

Supporting Information

The alternation of target cleavage pattern and off-target reduction of antisense oligonucleotides incorporating 2-N-(2-pyridyl)guanine

Takayuki Kanagawa,^a Aya Koyoma,^a Yoshiaki Masaki,^{*ab} Kohji Seio,^{*}

^a. Department of Life Science and Technology, Tokyo Institute of Technology, 4259-J2-16 Nagatsuta, Midoriku, Yokohama, 226-8501, Japan

^b. JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, JAPAN

Corresponding author's e-mail: kseio@bio.titech.ac.jp; ymasaki@bio.titech.ac.jp

Contents

| | |
|--|-----|
| Figure S1. RP-HPLC profiles of synthesized ASO5, 6, 7, 8 ----- | S2 |
| Figure S2. MALDI-TOF mass spectroscopy of ASO5 ----- | S3 |
| Figure S3. MALDI-TOF mass spectroscopy of ASO6 ----- | S3 |
| Figure S4. MALDI-TOF mass spectroscopy of ASO7 ----- | S4 |
| Figure S5. MALDI-TOF mass spectroscopy of ASO8 ----- | S4 |
| Figure S6. Representative melting curve and primary differential curve of hRluc targeting ASO (ASO1-4) ----- | S5 |
| Figure S7. Representative melting curve and primary differential curve of hMALAT1 targeting ASO (ASO5-8) ----- | S6 |
| Figure S8. Quantification of RNase H digestion product of hRluc targeting ASO (ASO1-4) ----- | S7 |
| Figure S9. Quantification of RNase H digestion product of hMALAT1 targeting ASO (ASO5-8) ----- | S8 |
| Figure S10. RNaseH cleavage assay of multiple pyG modification----- | S9 |
| Figure S11. On-target MALAT1 KD observed in microarray----- | S9 |
| Table S1. The number of in silico off-target genes searched by GGGenome----- | S10 |
| Table S2. Gene list whose KD was accelerated by pyG modification----- | S10 |
| Table S3. Gene list whose KD was suppressed by pyG modification----- | S11 |

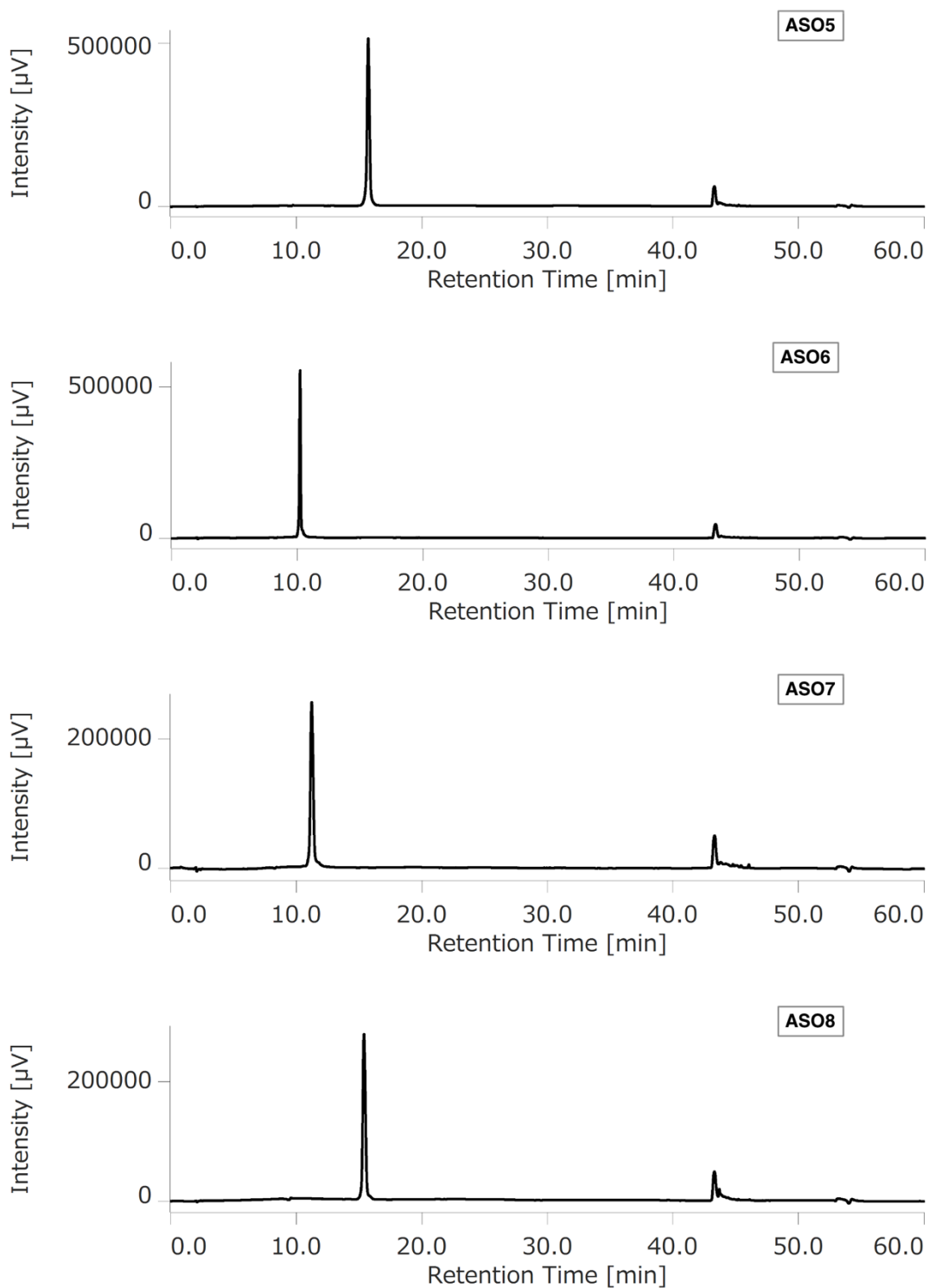


Fig. S1 RP-HPLC profiles of synthesized ASO5, 6, 7, 8. The peaks at around 43min was those eluted after the washing phase, and not of nucleic acid related compounds.

ASO5: calculated for $C_{166}H_{202}N_{61}O_{84}P_{15}S_{15}[M+H]^+$: 5342.3, found: 5342.2

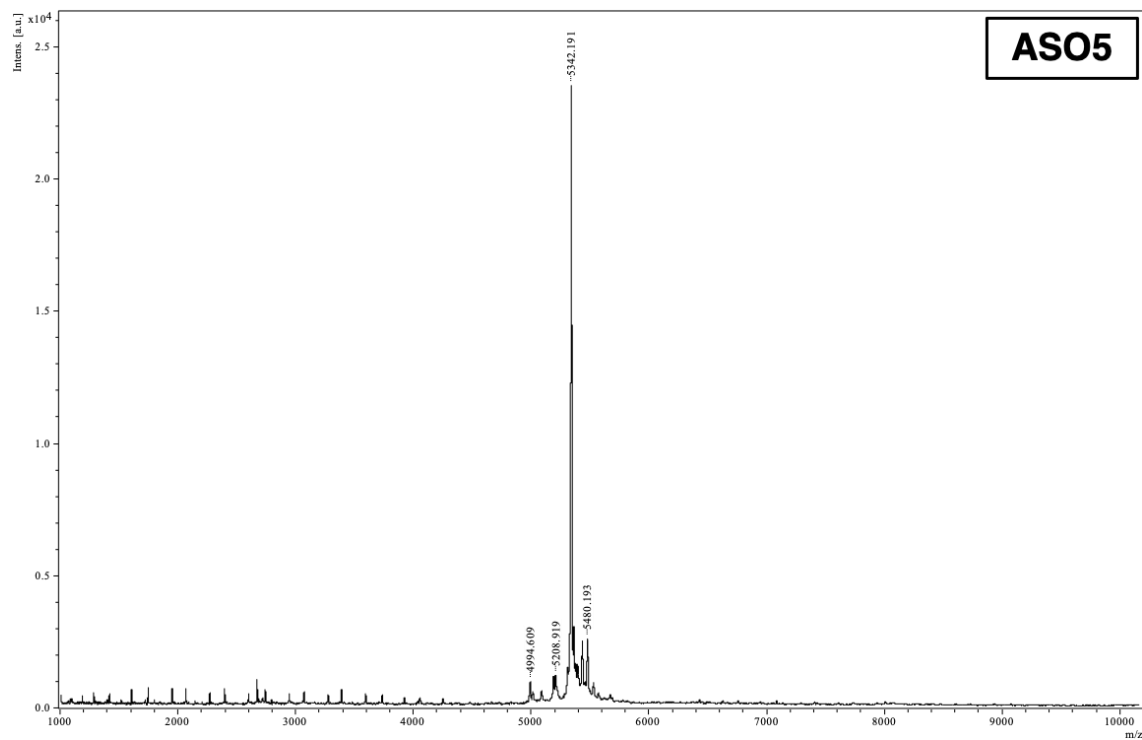


Fig. S2 MALDI-TOF mass spectroscopy of ASO5

ASO6: calculated for $C_{167}H_{203}N_{62}O_{85}P_{15}S_{15}[M+H]^+$: 5385.3, found: 5384.6

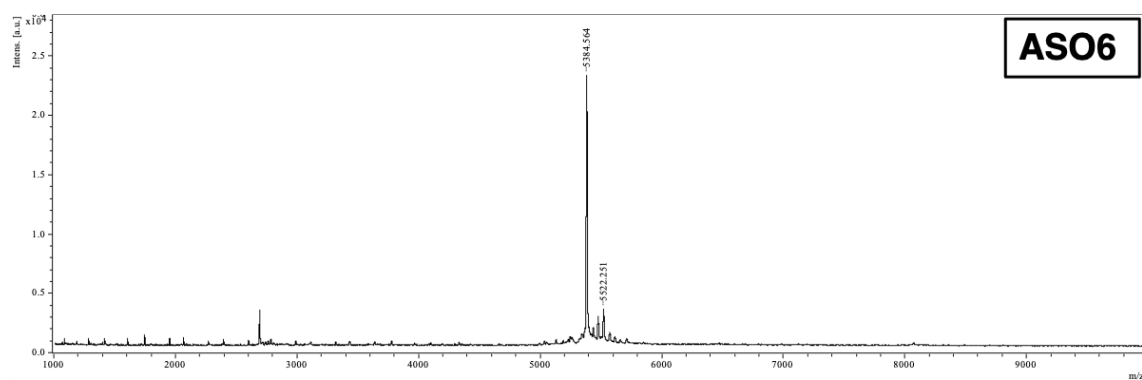


Fig. S3 MALDI-TOF mass spectroscopy of ASO6

ASO7: calculated for $C_{171}H_{205}N_{62}O_{84}P_{15}S_{15}$ $[M+H]^+$: 5419.4, found: 5419.1

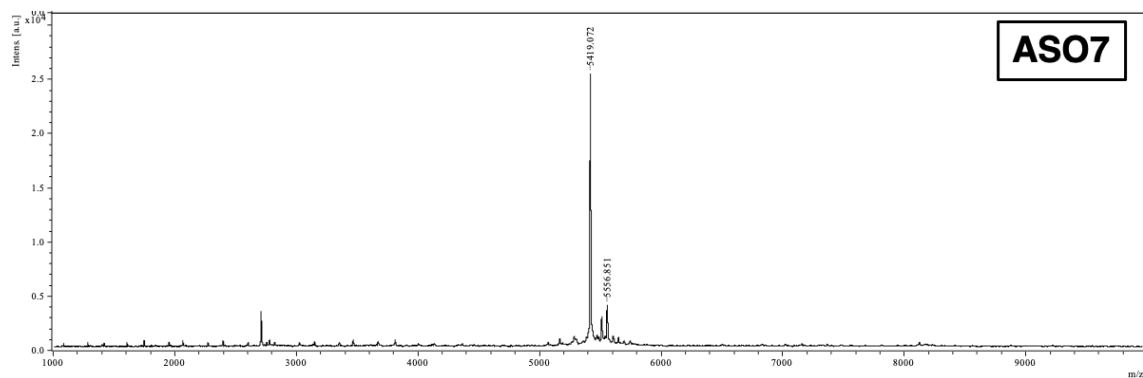


Fig. S4 MALDI-TOF mass spectroscopy of ASO7

ASO8: calculated for $C_{156}H_{190}N_{56}O_{79}P_{14}S_{14}$ $[M+H]^+$: 4997.0, found: 4998.0

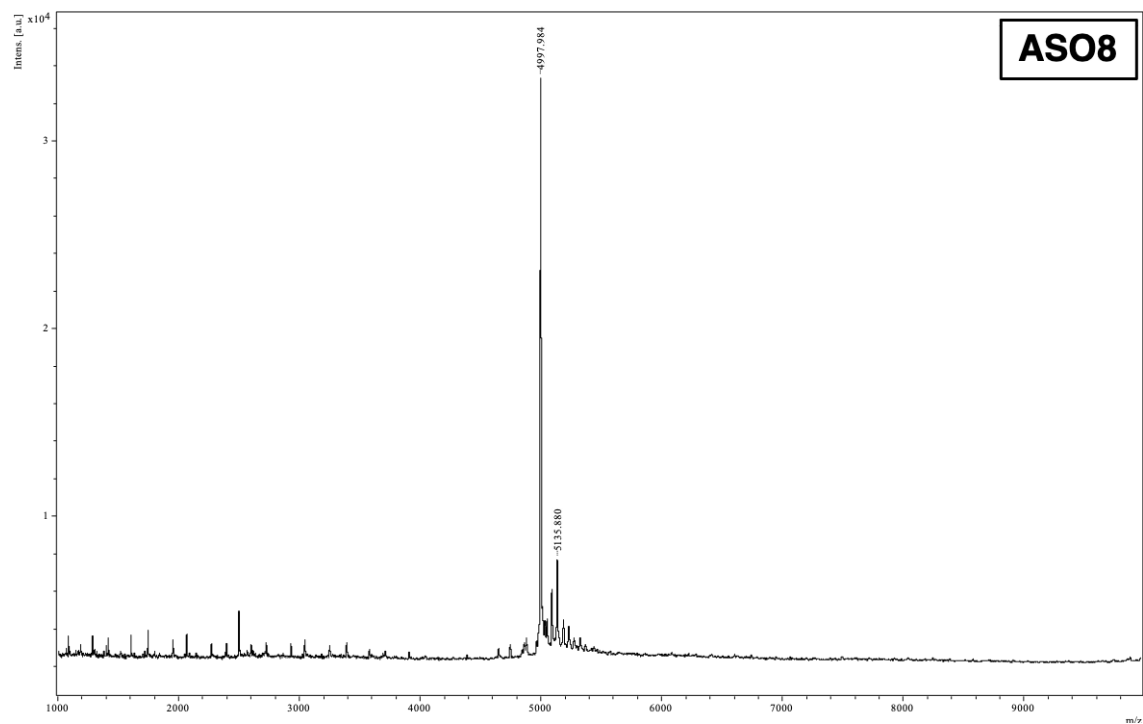


Fig.S5 MALDI-TOF mass spectroscopy of ASO8

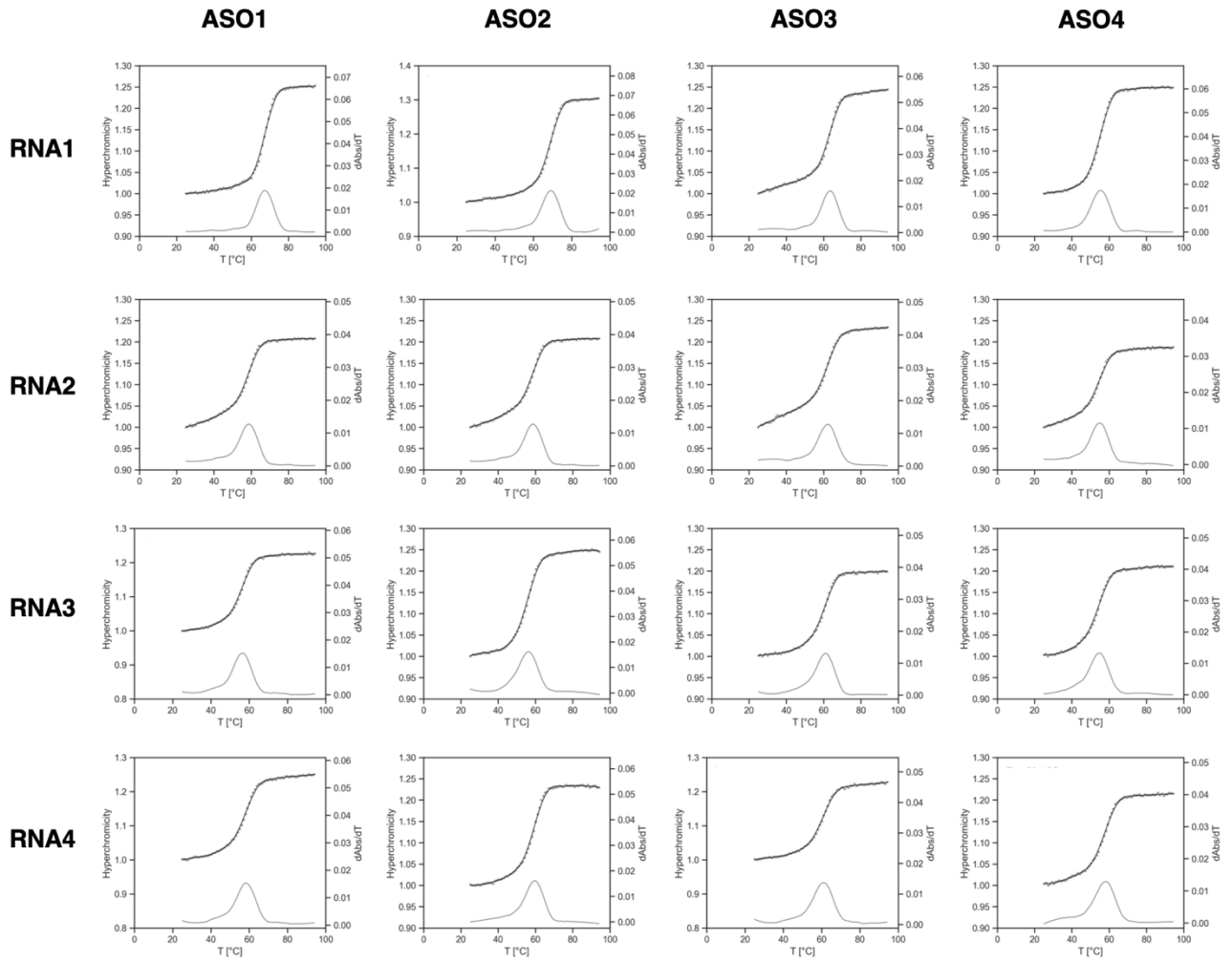


Fig.S6 Representative melting curve and primary differential curve of hRluc targeting ASO (ASO1-4)

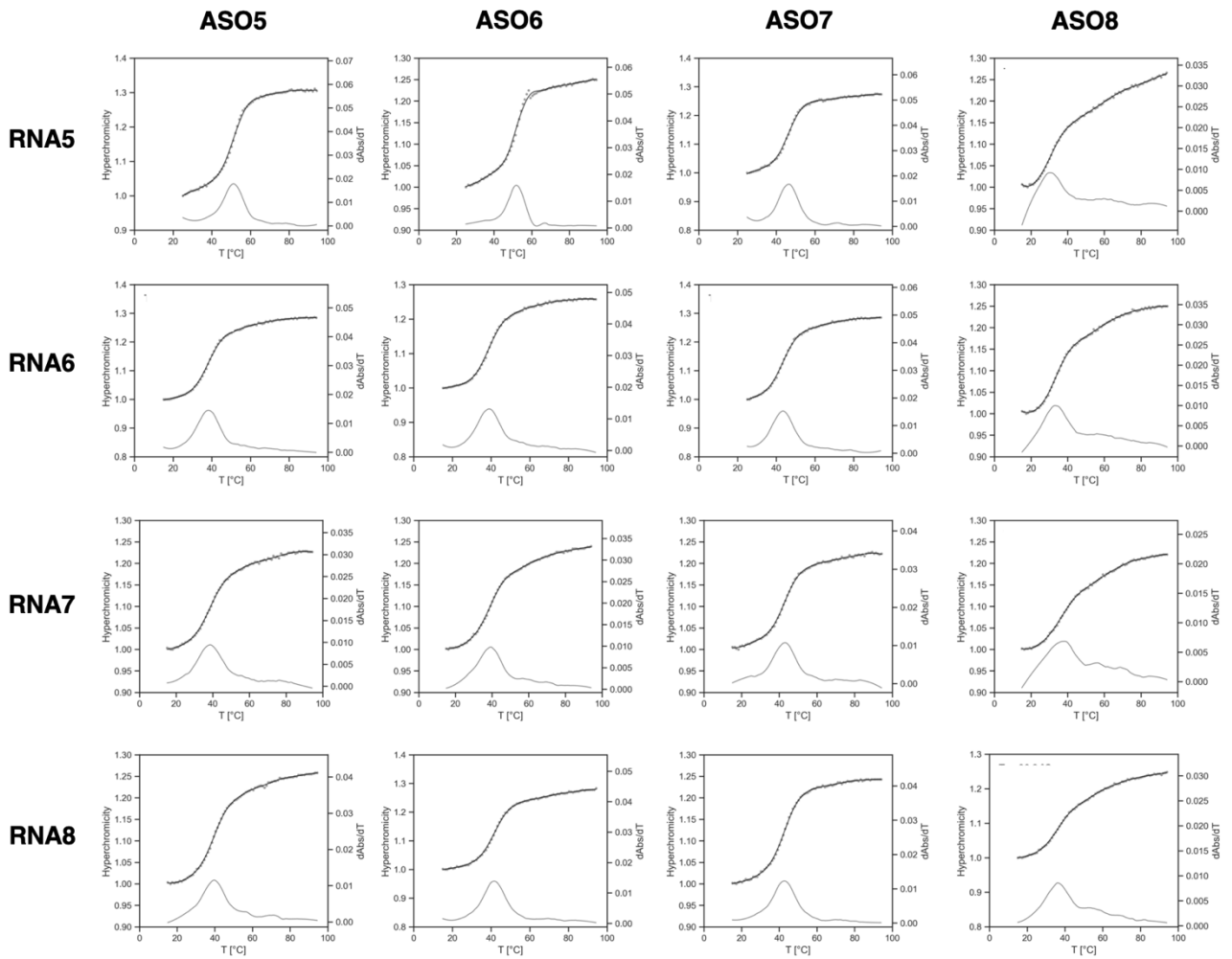


Fig.S7 Representative melting curve and primary differential curve of hMALAT1 targeting ASO (ASO5-8)

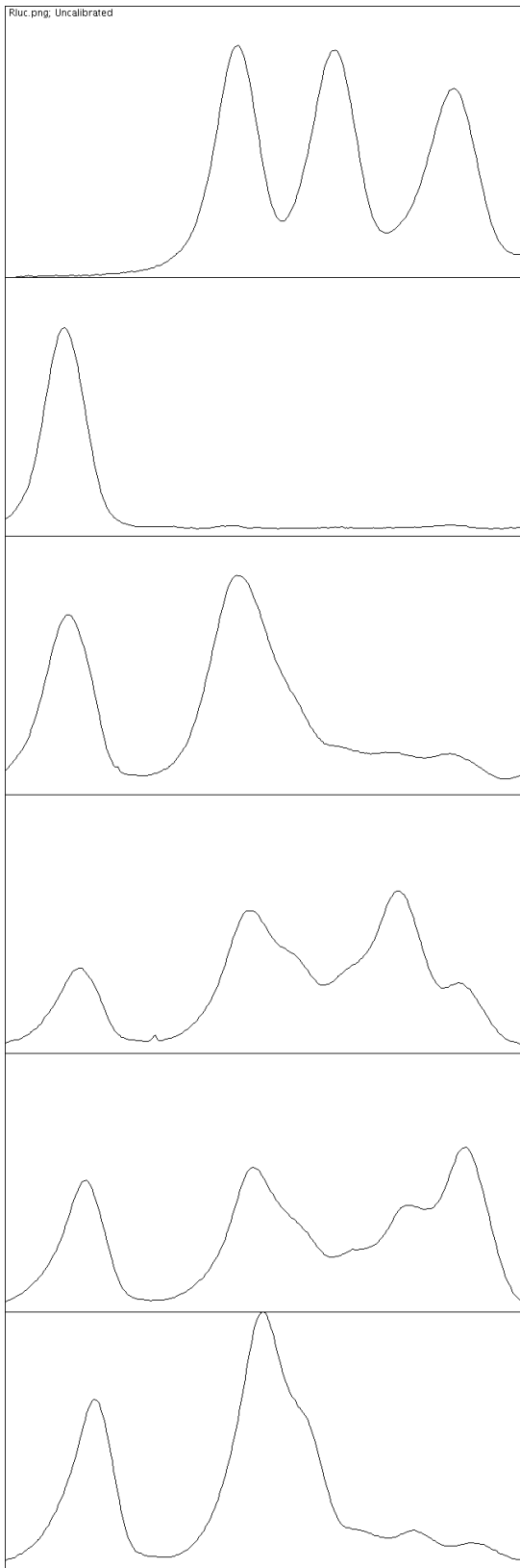


Fig. S8 Quantification of RNase H digestion product of hRluc targeting ASO (ASO1-4)

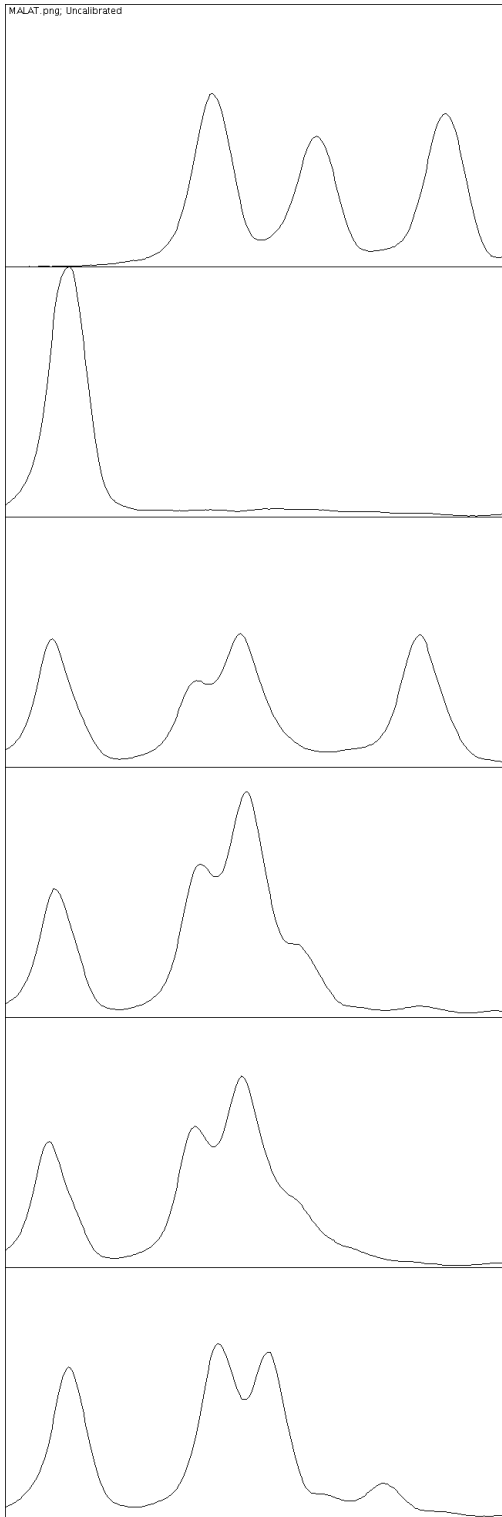
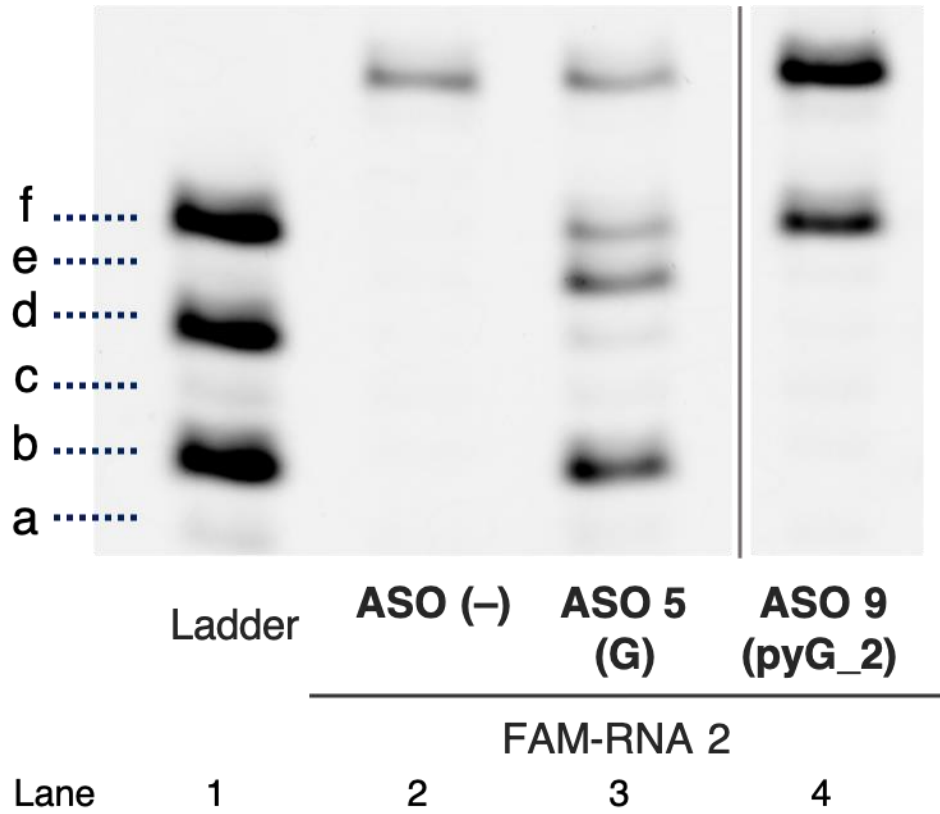


Fig. S9 Quantification of RNase H digestion product of hMALAT1 targeting ASO (**ASO5-8**)



ASO9: 5'-CATat^mcayataaTGT-3', N: LNA, y: pyG
 (calculated for C₁₇₆H₂₀₈N₆₃O₈₄P₁₅S₁₅ [M+H]⁺: 5496.5 , MALDI found: 5496.2)

Fig. S10 RNase H cleavage assay of multiple pyG modification

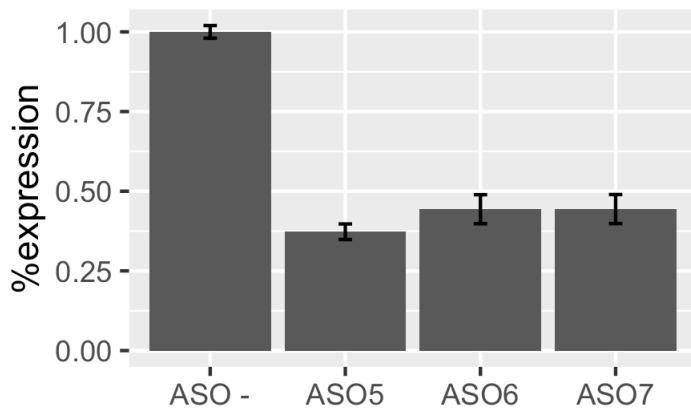


Fig. S11 On-target MALAT1 KD observed in microarray

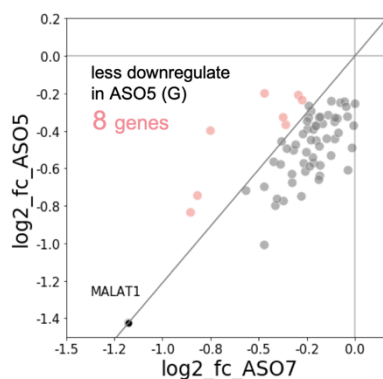
Table S1 The number of *in silico* off-target genes searched by GGGenome

| Distance | 0 | 1 | 2 | 3 |
|--------------------------------|---|----|------|-------|
| ATCATGCGGAAACTGG | 0 | 16 | 616 | 7853 |
| hRluc RNA target site | | | | |
| ATATTATCTGCATATG | 1 | 69 | 2476 | 10737 |
| hMALAT1 RNA target site | | | | |

“Human pre-spliced RNA, RefSeq curated on hg38.p12, D3G 21.01 (released Jan, 2021)” and “Human spliced RNA, RefSeq curated on hg38.p12, D3G 21.01 (released Jan, 2021)” were used as the target sequence of human unspliced RNA or spliced RNA and extract unique gene symbols.

Table S2 Gene list whose KD was accelerated by pyG modification

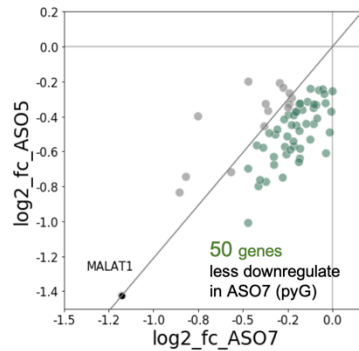
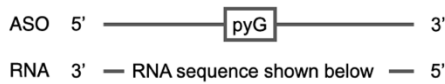
(The 8 genes in red in the figure below are the relevant genes)



| Gene Symbol | Pattern (3'-5') | RNA sequence (3'-5') |
|-------------|-----------------|----------------------|
| APPBP2 | =====DX===== | GTATAC-CCTATTACA |
| GPC5 | =X=D===== | GGA-ACGTCTATTACA |
| HS6ST3 | =====X===D=== | GTATACGTATAT-ACA |
| LOC284930 | =====X==X=== | GTATACGTCAATCACA |
| LRBA | =====X===D=== | GTATACGTATAT-ACA |
| LRP1B | X===D===== | ATAT-CGTCTATTACA |
| LRRIQ3 | ====D=====D=== | GTAT-CGTCTAT-ACA |
| RNF150 | ==X=====D===== | GTGTACGTCT-TTACA |

Table S3 Gene list whose KD was suppressed by pyG modification

(The 50 genes in green in the figure above are the relevant genes)



| Gene Symbol | Pattern (3'-5') | RNA sequence (3'-5') |
|--------------|-----------------|----------------------|
| ARHGAP10 | =====XX== | GTATACGTCTAAGACA |
| ARHGAP17 | =====I==I== | GTATACGTCTGATGTACA |
| ATF2 | =====X==D== | GTATACGTCCAT-ACA |
| BCAS3 | =====X==D= | GTATACGTCTGTTA-A |
| C15orf41 | =====I==X= | GTATACGTCTATCTAGA |
| CDH2 | =====X= | GTATACGTCTATTAGA |
| CDK14 | =X=====D===== | GAATACGT-TATTACA |
| CENPP | ====X=====X= | GTATATGTCTATTAAA |
| CHSY3 | =====DX== | GTATACGTCTA-AACA |
| CTIF | ==D===== | GT-TACGTCTATTACA |
| ERBB4 | ====X===== | GTATACTTCTATTACA |
| FAF1 | ==DX===== | GTA-GCGTCTATTACA |
| FAM133B | X===== | TTATACGTCTATTACA |
| FAM160A1 | =====X==X | GTATACGTCTATGACT |
| FAM239B | =====D=X= | GTATACGTCTAT-AAA |
| FBXO34 | =====XX== | GTATACGTCTATATCA |
| FHIT | =====D===== | GTATACGTC-ATTACA |
| FMNL2 | =====X==D== | GTATACGTATAT-ACA |
| HMBOX1 | =====D=X= | GTATACGTCTAT-AAA |
| IDE | =====X===== | GTATACGTCTAGTACA |
| LOC100288798 | =====IX== | GTATACGTCTATTTGCA |
| LOC389906 | =====D=X= | GTATACGTCTAT-AAA |
| LRMDA | ==X=====X= | GTTTACGTCTATTAAA |
| MATR3 | ==X===== | GTTTACGTCTATTACA |
| NBAS | ====D=====D== | GTATA-GTCTATT-CA |
| NR6A1 | =====D==X | GTATACGTCTAT-ACT |
| NRDC | =====I=====X | GTATACGTCTAATTACC |
| NT5C2 | =====D=D== | GTATACGTCT-T-ACA |
| PARD3 | =====X==X== | GTATACGTTTATGACA |
| PBX1 | ====X=====D== | GTATACATCTAT-ACA |
| PDE4B | ====DD===== | GTATA--TCTATTACA |
| PMF1-BGLAP | =====XX= | GTATACGTCTATTTAA |
| PMF1 | =====XX= | GTATACGTCTATTTAA |
| PPP1CB | =====X==X | GTATACGTCTAGTACT |
| RABL3 | ====X==D===== | GTACAC-TCTATTACA |
| SCFD2 | =====X==D= | GTATACGTCTTTTA-A |
| SENPI | =====I=====X | GTATACGTCTATGTACC |
| SLC4A7 | X==I===== | CTACTACGTCTATTACA |

Table continues from previous page

| Gene Symbol | Pattern (3'-5') | Target sequence (3'-5') |
|-------------|-----------------|-------------------------|
| SNTB2 | X===== | ATATACGTCTATTACA |
| SNX7 | =====D=X | GTATACGTCTAT-ACG |
| TAF3 | =====D=D= | GTATACGTCTAT-A-A |
| TBL1XR1 | =====I= | GTATACGTCTATTGACA |
| TMED5 | =====D=X | GTATACGTCTAT-ACT |
| TMEM135 | =====XX= | GTATACGTCTATGTCA |
| TTC28 | =====X=D= | GTATACGTCTTTTA-A |
| UBTD1 | X=====D= | TTATACGTCTAT-ACA |
| UST | ==D=====I= | GTA-ACGTCTATTGACA |
| VTI1A | ====D=====X= | GTATA-GTCTATTATA |
| ZDHHC14 | ====X=====X= | GTATAGGTCTATTAAA |
| ZNF839 | =====D=X= | GTATACGTCTAT-AAA |