Molecular Cell, Volume *27*

Supplemental Data

MicroRNA Targeting Specificity in Mammals:

Determinants Beyond Seed Pairing

Andrew Grimson, Kyle Kai-How Farh, Wendy K. Johnston, Philip Garrett-Engele, Lee P. Lim, David P. Bartel

Supplemental Discussion

When examining conserved miRNA sites for favorable UTR contexts, we used the signal above background (sites selectively maintained) instead of the signal:background ratio. Signal above background indicates the selection of miRNA sites per UTR or segment of UTR, and thus it is the relevant measure of whether sites in a given context are more frequently subject to selection than sites in other contexts. In contrast, signal:background ratio (frequently also called the signal:noise ratio) is a measure of how well sites under selection can be distinguished from those conserved by chance. In UTR contexts that are enriched for conserved miRNA sites, there is also typically an increase in the conservation of other sequences that do not correspond to miRNA sites. In these circumstances, the signal above background improves because the increase in conserved miRNA sites outpaces the increase in conserved sequences that do not correspond to miRNA sites, but the signal:background ratio can remain constant or drop, despite the higher density of evolutionary selection for miRNA targeting. As a consequence, miRNA sites under selection can be paradoxically more difficult to predict with confidence when in favorable contexts because they tend to be associated with more background conservation than miRNA sites in poor contexts. For instance, Lewis et al. (2005) showed that within more highly conserved UTRs, the number of conserved miRNA sites above background increases, but the signal:background ratio drops because of the increase in background conservation.

Two phenomena can explain the association of greater background conservation with favorable context determinants. First, the selection acting on a miRNA site also acts to preserve the favorable context of the site, causing greater conservation in the vicinity of the site, although more limited than that for the site itself. This effect is compounded when a UTR has multiple conserved miRNA sites, and most UTRs with a conserved site to one miRNA family do have conserved sites to one or more additional miRNA families. Second, some UTR context determinants that encourage miRNA effectiveness likely generalize to RNA-protein interactions (e.g., to improve site accessibility), and a UTR regulated by miRNAs might also preferentially be regulated by proteins. As a result, these context determinants are associated with the conserved sequences that do not match miRNAs but can match the control sequences used to estimate the background conservation.

Potential Correlations with Microarray Signal

We addressed the question of whether the level of mRNA expression, as measured by the intensity of the microarray signal, might correlate with and thereby confound interpretation of the specificity determinants. We examined the Spearman correlations for intensity vs. fold-change, considering each of the different canonical sites. The results were as follows:

8mer: rho = –0.093, $P = 0.0020$
7mer-m8: rho = –0.060, $P = 0.00098$
7mer-A1: rho = –0.037, $P = 0.042$
6mer: rho = –0.049, $P = 0.00013$

Thus downregulation was weakly correlated with higher intensity on the chip, presumably because the higher the intensity of the gene on the chip, the more likely that the gene is actually expressed in the cell, a prerequisite for downregulation. We minimized this effect by selecting for analysis only those genes that were expressed above median on the array, because genes expressed at levels lower than that would have a much smaller likelihood of going down. As a result, context determinants of our analysis were not correlated in a manor that would be a concern. For example, intensity and conservation are not significantly correlated. The other determinant that could have potentially been a concern was the local AU-effect. However, intensity and overall AU-richness were correlated in the wrong direction to explain the AU-effects (rho = –0.0454, indicating that AU-rich genes are expressed at slightly lower levels), and thus this was also not a concern.

Table S1. Vertebrate miRNA families. Listed are 73 human miRNAs, each representing a different family conserved in human, mouse, rat, dog and zebrafish.

Family	7mer-m8 site	Representative sequence	miRBase name	Accession
let-7	CUACCUC	UGAGGUAGUAGGUUGUAUAGUU	hsa-let-7a	MIMAT0000062
miR-1	ACAUUCC	UGGAAUGUAAAGAAGUAUGUA	hsa-miR-1	MIMAT0000416
miR-7	GUCUUCC	UGGAAGACUAGUGAUUUUGUUG	hsa-miR-7	MIMAT0000252
miR-9	ACCAAAG	UCUUUGGUUAUCUAGCUGUAUGA	hsa-miR-9	MIMAT0000441
miR-10	ACAGGGU	UACCCUGUAGAUCCGAAUUUGUG	hsa-miR-10a	MIMAT0000253
miR-15	UGCUGCU	UAGCAGCACAUAAUGGUUUGUG	hsa-miR-15a	MIMAT0000068
miR-17	GCACUUU	CAAAGUGCUUACAGUGCAGGUAGU	hsa-miR-17-5p	MIMAT0000070
miR-18	GCACCUU	UAAGGUGCAUCUAGUGCAGAUA	hsa-miR-18a	MIMAT0000072
miR-19	UUUGCAC	UGUGCAAAUCUAUGCAAAACUGA	hsa-miR-19a	MIMAT0000073
miR-21	AUAAGCU	UAGCUUAUCAGACUGAUGUUGA	hsa-miR-21	MIMAT0000076
miR-22	GGCAGCU	AAGCUGCCAGUUGAAGAACUGU	hsa-miR-22	MIMAT0000077
miR-23	AAUGUGA	AUCACAUUGCCAGGGAUUUCC	hsa-miR-23a	MIMAT0000078
miR-24	CUGAGCC	UGGCUCAGUUCAGCAGGAACAG	hsa-miR-24	MIMAT0000080
miR-26	UACUUGA	UUCAAGUAAUUCAGGAUAGGUU	hsa-miR-26b	MIMAT0000083
miR-27	ACUGUGA	UUCACAGUGGCUAAGUUCUGC	hsa-miR-27b	MIMAT0000419
miR-29	UGGUGCU	UAGCACCAUCUGAAAUCGGUU	hsa-miR-29a	MIMAT0000086
miR-30	UGUUUAC	UGUAAACAUCCUCGACUGGAAG	hsa-miR-30a-5p	MIMAT0000087
miR-31	AUCUUGC	GGCAAGAUGCUGGCAUAGCUG	has-mir-31	MIMAT0000089
miR-33	CAAUGCA	GUGCAUUGUAGUUGCAUUG	hsa-miR-33	MIMAT0000091
miR-34a	CACUGCC	UGGCAGUGUCUUAGCUGGUUGUU	hsa-miR-34a	MIMAT0000255
miR-92	GUGCAAU	UAUUGCACUUGUCCCGGCCUG	hsa-miR-92	MIMAT0000092
miR-93	AGCACUU	AAAGUGCUGUUCGUGCAGGUAG	hsa-miR-93	MIMAT0000093
miR-96	GUGCCAA	UUUGGCACUAGCACAUUUUUGC	hsa-miR-96	MIMAT0000095
miR-99	UACGGGU	AACCCGUAGAUCCGAUCUUGUG	hsa-miR-99a	MIMAT0000097
miR-101	GUACUGU	UACAGUACUGUGAUAACUGAAG	hsa-miR-101	MIMAT0000099
miR-103	AUGCUGC	AGCAGCAUUGUACAGGGCUAUGA	hsa-miR-103	MIMAT0000101
miR-122	ACACUCC	UGGAGUGUGACAAUGGUGUUUGU	hsa-miR-122a	MIMAT0000421
miR-124	GUGCCUU	UAAGGCACGCGGUGAAUGCCA	hsa-miR-124a	MIMAT0000422
miR-125	CUCAGGG	UCCCUGAGACCCUUUAACCUGUG	hsa-miR-125a	MIMAT0000443
miR-126	CGGUACG	UCGUACCGUGAGUAAUAAUGC	hsa-miR-126	MIMAT0000445
miR-128	CACUGUG	UCACAGUGAACCGGUCUCUUUU	hsa-miR-128a	MIMAT0000424
miR-129	GCAAAAA	CUUUUUGCGGUCUGGGCUUGC	hsa-miR-129	MIMAT0000242
miR-130	UUGCACU	CAGUGCAAUGUUAAAAGGGCAU	hsa-miR-130a	MIMAT0000425
miR-132	GACUGUU	UAACAGUCUACAGCCAUGGUCG	hsa-miR-132	MIMAT0000426
miR-133	GGGACCA	UUGGUCCCCUUCAACCAGCUGU	hsa-miR-133a	MIMAT0000427
miR-135	AAGCCAU	UAUGGCUUUUUAUUCCUAUGUGA	hsa-miR-135a	MIMAT0000428
miR-137	AAGCAAU	UAUUGCUUAAGAAUACGCGUAG	hsa-miR-137	MIMAT0000429
miR-138	CACCAGC	AGCUGGUGUUGUGAAUC	hsa-miR-138	MIMAT0000430
miR-139	ACUGUAG	UCUACAGUGCACGUGUCU	hsa-miR-139	MIMAT0000250
miR-140	AAACCAC	AGUGGUUUUACCCUAUGGUAG	hsa-miR-140	MIMAT0000431
miR-141	CAGUGUU	UAACACUGUCUGGUAAAGAUGG	hsa-miR-141	MIMAT0000432
miR-142	ACACUAC	UGUAGUGUUUCCUACUUUAUGGA	hsa-miR-142-3p	MIMAT0000434
miR-143	UCAUCUC	UGAGAUGAAGCACUGUAGCUCA	hsa-miR-143	MIMAT0000435
miR-144	AUACUGU	UACAGUAUAGAUGAUGUACUAG	hsa-miR-144	MIMAT0000436
miR-145	AACUGGA	GUCCAGUUUUCCCAGGAAUCCCUU	hsa-miR-145	MIMAT0000437
miR-146	AGUUCUC	UGAGAACUGAAUUCCAUGGGUU	hsa-miR-146a	MIMAT0000449
miR-148	UGCACUG	UCAGUGCACUACAGAACUUUGU	hsa-miR-148a	MIMAT0000243
miR-150	UUGGGAG	UCUCCCAACCCUUGUACCAGUG	hsa-miR-150	MIMAT0000451
miR-153	CUAUGCA	UUGCAUAGUCACAAAAGUGA	hsa-miR-153	MIMAT0000439
miR-181	UGAAUGU	AACAUUCAACGCUGUCGGUGAGU	hsa-miR-181a	MIMAT0000256
miR-182	UUGCCAA	UUUGGCAAUGGUAGAACUCACA	hsa-miR-182	MIMAT0000259
miR-183	GUGCCAU	UAUGGCACUGGUAGAAUUCACUG	hsa-miR-183	MIMAT0000261
miR-184	UCCGUCC	UGGACGGAGAACUGAUAAGGGU	hsa-miR-184	MIMAT0000454
miR-187	AGACACG	UCGUGUCUUGUGUUGCAGCCG	hsa-miR-187	MIMAT0000262

miR-192	UAGGUCA	CUGACCUAUGAAUUGACAGCC	hsa-miR-192	MIMAT0000222
miR-193	GGCCAGU	AACUGGCCUACAAAGUCCCAG	hsa-miR-193a	MIMAT0000459
miR-194	CUGUUAC	UGUAACAGCAACUCCAUGUGGA	hsa-miR-194	MIMAT0000460
miR-196	ACUACCU	UAGGUAGUUUCAUGUUGUUGG	hsa-miR-196a	MIMAT0000226
miR-199	ACACUGG	CCCAGUGUUCAGACUACCUGUUC	hsa-miR-199a	MIMAT0000231
miR-200b	CAGUAUU	UAAUACUGCCUGGUAAUGAUGAC	hsa-miR-200b	MIMAT0000318
miR-203	CAUUUCA	GUGAAAUGUUUAGGACCACUAG	hsa-miR-203	MIMAT0000264
miR-204	AAAGGGA	UUCCCUUUGUCAUCCUAUGCCU	hsa-miR-204	MIMAT0000265
miR-205	AUGAAGG	UCCUUCAUUCCACCGGAGUCUG	hsa-miR-205	MIMAT0000266
miR-210	ACGCACA	CUGUGCGUGUGACAGCGGCUGA	hsa-miR-210	MIMAT0000267
miR-214	CCUGCUG	ACAGCAGGCACAGACAGGCAG	hsa-miR-214	MIMAT0000271
miR-216	UGAGAUU	UAAUCUCAGCUGGCAACUGUG	hsa-miR-216	MIMAT0000273
miR-217	AUGCAGU	UACUGCAUCAGGAACUGAUUGGAU	hsa-miR-217	MIMAT0000274
miR-218	AAGCACA	UUGUGCUUGAUCUAACCAUGU	hsa-miR-218	MIMAT0000275
miR-219	GACAAUC	UGAUUGUCCAAACGCAAUUCU	hsa-miR-219	MIMAT0000276
miR-221	AUGUAGC	AGCUACAUUGUCUGCUGGGUUUC	hsa-miR-221	MIMAT0000278
miR-223	AACUGAC	UGUCAGUUUGUCAAAUACCCC	hsa-miR-223	MIMAT0000280
miR-338	AUGCUGG	UCCAGCAUCAGUGAUUUUGUUGA	hsa-miR-338	MIMAT0000763
miR-375	CGAACAA	UUUGUUCGUUCGGCUCGCGUGA	hsa-miR-375	MIMAT0000728

Table S2. Synthetic miRNA duplexes used in microarray transfection experiments.

miR-1
```
  5'-UGGAAUGUAAAGAAGUAUGUA-3'
     | ||||||||||||||||||
3'-AUAACUUACAUUUCUUCAUAC-5'
```

miR-7
```
  5'-UGGAAGACUAGUGAUUUUGUU-3'
     |••||||||||||||||||||
3'-UAAUUUUCUGAUCACUAAAAC-5'
```

miR-9
```
  5'-UCUUUGGUUAUCUAGCUGUAUGA-3'
     | ||||||||||||||||||||||
3'-UAACAAACCAAUAGAUCGACAUA-5'
```

miR-122
```
  5'-UGGAGUGUGACAAUGGUGUUUGU-3'
     |••||||||||||||||||||||||
3'-UAAUUUCACACUGUUACCACAAA-5'
```

miR-124
```
  5'-UAAGGCACGCGGUGAAUGCCA-3'
     ||||||||||||||||||||||
3'-UAAUUCCGUGCGCCACUUACG-5'
```

miR-128
```
  5'-UCACAGUGAACCGGUCUCUUUU-3'
     | ||||||||||||||||||||
3'-UAACUGUCACUUGGCCAGAGAA -5'
```

miR-132
```
  5'-UAACAGUCUACAGCCAUGGUCG-3'
     ||||||||||||||||||||||
3'-UAAUUGUCAGAUGUCGGUACCA-5'
```

miR-133
```
  5'-UUGGUCCCCUUCAACCAGCUGU-3'
     ||••||||||||||||||||||
3'-UAAAUUAGGGGAAGUUGGUCGA-5'
```

miR-142
```
  5'-UGUAGUGUUUCCUACUUUAUGGA-3'
     |•|||||||||||||||||||||
3'-UAAUAUCACAAAGGAUGAAAUAC-5'
```

miR-148
```
  5'-UCAGUGCAUCACAGAACUUUGU-3'
     | ||||||||||||||||||||
3'-UAACUCACGUAGUGUCUUGAAA-5'
```

miR-181
```
  5'-AACAUUCAACGCUGUCGGUGAGU-3'
     | ||||||||||||||||||||||
3'-UAUAGUAAGUUGCGACAGCCACU-5'
```

Table S3. Synthetic miRNA duplexes used in reporter experiments. miR-1, miR-124 and miR-133 duplexes are identical to those used in microarray transfection experiments (Table S2).

miR-25 5'-CAUUGCACUUGUCUCGGUCUGA-3' | |||||||||||||||||| 3'-UAGAAACGUGAACAGAGCCAGA-5'	miR-196 5'-UAGGUAGUUUCAUGUUGUUGGG-3' ||•|||||||||||||||| 3'-GAAUUCAUCAAAGUACAACAAC-5'

Table S4. Sequences of UTR fragments assayed in Figures 1G-I and 4C.
Reporter plasmids encoded Renilla luciferase and were constructed in pIS2. pIS2 was derived from pRL-SV40 (Promega) by insertion of a multiple cloning site (shown below) within the region corresponding to the 3' UTR of the luciferase mRNA. Listed are sequences of UTR fragments, annotating their GenBank accession number, the restriction sites used in cloning (5' site – 3' site), the reporter plasmid name (in brackets), and the miRNA target sites (underlined). To disrupt miR-124 sites, the seed match TGCCTT was changed to TcggTT, except for the sites disrupted in the rightmost set of Figure 1G, which were instead changed to TGCCaa. To disrupt miR-1 sites, the seed match CATTCC was changed to CtgaCC. To disrupt miR-133 site, the seed match GGACCA was changed to GctgCA.

>Figure 4C; NM_002508 AgeI-SpeI [pAG186]
accggtcgttgccctgacaacaccttgggagttgactgtatcgaacagaaatgaagacaagagtgccttat
ttcctttccaagtatttcacagcaacactctacttgaagcaacttggtccagattgaaaagtgtcctctgg
ctgagtggccactaggcccagacccagcccagcctgagccccaacaacttttccctcactgttccccaaaa
catgcaccctggacttctctaatagaaaagtctccacccctacacaaggacagaaccctccacccctaccc
ccaaccctcagacagacttatacacccctgagtgaggattacatgcccatcccagtgtcctaggacctttt
cccaatactagccccccagtggtgaacagaacctcccaaatttgagttgcacccttccctgtggccttatg
agctcagcctcgctttgaggtacccaccgtcctgtcagctccttgacctatgagccggggcctgactagga
aaagttgggagttaaggaggaaattagcattccttaatgttttgttttggtgctctgaatttcttctttat
tatagtcctatagttttactcctcagttcctcaccatcatcatcttgtctaagacccccattataatattc
atgcgctgctttttcatcaaaacctaccctgtcctagagatctatgggcatttggtggatgataatgagca
gcccctcccagatagaatgtcaatatttgagcagtaggatattggcatttgttagttaaaggcttaaatca
aaagaatgtccaatggtaggaatttcaaggtgtaggtcagatatttgagaataggggatttttttgatgtgg
ccttaaattataccaaagattactaattattcctctttgcccaaaatacttgcatccaaggttctagtctc
tgttgctgtgctggtctttagccccactgctggcactgatgtccctccttttcacgactagt

>Figure 1G, 1st set; NM_014397 SacI-SpeI [pAG184]
gcgtggatgcaccgtgccttatcaaagccagcaccactttgccttacttgagtcgtcttctcttcgagtgg
ccacctggtagcctagaacagctaagaccacagggttcagcaggttccccaaaaggctgcccagccttaca
gcagatgctgaaggcagagcagctgagggaggggcgctggccacatgtcactgatggtcagattccaaagt
cctttctttatactgttgtggacaatctcagctgggtcaataagggcaggtggttcagcgagccacggcag
ccccctgtatctggattgtaatgtgaatctttagggtaattcctccagtgacctgtcaaggcttatgctaa
caggagacttgcaggag
>Inter-site sequence for Figure 1G, 2nd set [pAG400]; otherwise as for pAG184
gtgccttacgtatatctgaactctgtaagccagcaccactttgccttа
>Inter-site sequence for Figure 1G, 3rd set [pAG404]; otherwise as for pAG184
gtgccttacgctttatctgtatatctgaactcttgaaacttatagcgtaagccagcaccactttgccttа
> Inter-site sequence for Figure 1G, 4th set [pAG420]; otherwise as for pAG184
gtgccttacagagttcagatatacgtaagccagcaccactttgccttа
> Inter-site sequence for Figure 1G, 5th set [pAG424]; otherwise as for pAG184
gtgccttacgctataagtttcaagagttcagatatacagataaagcgtaagccagcaccactttgccttа

>Figure 1I, 1st set; NM_005433 SacI-SpeI [pAG428]
gagctcttatcagcgtatttcagggtccaaacaaaatagagctaagatactgatgacagtgtgggtgacag
catggtaatgaaggacagtgaggctcctgcttatttataaatcatttcctttctttttttccccaaagtca
gaattgctcaaagaaaattatttattgttacagataaaacttgagagataaaaagctataccataataaaa
tctaaaattaaggaatatcatgggaccaaataattccattccagtttttaaagtttcttgcatttattat
tctcaaaagtttttctaagttaaacagtcagtatgcaatcttaatatatgctttctttgcatggacatg
ggccaggtttttcaaaaggaatataaacaggatctcaaacttgattaaatgttagaccacagaagtggaat
ttgaaagtataatgcagtacattaatattcatgttcatggaactgaaagaataagaacttttcacttcag
tccttttctgaagagtttgacttagaataatgaaggtaactagaaagtgagttaatcttgtatgaggttgc
attgatttttaaggcaatatataattgaaactactgtccaatcaaaggggaaatgttttgatctttagat
agcatgcaaagtaagacccagcattttactagt

>Figure 1I, 2nd set; NM_013438 SacI-SpeI [pAG434]

gagctctgagccaattgtttctgaagtgttttggtagttctattaagaaatagttaaatattgtgcttttc
agagcctcagagaaaggggacggggtgggggggtggggcagcggaatctgtcctggatggggccagctta
aataatactggcaaccaagattctgttaggatttctgtgcatatagtgtagtaaagaagtatcattcaggg
gtgaaaaacaaagagccgttttaatgatgttgagtacatttggctgttttatagccttttcttccctccc
ccaaagaattctgtttgcctaactcccaaactgttggggtggtacattcctttaggaccaattaaaacata
attgagggtcagtgatacatttggctgactctggttcagtattctcttaggtgattatattctctcatgta
cagttacaggaaattaaaatgttaaagtaacctaaaatgaattcagaccaataaaatcaagggaaatacaa
gttgattgcattacttctgtatgttgcttgctattaaaaaggttaagaggccaggttacccaccagtcctt
gcactgttctgacactttccccaggaggaaaacaagtacaaaggttacggtggaggcataagtaactagt

>pIS2 cloning site, pRL-SV40 sequence is in italics, restriction sites are underlined

*gaacaataattctag*gagctctataccggtctcgatatcactactagtgt*tctagagcggccgct*

Table S5. Sequences of UTR fragments assayed in Figure 6.
Reporter plasmids encoded Renilla luciferase and were constructed in pIS1. pIS1 was derived from pRL-TK (Promega) by insertion of a multiple cloning site (shown below) within the region corresponding to the 3' UTR of the luciferase mRNA. Listed are sequences of UTR fragments, annotating their name, the restriction sites used in cloning (5' site – 3' site), the reporter plasmid name (in brackets), and the miRNA target sites (underlined). To disrupt miR-25 sites, the seed match TGCAAT was changed to TcgtAT.

>Figure 6H; CDH10 SacI-SpeI [pAG651]
gagctctttcctgtaggatgtctcatggaatatatatgacattttatttaatcacttccaagagccaaagc
tatggaaatacagtgttgtccatcttagtaaataaaagataatttcagaaacatgaacaggatagttctcc
cttaagcaacctcacaaacaagccgcttctgttaggtacatgtcctgcccttgcaaatgaagctttaaaa
aggtgaagaaaaattttacagtatatcctgttctgtacattaaattaaaaaaacaaaaatgtacatgtgat
gttagtaggtgtgatatgcaacctggtatacagacatttgtgcaatttcatttcatcaaattctatctgct
aatgttttatattactagt
>Figure 6H; EVA1 SacI-SpeI [pAG652]
gagctctttcatgagcagtgacggatagtttagcttactatgtttccccccccaattcaatgatctataaca
acagagcaaagtctatgctcatttgcagactggaatcattaagtaatttaataaaaagattgtgaaacagc
atattacaagtttgaaaattcagggctggtgaaaaaaatcaactctaaatgatgataatttgtacagttt
tatataaaactctgagaactagaagaaattattaacttttttctttttaattctaattcacttgtttat
tttgggggaggaagactttggtatggagcaaagaaataccaaaactactttaaatggaataaaaccaactt
tattctttttttccccccatactggtagataaagcaaactttataagtgggctattgaaagaaaagttacaa
gcttaagatacagaagcatttgttcaaaggatagaaagcatctaaaagtttaggctcaagatcaatcttta
cagattgatattttcagtttttaatcgactggactgcagatgtttttttctttttaacaaactggaatttca
aacagattatctgtatttaaatgtatagaccttgatatttttccaatactatttttttaaaaaattgtatga
tttacatatgaacctcagttctgaaattcattacatatctgtctcattctgccttttatactgtctaaaaa
agcaaagttttaaagtgcaatttaaaactgtaaattacatctgaaggctatatatcctttaatcacattt
tatatttttttcttcacaattctaacctttgaaaatattataactggatatttcttcaaacagatgtcctgg
atgatggtccataagaataatgaagaagtagttaaaaatgtatggacagtttttccggcaaaatttgtagc
ttatgtcttggctaaatagtcaaggggtaatatgggcctgttgttactagt
>Figure 6H; DYRK2 SacI-SpeI [pAG653]
gagctcgtggataaatgggaatggaaacgtgtgtgttcctccaaattttctagtatgatcggtgagctgtt
ttgtaaagaagcctcatattacagagttgcttttgcacctaaatttagaattgtattccatgaactgttcc
tccctttttctctgcttttctcctctctgttcctcttttaataccacacgtctgttgcttgcatttagtttg
tcttcttccttcagctgtgtatcccagactgttaatacagaaaagagacatttcagctgtgattatgacca
ttgtttcatattccaattaaaaaaagaacagcagcctagctacttaaggtggggatttccatagttccaaa
gaagatttagcagattagagtgttcacactttcaggtgccactgtaaggttctctcagcctgggaaacta
tcaactctttctttaaaaagaaagagggttgaaaatcctctggacgaacagaagtcactttggctgttcag
taaggccaatgttaacaacacgtttagaggaggaaaagttcaacctcaagttaaatggtttgacttattct
tcgtatcattagaagaaccccagagatagcattcctctatttttatttacttctttttggattgcactgat
tgtttttgtgggaatgacactttatctggcaaagtaactgagagtttggtaaaagaatattttcttctctg
aataataattattttcacagtgaaaatttcagtattttatcactaatgtatgagcaatgatctatatcaat
ttcaaggcacgtgaaaaaaatttttttagtatgtgcaatttaatatagaaagatttctgcctgtttggacaa
taggttttgggtagtacagattaggataagtaagcttatatatgcacagagattattgtattacctgtaaa
ttgatttacaagtacttaaaagcgtggtccccagtgaggccaagaaagtttactagt
>Figure 6H; SLC37A3 SacI-NheI [pAG654]
gagctctgggaaagatcacactacatgcctgttgattggctcagtcactctgtgtctgatctaaacgtcat
tcagccctagaagcatgaattgcttcaaattattgtcaacttgttctcttccattttcattcaaattaact
ttgactcctggaagtattgagtcttcctttcaaggaccataaggtacatcagttatcttgaacattctgac
gtgtaaaggaccataaggtacatcatttatcttgaacattccgacgtgtaactcacagccgaagcactcca
tcccagatttgttgtctgggaacttaggatcctctaggaagctaatctgcttagtctatttttagaggatt
gatctctggcacaagcagcttcctggagttacctcagcagaagactagaattagagaaaagggaaagacct
tttcttctagtaagtgtcaagtacagatgctctttgacttatgatggggttatgtcccagtaaacccattg
gaagtcaaatgcatttaataccctgactggacaccacagcttagcctcgtccaccctaaatgtactcaga
acacttagaattgcccacagttgggcagaatcatctaacacaaactctatttataatcgagtgttgaata
tatcatgtaatttattgaatattgtacattatgttgaaattgcaaccatttcacaccattgtaaagtccaa

aaatcagaagctgaggactgtctgtacatgacattgaatcacatgaatggattactagggcctgaggggtg
tccgttctttagaaaaatgtagagtatatatatcttcagggctgtgaattccatactagttatcaatattt
taagatcattctgtatatacatattttaaattatgcaagcaaactaggaggagaatgtgcaatttgaggg
atacctaatttgcattcggttaggggatatttttcaacctcttgctttatactctgatatggggtgtgttt
atgaacttgttaatgttatttaatcatgcagatttccaatcagtgtattgctcatttactgacaaaaaggg
agctagt
>Figure 6H; NFIB SacI-SpeI [pAG656]
gagctcaacacaaaggcagaaaggaaagaaactgaagaaagaagataatagaccagcaattgcagcactta
caatcactaattcccttaaggttgaaactgtaatgacataaaaagggtcgatgatatttcactgatggtag
atcgcagcccctgcaacgtagcctttgttacatgaagtccgctgggaaatagatgttctgtctctatgaca
atatattttaactgactttctagatgccttaatatttgcatgataagctagttttattggtttagtattct
tgttgtttacgcatggaatcactattcctggttatctcaccaacgaaggctaggaggcggcgtcagaggtg
ctgggtgacagagccatgagccagccatttataagcactctgatttctaaaagttaaaaaaaatatatga
aatctctgtagcctttagttatcagtacagatttattaaatttcggcccttaacccagccttttccagtgt
gtaacccagtttgaaatcttaaaaaaagaaaaaatgaaaaaaaaaggaaaaaaagaaaaaaggaaaaaaac
agtttgaacacaaaggctctatggaagaaatgcctctatgtaggtgaagtgttctctctgcatgcaacagt
aaaaattaatataatattttccccacaaaagaaacacttaacagaggcaagtgcaatttataaatttatat
ctaaaggggaatcatgattataagtccttcagcccttggactctaaattgaggggattaaaaagaatttaa
aataattttgaacgaatttatttcccctcagttttgagggcattaaaaaggcattaaatcaagacaaat
catgtgcttgagaaaaataaaattaatgaaaacacagcacttatgttggtttagctgcagcctccttggag
gtagaatttatttatttaaaattactggttgcatcaagaacccatagggtgtacaaaaggttctataaaat
ctgcattatagagacaaagaggcaggcaaatccatgtcacaagggtaaagcttacagtttacaaactggga
acgccagggtgtaggatataaaaacgcactcttgagaaaacaaatgtaatcagggtgctgaaaacttgcat
ggtgctttcagacattagccttgttcaacaaatttcttgtattgacagatccatagtgtgcatgggcagac
acattttgcctctatgactagt
>Figure 6H; PITPNC1 SacI-SpeI [pAG657]
gagctctctgatagagaaaaagactgctttgtcactcaaacatgttccttcgacctttcagtgtgcatgtg
actcagtaacttcacatagaatatgattccctaagtatgctacacagcatcatattagatgtaagatgtaa
gacttgcaaaggacagaaggaatcttctgtaaccacatagctgtatgccagagaggaagccttgttattgg
gcatttgatgaggtttggcatggacttcaaggataaatgaatgaaaactttgcaccacttttgttacaagg
tacggtagaaaatagtgaagtcagtttcctctcatcaaatctaaaattctccaaaatactctcaggcataa
catacttagctgttaaattttgaactgctaattactaatacttgaataccaatagttactgagattcctat
tttgtggttagtctgactcaggatttggagcctaattaactctaaacttttgaaaattttaatcatcaagc
tatagaggctccaagtgcaattaataataaactagt
>Figure 6H; MGC23401 SacI-SpeI [pAG658]
gagctcccctggctaaggtctgtgtaaggcagacaagcgttattgatcatatcaagttccctacaatatcc
tgtcctcaaaaccggaagcaatgaacatgatcctcttcggttggataaatgaacttcctgtttggcctgct
tctaggccctgccagattctcataacatcatatacgtaagtatagttcctcaaagtgactgacatttattt
taattttgctttgttttttttatttctcccccattcctttattttgtgttattcctgactcacttgaca
ctctctgatgcctgagagattcctgtttgggatttaatatccagggctgtgtttacagtaaaaaaagcagg
cagtccctttagtttttccttttaaatttttttgagattcttcatttcaggatttaaaactatagcagt
ccatcttaaggaaagtgtaactgccatggccacaagtctgctagttgcacttgaatgctctatcagggttg
tttattacccttctacgttctggactccttgccgagactgtttaacttgaagattaaagaaactattgca
aatgccagtgcatcagaacctaagagtggtcaaatattatgtgcaatttttttgtaaagaaattttaattt
ataataaagtttaacagtttaaagaacagttaatatttgaacactagt
>Figure 6H; TESK1 SacI-SpeI [pAG659]
gagctcggacaccctgtaagaacagagcacacttgctggacagtgccagttccagatgggctgaccggctc
ttctccccgtgtaggggagccccagcatggactcaagggacagagcacttccagtcgacccccggctcgc
gttcccgtggggatcactgaaccagacacagcattgctgacacatgagactaacacgtgcaattatttaaa
aagatttcaataaaactgcctggcactagt
>Figure 6H; PHTF2 SacI-NheI [pAG660]
gagctcttctttgcttcaggagtctcagctagggagttgaagtgtttacatcagactgtcttgtgcaattc
ttatatttatttactggttcacttttttttacatttatttagtctttatatttttatttttaagcattg
atgtacttagttgttgaaagggtgatgaaactgatatccagatacttgagatcctggtaattggtcataaa
taattggcaaaataacaaattgtgaaaatagaagccattgctcagcaccgtttctccatcaatgccgtgaa
cttgccttacttgaggaaaaattctttaactttggaatattgcattgaactcagctatacacataaaacat
tttctttggtaaatcaagatccagtcagggtttctcttgaattattttggaacaatgccaggatccaaact

gattaagttacagtttaagcacccttcagtattaatatatacggtattatataacaggtcaacaagtgctc
tttgatgataaaacttgtaatagagcaataattgtaaatggttaccatactgtaagatattttgataaaaa
ttaactagtaatacttgtatttatttgaaacactgggctgtttgcacagctccaactgtgcatgctcaaaa
tgtgcacttttaaaattgttacttttaatgcgtatctttatatgggatctgttatagtatactagggcat
gatatggtatccttttgagtgaggtatatactcatctcacaagtgaagtgcctactgatattactaaagta
cattatgtttactcaagtaaataattttctccccatggtacactctagtgtaggctattcataccacactg
aaatgaacaactgaagaataaggctaagaaccaataaaatatttctctaattgctagttgtaaaactgtat
ccaaattttcagaaaagacagcttcagcttgcaaattctatcctctaaacttatctggtgcattctcccca
ccccacccccattatataagggctattttagatgcttttaacctccccaacaaataatttgccaagtgtcc
aatgagaacttatcatgttggtgtgttaggtaaatcgggcaaatatgatagtgtcttacattgggccttga
tgctagt

>Figure 6H; HBS1L SacI-SpeI [pAG661]

gagctccgtttctggatacagtgaaatagctaacctctgtttcaagaatgcagttattaagtcaaaggaac
aatgtgcaattgatatgtttttagatgagagagaaaaattaaagctaaaattagctgcaaagaagtattaa
taatcacctctgcaaaaattctaagttgccaactggcaaagaaagtctaatgttaaaaacaactttgcctt
tgaaacgttaataaatggatttactttgctaagatttatggcaagtgtcaaaaatagtatctgaagatact
gaatcatcatgaaatgaactctacttctggccaaagcacaatgtatttgcagttttctcttttgattcaat
tatactgcacatgttttaaggaaaagtaacttaattgggtttttcaggcagttgatatttgacctaagctt
ttttttttttttccagttaatgctaagaaaagatttggggaaggttataataaaagtattttgtggtgac
cataagaatgtccctccccaaaacaagtaaacttgtgaaagtttaatttggaattagtggaagctgttcct
ttgaaagccaagatattatttaagttgtaaagccagctaataaaatgccttagtttgagcataatacaaac
tgtgttttgttctttagacatttatgagattctctgccactaacagtaaatatgatttaggggttttgtgg
gatcttttcacctaagagtttcctagctctgtcacatgaattctttttatcctaatggcagatccaatt
tcatctcattcatttttttaaccataacttgcttttcccattctcacttttagaattatacctatgatggc
cctttctttatcttgcttaccttatagattataagtcccctgtgagtaagtaagagtcatgtttttcttta
aagtttcaacaacaaaccccaaggtttatttatttaacaacaggtattcgttgaggtgccttctgggtgca
aggagctttagagtaactagt

>Figure 6H; ACOX1 SacI-SpeI [pAG662]

gagctctaatctgcttcagaaagcgcctgtgtgcaactcaaattttgtggaatctttttcgaattcaaata
gctatagagcaaatgataaattgacccctttttataaatggagggaaaaaatgaacagatttcagagatta
aatgaaaaaaagcagatgttttaagtgcaattaacactgaaagagacctgttaaaccattcagaaaaagct
taagaaatgcgatatgacttccttttgtaatgctgctgatcccagtagactatgacttttgataattagca
gaatttaactactgagtagttgattattttcacattttaattgctaatcactggctatataagtgttttta
agcaaaggtatttttgaagtggtgtagaacccttccaagctttcctgctcagtgttctaccagacttaccc
tggggcctggcttaaaagcaggattgaagaaaagggactgggggaaggaaacttattggaaaacttgatgc
gaatgagtttctgcttggcacagtctctgcctgcttgctctcctttgctgatggattgcatttatcaaact
attcatgctagcatttttccaacgagggaacttattccgcacgggcctactgtaggaccattgtctcgtgt
aattaggaattttccatttgaaggattgctaaattgtcacagtagtaggaagtatagggaaacctctcagc
tgtggcactgttgtagctttggagtgactagt

>Figure 6H; SFRS3 SacI-SpeI [pAG663]

gagctcatgttttgtgagaatccttgggattaaagttttggttacaaattgttctttaacttgaaagcctg
tttttccttgcaaactcaaatctgtgagcttggtaccaagtccaggtataacattcctattggaagccata
cttatattttcttgtaaagtgcttttgaattaataaaatattagcataattgtgtatagtcagttgaaccc
actgttaccattgttcttatcccatgggaagcagttggttacacgattcttatttttataagaaacagctga
gaggcactatggattagtcttctgaagtgaaggaaatatagatgtcacctaagtgatagttaacccatttt
tttttttttaggcatagaagccagttcagggtccataatatttagtgaccaacattttaaagtatagcag
caacctggttcttaaacacaaagtaagttgcccattaacaaatggcttttatctttagcatgaaaactttc
cacaggtctaaaaattgcttccatttataatttgaggtgttgcatgggaattctaagctgatccatcatg
atgtaaaagttcacaatatggttcaaatgtaacagtgcagaattgaatatggaggcatgcataaccttcct
cttagaaaatggcaggtgttgtaatttcaaatttttgtgcaattagattaaatcataatgcaacagtcttg
tggcttagtttccttactagt

>Figure 6H; GAA SpeI-NotI [pAG664]

actagtcccaacgtgtctaggagagctttctccctagatcgcactgtgggccggggccctggagggctgct
ctgtgttaataagattgtaaggtttgccctcctcacctgttgccggcatgcgggtagtattagccacccc
ctccatctgttcccagcaccggagaaggggtgctcaggtggaggtgtggggtatgcacctgagctcctgc
ttcgcgcctgctgctctgccccaacgcgaccgctgcccggctgcccagagggctggatgcctgccggtccc

cgagcaagcctgggaactcaggaaaattcacaggacttgggagattctaaatcttaagtgcaattattttt
aataaaaggggcatttggaatcagcttctgcggccgc

>Figure 6I; TBL1X SacI-SpeI [pAG636]

gagctcaagaagaacctcgcatcttttggtgtgtatgcaccttggcagccagcaggaaagcagcgaagagc
cgcgcgccctctggcctcaagggagcattggcagaaccataaggtactgcagaagccgtcgaaaccgactg
ctcggctggattaggcgtgttggctacatctttcccacaaccgtataacgaccgatggtcatttctccaag
agcaggcttggcgagtccttgcagagctgacatagaggcgcccgcatgtcacttagtctaacgctgacaga
aatgaatgcagaaggaagattttcagtcctgaacgtgaattatagaggtagaacgtcgctaataatttctg
ccatctttataatttctttgtctcacagaagactaggagaaagatctttttaaataatctttttgctgtt
tttaaaaaattaaacaaggctttgtgttcctagaagagcttcatttcagtgaatctggtgacctccatctg
cttgctgtcataacccgacacggacttatttttgtcattagcaaggggaaaaggccaaaggacaagggcc
tcttctcccattggttttcctgtgggcagaagggctgaggaagatggcccagcccgtgggggctgctgggt
caccagcagcgggtagggtgcaatctggtgtgtgttccagcagtgagacggtgttattgtgaaggtggcat
tcatctgcggagccaaaacccagccatcggggaagggtcagggcttctgtggaacttggaacgtgccagga
ccacctgcaaaagccagggtgcgttgatcattctcagatcattgattggcctccacttgggtatgtgaatt
attcatgtcccagaagaccaaaaagtgctctggttctgagatgagtattttattcgtgttctgtttccgaa
acacttagcaaagaaggtcacagtgatgtggagtcgccgcacccatctttgaagatagccagtgtccctgg
atgaggtgatgatttcccgtcccaaggactctgtgaagtttagagtacagtttgttggggtccaaaagaca
ccatctctaccccacccaaataaaaatgcactcatctctgtagaacatctgctgtcaaaggccagcctgtc
gttagggcatggcttatgcttgacaaaccagtaacaactgtgggaactagt

>Figure 6I; SH3BP2 SacI-SpeI [pAG637]

gagctccaatccatccccttactttctgccatggagttccagcaggtcactctccctggcacaccttccag
gctggatttttaatgaaacagactcagggaggtaggggctggcagggaccctagaatccttgtgatttttc
ttagcaccttatgtcagggaaacctaaactgaggtcagcacttgggcccactgacagtgactgactggggg
agaaggtcctgcagccccctcccctgggtgtgttctggggacctgtggtttgctggcggaaacaagtgat
gaggctggttagcggatgtgggaggctgtgaccccaggggccatagggtgcggtggaactgcaggccctg
cagatgacggcagccagctgcttccaggaaccaggtgtccaaggccacctctgcaggggtttcctcttcag
cctgcctggggtgagaggtcagtgcaccacagccgaggctggagcacagggagcttctgttgttctgatct
atctctggaaaaccagccattcctcctccctgcagtcagaattctttgccctgtctgacctgaacttgctt
agggagtcatgccactccccactgtggccatagtttctcttcctgtaaaatttttattatttagtttttg
tttttgagatgtagtctcaccctgtcgcccaggctggagtgcaatgccgtgatctccgctcactgccacct
ccgcctctctagttcaagcgattttcctgcctcagcctcccgagtagctgggattccaggcgcccgccacc
acgcctggctaatttttttgtatttttagtagagacgggattttatcatgttggccaggctggtctcgaact
cctgacctcaggtgatctgcccaccttggcctcccaaagtgctgggattacaggcatgagccactgtgcct
ggccccttcctgtaaaatttttaaatggagaattgggtgcgagatgtggtttccagcctggtgcctggggt
gctgagctagtgagtggtgcagtccaggacacctttgctttatgtcacttacacggtcacctggagccggc
tcaagtggctaaagcatcctggggcccagagccaggtgatagtccctctggccaactggacagttgaggct
tgtggttaacccgaagcccagctggggccttggtccagcttcgcttcccagattctgcacctgctagcaca
gctgtccacgtctgtgtgagctgttctaggactagt

>Figure 6I; C9orf25 SacI-SpeI [pAG638]

gagctctacctctgtgtgttcatggagctggagaaaccaaaactgggtctctctccacccgggagctgagg
aggggtgggggccgtttgttcagttacctgggggaacctcacccagcaccctgccccgcttcccagggctg
gccaagagagaagactggacggcagggaggtcaacattaacgagtgggggatcacaaagccaggggcttg
gccccaccccagcgaagccatgaactcccaagccccggcctcagcctaaccaggaaaggggctcggaggga
gaaaggcccagggcagtggggacgggtcccagcacctcggagccccccttcacacgccccctaccggcccat
gtttctatttgcatttcccccctcctcagatgggaagcctggcaaagctgcccaaccgggtagtgcccctca
cccatcggatcactgccttcttcctcctcttgccccattcacccgcttccttgcactctgggagtggggca
gggtaaggtgggccttacagacaggctgaggtttgggtggtgtgatctgtcctaattggcccctcaccaat
gcatctatctgtctgccaccagcccaccccaccctgcccccaccctcacccacgcttctgctcattggtct
caatcccagcctggaagcagggagtaaggctggactccaatggcccattcccctcaccccagcccaagacg
tgggaggctgcttcccagtaagaatccaggagcaagctttggtgaatccagtgggctgagtggctctatgg
atgtggttgtttggatgtggtagggagtgcttagcagaatgtgtgtgtatctctgaggctgtcaggggtg
ggtgggtgtgtgtctctgttggacctgtctctcctcagtctattcctgtcatagatgtgggtgtccctggt
agggaagtgcctgtgtggtctgtttacagtgcagaaggcggctcagcagtgcaatgggctgacagccttgg
accgggagtcagaagcctgggttctggttgtagctctggcccctccctgtgatcttcaacaagccacttgc
tttctctcgtcttcagtttcctcaactagaaaatgaacagcagctgaaattatgttgaagatccttttata
cctgaaagagtttcttcatagaaatttctgtggtggggaggggcaatctctgtgtgtctcagtataccta

aaggttggggtggggcggagttgtgtgtctctgtagaaaggcgaaggggcagttcctatcaggactgtgta
tgtctgagcacatgtggctctgtttgggattacgtgtttgtctgtgaatgtgtgtgtgtgttggagggttg
tctattgtgtgtggctgtatagggtgtctgtagatcaagatgtgtatacagctgcttctgctattgctggt
ttgggggaggggactggaaagctgagactgaaaatcaagagggaagaggtgaggaaggcccacccctgggg
ctgggcagtgtagacctgccaccgtgtgaagaaaaatgaaagcaactagt

>Figure 6I; SOCS4 SacI-SpeI [pAG639]

gagctctaggcattacctcagggaaaagccttgacagtgaaaagaaactacttaatcgctaatagctagtc
tggaccagacacttaactgcgctttgggaatcatcctgggctaatcaggaaataaacttgacttttctatt
tttctctgaagctttacaggacatactcccactcccatttttttctaagcctacagcatcttgattcatag
tataattcctgaagtatgctgaatggacttttaaggttcctttagcatttttgtatagagttatatgtga
tctttctataaccatgccattcacttaaataattcatgaagggaatggtttaatttaaaggaagacctaat
aggaagaagattaaaggaagcttttagtcagcctgtggctagatttattgatttggtgatcaaacctgtgc
tatttcctgaatttcagctaatggccaactgtggtcaaagaatggctttcagttaaagttttgacttctta
ttataaataatataaagaatgtctcagtaattagaaatcccttaataagtagaacaaatgtggcaagtagg
actcctatcggtatacaaattccatttctttttataagaaaaacacataaaccattatcatttattcatga
gccaaagccattaagaagataaatccagttatcagcttcctgtctctgaagtttggatcattttatagatc
cacgcaatagagagcttccttcctatgaaacaaaaaccttgagatctctttttttttgagacagagtctcg
ctgttgtcagcccaggctggagtgcaatggcgcaatctcggttcactgcaacctctgccttccaggttcca
gcaattctcctacctcagctcccaagtagctgagatgacaggcgtgtgccaccacgcccagctaatttttg
tatttttagtagcgatggtgtttcactacattggccaggctggtctcgaactcctgacctcaggtgaccac
ccgcctcagcctcccaaagtgctgggattacaggcatgagccactgtgcccggccaagatctaactcacgc
ttacttttcattaaaatatttcttgtcgcttttagctattgtcagaatccaataaacagcttacgcattca
ttttggtaaagcaaatttcacaggaaaatagacaaaattataagagattcctggagacaacacaaggaaaa
caatgtttaaaagcaaaatagcctgtgtaatgttaggtaatgtaatgcccaagtgaataaaagatttttcg
tcaagcaaactactgtttctactttgggggaaaaactagt

>Figure 6I; LPL SacI-SpeI [pAG641]

gagctcgaattctggatctttcggactgaggccttctcaaactttactctaagtctccaagaatacagaaa
atgcttttccgcggcacgaatcagactcatctacacagcagtatgaatgatgttttagaatgattccctct
tgctattggaatgtggtccagacgtcaaccaggaacatgtaacttggagagggacgaagaaagggtctgat
aaacacagaggttttaaacagtccctaccattggcctgcatcatgacaaagttacaaattcaaggagatat
aaaatctagatcaattaattcttaataggctttatcgtttattgcttaatccctctctccccttctttttt
tgtctcaagattatattataataatgttctctgggtaggtgttgaaaatgagcctgtaatcctcagctgac
acataatttgaatggtgcagaaaaaaaaaaagaaaccgtaattttattattagattctccaaatgattttc
atcaatttaaaatcattcaatatctgacagttactcttcagttttaggcttaccttggtcatgcttcagtt
gtacttccagtgcgtctcttttgttcctggctttgacatgaaaagataggtttgagttcaaattttgcatt
gtgtgagcttctacagattttagacaaggaccgtttttactaagtaaaagggtggagaggttcctggggtg
gattcctaagcagtgcttgtaaaccatcgcgtgcaatgagccagatggagtaccatgagggttgctatttg
ttgtttttaacaactaatcaagagtgagtgaacaactatttataaactagatctcctatttttcagaatgc
tcttctacgtataaatatgaaatgataaagatgtcaaatatctcagaggctatagctgggaacccgactgt
gaaagtatgtgatatctgaacacatactagaaagctctgcatgtgtgttgtccttcagcataattcggaag
ggaaaacagtcgatcaagggatgtattggaacatgtcggagtagaaattgttcctgatgtgccagaacttc
gacccttttctctgagagagatgatcgtgcctataaatagtaggaccaatgttgtgattaacatcatcaggc
ttggaatgaattctctctaaaaataaaatgatgtatgatttgttgttggcatccccttttattaattcatta
aatttctggatttgggttgtgacccagggtgcattaacttaaaagattcactaaagcagcacatagcactg
gactagt

>Figure 6I; FANCC SacI-SpeI [pAG642]

gagctcctcaagtgagccttcggtactcctggtgcccgcccggccagaccgtcagcttgataattactaaa
gcaaaggcctgggtgggagaacaggtttctagtttttacccaagtcaagctgcacatctattatttaaaaa
ttcaaagtcttagaaccaagaatttggtcatgaaccattaaagaatttagagagaacttagctctttttag
actctttttaggagtcagggatctgggataaagccacactgtcttgctgtatggagaaattcttcaagggg
agtcagggtccctcaggcttcccttgtgtctccctggacctgcctgacaggccacaggagcagacagcaca
cccaagcccgggcctccggcacactctttccactctgtatttgctaaatgatgctaactgctaccaaaagg
cccttgggacatcagaggagccggcaggcgaaggtagaggatgtgttccagaaacattagaaggcaggatt
aattcagttagttagttctcttgttaaatggaaatgggaattggaaattcctgataaagaattggcctggc
tgggtgcagtggctcacacctgtgatcccagcactttgggaggccaaggcaggggattacttcagcccag
gagttccagactgcctggctaacatggcaataccctatctctactaaaaatacaaaaattatcggggtgca
atggcatgcatctgtaatcccagctattcaagaggctgaggcatgaggatctcttgaacccgggaggtggg

agttgtagtgagccgagatcatgacactgcactccagcctgggcaacagagcgagaccatctcttaaaaaa
aggcattgttagtgtaatctcaaggttaacatttatttcatgtcagtacagggtgctttttcctttcaggg
acattctggaattgtattggttgtacattcttttgtgtctattctgtttgtcaagtgagtcaagacttgct
tttgtccattttgatttgtgtgtattagtctgagtcttggctccgttttgaggtatgagcaaagttttgct
ggattagaagctaacctttagggaaattccttattttggtatgtggcaatgctaatagatccactgaagat
ctggaaaattccaggaacttttcacctgagcctttcttctgagaaaactagt

>Figure 6I; OPRL1 SacI-SpeI [pAG644]

gagctcatgggacaggtcaaagcattagggccacctccatggccccagacagactaaagctgccctcctgg
tgcagggccgaggggacacaaggacctacctggaagcagctgacatgctggtggacggccgtgactggagc
ccgtgcccctccctccccgtgcttcatgtgactcttggcctctctgctgctgcgttggcagaaccctgggt
gggcaggcacccggaggaggagcagcagctgtgtcatcctgtgccccccatgtgctgtgtgctgtttgcat
ggcagggctccagctgccttcagccctgtgacgtctcctcagggcagctggacaggcttggcactgcccgg
gaagtgcagcaggcagcttttctttggggtgggacttgccctgagcttggagctgccacctggaggacttg
cctgttccgactccacctgtgcagccggggccaccccaggagaaagtgtccaggtgggggctggcagtccc
tggctgcagaccccgagctggccctgggccagccgcacctctgaaggttttctgtgtgctgcacggtgcag
gcctcatccctgactgcagcttgactctgggcccaacccccatttcccttcaggagaccagcgagaggccc
tggcccattccctccagcggtgcaatgaactatcatgctgtggaccgtcaacccagccctgcttctcagtg
tggggcaggtgtctcaggacgaaggcgccgcgtgaccacatgggcagctctgttcacaaagtggaggcctc
gttttcctggtcttgactgctctgtttgggtgggagaagattctctgggggtccccacatcctcccaaggc
tcccctcacagcctctcctttgcttgaagccagaggtcagtggccgtgctgtgttgcgggggaagctgtgt
ggaaggagaagctggtggccacagcagagtcctgctctggggacgcctgcttcatttacaagcctcaagat
ggctctgtgtagggcctgagcttgctgcccaacgggaggatggcttcacagcagagccagcatgaggggtg
gggcctggcagggcttgcttgagccaaactgcaaaggctgtggtggctgtgaggacactgcgggggttggg
gggggggcgtctgtacctcaggggatgccccgctgtggtcacccagagaatcacccttcctggtctacaga
tggaagctgcaggttggtgactttgcaaatgcacttcctacagatgaactattaaaagacctgcaacattg
aaaaaactcatttttccaccaaaaccttggccaggtaacctaccttaggcacctgcaaagaacaggaagt
gatgactagt

>Figure 6I; OSBPL2 SacI-SpeI [pAG645]

gagctccttttaaactcgaaccagtgggggaaagatggatcttgaagctaatcctgcagagagttttatag
aggccagggattgccttctaaattatgataaaacagaagtgaagagtttcagagcatcagattgagtgaaa
agttgtcagattctgtatttttaacaatcttcaataatgtaaagattacttttaaaatatttaagttaaa
actacttgaatagtattttgctgaagagcaagatatgcattaatcaccggttttatactgtccaaaatgaa
gcatccccgtgacaaaccagagtgggcagaagcatcgagagcgtgacaggaaatcccaagactgcttccgc
ctcagaggcgtcccggctgcgattcgctgccctgttgtcagtgaggcctggctgtcaccgcacaccgcgtc
cgtgtctccaggggttcctttcttctcacacgtcgcgtgtacccatagcactcttgtgtttctgttttc
ccagtatgcatgtttaaaatagaagtgacaagaatcacatccggttgtgtcctgtgggagggtcagaggca
gaatctacttacagtggtgtaattaaagttatttaaccaaaaataggtatgtgtccatctcagcattcacc
tttatcaagtgactgatttttttttctttttctttcctttttttttttttttttgagacggagtttcactctt
gttgcccaggctggagtgcaatggcatgatctcggctcaccgcaacctccgcctcccgggttcaagcgatt
ctcctgcctcagcctcccaagtagctgggattacaggcacgcgccaccacgcctggctgattttgtatttt
tagtagacacgggttttcaccatgttggtcaggctggtctcaaactcccgacctcaagtagtctgcctgcc
tcaacctcccaaagtgctgggattacaggcgtgagccactgcgcctggccgtgactgatttttttcatgt
agaattgtcaacacgagagatcacagtggagcactttgaaagaccgtcggttgtgtgcacgcacgcacaca
ctcatgcacacgctgacacgcggttgcatggagtccaggttactcaggccggcacttctgagtgacaggtg
ccacctgcgtgtgtcttggcgtccacatcacacctgtgacggaagcacttctggaagtgaacactcgtttt
gactagt

>Figure 6I; SSB1 SacI-SpeI [pAG646]

gagctccaatctgctcggggtgggacgggtgcttcccacacctctgggagaaggctgcagccacctggggg
tcccagggtggtgggggtggcaggtggtaccacagctctgagagcagataccaggggtactaagaggtgct
tagacaagggctggtgcccggcccagggtgcccagcggggccatgccatggcagataaagctcaggacgtc
aaaaactcaccatggaccccaaggcagaaaccaagaactgtctgcaggcaaataagcacccagcacccatc
ctggctgccggtgccccgtaccctgtatttattcttttaacaataacaaaagccatttatttattccatct
agaaaggaaaccctgtttcagtcccctctctctggctgttctgttactttccttccacctgtgccctccct
gggatatgtatgcctcgcccgccctccctgggcacatgtgcacacgtgcccaggcacaagtatgtctctgg
gtcccttgccctgcagtttccaggggctctgctccaagttccctagcgggcccctcagggagaaatagcc
tcacgtgcaatctgggtgtcttcgggggcccgtctggaagggctgcagcaattcccctgtgtctccaggta
accagctaactcttgggctcaggcacccttgcacagggttgcatttctttagtcttctgtgggtcttttga

tgtgggtttgattttgcttttgcttttctagctgagatttcccaagggcatcctcagaagctctgggtgtg
ccagaggaccccccagaactaagaagggagggcgagtgggtctccattccccgagaagccaggggcagggtg
ggatggggaagaccaggagcagagtcgagcctcacagaagccagcgcgggtctctgctcagcaccccagcc
ggggctctggacccagggtaacagccccagttcatcccaacccctctcagagcctcaagaggggtagctcg
gctgccggaagagaggggtgccctatccctggcaacccctccacgtagcgtaccccagcacctgccaccgc
ctttgccatttctttgagcttgaagtactagt

>Figure 6I; EREG SacI-SpeI [pAG647]

gagctcaatcacagcccaaaatttgatggactaattattattttaaaatatatgaagacaataattctaca
tgttgtcttaagatggaaatacagttatttcatcttttattcaaggaagttttaactttaatacagctcag
taaatggcttcttctagaatgtaaagttatgtatttaaagttgtatcttgacacaggaaatgggaaaaaac
ttaaaaattaatatggtgtatttttccaaatgaaaaatctcaattgaaagcttttaaaatgtagaaactta
aacacaccttcctgtggaggctgagatgaaaactagggctcattttcctgacatttgtttatttttggaa
gagacaaagatttcttctgcactctgagcccataggtctcagagagttaataggagtattttgggctatt
gcataaggagccactgctgccaccacttttggattttatgggaggctccttcatcgaatgctaaacctttg
agtagagtctccctggatcacataccaggtcagggaggatctgttcttcctctacgtttatcctggcatgt
gctagggtaaacgaaggcataataagccatggctgacctctggagcaccaggtgccaggacttgtctccat
gtgtatccatgcattatataccctggtgcaatcacacgactgtcatctaaagtcctggccctggcccttac
tattaggaaaataaacagacaaaaacaagtaaatatatatggtcctatacatattgtatatatattcatat
acaaacatgtatgtatacatgaccttaatggatcatagaattgcagtcatttggtgctctgctaaccattt
atataaaacttaaaaacaagagaaaagaaaaatcaattagatctaaacagttatttctgtttcctatttaa
tatagctgaagtcaaaatatgtaagaacacattttaaatactctacttacagttggccctctgtggttagt
tccacatctgtggattcaaccaaccaaggacggaaaatgcttaaaaaataatacaacaacaacaaaaaata
cattataacaactatttacttttttttttttctttttgagatggagtctcgctctgttgcccaggttggag
tgcagtggcacgatctcggctcactgcaacctcacctcccgggttcaagagatcctcctgcctcagcctcc
tgagcagctgggactacaggcgcatgccaccatgcccagctaatttttgtatttttagtagaggcggggtt
tcaccatgttggccaggatggtctcaatctcctaactagt

>Figure 6I; AKR1D1 SacI-SpeI [pAG648]

gagctccctgaacagattttcactcccatgagtgccaagacggtgcaatgggtagtcccctagatgtgaa
aatgaagagagagggttttaccatcctgagaagaaataatgatggaaacatgtttaatgtttgtgcagtgt
aaatgactttgactcagtcacattgaagtaaaaatattaaaatctgttgaaataactcttaggaaattatc
aactaatttttcagatcagtatcttctagattccagacagaaaaaaattacacttcagaaaagacatcaa
aggcaacatatgacaacaagtaatttatgaatctgggtagtagcgttggtaatctgagttctttaagggtt
aacaggacaacgaagtgcatgtggcagtgtgctggcagtggccttgaggctttggaccattggttacaaaa
cagacacagccaagataagatccacacacacattattaacaaggaagtgatttgctgcaccttgagttgag
aggactacatgtagaaaagtcttaaaatagagctaaacaccacagtggtcaacaaagccatcataatgttg
gtgtttgtttccctccaatgtatgtatgtttagtttttatccaacctgaggaatgaaaacttaactggatc
tctcttgcatccttaaagggcctgagtctcaacatggctgctgatccatacttacacatcttactgtcaat
cttgcctacattgattatagaaccactattacgtgaaaaggcttgaaacaaccaacactagt

>Figure 6I; CNNM1 SacI-SpeI [pAG649]

gagctcacacatttcatgcatagcttggctcagaaggtgccattggcagacaggcacatgggaggctggag
taggaggtctgaagattagttcaggggatggaccaagaatttcccccagagctttaaagaagtgggactca
gccatgttggcgcgtgattgacattacagcacagaaaactgttagtgactggtttcctgttagataagggt
tccagcagcctggggcagtatgtctcagctggaatggaaagaatgtgagatggaacctcaagtcactgttt
ttaccagggacacatctgttttggctcccaatcagcagtcttcaatcgatcaataattctgctctggaaga
gaaggaacagggagcagagagacccaactgggagccagagatggaacttcaggtcttaagtgcaaatcaaa
gcaaaaaacaaacaaaacttacatggaaaaactgtaagtgctgaaagcaagtttagccatgacaaaccaaa
gagtgcccaggtcagccaagaaagatacataatctcatgggacttcagtgggagttacacaggaatgttga
agaatcattcttctttttcatgcatttgtccttctcccacccccttactacaccctagcagatcagctgag
tgtactttattccaagaacttactggatctctggtttttctcctgaagttggggcaggtgcaattccaagc
ataaccaccagatggcagagtgaccgcgcatacctgcttccaagaataaaacagttctgaaaagcaaccgc
aaagccgggcgcggtggctcacacctgtaatcccaacagtttggaagaccgaggcgggtggatcacttgaa
gtcaggagtttgagaccagcctggccaacatggtgaaaccccccatctctcctgggcatgtagtcccagcta
ctcgggaggctgaggcaggagaatcgcttgaacctgggaggcagaggttgcagtgagctgagatcacacca
ctgcacttcatcctaggcaacagagcgagactctgtctcccccctcaaaaaaaaaaaaagaaagaaagaaag
agaaaagaaaagaaaagcaaccacagccagccttagggaaaacttggaagtaagtgaaatttgtcttcaga
ataactatctcccttttctgatctgtctcctactctttagatgttctcagtcaagtactcactgaactcatt
gatcgagtgctgtctgctaaatctccaaaccattcccaactagt

>pIS1 cloning site, pRL-TK sequence is in italics, restriction sites are underlined
*gaacaataattctag*gagctctataccggtctcgatatcactactagtgt*tctagagcggccgct*

Table S6. Context score parameters for different miRNA target sites, with Pearson correlation coefficient (r) and corresponding *P* values indicating the confidence in a non-zero slope.

8mer, mean value -0.31

Determinant	Slope	y-intercept	r	*P* value
3' pairing	–0.0041	–0.299	–0.01	0.80
Local AU	–0.64	0.055	–0.23	$<10^{-13}$
Position	0.000172	–0.38	0.18	$<10^{-8}$

7mer-m8, mean value -0.161

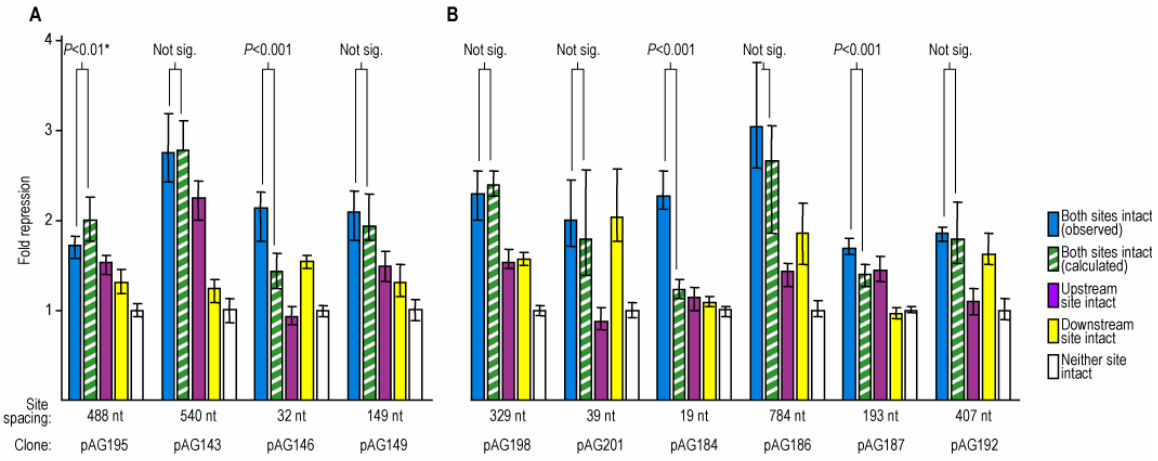
Determinant	Slope	y-intercept	r	*P* value
3' pairing	–0.031	–0.094	–0.07	$<10^{-3}$
Local AU	–0.50	0.108	–0.21	$<10^{-32}$
Position	0.000091	–0.198	0.11	$<10^{-8}$

7mer-A1, mean value -0.099

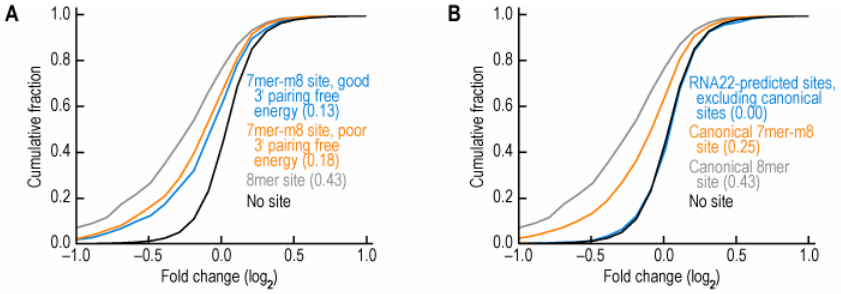
Determinant	Slope	y-intercept	r	*P* value
3' pairing	–0.0211	–0.0211	–0.06	$<10^{-2}$
Local AU	–0.42	0.137	–0.20	$<10^{-26}$
Position	0.000072	–0.131	0.10	$<10^{-7}$

6mer, mean value -0.015

Determinant	Slope	y-intercept	r	*P* value
3' pairing	–0.00278	–0.0091	–0.01	0.52
Local AU	–0.241	0.115	–0.14	$<10^{-26}$
Position	0.000049	–0.033	0.07	$<10^{-7}$



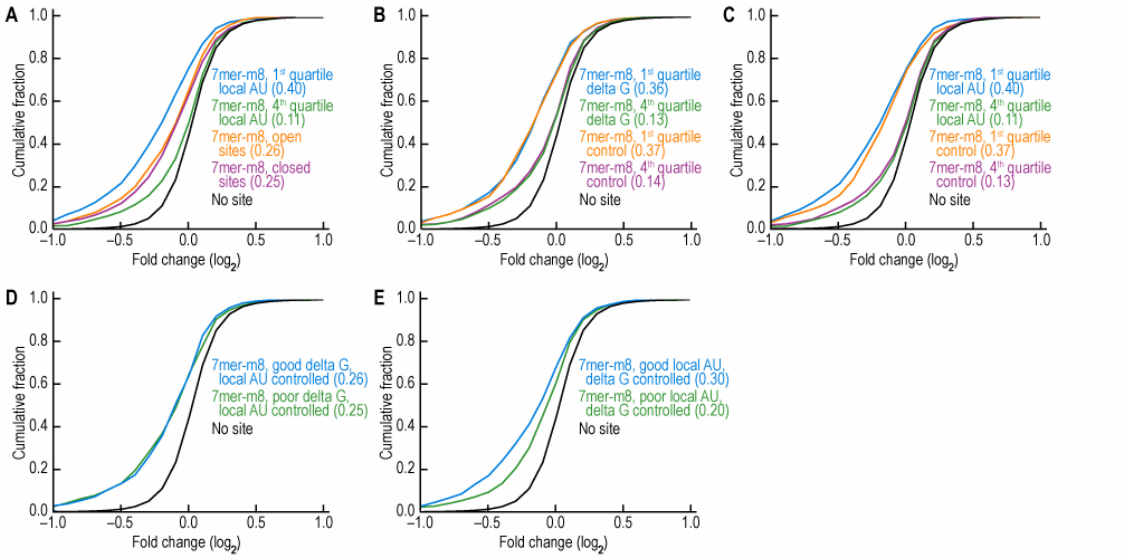
Supplemental Figure 1. MicroRNA-mediated repression of *Renilla* luciferase reporter genes fused to 3' UTR fragments containing two miR-1 (A) or miR-124 (B) sites, or mutant derivatives. After normalizing to the firefly transfection control, luciferase activity from HeLa cells cotransfected with each reporter construct and its cognate miRNA (A: miR-1; B: miR-124) was normalized to that from cotransfection of each reporter with a non-cognate miRNA (A: miR-124; B: miR-1). Plotted are the normalized values, with error bars representing the third largest and third smallest values among 12 replicates. *P* values (Wilcoxon rank-sum test) indicate whether repression from a reporter containing both sites (blue) was significantly greater than that expected from additive effects (green). For this additive model, repression expected from a reporter with two sites was the product of repression observed from otherwise identical reporters containing single intact sites (purple and yellow). The two UTR fragments containing the most closely spaced sites (pAG146 and pAG184) manifested significantly cooperative repression. pAG184 was selected for further study (Figure 1G). For six of the eight fragments containing less closely spaced sites, repression from constructs containing both sites (blue) did not statistically differ from that expected for additive repression (green). One exception was pAG187, which manifested significant cooperativity (pAG187), although the magnitude of the affect was relatively low. The other exception was pAG195, which manifested apparently negative cooperativity, although this result was of more borderline statistical significance. Parental clones are as described (Farh et al., 2005).



Supplemental Figure 2. Evaluation of previous algorithms designed to score productive pairing involving the 3' region of the miRNA.

(A) Performance of sites ranked using an energy-based rubric for predicting and scoring 3' supplementary pairing resembling those rubrics developed previously (Lewis et al., 2003; John et al., 2004; Krek et al., 2005). For each transfected miRNA, messages with single 7mer-m8 sites were identified together with the 20 UTR nucleotides upstream of each site, and folded with the miRNA using RNAhybrid (Rehmsmeier et al., 2004), without permitting the pairing of UTR nucleotides with each other or the pairing of miRNA nucleotides with each other. For each of the 11 miRNAs, messages were partitioned into four quartiles based on pairing free energy of their sites. Shown are the aggregate results for all 11 miRNAs for the quartile with the lowest deltaG values (most stable predicted pairing) and highest deltaG values. The quartile with the lowest deltaG values performed significantly worse than that with the highest deltaG values ($P = 0.0056$, two-sided K-S test). (Some previous methods that consider pairing to the 3' region of the miRNA normalize the predicted pairing free energy for each site to that of the fully paired miRNA. The results of this figure would have been the same if we had done similar normalization because for each of the 11 miRNAs considered, messages with sites were split evenly into the four equal quartiles before the sites for different miRNAs were combined.)

(B) Performance of sites with extensive Watson-Crick and G:U pairing along the length of the miRNA, but without canonical seed pairing, as proposed by Miranda et al. (2006). MicroRNA target predictions for the transfected miRNAs were obtained from the RNA22 web site. These predictions had been assigned to miRNAs using pairing parameters (G = 0, M = 14 and E = −25 Kcal/mol), which were more stringent than those used in the published work (Miranda et al., 2006), but we used them because the larger set described in the published work was not disclosed in November 2006. Predictions for miR-124 were not considered because these were for a miR-124 variant that was offset by one nucleotide from the miR-124 that we transfected. The predictions for the remaining 10 transfected miRNAs were filtered to remove those with canonical seed sites (1161 of the 1720 predictions for these 10 miRNAs had at least a 6mer site in their 3' UTR). The remainder consisted of sites with imperfect seed pairing and extensive 3' pairing. When considering their downregulation on the microarray, such sites performed no better than all the other genes with no seed site ($P = 0.096$, one-sided K-S test). The performance of canoncial 7mer-m8 and 8mer sites (Figure 1B) is shown for comparison. This result suggests that the many thousands of noncanonical predictions proposed by Miranda et al. (2006) did not include any more functional predictions than expected by chance.



Supplemental Figure 3. Evaluation of existing algorithms designed to score site accessibility by predicting UTR secondary structure.

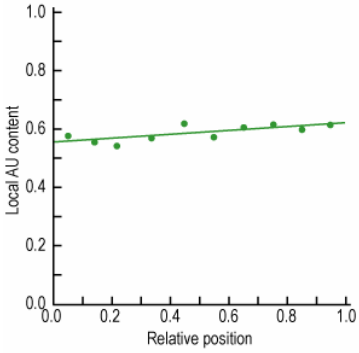
(A) Performance of sites with three or more nucleotides in open structure, as determined by the algorithm of Robins et al. (2005). Messages with a single 7mer-m8 site were split with respect to open structure (defined as at least 3 nt continuous unpaired within site). Shown are cumulative distributions of mRNA changes on the array for messages in the highest and lowest quartiles. A slight difference was observed between open and closed sites, but this difference was not significant ($P = 0.066$, 1-sided K-S test). Performance of sites in the highest and lowest quartiles with respect to local AU content (Figure 3) is shown for comparison.

(B) Performance of sites flanked by low or high predicted stability of UTR regions flanking the seed site, as determined by the algorithm of Zhao et al. (2005, 2007). Messages with a single 7mer-m8 site were split into quartiles with respect to average predicted stability of 70-nt segments immediately flanking both sides of the site. Shown are cumulative distributions of mRNA changes on the array for messages in the highest and lowest quartiles. To control for the effects of global nucleotide composition, 10 random 70-nt regions in the same UTR were folded and messages were again split into quartiles with respect to average predicted stability. Among sites that were significantly downregulated on the array ($P < 0.01$), there was no significant difference in stability when comparing the 70-nt UTR regions flanking the site and randomly selected 70-nt regions from the same UTR ($P = 0.120$, Wilcoxon rank-sum test). Thus, the success of this algorithm in predicting down regulated messages could be attributed to a correlation between predicted stability near the site and more global properties of the UTR.

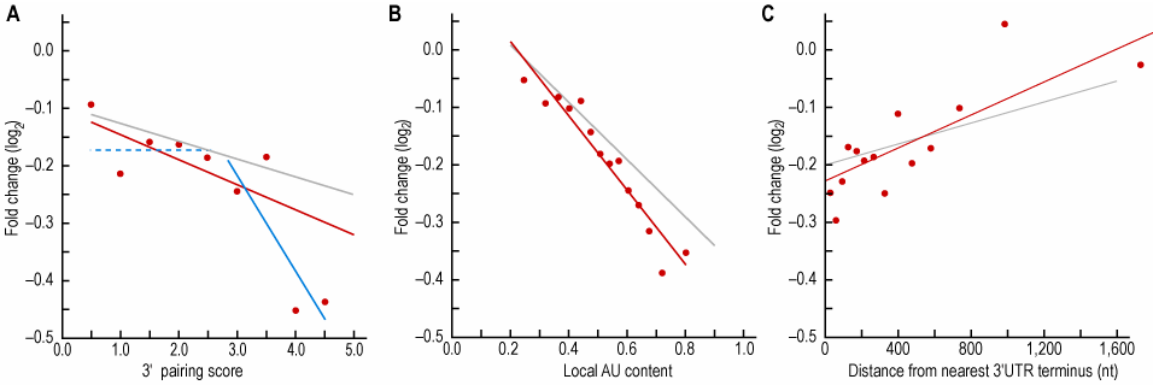
(C) Performance of sites in the highest and lowest quartiles with respect to local AU content, as scored in Figure 3B. To control for the effects of global nucleotide composition, 10 random positions in the same UTR were evaluated for local AU content. For sites that were significantly downregulated on the array ($P < 0.01$), the difference in AU content was significant when comparing the region immediately adjacent to the authentic site and randomly selected regions from the same UTR ($P = 0.0069$, Wilcoxon rank-sum test.) Although local AU content correlated with AU content throughout the remainder of the UTR, not all of the success of the algorithm could be attributed to a correlation with the more global property.

(D) Performance of sites flanked by low or high predicted stability of UTR regions flanking the seed site, after controlling for local and global AU content. To measure the residual effect of predicted secondary structure after controlling for local and global AU content, 1000 pairs of matched sites were picked randomly such that 1) their local AU content was within 5 percentiles of each other; 2) their global AU content was within 5 percentiles of each other; and 3) their average predicted folding stability in the regions flanking the sites, as determined by the algorithm of Zhao et al. (2005), was at least 30 percentiles apart. Repression was not significantly greater among sites with weaker predicted folding energy when AU content was held constant ($P = 0.088$, 1-sided K-S test). Thus, the specific predicted secondary structures scored by the algorithm of Zhao et al. were not informative after controlling for local AU content.

(E) Performance of sites with different local AU contents, after controlling for global AU content and predicted folding stability in the regions flanking the sites. 1000 pairs of matched sites were picked randomly such that 1) their global AU content was within 5 percentiles of each other; 2) their average predicted folding stability in the regions flanking the sites, as determined by the algorithm of Zhao et al. (2005), was within 5 percentiles of each other, and 3) their local AU content was at least 30 percentile apart. Repression was significantly greater among sites with high local AU content ($P < 10^{-7}$, 1-sided K-S test.). The Spearman correlation between predicted stability and global AU content was 0.6665, while the Spearman correlation between local AU content and global AU content was 0.6253. Because of the high co-correlation of these variables with global nucleotide content, random sampling of sequences in the same UTR was necessary to control for global AU and reveal underlying context determinants in the local neighborhood of effective sites.



Supplemental Figure 4. Evaluation of correlations between determinants used to generate target site context scores. We observed a modest but significant ($P = 0.0015$) Spearman correlation between local AU content and position within 3' UTR (evaluated as in Figure 3B and 5, respectively), indicating a slight increase in local AU content as the local window traversed from the 5' to the 3' termini of the 3' UTR. This correlation could not account for the contribution of position to final context score, because sites located close to either terminus of the 3' UTR were most effective, rather than sites near the 3' terminus of the 3' UTR.
We evaluated whether scores for the three determinants used to generate the final context score were significantly correlated by Spearman correlation, and found that they were not: position score and 3' pairing score, $P = 0.092$; position score and local AU score, $P = 0.51$; local AU score and 3' pairing score, $P = 0.40$.

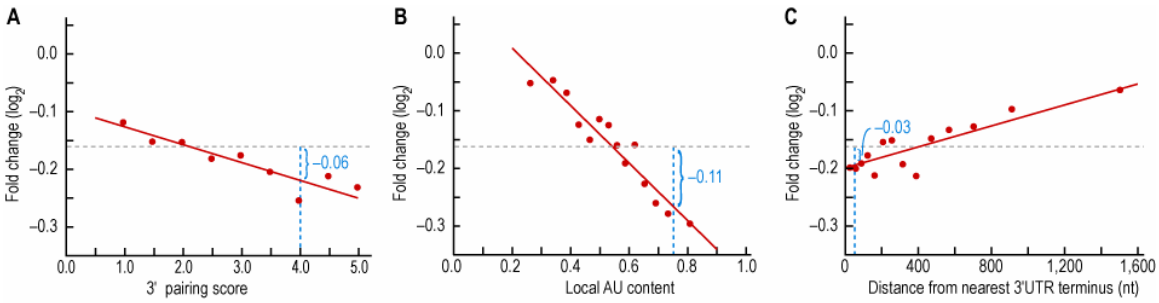


D

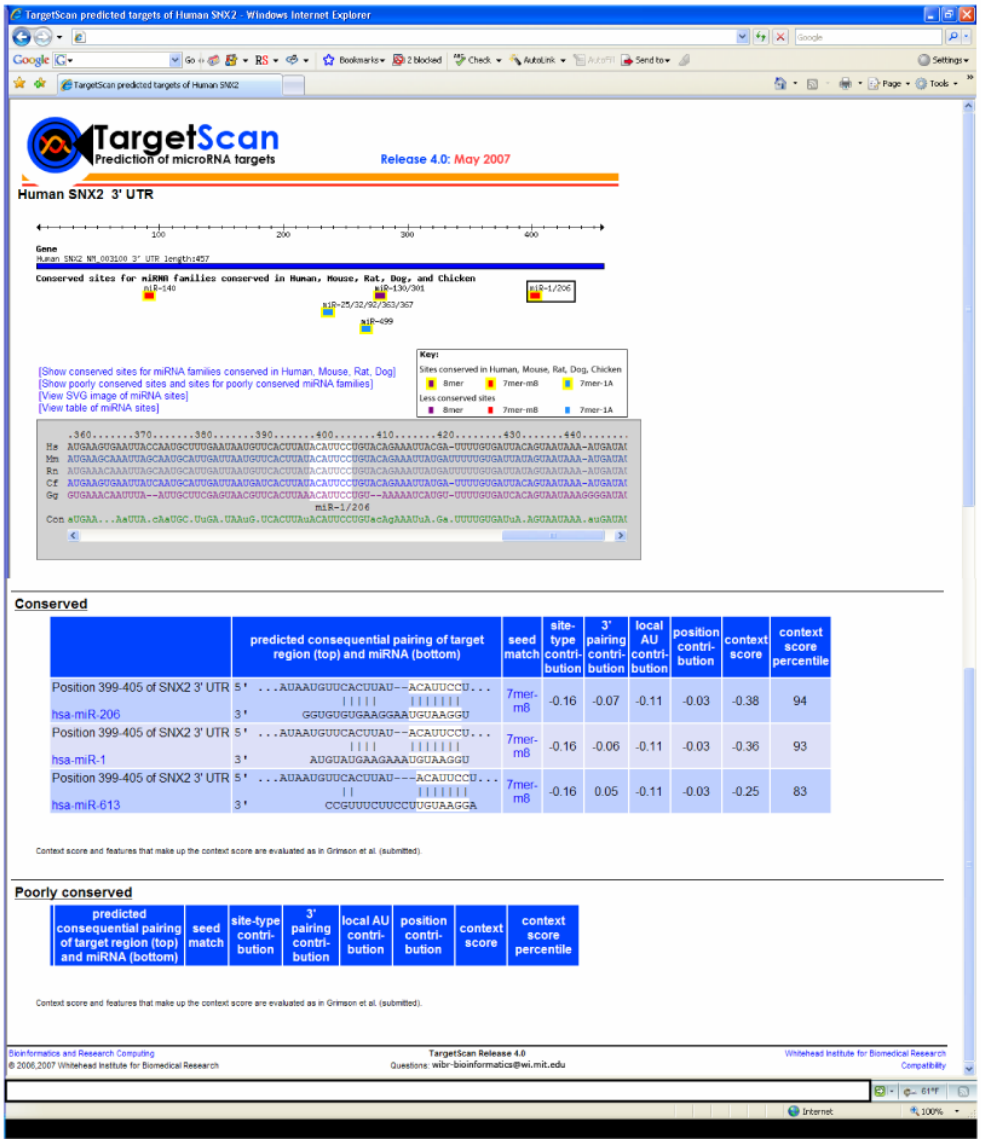
siRNA	7mer-m8 site	Sequence (sense strand)	Effective Strand
MAPK14-1	CUGCGGU	CCUACAGAGAACUGCGGUU-dTdT	-
MAPK14-2	AAUCACA	AUGUGAUUGGUCUGUUGGA-dTdT	+
MAPK14-3	CUAAAGU	UUCUCCGAGGUCUAAAGUA-dTdT	-
MAPK14-4	GACCUAA	UAAUUCACAGGGACCUAAA-dTdT	-
MAPK14-5	UCCUUAU	CCAGUGGCCGAUCCUUAUG-dTdT	-
MAPK14-6	AGUAGGC	UGCCUACUUUGCUCAGUAC-dTdT	+
MAPK14-7	CUGAUGA	GUCAUCAG CUUUGUGCCAC-dTdT	+
MAPK14-8	GGGAACU	GGCCUUUUCACGGGAACUC-dTdT	-
IGF1R-1	UUACCGA	GCUCACGGUCAUUACCGAG-dTdT	-
IGF1R-2	UCCUCAG	CCUGAGGA ACAUUACUCGG-dTdT	+
IGF1R-3	GGUCAGC	UGCUGACCUCUGUUACCUC-dTdT	+
IGF1R-4	CCGUGUC	CGACACGGCCUGUGUAGCU-dTdT	+
IGF1R-5	UGGCCGG	GAUGAUUCAGAUGGCCGGA-dTdT	-
IGF1R-6	GCUGCAA	CUUGCAGCAACUGUGGGAC-dTdT	+
IGF1R-7	CCGUGAG	CCUCACGGUCAUCCGCGGC-dTdT	+
IGF1R-12	UCAGCAU	AAUGCUGACCUCUGUUACC-dTdT	+
IGF1R-13	CGGUAAU	CAUUACCGAGUACUUGCUG-CU	+
IGF1R-16	CCUCGGA	GGCCUCGAGAGCCUCGGAG-AC	-

Supplemental Figure 5. Evaluation of scoring schemes in an independent dataset. We examined correlations between fold change in mRNA level and 3' pairing score (A), local AU content score (B) and position (C), observed in a dataset not used in development of scoring rubrics (Jackson et al., 2003). For each of the transfection experiments, motif analysis was performed (Farh et al., in preparation) to determine the motif most significantly associated with downregulation. Names and sequences of siRNA sense strands from Jackson et al. (2003) transfection experiments are shown (D); in general, the seed site for one of the strands (indicated as +/-) predominated in accordance with a strand preference for the 5' end with the weaker pairing energy (Schwartz et al., 2003). Results of 18 experiments, each transfecting a duplex for a different siRNA, were consolidated. Messages with a single 7mer-m8 site were scored for 3' pairing, local AU content and position using our previously developed rubrics (Figures 2F, 3B, and 5, respectively) and analyzed as in Figure 6A-C (shown in red; $P < 10^{-4}$, r = -0.08, $P < 10^{-31}$, r = -0.24, $P < 10^{-6}$, r = 0.10, Pearson correlations). For comparison, the regression lines derived in Figure 6A-C are also shown (grey). The fold changes were larger in the set of siRNA transfections than in the miRNA transfections, and thus the slopes of the regression lines are correspondingly steeper.

For the 3' pairing feature (A), messages with sites scoring >3.5 were markedly downregulated, a result consistent with the miR-155 knockout data (Figure 7E) but in contrast to the linear trend observed for the miRNA transfections (Figure 6A). Perhaps sites with low scores (<3.0), which includes most of the sites (81%), are negligibly affected by 3' pairing, and those with higher scores are affected more than anticipated by fitting all the data to a linear model (red line). To better model this behavior, sites with low scores were excluded (average fold change indicated by dashed line), and the remaining data was modeled by linear regression (blue line). Although more complex, this approach also appears to better reflect the miR-155 in vivo results (Figure 7E). However, because of their rarity, more data will be required to accurately quantify the efficacy of sites with exceptional 3' pairing (scores >4.0), which comprise only ~1% of the population.


A
B
C
Fold change (log2)
0.0
−0.1
−0.2
−0.3
−0.06
3' pairing score
0.0
1.0
2.0
3.0
4.0
5.0
−0.11
Local AU content
0.0
0.2
0.4
0.6
0.8
1.0
−0.03
Distance from nearest 3'UTR terminus (nt)
0
400
800
1,200
1,600

D


TargetScan predicted targets of Human SNX2 - Windows Internet Explorer
TargetScan
Prediction of microRNA targets
Release 4.0: May 2007
Human SNX2 3' UTR
Gene
Human SNX2 NM_003100 3' UTR length=457
Conserved sites for miRNA families conserved in Human, Mouse, Rat, Dog, and Chicken
miR-140
miR-130/301
miR-1/206
miR-25/32/92/363/367
miR-499
[Show conserved sites for miRNA families conserved in Human, Mouse, Rat, Dog]
[Show poorly conserved sites and sites for poorly conserved miRNA families]
[View SVG image of miRNA sites]
[View table of miRNA sites]
Key:
Sites conserved in Human, Mouse, Rat, Dog, Chicken
8mer
7mer-m8
7mer-1A
Less conserved sites
8mer
7mer-m8
7mer-1A
Hs
Mm
Rn
Cf
Gg
Con
miR-1/206
Conserved
predicted consequential pairing of target region (top) and miRNA (bottom)
seed match
site-type contribution
3' pairing contribution
local AU contribution
position contribution
context score
context score percentile
Position 399-405 of SNX2 3' UTR
hsa-miR-206
7mer-m8
-0.16
-0.07
-0.11
-0.03
-0.38
94
Position 399-405 of SNX2 3' UTR
hsa-miR-1
7mer-m8
-0.16
-0.06
-0.11
-0.03
-0.36
93
Position 399-405 of SNX2 3' UTR
hsa-miR-613
7mer-m8
-0.16
0.05
-0.11
-0.03
-0.25
83
Context score and features that make up the context score are evaluated as in Grimson et al. (submitted).
Poorly conserved
predicted consequential pairing of target region (top) and miRNA (bottom)
seed match
site-type contribution
3' pairing contribution
local AU contribution
position contribution
context score
context score percentile
Context score and features that make up the context score are evaluated as in Grimson et al. (submitted).
Bioinformatics and Research Computing
© 2006,2007 Whitehead Institute for Biomedical Research
TargetScan Release 4.0
Questions: wibr-bioinformatics@wi.mit.edu
Whitehead Institute for Biomedical Research
Compatibility
Internet
100%

Supplemental Figure 6. Deriving and annotating the context score for a miR-1 target site within the *SNX2* 3'UTR, using the regression parameters of Supplemental Table 6.

(A) Regression relating repression on the array and 3' pairing score, identical to Figure 6A but illustrating how the 3' pairing contribution of the context score was determined for the miR-1 site of *SNX2*. The miR-1 site of *SNX2* had a 3' pairing score of 4 (scored as in Figure 2F). Using the slope of the regression line (–0.031, see Supplemental Table 6 for all parameters used to calculate context scores), y-intercept value (–0.094), and mean fold change for a 7mer-m8 site (–0.161), a score of 4 corresponded to an expected fold change of –0.06 ($\log_2$) over that of an average 7mer-m8 site, calculated as follows: 4(–0.031) – 0.094 + 0.161 = –0.06.

(B) Regression relating expected repression on the array and local AU content score, identical to Figure 6B but illustrating how the local AU content contribution of the context score was determined for the miR-1 site. The miR-1 site of *SNX2* had a local AU content score of 0.76 (scored as in Figure 3B). Using the slope of the regression line (–0.50), y-intercept value (0.108), and mean fold change for a 7mer-m8 site (–0.161), a score of 0.76 corresponded to an expected fold change of –0.11 ($\log_2$) over that of an average 7mer-m8 site, calculated as follows: 0.76(–0.50) + 0.108 + 0.161 = –0.11.

(C) Regression relating expected repression on the array and position within 3'UTR, identical to Figure 6C but illustrating how the position contribution of the context score was determined for the miR-1 site. The miR-1 site of *SNX2* was located 52 nt from the closest UTR terminus. Using the slope of the regression line (0.000091), y-intercept value (–0.198), and mean fold change for a 7mer-m8 site (–0.161), a distance of 52 nt corresponded to an expected fold change of –0.03 ($\log_2$) over that of an average 7mer-m8 site, calculated as follows: 0.000091(52) – 0.198 + 0.161 = –0.03.

(D) TargetScan 4.0 web page illustrating how context scores derived from the newly identified specificity determinants have been annotated for predicted sites in mammals (Targetscan.org). Shown is the context score of the miR-1 target site in *SNX2*, with the contributions from the site type and three context determinants. The combined context score for the miR-1 target site in *SNX2* was equal to the mean repression for the 7mer-m8 site plus the three contributing scores (–0.161 + –0.06 + –0.11 + –0.03 = –0.36). This context score corresponded to a predicted repression in the top 7th percentile of miR-1 sites. To account for the other two context determinants, sites falling within 15 nt of a stop codon are flagged, and sites within optimal distance for cooperative action are displayed.