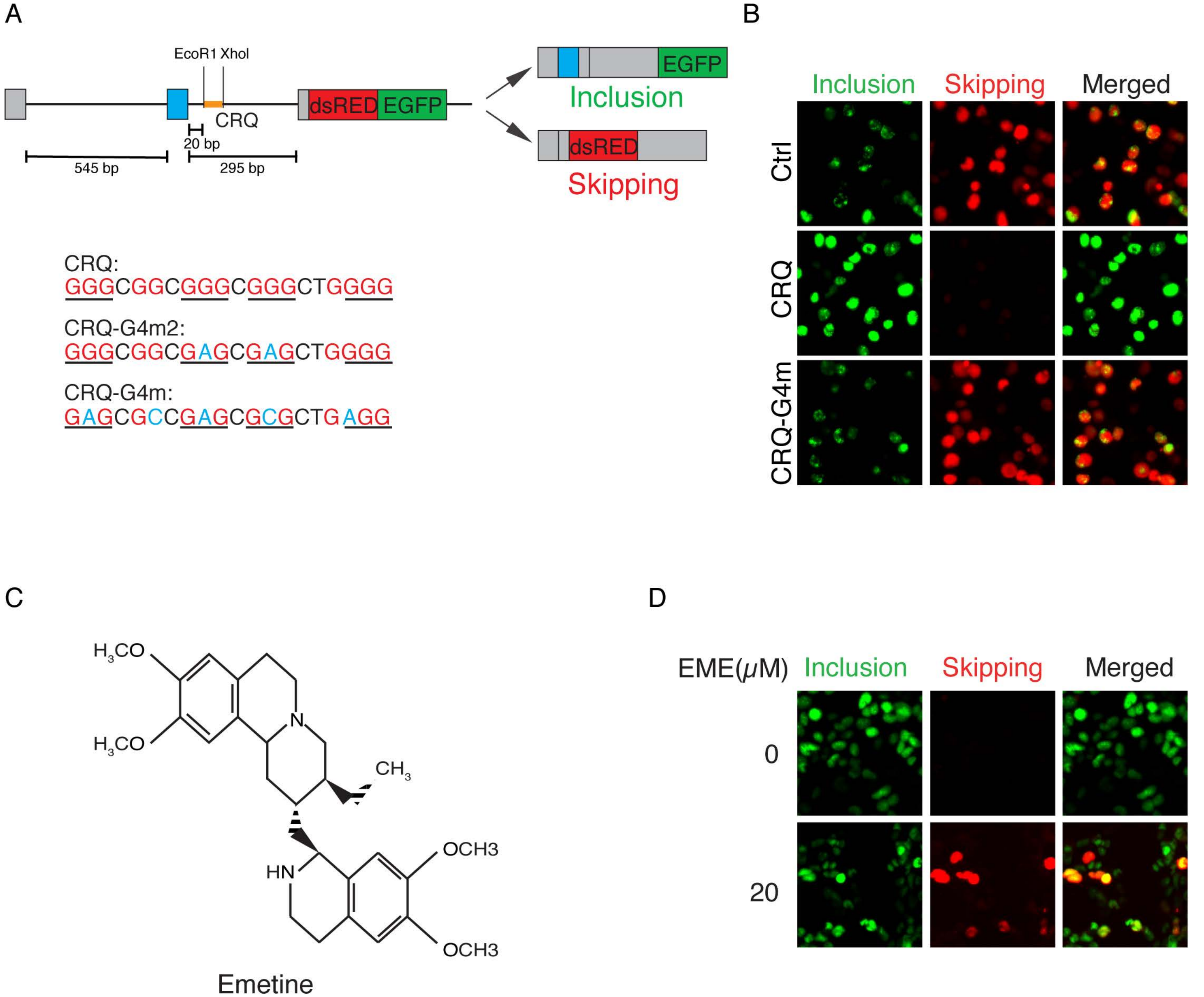
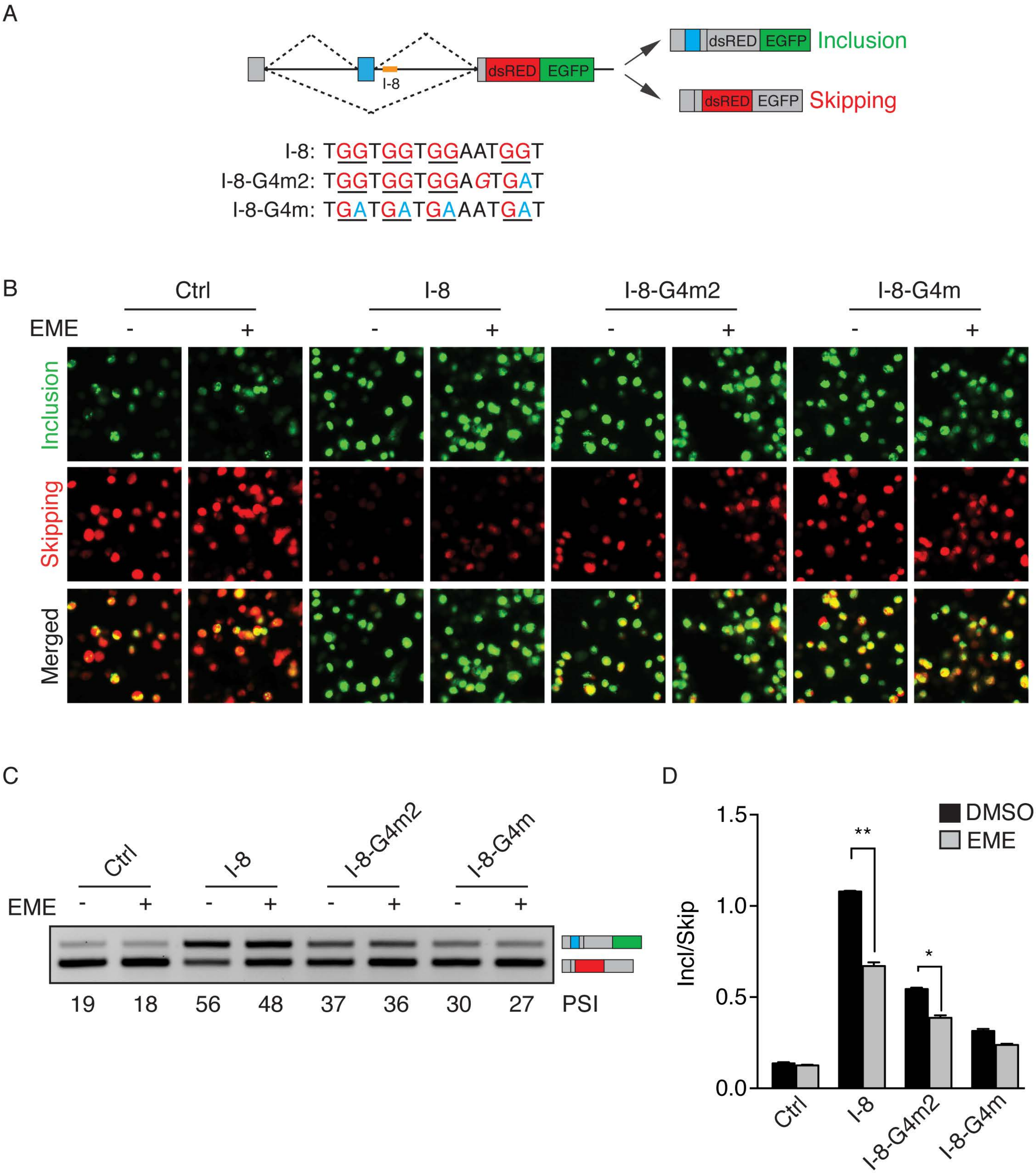
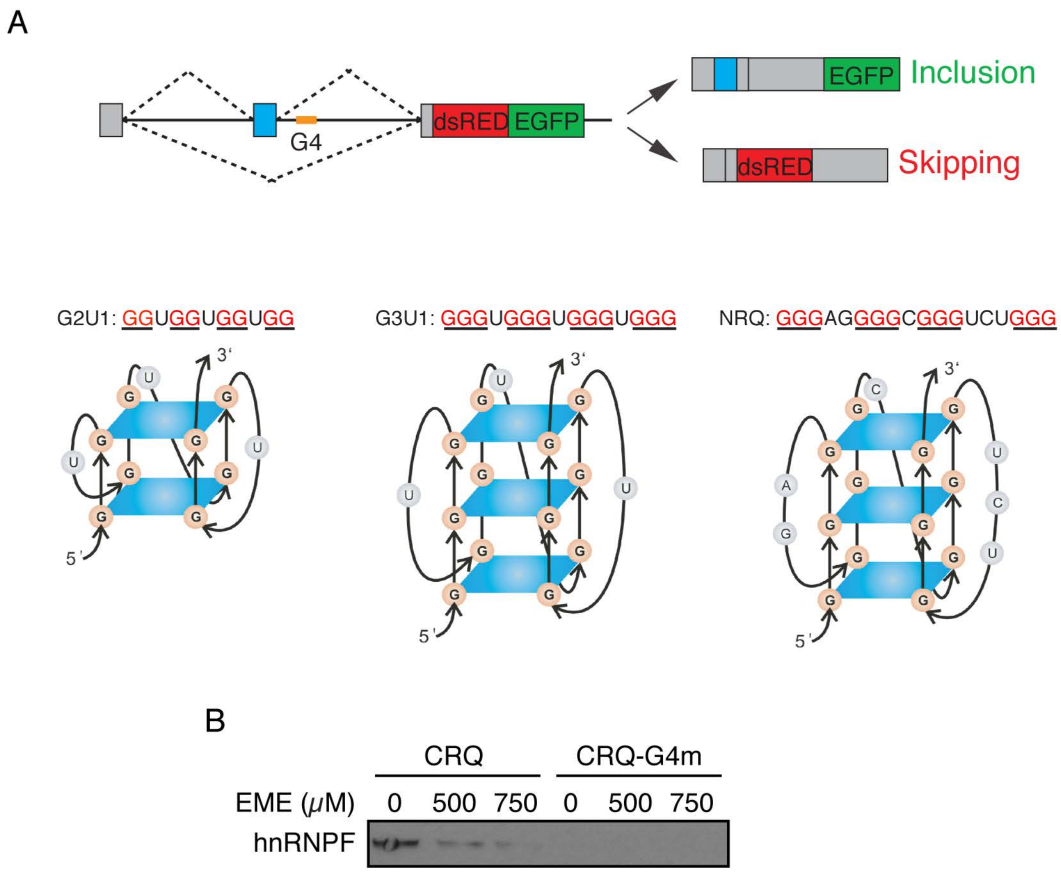
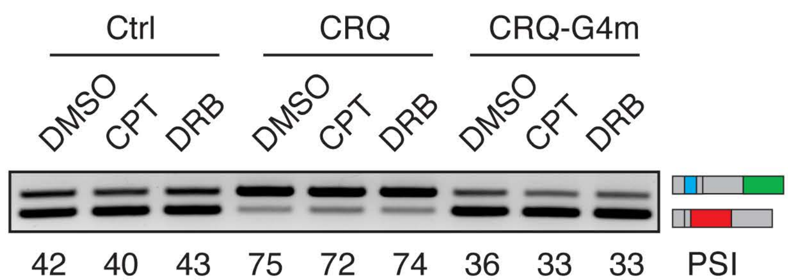
Zhang et. al._Supplementary Figure S1



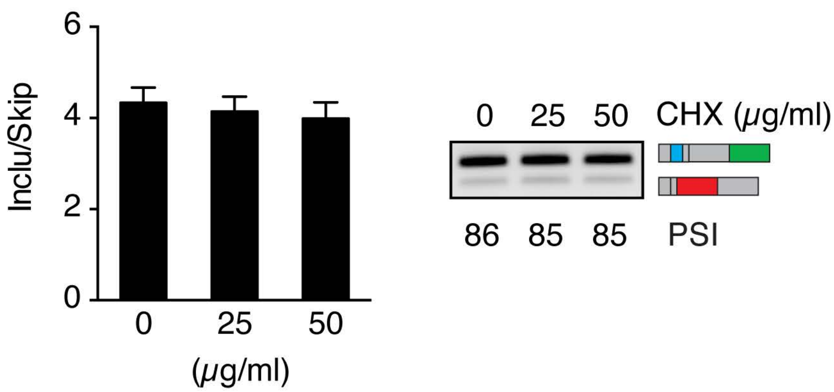

A
dsRED
EGFP
I-8
dsRED EGFP Inclusion
dsRED EGFP Skipping
I-8: TGGTGGTGGAATGGT
I-8-G4m2: TGGTGGTGGAGTGAT
I-8-G4m: TGATGATGAAATGAT
B
Ctrl
I-8
I-8-G4m2
I-8-G4m
EME
- +
Inclusion
Skipping
Merged
C
Ctrl
I-8
I-8-G4m2
I-8-G4m
EME
- + - + - + - +
19 18 56 48 37 36 30 27 PSI
D
DMSO
EME
Incl/Skip
1.5
1.0
0.5
0.0
**
*
Ctrl
I-8
I-8-G4m2
I-8-G4m

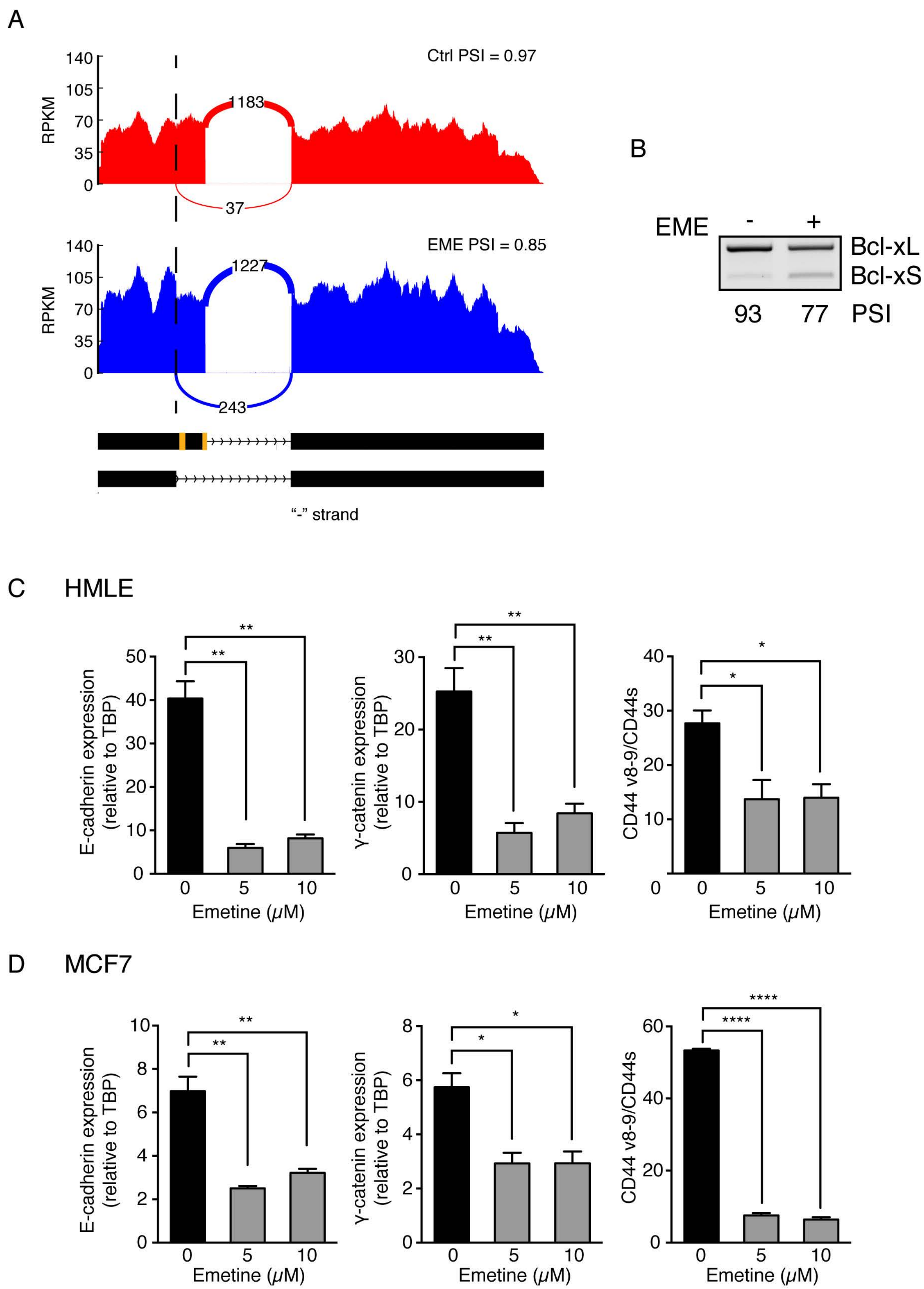

A
G4
dsRED
EGFP
EGFP
Inclusion
dsRED
Skipping
G2U1: GGUGGUGGUGG
G3U1: GGGUGGGUGGGUGGG
NRQ: GGGAGGGGCGGGUCUGGG
3'
5'
B
CRQ
CRQ-G4m
EME (μM)
0
500
750
0
500
750
hnRNPF

A

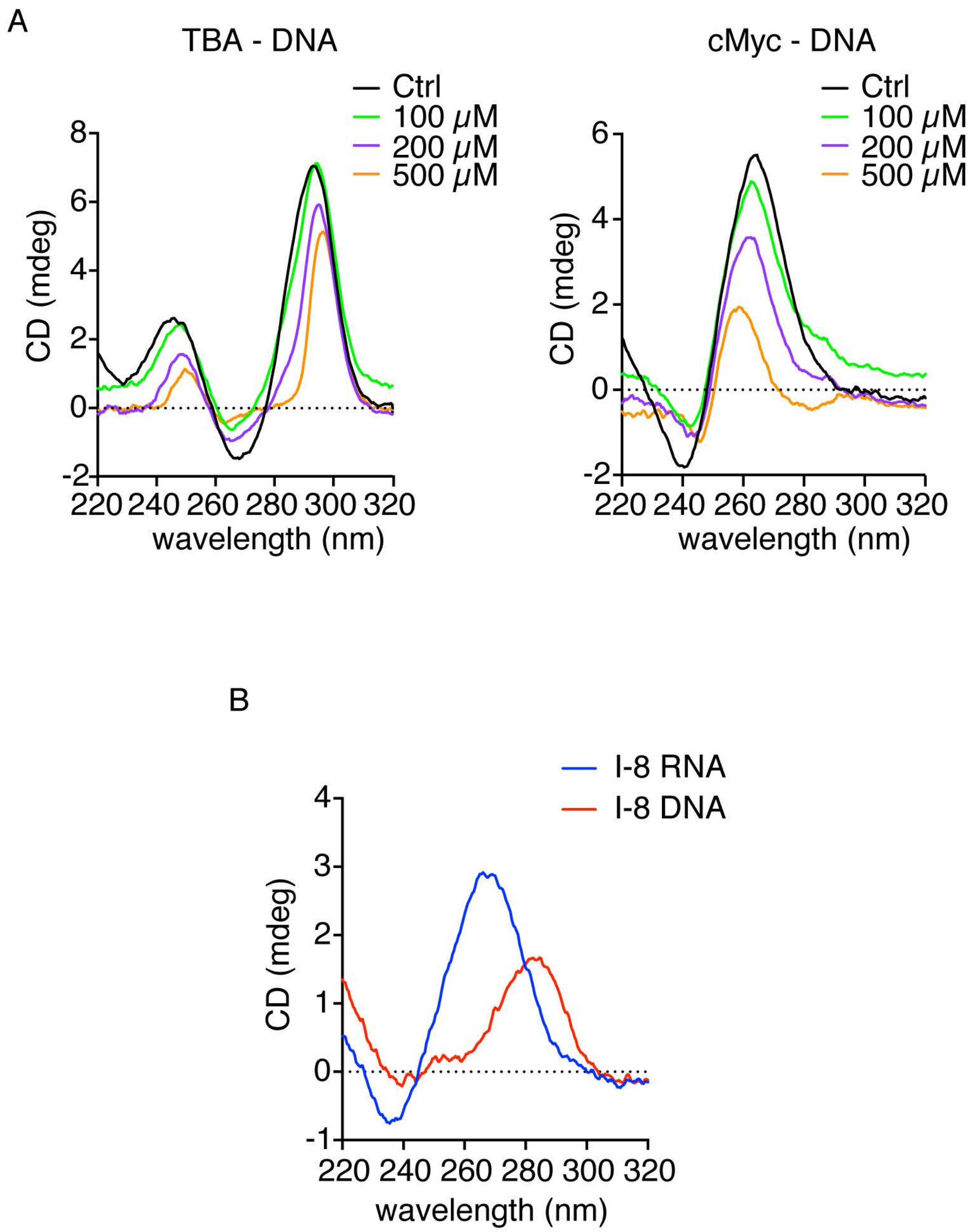

Ctrl
CRQ
CRQ-G4m
DMSO
CPT
DRB
DMSO
CPT
DRB
DMSO
CPT
DRB
42
40
43
75
72
74
36
33
33
PSI

B


Inclu/Skip
6
4
2
0
0
25
50
(μg/ml)
0
25
50
CHX (μg/ml)
86
85
85
PSI


A
Ctrl PSI = 0.97
RPKM
140
105
70
35
0
1183
37
EME PSI = 0.85
RPKM
1227
243
"-" strand
B
EME - +
Bcl-xL
Bcl-xS
93 77 PSI
C HMLE
E-cadherin expression (relative to TBP)
γ-catenin expression (relative to TBP)
CD44 v8-9/CD44s
Emetine (μM)
0 5 10
**
*
D MCF7
E-cadherin expression (relative to TBP)
γ-catenin expression (relative to TBP)
CD44 v8-9/CD44s
Emetine (μM)
****

Zhang et. al._Supplementary Figure S6



SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Diagram of the minigene splicing reporters and the structure of emetine. (A) Diagram of the minigene splicing reporters that contain the CRQ and its mutated sequences in the intron downstream of the variable exon. The size of introns flanking the variable exon and the cloning sites to insert the CRQ sequences are shown. (B) Fluorescent images of HEK 293FT cells transfected with CRQ and CRQ-G4m constructs. (C) Chemical structure of emetine. (D) Fluorescence images of HEK 293A CRQ cells after 48 hours of emetine (20 µM) treatment.

Supplementary Figure S2. The effect of emetine on the alternative splicing of I-8 and its mutant minigene constructs. (A) Diagram of the minigene splicing reporters where I-8 and its mutated sequences are inserted in the intron downstream of the variable exon. (B) Fluorescent images of HEK 293FT cells transfected with I-8 or its mutant reporters with or without the presence of emetine. Emetine promotes exon skipping to produce dsRed only with the I-8 minigene reporter but not with the control or mutated minigenes. (C, D) Semi-quantitative PCR (C) and q-PCR (D) of cells transfected with I-8 minigene showing significant increase in exon skipping in the presence of emetine. Error bars represent S.E.M. *, $p < 0.05$, **, $p < 0.01$; Student's *t*-test.

Supplementary Figure S3. The formation of G-quadruplexes. (A) Diagram of the minigene splicing reporters where different G-quadruplex sequences are inserted (Top). Sequences and proposed G-quadruplex secondary structures of G2U1, G3U1, NRQ are shown in the bottom panel. (B) RNA pulldown followed by western blot analysis showing that the addition of emetine disrupts the interaction between hnRNPF and CRQ.

Supplementary Figure S4. Inhibition of transcriptional elongation or translation exhibits no significant changes on G-quadruplex-dependent alternative splicing. (A) Semi-quantitative RT-PCR of HEK 293FT cells showing that transcriptional inhibitors CPT and DRB have no significant effects on the alternative splicing of CRQ and its mutant minigenes. (B) qRT-PCR (Left panel) and semi-quantitative RT-PCR (Right panel) of HEK 293A CRQ cells showing no detectable effects on CRQ alternative splicing upon the treatment of the translation inhibitor CHX.

Supplementary Figure S5. Emetine regulates alternative splicing of Bcl-x and promotes an EMT cell state. (A) Genome browser tracts of RNA sequencing data showing emetine treatment promotes the usage of the Bcl-x exon 2 proximal 5' splice site, resulting in production of the short isoform Bcl-xS. The location of previously reported two G-quadruplexes are shown in yellow boxes. (B) Semi-quantitative RT-PCR validation of emetine's effect on endogenous Bcl-x exon 2 alternative 5'-splice site usage are shown. (C) qRT-PCR of HMLE cells showing that emetine treatment resulted in a decrease in expression of epithelial markers E-cadherin and γ-catenin, and a decrease in the ratio between CD44v and CD44s. (D) qPR-PCR of MCF7 cells showing consistent results in response to emetine. Error bars represent S.E.M. *, $p < 0.05$, **, $p < 0.01$, **** $p < 0.0001$; Student's *t*-test.

Supplementary Figure S6. The effect of emetine on DNA and RNA oligonucleotides. (A) CD spectra analysis of DNA oligonucleotides TBA (Left panel) and cMyc (Right panel) in the presence of different concentrations of emetine. (B) CD spectra analysis of I-8 RNA and I-8 DNA oligonucleotides in the presence of 100 mM K^+. The I-8 DNA forms a positive peak at 284.5 nm, but not at 290 nm - 295 nm, indicating that I-8 DNA does not form antiparallel DNA G-quadruplex structure.

Supplementary Table S1 primers for plasmid construction

primer	Sequence	comment
RG6EX_Ran 15nt for	5'-AATTCGACTACCAGAGAAGAC-3'	Control for 15 nt inserts
RG6EX_Ran 15nt rev	5'-TCGAGTCTTCTCTGGTAGTCG-3'	
RG6EX_Ran 21nt for	5'-AATTCGACTTCATCGTTCCGTCAAGGC-3'	Control for 21 nt inserts
RG6EX_Ran 21nt rev	5'-TCGAGCCTTGACGGAACGATGAAGTCG-3'	
RG6EX_I-8 15 nt_F	5'-AATTCTGGTGGTGGAATGGTC-3'	I-8 15 nt and mutations
RG6EX_I-8 15 nt_R	5'-TCGAGACCATTCCACCACCAG-3'	
RG6EX_I-8 G2m (15nt)_F	5'-AATTCTGGTGGTGGAGTGATC-3'	
RG6EX_I-8 G2m (15nt)_R	5'-TCGAGATCACTCCACCACCAG-3'	
RG6EX_I-8 15nt G4m_F	5'-AATTCTGATGATGAAATGATC-3'	
RG6EX_I-8 15nt G4m_R	5'-TCGAGATCATTTCATCATCAG-3'	
G2U1-F	5'-AATTCGGTGGTGGTGGC-3'	G2U1
G2U1-R	5'-TCGAGCCACCACCACCG-3'	
G3U1-F	5'-AATTCGGGTGGGTGGGTGGGC-3'	G3U1
G3U1-R	5'-TCGAGCCCACCCACCCACCCG-3'	
NRQ-F	5'-AATTCGGGAGGGGCGGGTCTGGGC-3'	NRQ
NRQ-R	5'-TCGAGCCCAGACCCGCCCCTCCCG-3'	
CRQ-F	5'-AATTCGGGCGGCGGGCGGGCTGGGGC-3'	CRQ and mutations
CRQ-R	5'-TCGAGCCCCAGCCCGCCCGCCGCCCG-3'	
CRQ G4m2-EcoR1-F1	5'-AATTCGGGCGGCGAGCGAGCTGGGGC-3'	
CRQ G4m2-Xho1-R1	5'-TCGAGCCCCAGCTCGCTCGCCGCCCG-3'	
CRQ-G4m-EcoR1-F1	5'-AATTCGAGCGCCGAGCGCGCTGAGGC-3'	
CRQ-G4m-Xho1-R1	5'-TCGAGCCTCAGCGCGCTCGGCGCTCG-3'	
Semi Primer F	5'-CAAAGTGGAGGACCCAGTACC-3'	Semi-RT PCR
Semi Primer R	5'-GCGCATGAACTCCTTGATGAC-3'	
qRT-PCR Primer F inclusion	5'-GATTACAAGGATGACGATGACAAG-3'	qRT-PCR
qRT-PCR Primer R inclusion	5'-CCTGATCCTCCTGACCTCAAT-3'	
qRT-PCR Primer R skipping	5'-CCTGATCCTCCTGACCTCTAG-3'	

Supplementary Table S2: A high-throughput screen identifying 38 hits that increased percentage of red-colored cells

Compound Name	Total Number of Cells	Percentage of red cells (%)	Formula	MolWt
HOMIDIUM BROMIDE	635	93.858269	C21H20BrN3	394.32
DAUNORUBICIN	867	89.042679	C27H29NO10	527.53
RUTILANTINONE	985	84.060913	C22H20O9	428.40
EPIRUBICIN HYDROCHLORIDE	716	80.586594	C27H30ClNO11	579.99
SANGUINARINE SULFATE	678	75.663719	C20H15NO8S	429.41
DALBERGIONE	702	69.943024	C15H12O2	224.26
DALBERGIONE, 4-METHOXY-4'-HYDROXY-	943	53.340405	C16H14O4	270.29
1,4-NAPHTHOQUINONE	1276	42.554859	C10H6O2	158.16
ANDROGRAPHOLIDE	682	32.111439	C20 H30 O5	350.452
FLUBENDAZOLE	607	28.336079	C17 H13 F N2 O3	312.299
PIROCTONE OLAMINE	756	26.455027	C14 H23 N O2 . C2 H7 N O	298.424
DORAMECTIN	816	22.058823	C50 H74 O14	899.121
CHICAGO SKY BLUE	4216	19.568312	C34 H24 N6 O16 S4 . 4 Na	992.814
DOCETAXEL	735	17.959183	C43 H53 N O14 . 3 H2 O	861.929
MECHLORETHAMINE	643	16.951788	C5H11Cl2N	156.06
PIPLARTINE	614	16.775244	C17H19NO5	317.34
OBTUSAQUINONE	680	16.029411	C16H14O3	254.29
EMETINE	810	15.951807	C29H42Cl2N2O4	553.58
STROPHANTHIDINIC ACID LACTONE ACETATE	621	12.077294	C25H32O7	444.53
EVANS BLUE	3889	11.339676	C34H24N6Na4O14S4	960.82
BISANHYDRORUTILANTIN ONE	4450	10.516854	C22H16O7	392.37
6-METHOXYHARMALAN	718	10.16713	C13 H14 N2 O	214.267
CADMIUM ACETATE	664	9.939759	C4 H6 Cd O4	230.497
NIFUROXAZIDE	1127	9.760426	C12 H9 N3 O5	275.219
MYCOPHENOLIC ACID	1174	9.625213	C17H20O6	320.35
DIGOXIN	915	8.74317	C41H64O14	780.96
COLCHICEINE	1276	8.699059	C21 H23 N O6	385.414
TRIHYDROXYBENZALDEH YDE	1098	8.652095	C7 H6 O4	154.12
FLUVASTATIN	1903	8.407777	C24 H26 F N O4	411.47
MEPARTRICIN	4197	8.386943	C60H88N2O19	1141.37
PITAVASTATIN CALCIUM	1926	8.359294	C25 H23 F N O4 . Ca	460.536
DIGOXIGENIN	656	8.231708	C23H34O5	390.52
PARTHENOLIDE	638	8.15047	C15H20O3	248.32
3-DESMETHYL-5-DESHYDROXYSCLEROIN	1378	7.764877	C14 H12 O4	244.245
POMIFERIN	1298	7.318952	C25H24O6	420.47
PATULIN	676	7.248521	C7H6O4	154.12
GAMBOGIC ACID	811	7.151665	C38H44O8	628.77
ACETYL ISOGAMBOGIC ACID	738	7.046071	C40H46O9	670.81

Supplementary Table S3: Dose reponse assay

Compound Name	Dose of Compound (μM)	Percentage of red cells (%)
EMETINE	20	42.04355
	10	19.04033
	5	9.421266
	2.5	7.843137
	1	0.823924
	0.5	0.337675
MECHLORETHAMINE	20	31.929825
	10	13.743218
	5	7.5
	2.5	3.464419
	1	0.725875
	0.5	0.925069
1,4-NAPHTHOQUINONE	20	85.863266
	10	69.632492
	5	0.388749
	2.5	0.376857
	1	0.680426
	0.5	0.45106
OBTUSAQUINONE	20	40.463917
	10	37.727272
	5	8.5
	2.5	2.734618
	1	0.521119
	0.5	0.577399
PARTHENOLIDE	20	27.855711
	10	17.827297
	5	0.393701
	2.5	0.588865
	1	0.371103
	0.5	0.425758
SANGUINARINE SULFATE	20	72.256096
	10	79.487183
	5	68.253967
	2.5	44.699139
	1	1.194624
	0.5	0.409165

The full sequence of the CRQ minigene:

Capital letters denote exon sequences. The variable exon is colored in cyne. The CRQ sequence is marked in orange flanked by restriction enzyme sites colored in yellow. The sequences encoding the dsRED and EGFP proteins are marked in red and green, respectively. The splice sites are bolded.

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg
ctccctgcttgtgtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaat
tgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga
ctagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatgg
cccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggga
ctttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgc
cccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagt
acatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcac
ggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgt
aacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaac
TAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGA
CCCAAGCTGGCTAGCGTTTAAACTTAAGCTTCCATGGATTACAAGGATGACGATGA
CAAGGGGGTACCTGCCCCAAAAAAAAAACGCAAAGTGGAGGACCCAGTACCCGGA
TCTAGAG**gt**aggtgatcctcctgctgctttggttcagggttttgcttgagggggggggtggtgatttccttgccatgggca
gactgagcagaaaaggccattgggaccatgttctgaatgcctccacctcaaccaccggccggtaggaccaaagccacc
ccgtgttttctcaggatctcttttcccagggagatccctcggcccaaagagggagatggcaatgctggatgtgtgcacaata
attcaacaggcattggaacttcagcatcgatgctgaatgcaattaacaatgctcaagcagaacccccggctccatcagca
cagtgcaggaccaaaccccatgctgcagcagtggggctgtctgtacggggtgggcaatgggaaccggggtctgctggg
gctcctgctgcttcagtgctgccatgcagccacacatcctgagagctgaaagggtcggcgtcctcacctggtgcacaccgt
agcggccgccccacagctttaaggcacctggctaacctctgcgcttcttcccttccctcctccctggctc**ag**GATCCAG
GCGATATCCGGAAGAATTGAG**gt**agttactgcacctttcttgaattcgggcggcgggcgggctggggctcga
ggggggcacagctggggagggaacaagggacaaaaccaggagggggctccgagtccttggatttattccccctcatcc
atgccttaccttcaggtaagggcctgaacagagccctttacttcctgcttctttctcccatagctccctctccttcgggtctcctgg
actcagtgccacggttgtcccattctgggggtctgtagggagccagcaggagctgcggccgtcctactgaccctgtccttatt
gcac**ag**GTCAGGAGGATCAGGAGGACGAGGAGGAAGAGGAGACCGGTGTGAGCAA
GGGCGAGGAGGACAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCAC
ATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCC
CCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCC
TACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGG
GCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGA
CCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCG
GCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGG
AGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCA
AGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCAGAGGTCAAGACCA
CCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAA
GTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGC
GCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAACCGCG
GTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC

TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTA
CAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG
AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAgggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccct
cccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctga
gtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatg
ctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgt
agcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctc
ctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttt
tcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctatt
cttttgatttataagggattttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattct
gtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaat
tagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcag
caaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgac
taattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggccta
ggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgatt
gaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaa
tcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgcc
ctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacg
ttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgc
cgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaag
cgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatca
ggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatgg
cgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggacc
gctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttac
ggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaa
tgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcgga
atcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttatt
gcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggttt
gtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttc
ctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaat
gagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatga
atcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcg
ttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaa
agaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctcc
gcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca

ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcg
ggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcac
gaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcg
ccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggc
ctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtag
ctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaagg
atctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgag
attatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggt
ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtc
gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccgg
ctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatc
cagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggc
atcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgt
tgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatgg
cagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga
atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagt
gctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactc
gtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaa
aaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattg
tctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgcca
cctgacgtc