

Supplementary Material:

Logistic Regression Model: The model is a generalized linear model in which we do not directly fit the features (the terminal nucleotide and thermodynamic asymmetry) linearly with the outcomes (silencing efficiency) but attempt to estimate the probability of the siRNA being highly efficient or inefficient in silencing. The reasons for this are two-fold: (1) we include both continuous (energy) and discrete (terminal nucleotide) features, and (2) the relationship between features and outcomes is not linear⁽³⁷⁾.

Therefore, we apply the logistic regression model, a generalized linear model that assumes the logit (log-odds) of the siRNA being highly efficient are determined by the features:

$$\log\left(\frac{P(\text{high}|X)}{P(\text{med}|X)}\right) = \beta_0 + \beta^T X \quad [1]$$

in which $P(\text{high}|X)$ is the probability that the siRNA (given its feature vector X) is highly active, while $P(\text{med}|X)$ is the average siRNA silencing efficiency, so $P(\text{high}|X)/P(\text{med}|X)$ represents the “Odds” that the siRNA stands out as a “good” siRNA. In a logistic regression model, the odds are associated with an exponential function of the linear combination of the features.

Silencing is a discrete Poisson process, and the frequencies of silencing are subject to an exponential distribution. Therefore, we can relate the odds of good silencing and the sequence features with exponential functions.

Similarly, the log-odds of the siRNA being inefficient are determined as:

$$\log\left(\frac{P(\text{low}|X)}{P(\text{med}|X)}\right) = \beta'_0 + \beta'^T X \quad [2]$$

The model predicts the posterior probability of siRNA being highly efficient and inefficient:

$$P(\text{high}|X) = \frac{\exp(\beta_0 + \beta^T X)}{1 + \exp(\beta_0 + \beta^T X) + \exp(\beta_0' + \beta'^T X)} \quad [3]$$

$$P(\text{med}|X) = \frac{1}{1 + \exp(\beta_0 + \beta^T X) + \exp(\beta_0' + \beta'^T X)} \quad [4]$$

$$P(\text{low}|X) = \frac{\exp(\beta_0' + \beta'^T X)}{1 + \exp(\beta_0 + \beta^T X) + \exp(\beta_0' + \beta'^T X)} \quad [5]$$

The siRNAs in the datasets are divided into three groups (high/med/low) based on their activity (discretization with equal frequency), and we use Maximum Likelihood Estimation to learn the parameters in the logistic regression model.

EGFP sequence 5' to 3' (targeted regions highlighted; darker highlights indicate overlapping target regions):

atggtgagcaagggcgaggagctgtcaccgggggtggtgccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgt
gtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagtcatctgcaccaccggcaagctgcccgtgccttgccc
acctcgtgaccaccctgacctacggcgtgcagtgttcagccgctaccccgaccacatgaagcagcagcacttctcaagtcgccatgc
ccgaaaggctacgtccaggagcgcaccatcttcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgagggcgacac
cctggtgaaccgcatcgagctgaaggcgcacgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcgcacgaggtgaactcaagatccgccacaacatcgaggacggcagcgtgcag
ctgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcaccagtcgccct
gagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgggatcactctcggcatggacgagctgt
acaagaagcttagccatggcttcccgggaggtggaggagcagatgatggcacgctgcccattgtcttgcccaggagagcgggatg
gacctcaccctgcagcctgtgcttctgctagatcaatgtgtag

PKR sequence 5' to 3' (targeted regions highlighted; darker highlights indicate overlapping target regions):

aagctccaaccaggatcacgggaagaagaatggctggtgatcttcagcaggtttctcatggaggaacttaatacataccgtcagaagca
gggagtagtacttaaatcaagaactgcctaattcaggaccccatgataggaggtttacattcaagttataatagatggaagagaatttc
cagaa**ggtgaaggtagatcaaga**aggaagcaaaaaatgccgcagccaaattagctggtgagatacttaataaggaaaagaaggcagta
gtcctttatttgacaacaacgaattctcagaaggattatccatggggaattacat**aggccttatcaatagaatt**gcccagaagaaaagacta
actgta**aattatgaacagtgtgcat**ggggggtgcatg**ggccagaaggatttcatt**ataaatgcaaaatgggacagaaagaatatagtattggt
acaggttctactaaacaggaagcaaaacaattggccgctaaacttgcatatcttcagatattatcagaagaaacctcagtgaatctgactacc
tgcctctggtcttttgcactacgtgtgagcccaagcaactcttagtgaccagcacactcgtcttgaatcatca**ctgaagggtgactctc**
agcagatacatcagagataaattctaacagtgacagtttaaacagttctcgttgcttatgaatggtctcagaaataatcaagggaaggcaaaa
agatctttggcaccagattgacfttctgacatgaaagaacaagaactgactgtggacaagag**gtttggcatggattttaa**gaaatagaat
taattggctcaggtggatttggccaagtttcaaagcaaaacacagaattgacggaaagacttacgttattaacgtgtaaatataataac**gag**
aaggcggagcgtgaagtaaaagcatt**ggcaa**aactgatcatgtaaatattgttactacaatggctgtgggatggattgattatgatcctga
gaccagtgatgattctctgagagcagtgattatgatcctgagaacagcaaaaatagttcaaggtcaagactaagtgcctttatccaaatg
gaattctgtgataaaggaccttggacaatggattgaaaaagaagaggcgagaaactagacaaagtt**ttggctttggaactcttt**gaaaa
ataacaaaaggggtggattatatacattcaaaaaaattatcatagagatcttaagccaagtaataatcttagtagatacaaaacaagtaaa
gattggagactttggacttgaacatctctgaaaaatgatggaaagcgaacaaggagtaagggaactttgcgatacatgagcccagaacag
attcttcgaagactatggaaaggagtgacctctacgctttggggctaattcttctgactcttcatgtatgtgacactgcttttgaacat
caaagttttcacagacctacgggatggcatcatctcagatatattgataaaaaagaaaaactctctacagaaattactcacaagaaacct
gaggatcgaactaacacatctgaaataactaaggacctgactgtgtggaagaaaagcccagagaaaaatgaacgacacacatgtagagc
ccttctgaaaaagtatcctgcttctgatatgcagtttcttaaatatctaaaatctgtagggaatatcaatagatatttacttttatttaattg
ctttaaattttactattttactaatctttctgcagaaacagaaaggtttctttttgcttcaaaaacattcttacattttacttttctggtcactct
ttattctttcttttttaaaagacagagctcgtctgttggccaggetggagtgcaatgacacagcttggctcactgcaactctgctcttggg
ttcaagtattctcctgcctcagcctcctgagtagctggattacaggeatgtgccaccaccactaatttttgttttaataaaagacaggggt
tcaccatgttggccaggtggtctcaactcctgacctcaagtaatccacctgcctcggcctccaaagtgtgggattacagggatgagcc
accgcgccagcctcatctctttgtctaaagatgaaaaaccacccccaaatftttttatactattaatgaatcaatcaattcatatctattatt
aaatttctaccgcttttagccaaaaaaatgtaagatcgttctctgcctcacatagcttacaagccagctggagaaatatggtactcattaaaa
aaaaaaaaaaaa

Dharmacon EGFP selection output (sorted by Dharmacon “Score”):

Sense Strand Sequence	Start Pos	GC %	Score	Algorithm Rank
GGCACAAGCTGGAGTACAA	416	53	91	17
AGGTGAAGTTCGAGGGCGA	335	58	87	140
CCGCCATGCCCGAAGGCTA	260	68	83	31
TGGACGAGCTGTACAAGAA	701	47	83	107
ACTTCAAGGAGGACGGCAA	389	53	81	279
GCAAAGACCCCAACGAGAA	626	53	81	120
TCTTCAAGGACGACGGCAA	299	53	80	313
GCATCGACTTCAAGGAGGA	383	53	79	100
GCATGGACGAGCTGTACAA	698	53	79	66
CGCCGGAGGTGGAGGAGCA	737	74	79	14
AGGTGGAGGAGCAGGATGA	743	58	79	37
GCGCCGAGGTGAAGTTCGA	329	63	78	13
AGAACGGCATCAAGGTGAA	476	47	77	188
TGGTCGAGCTGGACGGCGA	47	68	76	178
GGCCACAAGTTCAGCGTGT	73	58	76	135
CCGGCGAGGGCGAGGGCGA	92	84	75	98
TGGCCGACAAGCAGAAGAA	461	53	75	67
CTGCTAGGATCAATGTGTA	827	42	75	n/a
ACATGAAGCAGCACGACTT	233	47	74	302
CCATCTTCTTCAAGGACGA	293	47	74	133
CGGAGGTGGAGGAGCAGGA	740	68	74	33
GCCACAACGTCTATATCAT	443	42	73	180
ACGAGAAGCGCGATCACAT	638	53	73	321
GTACAAGAAGCTTAGCCAT	711	42	73	551
ACGTAAACGGCCACAAGTT	65	47	72	232
GTAACCGGCCACAAGTTCA	67	47	72	108
GCACCATCTTCTTCAAGGA	290	47	72	45
GCAAGGGCGAGGAGCTGTT	8	63	71	216
CGAGCTGGACGGCGACGTA	51	68	71	6
CCTACGGCGTGCAGTGCTT	197	63	71	280
GGGCACAAGCTGGAGTACA	415	58	71	1
TCTATATCATGGCCGACAA	452	42	71	223
AAGAACGGCATCAAGGTGA	475	47	71	157
GCAAGCTGACCCTGAAGTT	122	53	70	211
ACAAGCTGGAGTACAATA	419	42	70	105
CCGACCACTACCAGCAGAA	539	58	70	88
ACATGGTCCTGCTGGAGTT	653	53	70	304
TGAGCAAGGGCGAGGAGCT	5	63	69	577
AGCTGGACGGCGACGTAAA	53	58	69	35

TGAACCGCATCGAGCTGAA	362	53	69	208
GCAACATCCTGGGGCACAA	404	58	69	93
ACAACAGCCACAACGTCTA	437	47	69	114
AAGCAGAAGAACGGCATCA	469	47	69	55
GCGTGCAGCTCGCCGACCA	527	74	69	40
TCGGCATGGACGAGCTGTA	695	58	69	68
GCGTGTCCGGCGAGGGCGA	86	79	68	79
GTGCAGTGCTTCAGCCGCT	205	63	68	276
AGGACGACGGCAACTACAA	305	53	68	63
AGTACAACACTACAACAGCCA	428	42	68	229
ACGTCTATATCATGGCCGA	449	47	68	141

Ambion EGFP Selection Output (sorted by ascending target position):

Target	Position	% GC	Algorithm Rank
Target sequence 1: AAGGGCGAGGAGCTGTTACCC	12	61.9	369
Target sequence 2: AAACGGCCACAAGTTCAGCGT	71	52.4	342
Target sequence 3: AAGTTCAGCGTGTCCGGCGAG	81	61.9	481
Target sequence 4: AAGCTGACCCTGAAGTTCATC	126	47.6	418
Target sequence 5: AAGTTCATCTGCACCACCGGC	138	57.1	721
Target sequence 6: AAGCTGCCCCGTGCCCTGGCCC	159	76.2	700
Target sequence 7: AAGCAGCACGACTTCTTCAAG	240	47.6	263
Target sequence 8: AAGTCCGCCATGCCCGAAGGC	258	66.7	602
Target sequence 9: AAGGCTACGTCCAGGAGCGCA	274	61.9	38
Target sequence 10: AAGGACGACGGCAACTACAAG	306	52.4	239
Target sequence 11: AACTACAAGACCCGCGCCGAG	318	61.9	499
Target sequence 12: AAGACCCGCGCCGAGGTGAAG	324	66.7	269
Target sequence 13: AAGTTCGAGGGCGACACCCTG	342	61.9	510
Target sequence 14: AACCGCATCGAGCTGAAGGGC	366	61.9	587
Target sequence 15: AAGGGCATCGACTTCAAGGAG	381	52.4	275
Target sequence 16: AAGGAGGACGGCAACATCCTG	396	57.1	389
Target sequence 17: AACATCCTGGGGCACAAAGCTG	408	57.1	569
Target sequence 18: AAGCTGGAGTACAACACTACAAC	423	42.9	398
Target sequence 19: AACTACAACAGCCACAACGTC	435	47.6	628
Target sequence 20: AACAGCCACAACGTCTATATC	441	42.9	331
Target sequence 21: AACGTCTATATCATGGCCGAC	450	47.6	391
Target sequence 22: AAGCAGAAGAACGGCATCAAG	471	47.6	264
Target sequence 23: AAGAACGGCATCAAGGTGAAC	477	47.6	541
Target sequence 24: AACGGCATCAAGGTGAACTTC	480	47.6	206
Target sequence 25: AAGGTGAACTTCAAGATCCGC	489	47.6	658
Target sequence 26: AACTTCAAGATCCGCCACAAC	495	47.6	503

Target sequence 27: AAGATCCGCCACAACATCGAG	501	52.4	470
Target sequence 28: AACATCGAGGACGGCAGCGTG	513	61.9	488
Target sequence 29: AACACCCCATCGGCGACGGC	558	71.4	630
Target sequence 30: AACCACTACCTGAGCACCCAG	597	57.1	416
Target sequence 31: AAAGACCCCAACGAGAAGCGC	630	57.1	757
Target sequence 32: AACGAGAAGCGCGATCACATG	639	52.4	237
Target sequence 33: AAGCGCGATCACATGGTCCTG	645	57.1	270
Target sequence 34: AAGAAGCTTAGCCATGGCTTC	717	47.6	543
Target sequence 35: AAGCTTAGCCATGGCTTCCCG	720	57.1	722

Dharmacon PKR selection output (sorted by Dharmacon "Score"):

Sense Strand Sequence	Start Pos	GC %	Score	Algorithm Rank
CAAATTAGCTGTTGAGATA	234	32	104	481
GATCAAAGAAGGAAGCAAA	203	37	96	506
GAAGAAAAGCCCAGAGAAA	1647	42	96	418
AGAAGAAACCTCAGTGAAA	531	37	95	556
CCACATGATAGGAGGTTTA	136	42	94	30
TATAATAGATGGAAGAGAA	165	26	92	1165
GATCTTAAGCCAAGTAATA	1270	32	92	28
CAAGCCAGCTGGAGAAATA	2346	47	92	162
GAAATTACTCTCAAAGAAA	1578	26	91	562
CTACTAAACAGGAAGCAAA	473	37	90	723
GTTCAAGGTCAAAGACTAA	1091	37	90	313
AGAAATGGCTGGTGATCTT	27	42	89	1159
ATGAATGGTCTCAGAAATA	715	32	89	47
CAAAGTATACTGTGGACAA	794	37	88	920
CAAATGGAATTCTGTGATA	1123	32	88	480
GCGAGAACTAGACAAAGT	1178	42	88	566
GAAATATGGTACTCATTAA	2359	26	88	
GGAAAGCGAACAAGGAGTA	1357	47	87	92
GAAAGGAAGTGGACCTCTA	1427	47	87	419
AGAAATTACTCTCAAAGAA	1577	26	87	528
AGTTATAATAGATGGAAGA	162	26	86	401
CAGATACATCAGAGATAAA	659	32	86	139
CATCAGAGATAAATTCTAA	665	26	86	551
CGGAAAGACTTACGTTATT	897	37	86	126
ACGAGAAGGCGGAGCGTGA	935	63	86	308
AAGTAAAAGCATTGGCAAA	953	32	86	774
AGTTCAAGGTCAAAGACTA	1090	37	86	393
TGGAGGAACTTAATACATA	62	32	85	125
CATGAAAGAAACAAAGTAT	783	26	85	1233

AGGCGAGAACTAGACAAA	1176	42	85	83
GTAAGGGAACCTTTGCGATA	1373	42	85	270
TCTCAGATATATTTGATAA	1538	21	85	404
CGACCTAACACATCTGAAA	1606	42	85	48
TGTTAGAGCCCTTCTGAAA	1681	42	85	758
GAGAATTTCCAGAAGGTGA	179	42	84	213
GCAAAATGGGACAGAAAGA	434	42	84	93
CAGGATACGGGAAGAAGAA	12	47	84	296
TTTTAAAGAAATAGAATTA	828	11	84	407
GAATTGACGGAAAGACTTA	890	37	84	168
GGATACGGGAAGAAGAAAT	14	42	84	906
GCTAATTCTTGCTGAACTT	1455	37	84	824
ATTCATATCTATTTATTAA	2269	11	84	97
AATAGATGGAAGAGAATTT	168	26	83	1122
CTGTTGAGATACTTAATAA	242	26	83	150
AAGACTAACTGTAAATTAT	366	21	83	966
GGACAGAAAGAATATAGTA	442	32	83	36
GCAGATACATCAGAGATAA	658	37	83	77
CTCCTGACATGAAAGAAA	775	37	83	524
AGGCGGAGCGTGAAGTAAA	941	53	83	52
GCAAGACTATGGAAAGGAA	1416	42	83	563

Ambion PKR Selection Output (sorted by ascending target position):

Target	Position	% GC	Algorithm Rank
Target sequence 1: AAGCTTCCAACCAGGATACGG	1	52.4	1581
Target sequence 2: AACCAGGATACGGGAAGAAGA	9	47.6	164
Target sequence 3: AAGAAGAAATGGCTGGTGATC	23	42.9	1731
Target sequence 4: AAGAAATGGCTGGTGATCTTT	26	38.1	970
Target sequence 5: AAATGGCTGGTGATCTTTCAG	29	42.9	1394
Target sequence 6: AACTTAATACATACCGTCAGA	68	33.3	709
Target sequence 7: AATACATACCGTCAGAAGCAG	73	42.9	2295
Target sequence 8: AAGCAGGGAGTAGTACTTAAA	88	38.1	22
Target sequence 9: AAATATCAAGAACTGCCTAAT	106	28.6	1062
Target sequence 10: AAGAACTGCCTAATTCAGGAC	113	42.9	1878
Target sequence 11: AACTGCCTAATTCAGGACCTC	116	47.6	1595
Target sequence 12: AATTCAGGACCTCCACATGAT	124	42.9	1416
Target sequence 13: AAGTTATAATAGATGGAAGAG	161	28.6	1673
Target sequence 14: AATAGATGGAAGAGAATTTCC	168	33.3	2265
Target sequence 15: AAGAGAATTTCCAGAAGGTGA	177	38.1	214
Target sequence 16: AATTTCCAGAAGGTGAAGGTA	182	38.1	406
Target sequence 17: AAGGTGAAGGTAGATCAAAGA	191	38.1	61
Target sequence 18: AAGGTAGATCAAAGAAGGAAG	197	38.1	1778

Target sequence 19: AAAGAAGGAAGCAAAAAATGC	207	33.3	1962
Target sequence 20: AAGGAAGCAAAAAATGCCGCA	211	42.9	437
Target sequence 21: AAGCAAAAAATGCCGCAGCCA	215	47.6	607
Target sequence 22: AAAAAATGCCGCAGCCAAATT	219	38.1	1150
Target sequence 23: AAAATGCCGCAGCCAAATTAG	221	42.9	1887
Target sequence 24: AATGCCGCAGCCAAATTAGCT	223	47.6	1354
Target sequence 25: AAATTAGCTGTTGAGATACTT	235	28.6	1172
Target sequence 26: AATAAGGAAAAGAAGGCAGTT	256	33.3	1713
Target sequence 27: AAGGAAAAGAAGGCAGTTAGT	259	38.1	884
Target sequence 28: AAAAGAAGGCAGTTAGTCCTT	263	38.1	1249
Target sequence 29: AAGAAGGCAGTTAGTCCTTTA	265	38.1	55
Target sequence 30: AAGGCAGTTAGTCCTTTATTA	268	33.3	2
Target sequence 31: AACAACGAATTCTTCAGAAGG	294	38.1	1517
Target sequence 32: AACGAATTCTTCAGAAGGATT	297	33.3	684
Target sequence 33: AATTCTTCAGAAGGATTATCC	301	33.3	2217
Target sequence 34: AAGGATTATCCATGGGGAATT	311	38.1	484
Target sequence 35: AATTACATAGGCCTTATCAAT	328	28.6	1553
Target sequence 36: AATAGAATTGCCCAGAAGAAA	346	33.3	693
Target sequence 37: AATTGCCCAGAAGAAAAGACT	351	38.1	1114
Target sequence 38: AAGAAAAGACTAACTGTAAAT	361	23.8	1036
Target sequence 39: AAAAGACTAACTGTAAATTAT	364	19	961
Target sequence 40: AAGACTAACTGTAAATTATGA	366	23.8	105

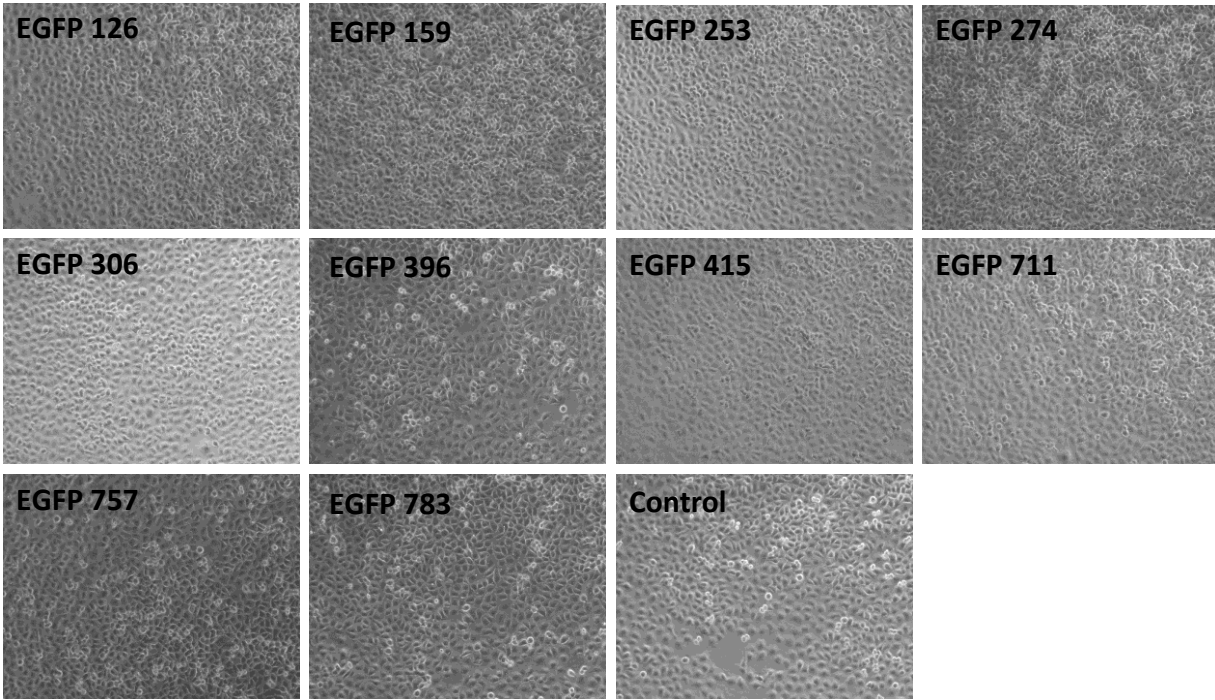
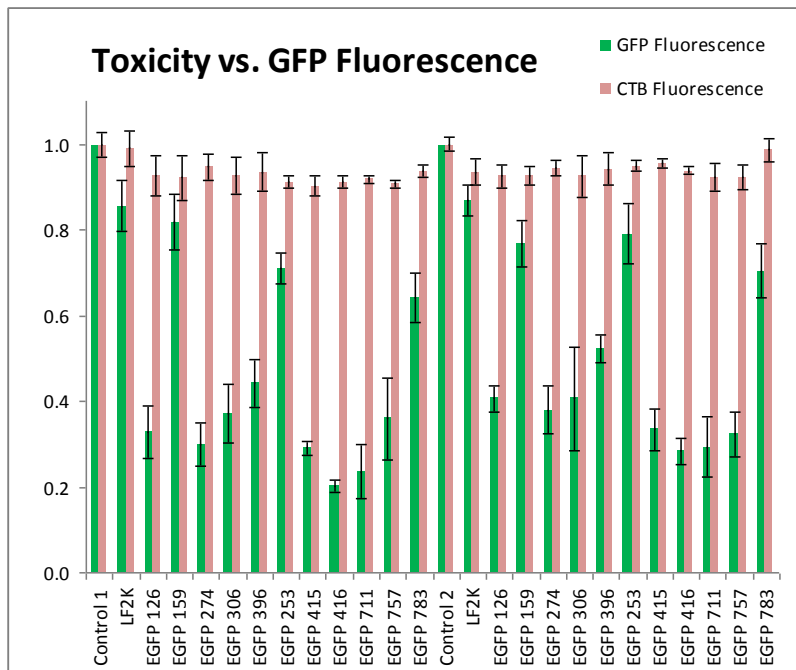


Figure S1: Images and Cytotoxicity Data for siRNA Transfection Experiments.

Treated cells showed no toxicity compared to control cells by microscopic inspection either for GFP experiments (above) or PKR (no images taken). Similar results were obtained when cell toxicity was assayed by CellTiter-Blue (right).



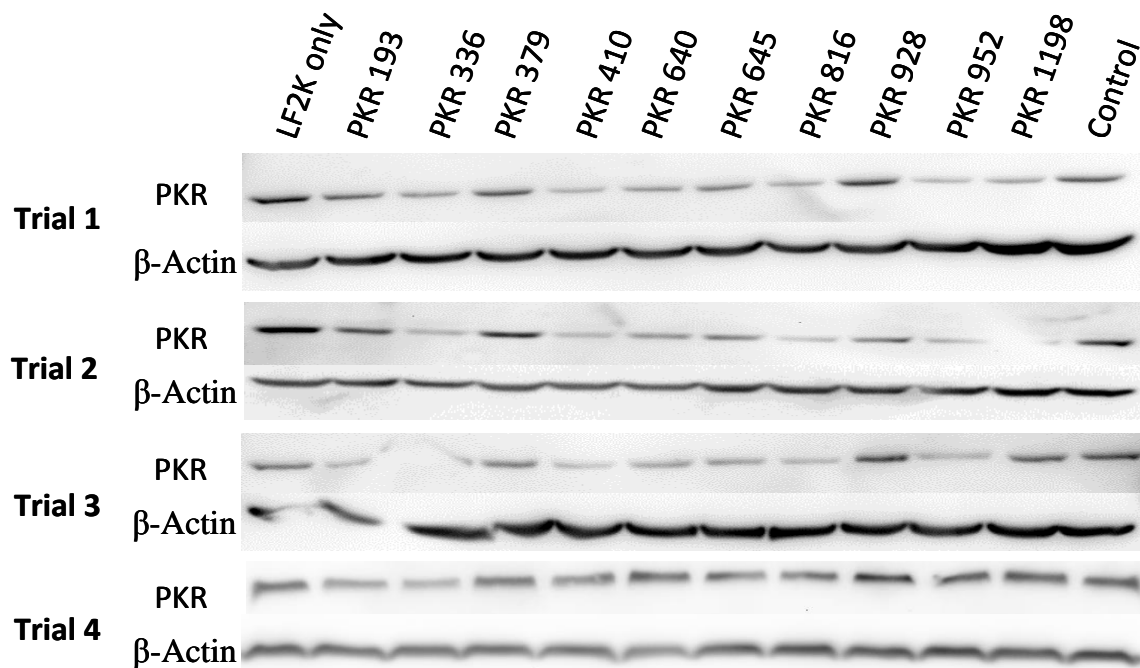


Figure S2: Western Blots for PKR Silencing. The results of the western blotting raw data (above) were quantified and correspond to Figure 4 in the manuscript.