Supporting Information S5. The dataset Test96 consists of a positive dataset and a negative dataset. The positive dataset contains 13 true m5C sites containing sequences with the m5C sites in the center, while the negative dataset contains 83 negative m5C sites containing sequences. Each of these segments is 41-bp long.

I. 13 true m5C sites containing sequences

>P174
ACUUAAGAGUAUCUGUAUCGCGGCUCGUUGGUCUAGGGGUA
>P498
AAGGCGAUGGACUAGAAAUCCAUUGGGGUUUCCCCGCGCAG
>P779
UACUCAGGUAUCGUAAGAUGCAAGAGUUCGAAUCUCUUAGC
>P781
CUGUUGGACGGUUGUCCGCGCGAGUUCGAACCUCGCAUCCU
>P785
CACACUAGUAAUGUGGAGAUCAUCGGUUCAAAUCCGAUUGG
>P788
GCGCUUGUCGCGUGCCAGAUCGGGGUUCAAUUCCCCGUCGC
>P789
CAGAUUGCAAAUCUGUUGGUCCUUAGUUCGAUCCUGAGUGC
>P796
AGGGCUCAUAACCCUGAUGUCCUCGGAUCGAAACCGAGCGG
>P802
AAGCAAGGAUACCGAAAUGUCAGGGGUUCGAGCCCCCUAUG
>P813
UUUAAAAGCAACGCGACCGUCGUGGGUUCAAACCCCACCUC
>P814
UCACUACGAAAUCUUGAGAUCGGGCGUUCGACUCGCCCCCG
>P819
GGUCAAGAUAUCUGGAGGUCCUGUGUUCGAUCCACAGAGUU
>P836
AUCGUUGUGGCCGAUGAAACCCUGGUUCGAUUCUAGGAGAU

II. 83 false m5C sites containing sequences

>N10
CGACAGCAGUGUCUCCUUUACCCUCCUGAAGGAGUGUGAGG
>N15
UGAGGGACGGCCCUUGUCUGCGGUCAUCCCUCCCCUCCUGA
>N19
CGGGGAAGCGUUCGCAGAAACUACAGGGUGGUUAUUUAGUU
>N53
GGGGCCCCUUCCUCUGCCGCCCUGGACCUGGGCUGUGGCAC

>N60
GCUGUGGCAGGGAGUAUUUUCUCUAAAAACAGGAUGGAUGU
>N79
GGAACCACGUUGGGGGCCACCACAGGGCAGAUGGGUCCCGG
>N102
UCCCCACCCAACAGGAGGGGCUCCCUGGGAUGACACCGUGG
>N129
AACAGUGGGACCUGCACAUCCGAGAGUCCGAGCUUUACUGU
>N148
AGGUCUGAGCGAGGACAGCUCCCCACCACGGUCAGGUCUGA
>N152
CACCAGGUCUCACUGGAGGACGGGCCGGCUUCUCCCAUGCC
>N153
CUCUGUCCUAAAAAUCCCCUCUGCUCCGCCCACUCCUCCCU
>N156
GUGUCCAGGCUGUGCAGUUGCUGGGGCUAGGUCCGCUCGCU
>N175
AGAAUCCUGACUGUCGAAGCCCCCAAAUAGUUCUGUAGGAU
>N201
AAAAUACAGAGUUAUUCACCCCUGAGCUUCUGUGCACACAA
>N216
GCUUUCGAGUUGUCAUUCUCCCGCGUGGUCUGUGGGGAGAA
>N219
AAAUCCCCGGCUGAAGUUGGCGGGUAGCACUGCCCUGCACU
>N244
UGCCGCAUCCCCAGCCCAAGCUGCCUGGGCCCCCCUGAUGC
>N348
GAGUGACUGGCACAUGGAGGCUGAUGAUUAAAUGUUUGUUG
>N359
GAUUUAACCAUUCCUACCUUCCCUCUUGGGUAUGCAGCAAA
>N361
CAGGCUAGAGUGCAGUGGUGCCAUCUCGGCUUACUGCAACC
>N363
UGAAAAUCAGAGUAGUGAGCCUUUCAAGGGCUAAGGUCCAU
>N391
CCCAUAGUAAAGUUUAUCUCCUCUCCUCUCCUCUCCUUUCC
>N404
UCUCGAGGGAAUGAGUGGGGCAUAAAAAUUCUGUUUCUCAG
>N409
UAACUUUUCCUCUGCCUCCCCACUUAGUCGUAACCCUUGUG
>N421
UCUUCAUUUAAAGGGAAAAUCUGUAAAUGUAUACUCUGUCU
>N453
GGAAUUAAGAAAGGCCUUUUCAGGCUGGGCAUGGUGGCUCA

>N482
AAAGUAGUGGGCUGUGAAUUCAGGGCUACUUGGAGCCCUUA
>N485
CCUCAGAGGAAGGAGUGCUUCUUAGCAGUGAGCAUUGUGGA
>N499
GCCUCCCCAGGGAUUCUCAUCUGGGUUUUCAAUACUCUACU
>N521
CUUAGAGUAGACCCUAUCACCUCCUGCUGUCACCAGUGAGU
>N531
AGCCAGGGUGACCUUGACCUCCUGACCUCAUGAUCCACCCG
>N544
GAGUAUAGCCGCCCCAUGGGCCCAACAGACAGGGCCCAGUG
>N551
GGCUUGCGAUAGGCUUUUGGCUCAUCGAGUGGAAACCAAAA
>N561
UCAUGCCACUGCACUACAGCCUGGGCAGCAAGAGCGAGAUA
>N568
CCUCUUGCUUACCCCCUCACCGUAGGGAGAACCUCCUCCUG
>N575
GCUCCUCCUGGAACCUGUAGCUGGAGGCUGCUGGAGGAUGC
>N580
AGCAAGGACCCCAGAAUGGACUAGUCCAGCUGUCAUUUGGG
>N594
UCGUUUUUUCUUUGCAUUCCCAGAAAAUAUUUACAGAUUUG
>N596
GAUUCUUGUCAGAUACAUUGCAAAUUUUUUUCUCUUUUUGU
>N611
CAACGUUCCUGUCCCACAGACGCCUUCCUGGUGUGUGCUGC
>N640
AGGGGCUACUAGUCGGCCAUCUUUUUGAUGUCAUUUGUGUU
>N643
CCUUCUAAGACCCACUUUGCCCCCCGUGGACAGUGGGUUGU
>N649
GUGCCUCGGCUCAGCCCCUGCCCUGUUUCCCUGGGGUGUGG
>N678
UUCACAUGCUGUCCUAAUUACUUUGUAUUUCAUGUUGUUUU
>N746
CACUGGCCACACUGGGUCUCCUCUGGCCACAAUCCGUCCGC
>N757
GGUAAUCCCUACUCAGCUUCCACACUCACCCUUGCAUUUCA
>N763
UCUCUGGGGACACCCAGGACCCAACGGUGCAAUGUCAGGAG
>N774
UGGAGCUGGAAGUCUGAGAUCCAGGCGGGCAGGGGAUAGAC

>N778
GUGGGUGGAACCAGAGCCUUCUGUGAAGCCCCCCAUGGCUU
>N787
GGGCUGCUGUUCCAAACCGCCACAAGCUGGGGAGCUUAUAC
>N789
UGGUCCCCACUGGCCACACUCAGCCUCUGCCCUCUCCACUC
>N821
AAUGCAGCCAGCUGGAGACUCCAGGCCCCCAGGGAAGAUCU
>N834
GAGCUCCACAACUGCUGCCUCUGAGAGGUCCCUUCAGGGGC
>N845
CGCCCCCAGCCCCGCCCACCCUGAGCCCCGCACUCCACCCA
>N886
CUCAGCACCCUCCGUGAUGCCAUGCUGUUUUCUUUCCAGCC
>N889
CUGCCGCUGCCGCCCGCUGCCCCCGCGCCCUCAGCAUCCAC
>N904
CCGCGUCCAGACUCCACCCUCGGUCCUGGAGGGCCAUGGGA
>N912
AUGGACGAGGCUUCCAUGCACUGACAGCUGGGCUCACGGUG
>N934
UACUUAGGAAUCGAGAAUUUCUAGUUCAGAACUGCAAAAAG
>N936
CUCCGCCUCCCAGGUUCAAGCAGUUCUCUGCCUCAGCCUCC
>N984
UUUGAAAAUGUUCUUUUCUUCCCUUAUGGACAAGAUUAGAA
>N994
UAAAAUAUUAUGACCUGUGACAUCAACAUGGAUUAACAAGU
>N1008
UAACGUUUUAAAAUGAAAAACAGUUUUGCAAGUCUCUCUUU
>N1053
UCUUUGUCAUUUUUUUUUCUCAUUAUUUGGAGACUUUUGAC
>N1075
UUUUUUUUUAAACCCAGUUGCAUUGUAUUUAUGUGAUAAAC
>N1099
UACAGGCACCUGCCACUAUGCUCACCUAAUUUUGGAAUUUU
>N1116
UAAUGUUAAUCAGUGGAGGUCAUAGUUUCAUUUUUUAACAU
>N1124
CACUCCCUCCUCGCCCGCUCCAAGGAAGCACCCGCCGCCUC
>N1127
CCCUGCCGCUUCCGGCCCCGCAACACAGGCAGUUCGCACCG
>N1131
ACUUCAUGGAACUCUCCAGACAUCCUGCUCCCUGCCCCUAA

>N1133
CAGGGAAGUGGCCCUGCGGCCAUGACUCAUCUCAGCGCUCA
>N1169
AUCCCCCCAGUAGGGUCCUGCCCACACUGGCUCCACCCUGA
>N1172
GAGAAGGAACCGGCCCACCCCGGCUGCAUCCCUGGGGCCGG
>N1192
UGAGGUUCCUGCCGAAGCCCCAACGUGUGAUUAGGAAAGCG
>N1194
GGGGUGGGAGUUCUAAUAAUCACUCAUUUGCUGCUUUGCAG
>N1195
GAGCUUAAGAGCAGUACACACUGGAAGGCAGGGAGGGAUUG
>N1232
UGACGGACAGAAAGUCAUCCCCAUGGUGUGGUCACUGUGGG
>N1256
UUCAGAAAGGUUACGUAGCUCAGCCAAAGUCACACAGCUGG
>N1327
CUGACCCCUCAUGAACAAAUCUCUGCCAGCUUCAAACUUUC
>N1368
CUAAACUGAACCAUUUCUAGCUUUUGAUUUAAAGUGACAGA
>N1406
UAUCUUGGCUUUCAACAAGCCUUCCUCACUACACUAAAUCA
>N1413
GAGCGGGGAGAAGAGAGAGGCUGGGAAAGGCUGGUGGGCGG
>N1423
UUCUGCACAACUGGCCGCAGCUUCUCCAUCCGUACUUGCUG