Supplementary Results

A novel long non-coding RNA regulates the integrin, ITGA2 in breast cancer

Breast Cancer Research and Treatment

Tristan Joseph Verhoeff[1], Adele F. Holloway[2, †] and Joanne L. Dickinson[1, †,*]

[1] Menzies Institute for Medical Research, College of Health and Medicine, University of Tasmania, Hobart, 7000, Australia

[2] Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, 7000, Australia

0000-0001-5103-8592 Tristan J Verhoeff

0000-0003-2213-1717 Adele F Holloway

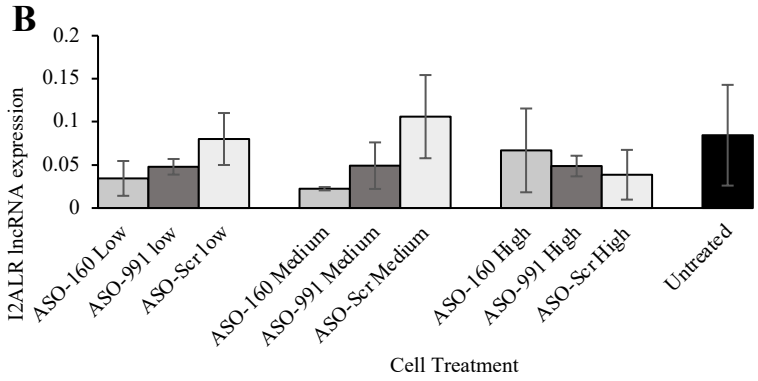
0000-0003-4621-1703 Joanne L Dickinson

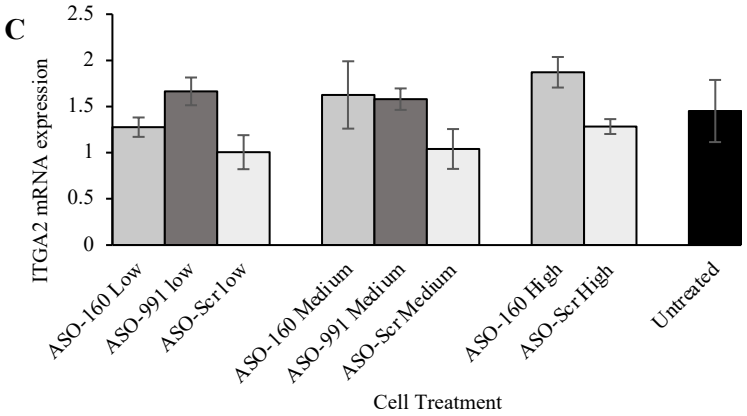
* To whom correspondence should be addressed. Tel: +61 3 62267622; Fax: +61 3 62267704; Email: jo.dickinson@utas.edu.au

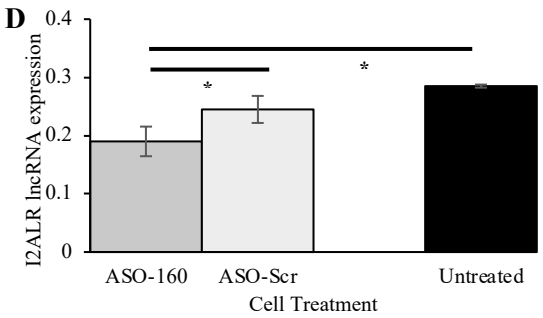
Supplementary Results Figure 1

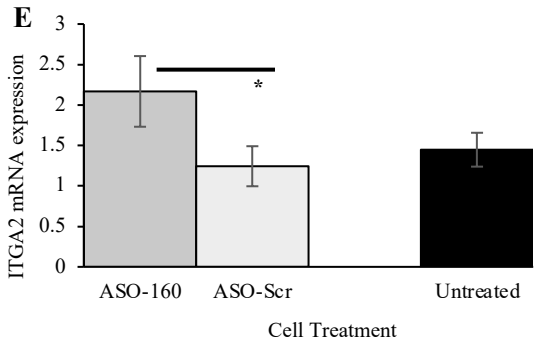
A

Exon 1 >hg38 range=chr5:52990090-52990278
ATTGCGGCATCTAGGATCCAGAAGTGAAAAGAACTTTTCCTAGCCTAAGT
TTACATTCGTCGAATAGACAATGTTACAACAATAAGAGGAAGGCAGAAGG
GAAGGAAGAATTCCCCTCTGCCAAGTATTTAAGTGGCGCTTGGTTTATTC
CAAGTCCCCCCGGGAAGAAACACAGCGCTTCGTCCTCAA
Exon 2 >hg38 range=chr5:52988204-52988276
ATCTTCATGGAATGCTACTATGAGCTGAATGGGATGCCTGAGATACTGCC
TCTAGGAAGCAAAGTGGCTCAAG
Exon 3 >hg38 range=chr5:52978252-52978399
AGGCTGGCTGGCAAGATGGCCGAATAGGAACAGCTCCAGTCTGCAGCTCC
TAGTGAGATCAACGCCAAAGGCGGGTGACTTCTGCATTTCCAACTGAGAT
ACCCAGCTCATCTCACTGGGACTGGTTAGACA **GTGGGTGCAGCCCATG**
Exon 4 >hg38 range=chr5:52948825-52948907
AGGTACAAGGCCTGAAAGAACCTATCCAAATATCAGCTCTGGTTTTCCTC
TTCTCCTGCCAGCGGTGTTCTAAAACCATAGAG
Exon 5 >hg38 range=chr5:52932419-52932536
ATTGAGAAGACTGATCAGCTTCCACAGGATGAGAACGATGACCAAGAAAT
ATGCTGGCGGCCTCTAAGAAAATATCGTGTTTTGCGAAGGACTTCTGGGT
CTTCAACTTGCATTCGGG
U3R >hg38 range=chr5:52931419-52932418
aattaagactggacagcaaatgcaatccggtatcactaacacctaacact
ttgggttatttagtataatggttctcaaacattaggatgcatgagaatta
cctggaaggcttgtttaaccccatcctccgagtttctgatttggtaggtt
tctggtgggggccagagaatttgcctaacaagtttccaggtgatgctgatg
cccttggccaggcagggaccacactttcagaaccactgggttaatagaaa
ggcatccacatataaaaatgtagttttctcctacctgtaaaactgtagtcca
acttcatattttaccgggatagaaatgtttactacattatcagaaagggt
ttcaggaggttcttcgctgtcactagaaaatacacggcacattcagacaa
accattaaaatcttgactatttaaagacatttctaagtgtttgaaagaga
agcttatcctggcaactaaataaaagcatgtaagagtataatttttaatc
a-poly(A)





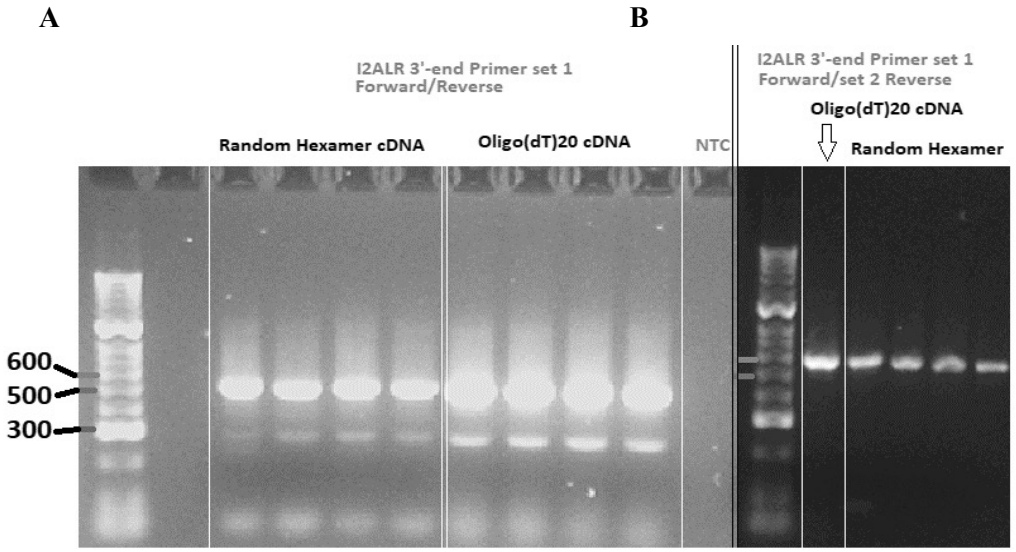




Supplementary Fig. 1. (**A**) Sequence for *I2ALR* combined exons (*ENST00000505701.5* + *ENST00000503559.1*, orange) and downstream U3R including poly-A signal (blue, signal highlighted red), the recognition sites for ASOs (ASO-160 and ASO-991) are highlighted light-blue and yellow respectively, with primers used to amplify the U3R underlined. (**B**) MDA-MB-453 cell I2ALR expression following ASO-160 or ASO-991 transfection after 24-hours at 10 nM (low), 25 nM (medium) and 50 nM (high). (**C**) *ITGA2* expression after 24-hours transfection at ASO concentrations low, medium, and high. (**D**) *I2ALR* expression in MD-MB-453 cells at 48-hours following transfection at 10nM ASO-160 (**E**) *ITGA2* expression after 48-hour following transfection with ASO-160 at 10nM. (**B, C**) n = 1 biological replicates, statistical significance not calculated; (**D, E**) n = 3 biological replicates, expression absolute copy number normalized to *GAPDH*, ±SEM. Statistical significance (*P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.

Supplementary Results Figure 2 : Characterisation of I2LAR

Sequence analysis of the most 3' *I2ALR* exon revealed two putative poly-adenylation motifs. A classical motif (*fide* Proudfoot, 2011) with 5'-AATAAA signal (469−474 bp downstream from the exon) with a 5'-CA cleavage site 26 bp further downstream. In addition, a poly-A enhancing G/U rich region was identified 3' to the cleavage site and a uracil rich region 5' to the motif (Figure 5 B). A non-classical poly-A signal was identified 404−429 bp downstream from the exon. To characterise the 3'- poly-A signal, a DNA fragment was amplified employing a primer anchored in the last known exon to several sites adjacent to the poly-A signals. A 511 bp fragment was amplified ('*I2ALR* 3'-end primer set 1') from cDNA generated using oligo$(dT)_{20}$ (Supplementary Results Figure 2A). Sanger sequencing indicated that exon 5 of *I2ALR* extended at least 405 bp further downstream without additional splicing (Figure 5 B). To distinguish between the classical and non-classical poly-A motifs, the 'I2ALR 3'-end set 1' forward primer was combined with the 'I2ALR 3'-end set 2' reverse primer (primer sequences Supplementary Methods Table 2) so that the reverse primer binding site was located between the two motifs (as shown Figure 5 B). This was performed using cDNA generated with oligo$(dT)_{20}$ or random hexamers, and a product was generated, matching the expected 565 bp size (Supplementary Results Figure 2 B, below), suggesting that the classical poly-A signal was functional, and *I2ALR* was polyadenylated with its 3' most exon 619 bp in length (designated as U3R). The I2ALR transcription start site (TSS) and promoter are located 500−600 bp downstream from the ITGA2 first exon, however boundaries of the promoter elements and precise lncRNA TSS are not defined.



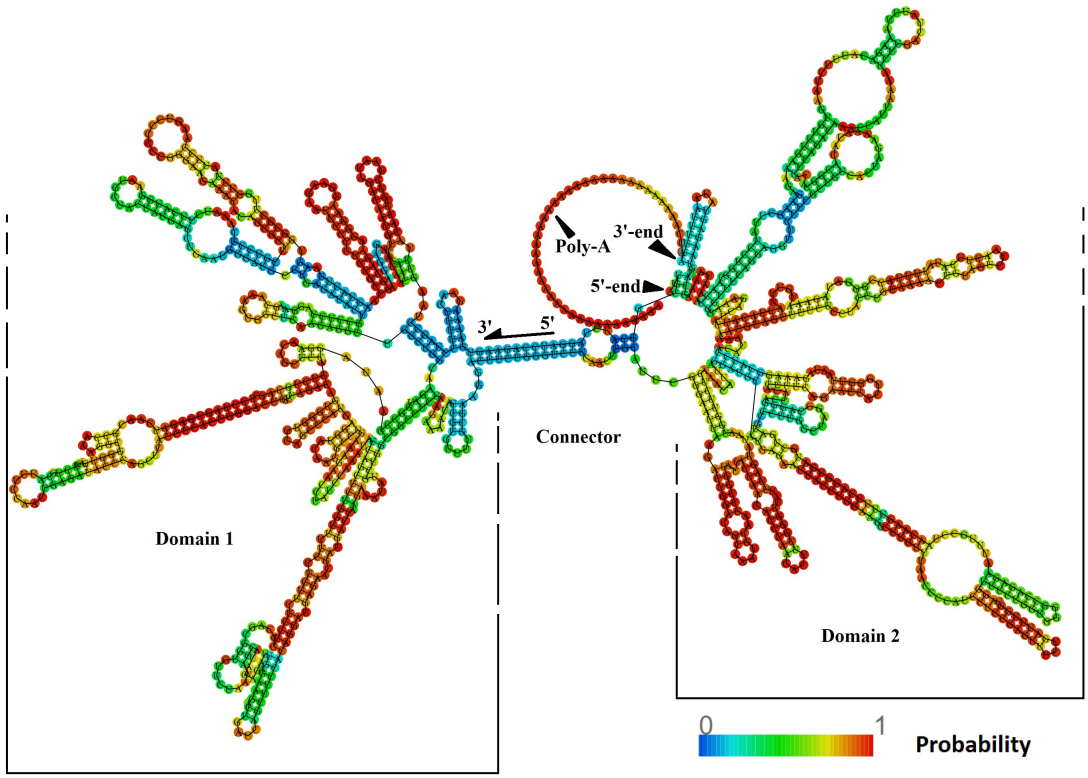
Supplementary Figure 2 : Characterisation of the *I2ALR* poly-adenylation region from the last exon to regions near putative poly-A signals by PCR amplification. **A**) Using the 'I2ALR 3'-end primer set 1' a ~500 bp product from poly-A transcripts was generated. (**B**) Combining primer set 1 − set 2 reverse primer, between a classical and non-classical poly-A signal, generated the expected ~560 bp product from poly-A transcripts. Gel ladder with 300, 500 and 600 bp bands marked.).

Supplementary Results Figure 3

A

```
Energy = -19.26 kcal/mol (9 ideal base pairings)
      32                                                          89
       |                                                          |
5'-AGAGAAGCCCCUCUGGACAGCUUCU-AGAGUGUGCAGGUUCUCGUAUCCCCUCGGCCAAGGGUAUCCUCUG-3' ITGA2
       |:|||  |||     |||| |:|||||   ||| |:|     ||| |||||||||:|||
3'-GAUAAUUGGGUCACC-----AAGACUUUCACA---CCA-GGGAC---GGA-CCGGUUCCCGUAGUCGUA-5' LncRNA I2ALR
       |                                                          |
     853                                                         808

Energy = -21.37 kcal/mol (6 ideal base pairings)
    3837                                                      3885
       |                                                      |
5'-AAUGAAGAAAUUGUGGGGGGUGGGGGAGGUGCGGGGGG--CAGGUAGGGAA-3' ITGA2
       |||||:|  ||:||:|||||  ::  |::|     ||||||:
3'-UUUAGUCUUUGAG-CCUCCUACCCCAAUUU-GUUCGGAAGGUCCAUUAAGA-5' LncRNA I2ALR
       |                                                      |
     748                                                     710

Energy = -14.81 kcal/mol (9 ideal base pairings)
    4548                                          4589
       |                                          |
5'-UAGGAUUUGUUUGGCUGACUGGCAG-UAACCUAGUGAAUUUCUGAAAGAUGAG-3' ITGA2
     :  |||:| ||| || ||  |   |||||:||||  |||||||||
3'-UGUAAAAAUACACCUACGGAAAGAUAAUUGGGUCACC--AAGACUUUCACACC-5' LncRNA I2ALR
       |                                          |
     875                                         835

Energy = -14.01 kcal/mol (7 ideal base pairings)
    5506            5518
       |            |
5'-GACUACCCUGGCCUGGCCCGUGG-3' ITGA2
       ||||| ||||||| :
3'-CACCAGGGAC-GGACCGGUUCCC-5' LncRNA I2ALR
       |            |
     828            817

Energy = -15.24 kcal/mol (8 ideal base pairings)
    7305            7319
       |            |
5'-UACAUUCAGGCAUACCCAUUUUAAU-3' ITGA2
     :|||||||| ||||||:
3'-CAUAGAGUCCGUA-GGGUAAGUCGA-5' LncRNA I2ALR
       |            |
     230            217
```

B



Supplementary Figure 3. Bioinformatic prediction of I2ALR lncRNA interactions with ITGA2 mRNA and I2ALR secondary structure**.** (**A**) "IntaRNA" predicted complementary interactions between *ITGA2* mRNA (7869 bp) and 1112 bp *I2ALR* sequence (combined exons of ENST00000505701.5 + ENST00000503559.1 and uncharacterized 3' region or U3R). Six potential interactions were predicted with complementary areas of > 6 perfect base pairings, with associated energy scores indicated, '|' and ':' represent Watson-Crick (ideal) and wobble base pairings respectively. (**B**) "RNAfold" predicted *I2ALR* secondary structure (minimum free energy), predicted structure from the 5 known exons and extended region down to the poly-A signal, base pair probabilities range from 0−1 (dark blue−red), when unpaired the colour denotes probability of being unpaired.