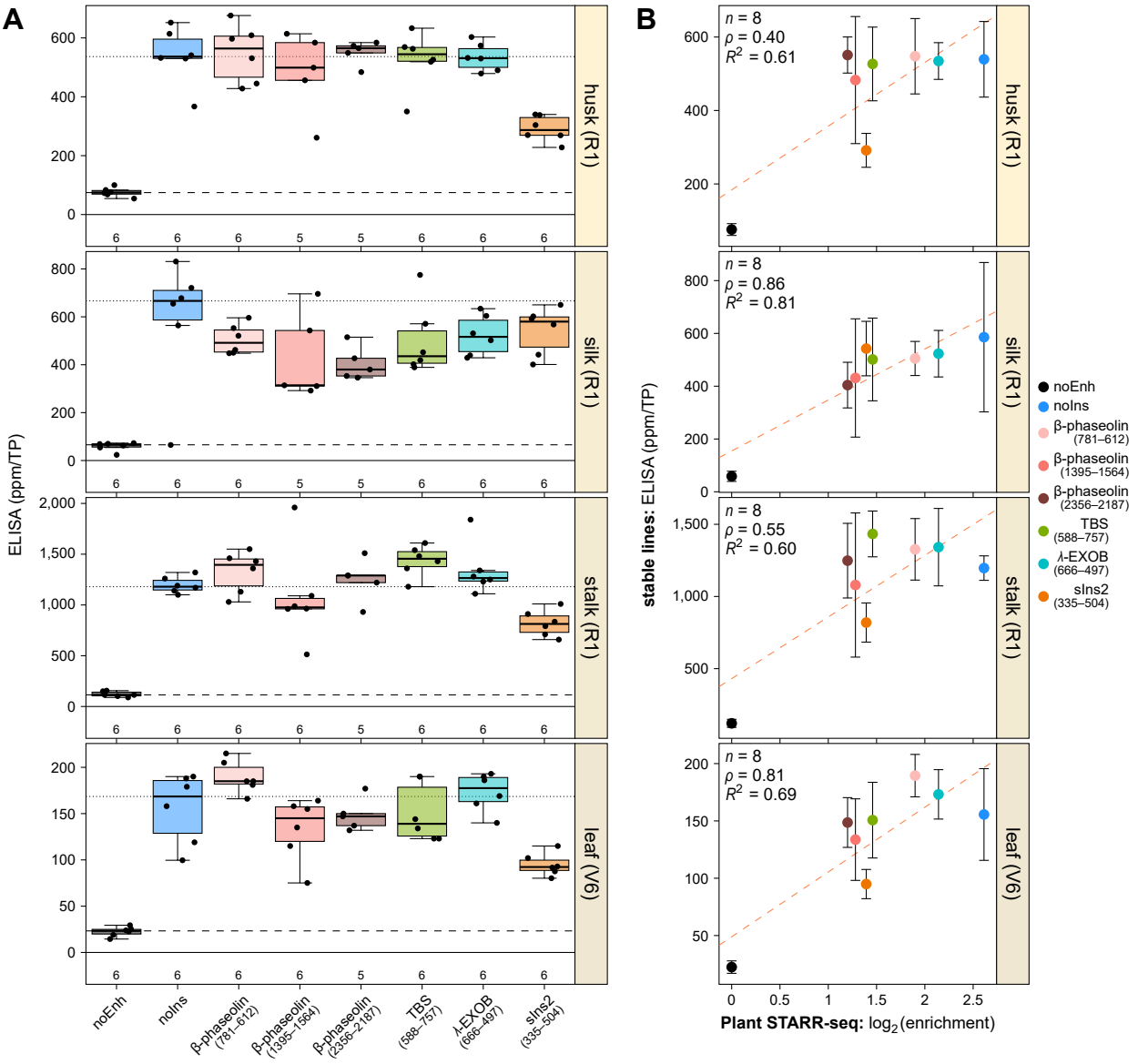

**Figure 5.** Silencer activity depends on enhancer strength. **A)** Selected insulators and insulator fragments were cloned in between (insulator construct) or upstream of (silencer construct) an enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. Eight different 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). **B)** Strength of the eight enhancers in constructs without an insulator. Reporter mRNA enrichment was normalized to a control construct without an enhancer (none; log2 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. **C)** 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. **D)** Correlation between the slope of the regression lines from (**C**) and the strength of the corresponding enhancer from (**B**). Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of constructs are indicated. A linear regression line is shown as a dashed line.



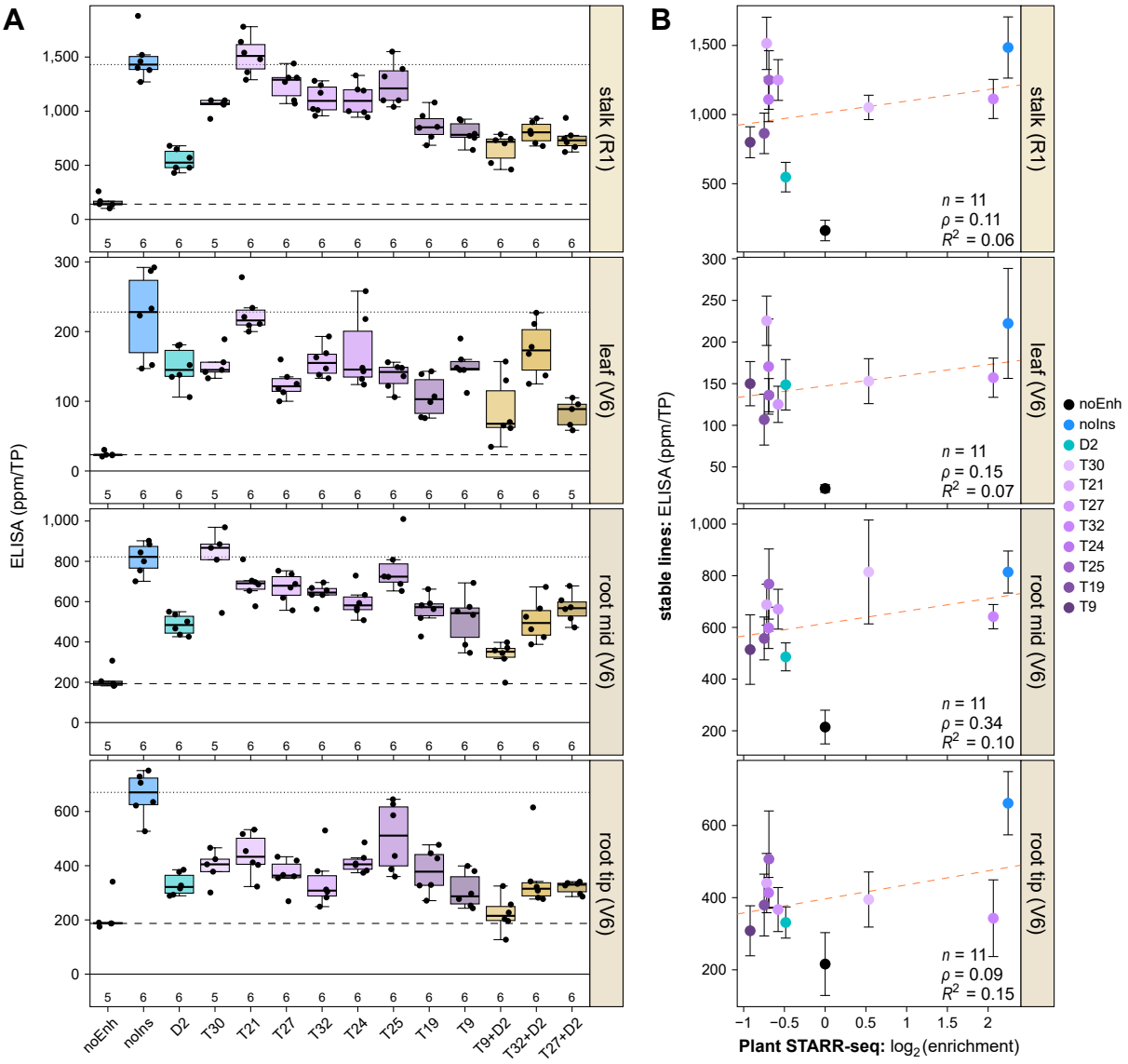

**Supplementary Figure S1.** Plant STARR-seq detects activity of enhancer-blocking insulators. **A)** Full-length insulators were cloned in the forward (fwd) or reverse (rev) orientation between a 35S enhancer and a 35S minimal promoter driving the expression of a barcoded GFP reporter gene. **B)** In all experiments, control constructs as in (**A**) but without an insulator (noIns) or without an insulator and without an enhancer (noEnh) were added to the library. **C)** 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; log2 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.



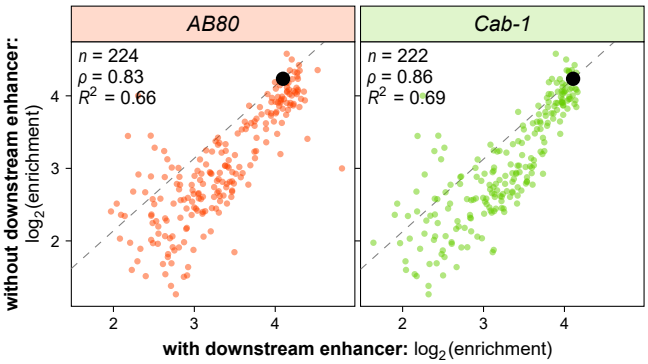

**Supplementary Figure S2.** Plant STARR-seq yields highly reproducible results. **A–G)** Correlation between biological replicates of Plant STARR-seq for the full-length insulator library used in Supplementary Figure S1 (**A**), the insulator fragment library used in Figure 1 (**B**), the insulator fragment combination library used in Figure 3 (**C**), the downstream enhancer library (**D**) and the insulator/silencer library (**E**) used in Figure 4, and the enhancer-insulator combination library used in Figure 5 (**F**). Experiments were performed in tobacco leaves (tobacco) or maize protoplasts (maize) as indicated. Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of constructs are indicated. The color in the hexbin plots in (**C**) represents the count of points in each hexagon.



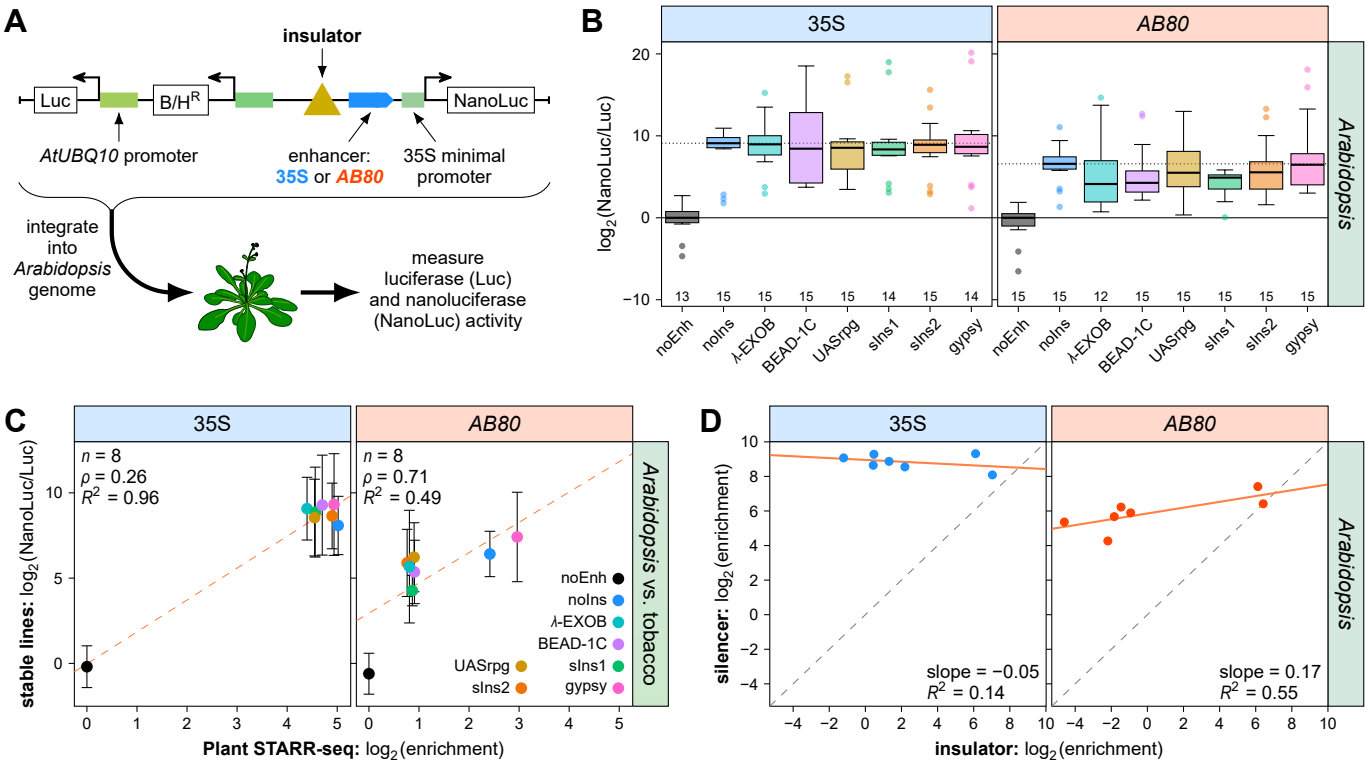

**Supplementary Figure S3.** Activity of insulator fragments in different maize tissues. **A, B)** Transgenic maize lines were created using constructs as in Figure 2H. The activity of insulator fragments was measured in the indicated tissues (**A**) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (**B**). In (**A**), box plots are as defined in Figure 2, and the dotted and dashed lines represent the median enrichment of control constructs without an insulator or without an enhancer, respectively.




**Supplementary Figure S4.** Activity of insulator fragment combinations in different maize tissues. **A, B)** Transgenic maize lines were created using insulator fragment combinations in constructs as in Figure 2H. The activity of insulator fragments was measured in the indicated tissues (**A**) and compared to the corresponding results from Plant STARR-seq in maize protoplasts (**B**). In (**A**), box plots are as defined in Figure 2, and the dotted and dashed lines represent the median enrichment of control constructs without an insulator or without an enhancer, respectively.




**Supplementary Figure S5.** Enhancers downstream of insulator fragments slightly reduce their activity. 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 $\rho$, and number ($n$) of constructs are indicated.




**Supplementary Figure S6.** Enhancer-dependent silencer activity in stable transgenic plants. **A)** 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; log2 set to 0). **B, C)** The activity of full-length insulators was measured in *Arabidopsis* lines (**B**) and compared to the corresponding results from Plant STARR-seq in tobacco leaves (**C**). Box plots in (**B**) are as defined in Figure 2 and the dotted line indicates the median activity of a control construct without an insulator. In (**C**), the dashed line represents a linear regression line and error bars represent the 95% confidence interval. Pearson's $R^2$, Spearman's $\rho$, and number ($n$) of constructs are indicated. **D)** Comparison of the mean NanoLuc/Luc ratio of full-length insulators in insulator (Figure 2B) or silencer constructs (**B**). A linear regression line is shown as a solid line and its slope and goodness-of-fit ($R^2$) is indicated.