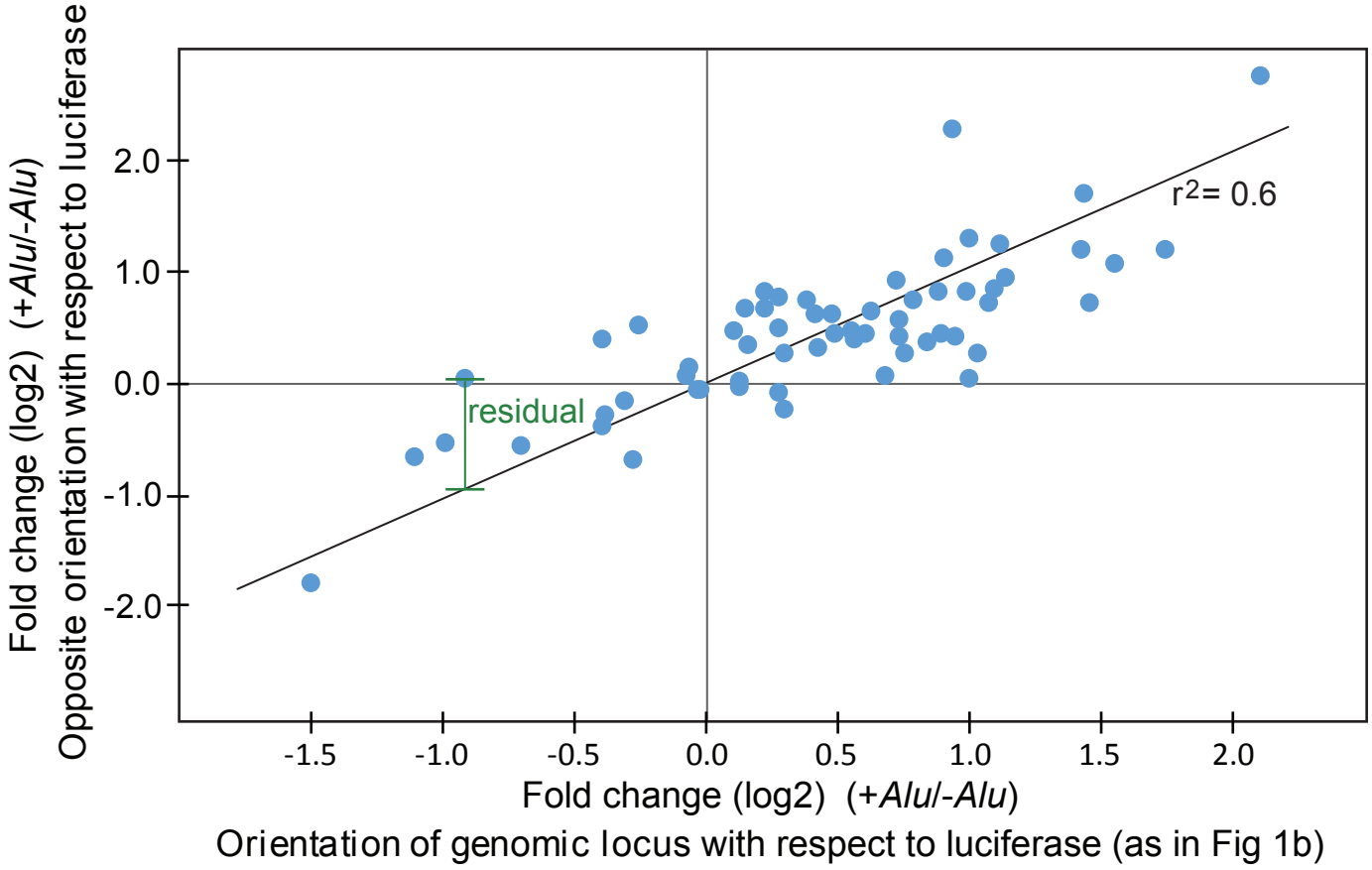
A


Fold change (log2) (+Alu/-Alu)
Opposite orientation with respect to luciferase
2.0
1.0
0.0
-1.0
-2.0
r2= 0.6
residual
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Fold change (log2) (+Alu/-Alu)
Orientation of genomic locus with respect to luciferase (as in Fig 1b)

B

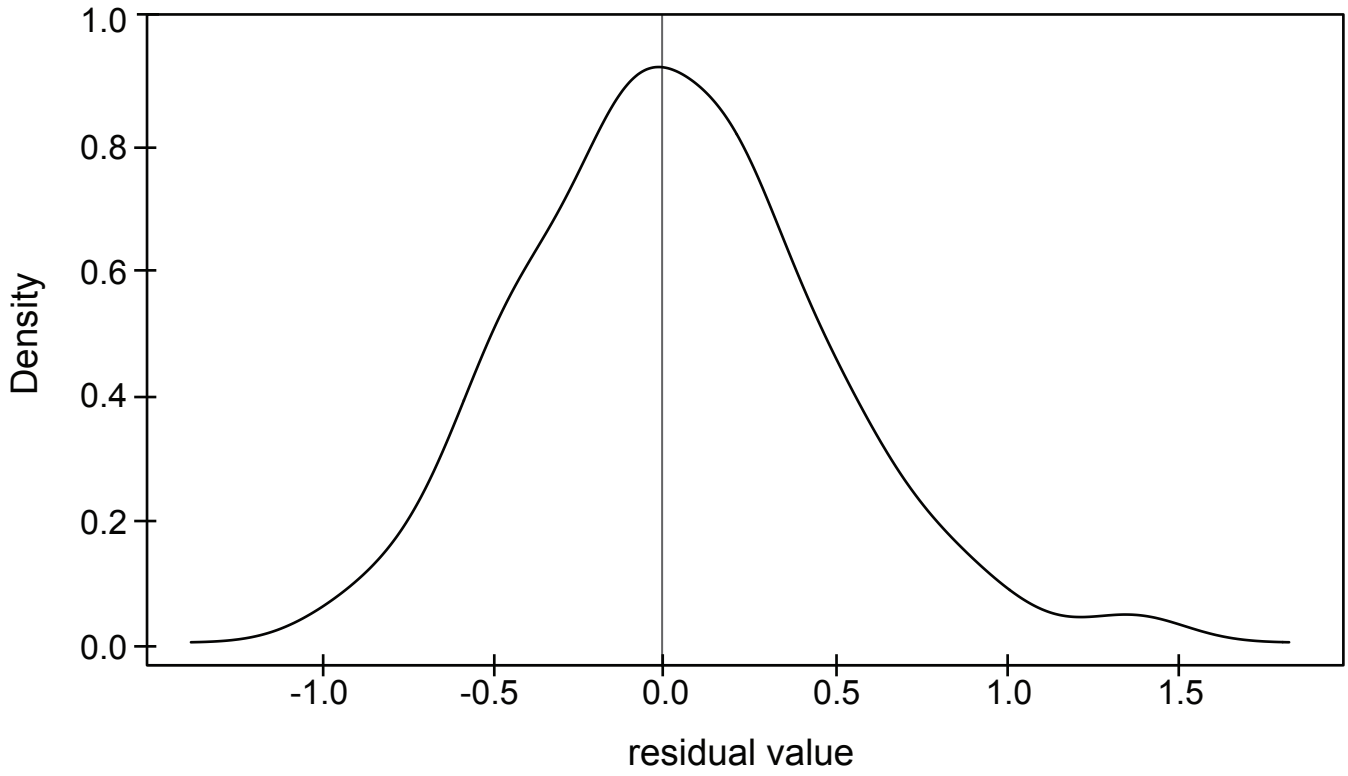

Density
1.0
0.8
0.6
0.4
0.2
0.0
-1.0
-0.5
0.0
0.5
1.0
1.5
residual value

Figure S1: *Orientation of the genomic sequence with respect to luciferase has relatively minor effects on ectopic assay results.* A) Scatter plot showing luciferase assay results for each of the loci that were evaluated in both orientation with respect to luciferase (n=62). The line is the null hypothesis, that the orientation of the genomic locus with respect to luciferase has no effect on luciferase expression, y=x. From this, the residual, the difference between the dependent variable (y) and the predicted value, was calculated (green). D) A density plot of the residuals, indicates that similar, although not identical, results were obtained for loci evaluated in both orientations with respect to luciferase.

Figure S2

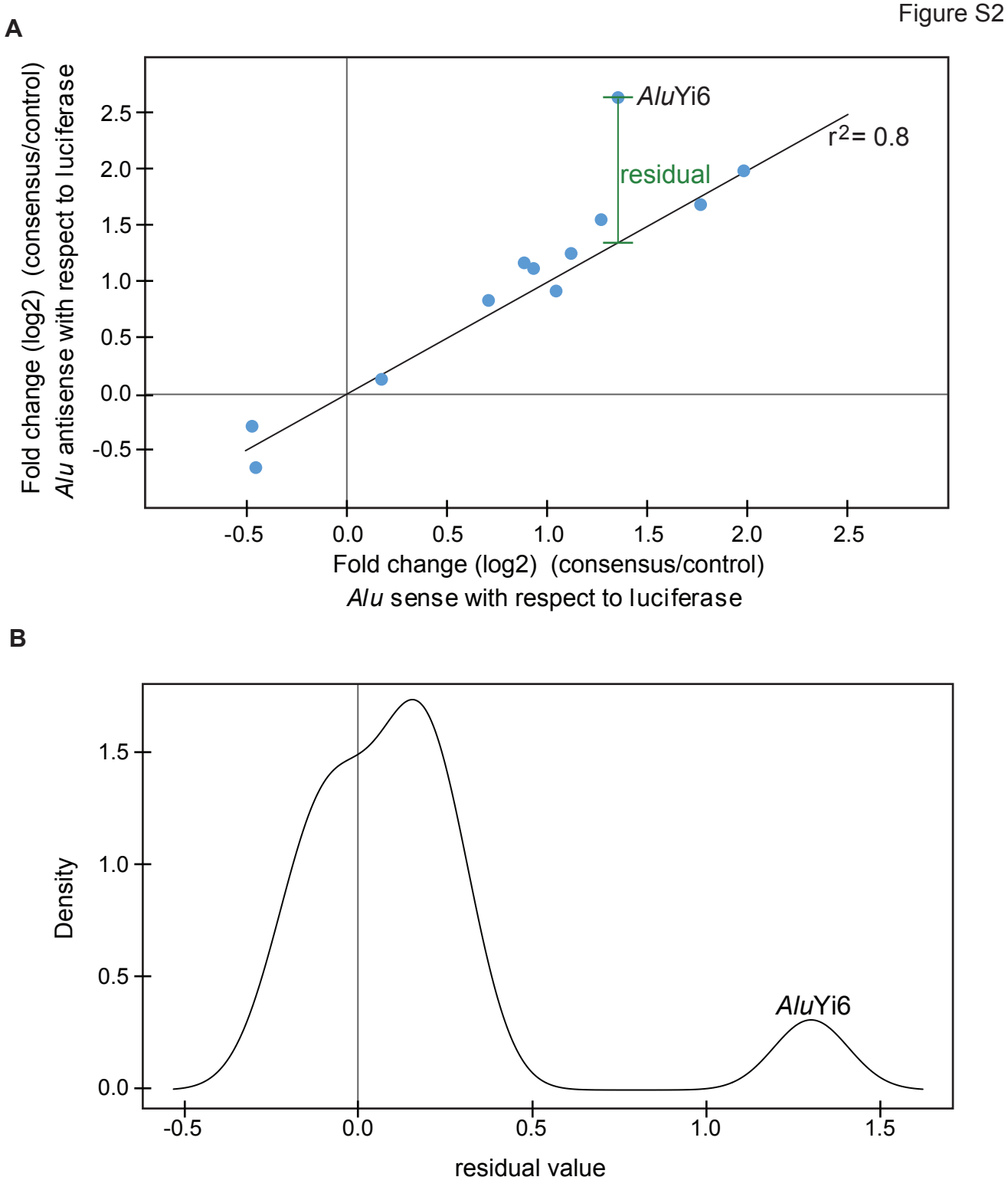
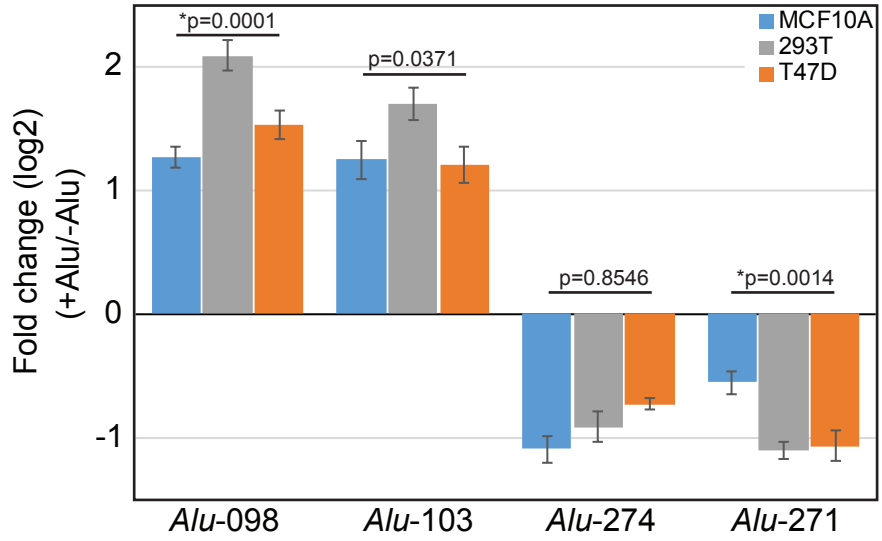

A
Fold change (log2) (consensus/control)
Alu antisense with respect to luciferase
AluYi6
residual
r2= 0.8
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
Fold change (log2) (consensus/control)
Alu sense with respect to luciferase
B
Density
1.5
1.0
0.5
0.0
AluYi6
-0.5
0.0
0.5
1.0
1.5
residual value

Figure S2: *Orientation of isolated Alu consensus sequences has minimal effects in the ectopic assay.* A) Scatter plot where each point is one *Alu* consensus sequence with the luciferase expression shown for when the *Alu* is sense with respect to luciferase versus when the *Alu* is anti-sense with respect to luciferase. Line represents the null hypothesis, that *Alu* orientation with respect to luciferase has no effect on luciferase expression, y=x. An example residual value is shown (green). B) Density plot of residual values with the outlier identified as *Alu*Yi6.

Figure S3

A Ectopic Assay



B Endogenous Locus Analysis

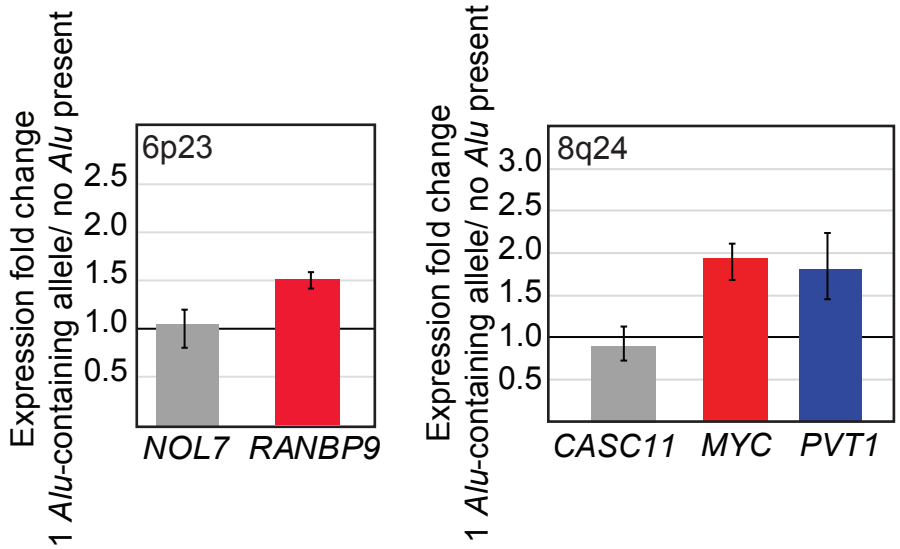


Figure S3: *Evaluation of Alu effects in breast-derived cell lines. A) Ectopic assay results for breast cancer risk loci are similar across multiple cell lines.* Luciferase assays for 4 breast cancer risk loci in two breast-derived cell lines, MCF10A and T47D are compared to original results in 293T cells. While results are statistically different for some loci (*, ANOVA), the overall results are similar. *B) Alu variants cause differential gene expression at the endogenous locus in breast-derived cell lines.* qRT-PCR was performed on the T47D parental line which is heterozygous for the presence of *Alu*-098 (6p23, right) and *Alu*-103 (8q24, left) and CRISPR edited lines homozygous for no *Alu*-098 or *Alu*-103. Results are shown as the expression fold change of the heterozygous versus homozygous lines. Similar gene expression patterns were detected in 293T cells (Fig 6).

Figure S4

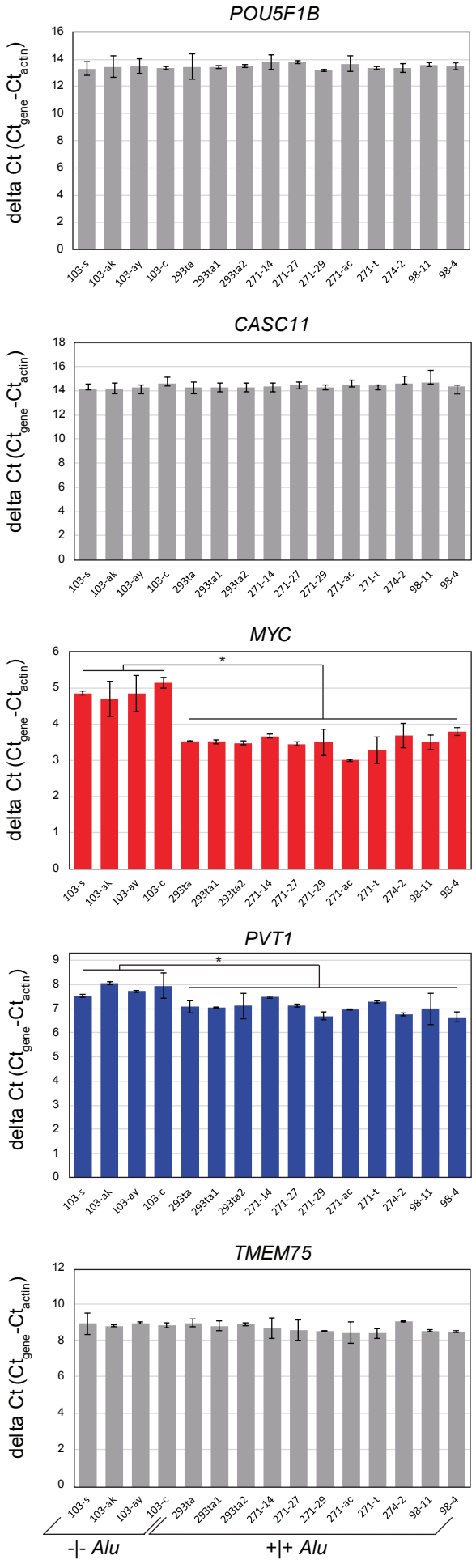

POU5F1B
CASC11
MYC
PVT1
TMEM75
delta Ct (Ct_gene-Ct_actin)
103-s
103-ak
103-ay
103-c
293ta
293ta1
293ta2
271-14
271-27
271-29
271-ac
271-t
274-2
98-11
98-4
*
-|- Alu
+|+ Alu

Figure S4: *Alu genotype dependent expression of 8q24 genes is consistent across 15 cell subclones.* qRT-PCR was used to measure gene expression in all available 293T-derived cell lines. These include: four lines edited at 8q24 to remove the polymorphic *Alu* element that are homozygous for no *Alu*-103 presence (left); and 11 cell lines that are homozygous for the *Alu*-103 presence (right). Of the 11, 3 are subclones of the 293T parental line and 8 lines were edited at other breast cancer risk loci but unedited at 8q24. Differential expression based on genotype is evident at *MYC* (red) and *PVT1* (blue) (*$p<0.01$, ANOVA). Error bars are the standard deviation from 6 datapoints (technical triplicates of experimental duplicates).

Figure S5

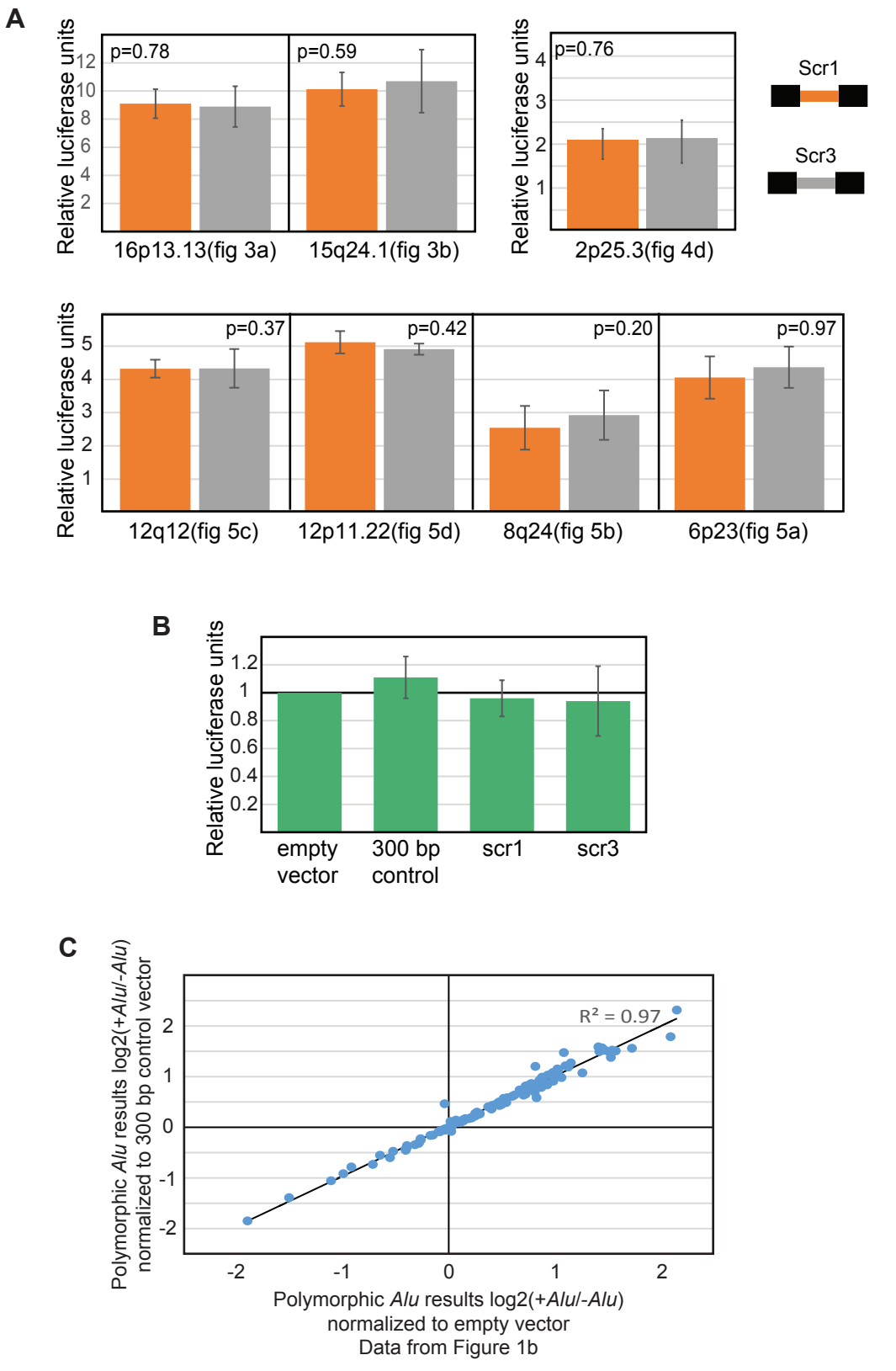

A
Relative luciferase units
p=0.78
p=0.59
16p13.13(fig 3a)
15q24.1(fig 3b)
p=0.76
2p25.3(fig 4d)
Scr1
Scr3
p=0.37
p=0.42
p=0.20
p=0.97
12q12(fig 5c)
12p11.22(fig 5d)
8q24(fig 5b)
6p23(fig 5a)
B
Relative luciferase units
empty vector
300 bp control
scr1
scr3
C
Polymorphic Alu results log2(+Alu/-Alu) normalized to 300 bp control vector
R² = 0.97
Polymorphic Alu results log2(+Alu/-Alu) normalized to empty vector
Data from Figure 1b

Figure S5. *Ectopic enhancer assay controls and normalization demonstrate the robustness of the reporter system.* A) Similar results were observed for 2 different scrambled (scr) sequences (scr1 in orange and scr3 in gray) tested in the genomic context (black in cartoon) of 7 different polymorphic *Alu* loci. Because the results are not statistically different between scr1 and scr3 evaluated at specific locus (t-test p-values shown), results were combined for the 2 scrambled sequences for each locus and reported as a single dataset in the main figure (corresponding figure number is shown). B) Scrambled sequences (scr) tested in isolation within the luciferase reporter vector performed similar to the "empty" luciferase vector with no additional sequences and the luciferase vector with a 300 bp control sequence with no known regulatory sequences present. Relative luciferase units shown normalized to the empty vector. C) There is significant correlation (R^2=0.97) for polymorphic *Alu* loci ectopic assay results normalized to the empty luciferase vector (x-axis, as shown in Figure 1) or normalized to the luciferase vector with the 300 bp control sequence (y-axis).