Small activating RNA binds to the genomic target site in a seed-region-dependent manner

Xing Meng[1], Qian Jiang[1], Nannan Chang[1], Xiaoxia Wang[1], Chujun Liu[1], Jingwei Xiong[1], Huiqing Cao[1,*], and Zicai Liang[1,2,*]

[1]Institute of Molecular Medicine, Peking University, Beijing, 100871, PR China
[2]Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin, 300072, PR China

* To whom correspondence should be addressed.

Zicai Liang, Tel: +86 10 62769862; Fax: +86 10 62769862; E-mail: liangz@pku.edu.cn

Correspondence may also be addressed to Huiqing Cao, Tel: +86 10 62769861; Fax: +86 10 62769861; E-mail: caohuiqing@pku.edu.cn

Present Address: Institute of Molecular Medicine, Peking University, 5 Yiheyuan Rd, Beijing, 100871, PR China.

MIQE checklist and additional information regarding the qPCR experiments: Information of the performed qPCR experiments of PR mRNA and AT-2 expression based on the MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines.

ITEM TO CHECK	IMPORTANCE	CHECKLIST
EXPERIMENTAL DESIGN		
Definition of experimental and control groups	**E**	√
Number within each group	**E**	√
Assay carried out by core lab or investigator's lab?	D	√
Acknowledgement of authors' contributions	D	
SAMPLE		
Description	**E**	√
Volume/mass of sample processed	D	√
Microdissection or macrodissection	**E**	NA
Processing procedure	**E**	√
If frozen - how and how quickly?	**E**	√
If fixed - with what, how quickly?	**E**	NA
Sample storage conditions and duration (especially for FFPE samples)	**E**	√

Experimental and control groups were defined in the figure legends of each figure, as well as the number of experimental repeats. All qPCR experiments were carried out by the investigator's lab. RNA samples were total RNA extracted from MCF-7, HEK293A or mutated cells treated with saRNA or non-specific small RNA (NC). The total volume of each sample ranges from 30 μl to 50 μl, and the total mass ranges 5 μg to 25 μg. After extracted, the RNA samples were frozen in -80°C and reverse transcribed in 2 days.

NUCLEIC ACID EXTRACTION		
Procedure and/or instrumentation	**E**	√
Name of kit and details of any modifications	**E**	√
Source of additional reagents used	D	√
Details of DNase or RNAse treatment	**E**	√
Contamination assessment (DNA or RNA)	**E**	√
Nucleic acid quantification	**E**	√
Instrument and method	**E**	√
Purity (A260/A280)	D	√
Yield	D	√
RNA integrity method/instrument	**E**	√
RIN/RQI or Cq of 3' and 5' transcripts	**E**	√

Electrophoresis traces	D	√
Inhibition testing (Cq dilutions, spike or other)	**E**	√
REVERSE TRANSCRIPTION		
Complete reaction conditions	**E**	√
Amount of RNA and reaction volume	**E**	√
Priming oligonucleotide (if using GSP) and concentration	**E**	√
Reverse transcriptase and concentration	**E**	√
Temperature and time	**E**	√
Manufacturer of reagents and catalogue numbers	D	√
Cqs with and without RT	D*	√
Storage conditions of cDNA	D	√

Total RNA was isolated using 0.5 ml per sample (1×10^5-1×10^6 cells) TRI Reagent (Sigma-Aldrich) following the manufacturer's protocol. The primer pairs used in qPCR were designed to target different exons of mature mRNA, which ensured that genomic DNA could not be detected. We checked the PCR product by gel, melting curve analysis and sequencing to make sure there was only one expected amplicon. These ensured that DNA template won't be detected even if there was residual DNA in total RNA. Therefore, it is unnecessary in our system to digest total RNA with DNase. RNA concentrations and qualities were determined using a NanoDrop spectrophotometer (ND-1000, Nanodrop). RNA samples with 260/280 ratio from 1.9 to 2.1 and 260/230 ratio from 1.8 to 2.2 were considered qualified in our experiment. The integrity of RNA was determined by the electrophoresis of 0.5 μg RNA on 1% agarose gel. Different dilutions of cDNA were demonstrate by different primers and observed decreases in Cqs that consistent with the anticipated result.

Reverse transcription was performed using oligo dT primer (Transgene) and TransScript reverse transcriptase (Transgene) following the manufacturer's protocol. A 20 μl reverse transcription reaction contained 1 μg of total RNA, 4 μl 5×RT-Buffer, 1 μl oligo dT primer (0.5 μg/ul), 1 μl dNTP Mix (25 mM, Promega), 0.5 μl RNasin (40 U/μl, Tiangen), 1 μl TransScript reverse transcriptase (200 U/μl)and RNase-free water. The conditions of reverse transcription were 42°C for 30 min, followed by 85°C for 5 min. cDNA was stored in -80°C. In samples without RT, no amplification was detected.

qPCR TARGET INFORMATION		
If multiplex, efficiency and LOD of each assay.	**E**	NA
Sequence accession number	**E**	√
Location of amplicon	D	√
Amplicon length	**E**	√

In silico specificity screen (BLAST, etc)	**E**	√
Pseudogenes, retropseudogenes or other homologs?	D	√
Sequence alignment	D	√
Secondary structure analysis of amplicon	D	
Location of each primer by exon or intron (if applicable)	**E**	√
What splice variants are targeted?	**E**	NA
qPCR OLIGONUCLEOTIDES		
Primer sequences	**E**	√
RT Primer DB Identification Number	D	NA
Probe sequences	D**	NA
Location and identity of any modifications	**E**	NA
Manufacturer of oligonucleotides	D	√
Purification method	D	√

Detail information of targets and amplicons for each target gene was list in the table below:

Gene name	Amplicon length	NCBI accession No.	Primer sets	Primers locations according to Gene
PR (PR-B & PR-A)	196 bp	NM_000926.4 & NM_001202474.3	PR-3/4-RT-F	exon 3
			PR-3/4-RT-R	exon 4
AT-2	150 bp	NR_073144.1	AT2-2/3-F	exon 1/2 junction
			AT2-2/3-R	exon 3
GAPDH	115 bp	NM_002046.5	GAPDH-F	exon 1
			GAPDH-R	exon 2/3 junction

The primer sets of each target were analyzed with NCBI Blast to ensure specificity. The sequences of primers were listed in Table S1. DNA oligonucleotides were from Biosune (Beijing) and purified with PAGE method.

qPCR PROTOCOL		
Complete reaction conditions	**E**	√
Reaction volume and amount of cDNA/DNA	**E**	√
Primer, (probe), Mg++ and dNTP concentrations	**E**	√
Polymerase identity and concentration	**E**	√
Buffer/kit identity and manufacturer	**E**	√
Exact chemical constitution of the buffer	D	
Additives (SYBR Green I, DMSO, etc.)	**E**	√
Manufacturer of plates/tubes and catalog number	D	√

Complete thermocycling parameters	**E**	√
Reaction setup (manual/robotic)	D	√
Manufacturer of qPCR instrument	**E**	√
qPCR VALIDATION		
Evidence of optimisation (from gradients)	D	
Specificity (gel, sequence, melt, or digest)	**E**	√
For SYBR Green I, Cq of the NTC	**E**	√
Standard curves with slope and y-intercept	**E**	√
PCR efficiency calculated from slope	**E**	√
Confidence interval for PCR efficiency or standard error	D	√
r^2 of standard curve	**E**	√
Linear dynamic range	**E**	√
Cq variation at lower limit	**E**	√
Confidence intervals throughout range	D	
Evidence for limit of detection	**E**	√
If multiplex, efficiency and LOD of each assay.	**E**	NA

qPCR analyses were performed in optical 96-well plates (Axygen) with StepOnePlus™ Instrument (ABI). The cycling conditions were 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, and 65°C for 15 s, 68°C for 20 s, and followed by a melting curve analysis (95 °C for 15 s, 55 °C for 15s, 1 °C increases for 30 s up to 95 °C). A 20 µl reaction contained 1 µl of 5×SYBR Green (ThermoFisher), 0.25 µl of each primer (20 µM), 1 µl of dNTP mix (2.5 mM, Promega), 0.25 µl of HotMaster Taq (5 U/µl, Tiangen) and 2 µl of cDNA template (diluted from 20 µl to 50 µl). Reactions were set up manually in a sterile bench using designated equipment.

The primer sets was checked for the specificity by gel, melting curve analysis and sequencing. Primers used in qPCR were demonstrated to have good linear correlation and equal priming efficiency for the different dilutions compared to the Ct values. We considered a PCR efficiency range from 90%-110% as acceptable in our study. The informations in detail were listed in the table below.

Gene	Efficiency	Slope	R^2	Cq of lower limit	NTC
PR	105.35%	-3.16	0.9996	28.27	NA
AT-2	96.60%	-3.43	0.9964	31.82	NA
GAPDH	96.49%	-3.40	0.9942	17.82	NA

DATA ANALYSIS		
qPCR analysis program (source, version)	**E**	√
Cq method determination	**E**	√
Outlier identification and disposition	**E**	√

Results of NTCs	**E**	√
Justification of number and choice of reference genes	**E**	√
Description of normalisation method	**E**	√
Number and concordance of biological replicates	D	√
Number and stage (RT or qPCR) of technical replicates	**E**	√
Repeatability (intra-assay variation)	E	√
Reproducibility (inter-assay variation, %CV)	D	
Power analysis	D	
Statistical methods for result significance	**E**	√
Software (source, version)	E	√
Cq or raw data submission using RDML	**D**	

Raw data of qPCR was analyzed using StepOne™ Software v2.2.2 (ABI) and the determination of Cqs and outlier was performed automatically. The relative expression of each target gene was determined using the formula $2^{-\Delta\Delta Ct}$. For NTC, we did not observed any specific amplicon. GAPDH was chose as reference gene and used as the internal control. Numbers of biological replicates were stated in figure legends, and three technical replicates of qPCR were performed for each sample. qPCR assay in our study has good repeatability (SD=0.19). All data are shown as mean ± SEM. Statistical analysis was performed with Student's t-test to evaluate single-factor differences between two sets of data, or with ANOVA followed by the Bonferroni post-hoc test for multiple comparisons.

Supplementary Data

Table S1

Primers used in these studies.

Primer	Sequence
PR-3/4-RT-F	TGGAAGAAATGACTGCATCG
PR-3/4-RT-R	TAGGGCTTGGCTTTCATTTG
GAPDH-RT-F	GCTCTCTGCTCCTCCTGTTC
GAPDH-RT-R	ACGACCAAATCCGTTGACTC
1611T7E1-F1	GGCATTAGCCATCAGGAAAA
1611T7E1-R1	AATCTCCCAGCAGCCACTAA
1611T7E1-F2	GGCAATGCTGTCCTGATTTT
1611ST7E1-R2	AAATCTCCCAGCAGCCACTA
1611pulldown-6-F	CATGTGGAACATGGATGAACCT
1611pulldown-6-R	AATCTCAATACCCACTAGCAGTT
1611pulldown-36-F	AATAGCCAAATCACAGGTACAAAA
1611pulldown-36-R	TGACTCTAAAATCTCAATACCCACT
ATEx2/3-F	CGGACCTCAAGGTCTAGCTG
ATEx2/3-R	CCAAGAGGAGCACAAGACATC

Sequences are listed 5’ to 3’.

Table S2

Wild type saRNAs used in these studies.

RNA	site	strand sequence
PR+1	+1/ +19	AGUCCACAGCUGUCACUAA
PR+2	+2/ +20	GUCCACAGCUGUCACUAAU
PR+4	+4/ +22	CCACAGCUGUCACUAAUCG
PR+23	+23/ +41	GGGUAAGCCUUGUUGUAUU
PR+56	+56/ +74	GCAUUCUCAAUGAGAACUA
PR-1	–1/ +18	CAGUCCACAGCUGUCACUA
PR-38	-38/ -19	CCCUAGAGGAGGAGGCGUU
PR-120	-120/-101	CAGAAUAACGGGUGGAAAU
PR-214	-214/-195	CCAAUUAGAUAGGGCAUAA
PR-520	-520/-501	GCACACAACAUCCAAACUU
PR-532	-532/-513	AAUCGAAGUGUGGCACACA
PR-549	-549/-530	ACACAUUCCUUAAGACAAA
PR-808	-808/-789	CCCAAUGGUCUUGGGUCAA
PR-891	-891/-872	GCUAGACCUUGAGGUCCUA
PR-1042	-1042/-1023	CCUAACAUGAUAAAGGUUU
PR-1056	-1056/-1037	CCAAAUCCUAGUAACCUAA
PR-1060	-1060/-1041	AACUCCAAAUCCUAGUAAC
PR-1382	-1382/-1363	UUGAGUGAGCUGUAUGAUA
PR-1484	-1484/-1465	GCUAGUGGGUAUUGAGAUU
PR-1596	-1596/-1577	AGUACUGAUUCAUGUGGAA
PR-1597	-1597/-1578	AAGUACUGAUUCAUGUGGA
PR-1598	-1598/-1579	GAAGUACUGAUUCAUGUGG
PR-1599	-1599/-1580	UGAAGUACUGAUUCAUGUG
PR-1600	-1600/-1581	GUGAAGUACUGAUUCAUGU
PR-1601	-1601/-1582	AGUGAAGUACUGAUUCAUG
PR-1602	-1602/-1583	GAGUGAAGUACUGAUUCAU
PR-1603	-1603/-1584	GGAGUGAAGUACUGAUUCA
PR-1604	-1604/-1585	AGGAGUGAAGUACUGAUUC
PR-1605	-1605/-1586	AAGGAGUGAAGUACUGAUU
PR-1606	-1606/-1587	GAAGGAGUGAAGUACUGAU
PR-1607	-1607/-1588	UGAAGGAGUGAAGUACUGA
PR-1608	-1608/-1589	GUGAAGGAGUGAAGUACUG
PR-1609	-1609/-1590	AGUGAAGGAGUGAAGUACU
PR-1610	-1610/-1591	GAGUGAAGGAGUGAAGUAC
PR-1611	-1611/-1592	GGAGUGAAGGAGUGAAGUA
PR-1612	-1612/-1593	AGGAGUGAAGGAGUGAAGU
PR-1613	-1613/-1594	AAGGAGUGAAGGAGUGAAG
PR-1614	-1614/-1595	AAAGGAGUGAAGGAGUGAA
PR-1615	-1615/-1596	AAAAGGAGUGAAGGAGUGA
PR-1616	-1616/-1597	AAAAAGGAGUGAAGGAGUG
PR-1617	-1617/-1598	UAAAAAGGAGUGAAGGAGU

PR-1618	-1618/-1599	AUAAAAAGGAGUGAAGGAG
PR-1619	-1619/-1600	AAUAAAAAGGAGUGAAGGA
PR-1620	-1620/-1601	CAAUAAAAAGGAGUGAAGG
PR-1621	-1621/-1602	UCAAUAAAAAGGAGUGAAG
PR-1622	-1622/-1603	GUCAAUAAAAAGGAGUGAA
PR-1623	-1623/-1604	UGUCAAUAAAAAGGAGUGA
PR-1624	-1624/-1605	UUGUCAAUAAAAAGGAGUG
PR-1625	-1625/-1606	UUUGUCAAUAAAAAGGAGU
PR-1626	-1626/-1607	AUUUGUCAAUAAAAAGGAG

Only the sense strands are shown (The sense strand is in the same orientation of the coding strand of target gene). Sequences are listed 5' to 3'. The other strands of the RNA duplex are complementary and not include two thymidine bases at the 3' termini as indicated.

Table S3

Mutated saRNAs used in these studies.

RNA	Sense strand	Antisense strand
PR-11M2	GCUGUCUGGCCAGUCCAGA	UCUGGACUGGCCAGACAGC
PR-11M4	GCUGUCUGGCCAGUCGACA	UGUCGACUGGCCAGACAGC
PR-11M6	GCUGUCUGGCCAGGCCACA	UGUGGCCUGGCCAGACAGC
PR-11M8	GCUGUCUGGCCCGUCCACA	UGUGGACGGGCCAGACAGC
PR-11M10	GCUGUCUGGGCAGUCCACA	UGUGGACUGCCCAGACAGC
PR-11M12	GCUGUCUUGCCAGUCCACA	UGUGGACUGGCAAGACAGC
PR-11M14	GCUGUGUGGCCAGUCCACA	UGUGGACUGGCCACACAGC
PR-11M16	GCUUUCUGGCCAGUCCACA	UGUGGACUGGCCAGAAAGC
PR-11M17	GCGGUCUGGCCAGUCCACA	UGUGGACUGGCCAGACCGC
PR-11M18	GGUGUCUGGCCAGUCCACA	UGUGGACUGGCCAGACACC
PR-1611M2	GGAGUGAAGGAGUGAAGGA	UCCUUCACUCCUUCACUCC
PR-1611M4	GGAGUGAAGGAGUGACGUA	UACGUCACUCCUUCACUCC
PR-1611M6	GGAGUGAAGGAGUUAAGUA	UACUUAACUCCUUCACUCC
PR-1611M8	GGAGUGAAGGAUUGAAGUA	UACUUCAAUCCUUCACUCC
PR-1611M10	GGAGUGAAGUAGUGAAGUA	UACUUCACUACUUCACUCC
PR-1611M12	GGAGUGACGGAGUGAAGUA	UACUUCACUCCGUCACUCC
PR-1611M14	GGAGUUAAGGAGUGAAGUA	UACUUCACUCCUUAACUCC
PR-1611M16	GGAUUGAAGGAGUGAAGUA	UACUUCACUCCUUCAAUCC
PR-1611M17	GGCGUGAAGGAGUGAAGUA	UACUUCACUCCUUCACGCC
PR-1611M18	GUAGUGAAGGAGUGAAGUA	UACUUCACUCCUUCACUAC
PR-1611senseM4	GGAGUGAAGGAGUGACGUA	UGUGGACUGGCCAGACAGC
PR-1611senseM6	GGAGUGAAGGAGUUAAGUA	UGUGGACUGGCCAGACAGC
PR-1611senseM8	GGAGUGAAGGAUUGAAGUA	UGUGGACUGGCCAGACAGC
PR-1611antisenseM4	GGAGUGAAGGAGUGAAGUA	UACGUCACUCCUUCACUCC
PR-1611antisenseM6	GGAGUGAAGGAGUGAAGUA	UACUUAACUCCUUCACUCC
PR-1611antisenseM8	GGAGUGAAGGAGUGAAGUA	UACUUCAAUCCUUCACUCC

Sequences are listed 5' to 3'. Two thymidine bases at the 3' termini are not shown as indicated. Mutate bases are shown in red.

Table S4

Targets of gRNAs used in these studies.

	gRNA target sequence
PR-G1	AATATTATTTGTCAATAAAA
PR-G2	TTGTCAATAAAAAGGAGTGA
PR-G3	AAGGAGTGAAGTACTGATTCATG
PR-G4	GAAGTACTGATTCATGTGGAACA

Sequences are listed 5’ to 3’.

Table S5

siRNA used in these studies.

RNA	target mRNA	strand sequence
PR-si1	PR	GCACAAUUACCCAAGAUAU
AT2-si1	AT-2	CCAGGACACCACAUUUAAU
AT2-si2	AT-2	CCUGCCUUGACGUUCGAAA

Only the sense strand is shown (The sense strand is in the same orientation of the coding strand of target gene). Sequences are listed 5' to 3'. The other strand of the RNA duplex is complementary and not includes two thymidine bases at the 3' termini as indicated.

FIGURES LEGENDS

Figure S1. The activity of mutated PR-1611 in reporter assay.

(**A**) Schematic representation of the construction of luciferase reporter gene. A new start codon (red) and restriction enzyme cutting sites (blue) were inserted adjacent to the luciferase gene (yellow). Targets of indicate small RNAs (light blue) are inserted between the restriction enzyme cutting sites to form a modified luciferase gene. (**B**) Schematic representation of the working model of the Dual luciferase assay. Targets of sense (green) or antisense strand (red) of indicate small RNA are cloned in the reporter, and sense strand-loaded Ago2 or antisense strand-loaded Ago2 can recognize and inhibit the expression of reporters containing their targets. Thus the luciferase readout will reflect the Ago2 loading level of each strand. (**C** and **D**) HEK293A cells were co-transfected with reporter carrying the target of antisense strands (C) or sense strand (D) of PR-1611 and mutated PR-1611. The relative luciferase levels of reporters were assessed by Dual-Luciferase Reporter Assay System.

Figure S2. Sequencing diagram of Clone-5.

Sequencing diagram of Clone-5 with an 8-bp deletion in one allele and a 1-bp insertion in the other allele of the PR-1611 target site in the PR promoter region.

Figure S3. Semi-quantitative analysis of promoter DNA binding with saRNA.

(**A** and **B**) PR promoter DNA was demonstrated by semi-quantitative PCR (31 cycles) after streptavidin purification from WT HEK293A (A) or Clone-5 cells (B).

Figure S4. Correlation between PR mRNA and AT-2.

(**A** and **B**) Plots of PR mRNA (A) and AT-2 (B) expression in MCF-7 cells transfected with PR-si1 for 24 h. (**C** and **D**) Plots of PR mRNA (C) and AT-2 (D) expression in MCF-7 cells treated with AT2-si1 or AT2-si2 for 24 h. (**E**) Plot of AT-2 mRNA expression in MCF-7 cells with co-transfection of siRNAs of AT-2 and saRNAs of PR. mRNA expression level were assessed by qPCR. In the co-transfection experiments, cells were harvested 72 h after the second transfection. **P <0.01; ***P <0.001.

Figure S5. A working model of RNAa induced by PR-1611.

A promoter-targeted saRNA (PR-1611) is loaded in an Ago2 protein, using the antisense strand (red) as the guide strand. The saRNA-loaded Ago2 targets the complementary sequence of the saRNA at distal PR promoter. The bind between saRNA-loaded Ago2 and promoter DNA results in the recruitment of various chromatin modifying or remodeling complexes, which together mediate the transcriptional activation of PR gene.

FIGURES

Figure S1

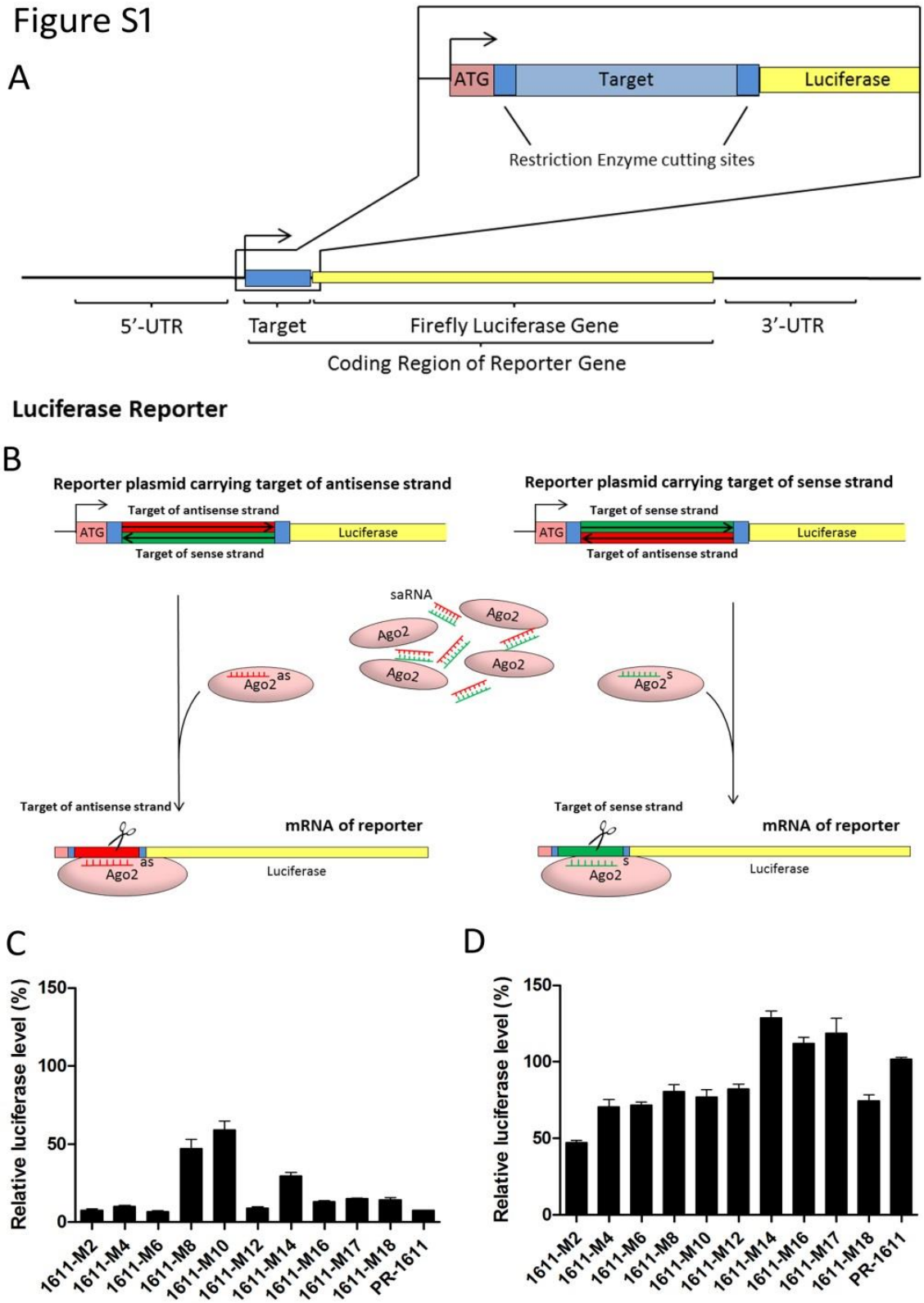


Figure S2

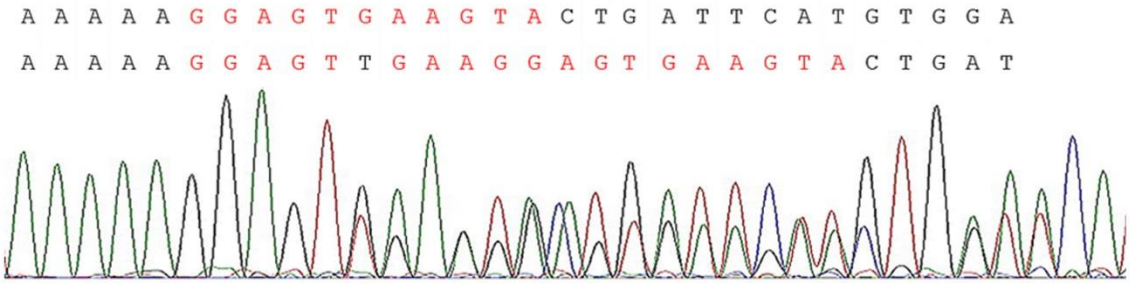

A A A A A G G A G T G A A G T A C T G A T T C A T G T G G A
A A A A A G G A G T T G A A G G A G T G A A G T A C T G A T

Figure S3

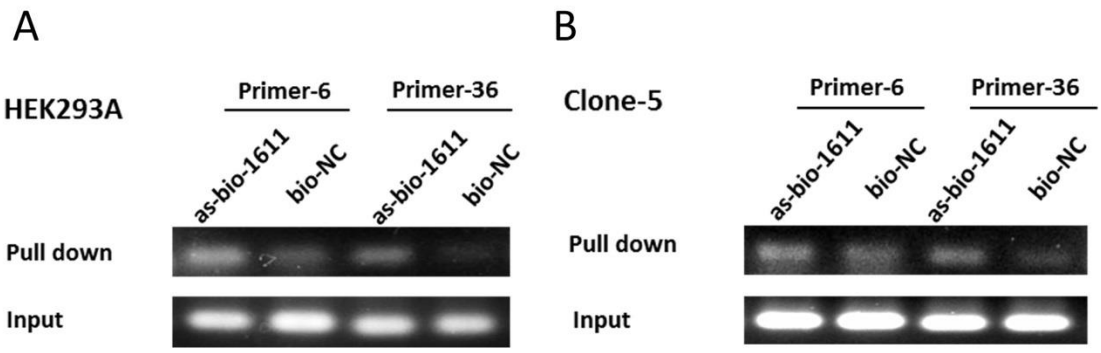

A
HEK293A
Primer-6
Primer-36
as-bio-1611
bio-NC
as-bio-1611
bio-NC
Pull down
Input
B
Clone-5
Primer-6
Primer-36
as-bio-1611
bio-NC
as-bio-1611
bio-NC
Pull down
Input

Figure S4

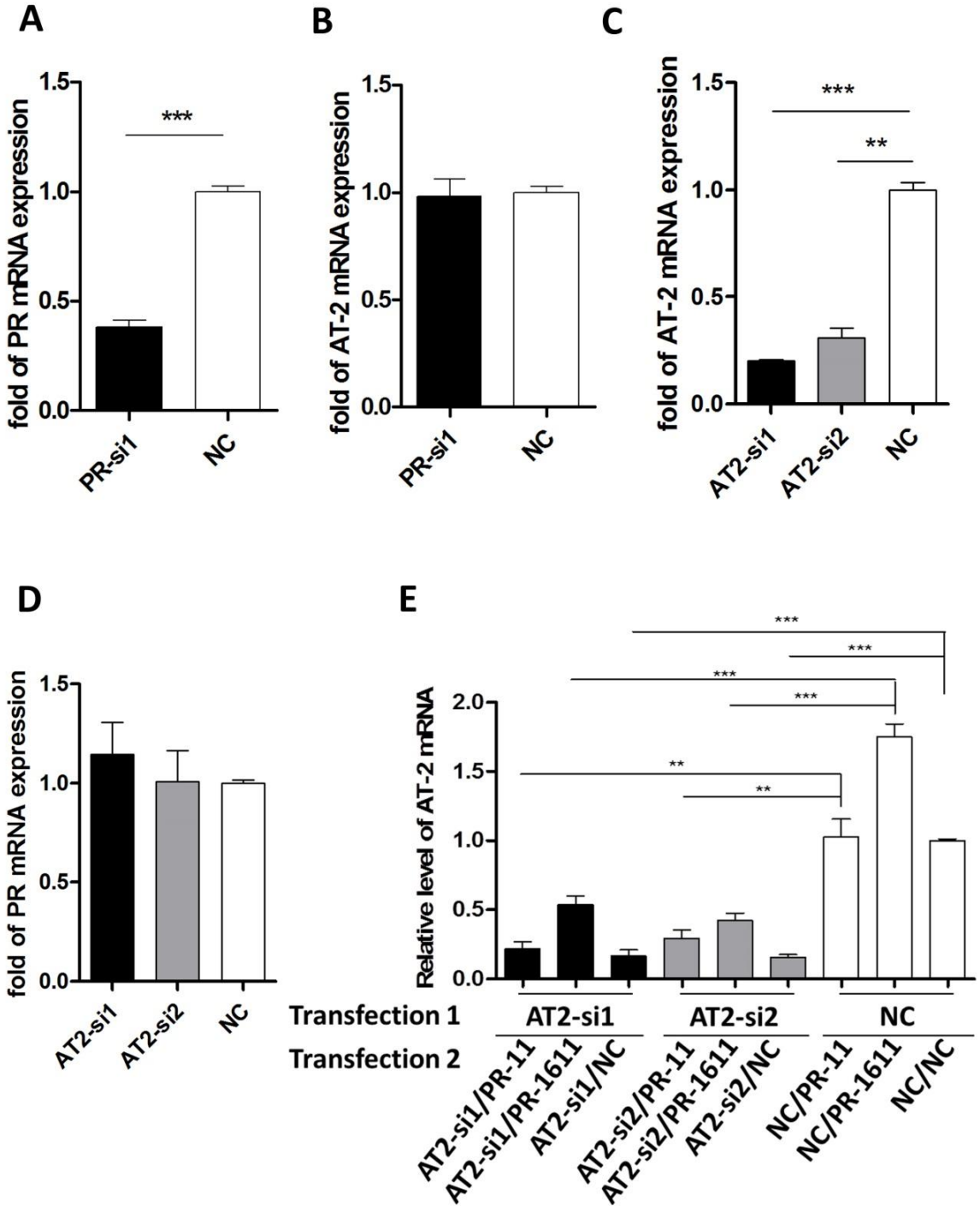
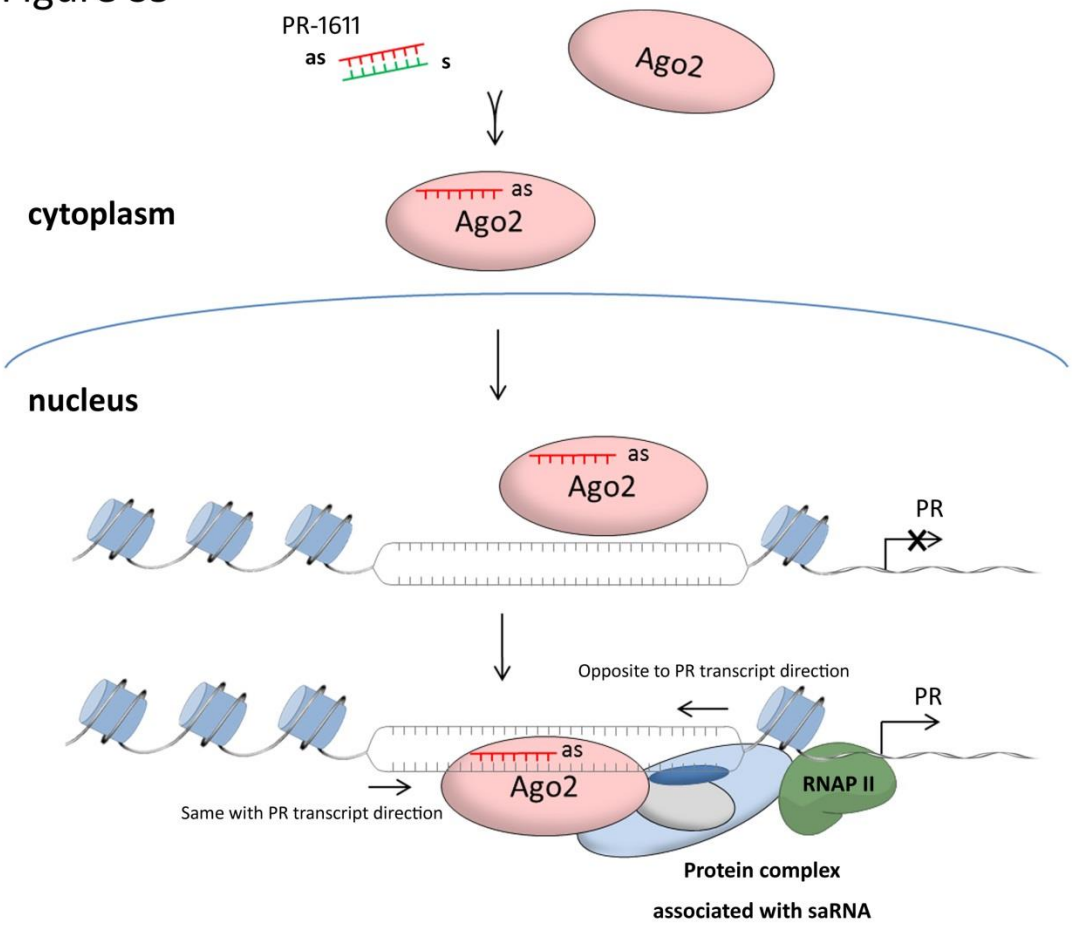

A
B
C
fold of PR mRNA expression
fold of AT-2 mRNA expression
1.5
1.0
0.5
0.0
***
**
PR-si1
NC
AT2-si1
AT2-si2
D
E
Relative level of AT-2 mRNA
2.0
Transfection 1
Transfection 2
AT2-si1
AT2-si2
NC
AT2-si1/PR-11
AT2-si1/PR-1611
AT2-si1/NC
AT2-si2/PR-11
AT2-si2/PR-1611
AT2-si2/NC
NC/PR-11
NC/PR-1611
NC/NC

Figure S5


PR-1611
as
s
Ago2
cytoplasm
as
Ago2
nucleus
as
Ago2
PR
Opposite to PR transcript direction
PR
as
Ago2
RNAP II
Same with PR transcript direction
Protein complex
associated with saRNA