

**Supporting Information S4. The independent dataset S(4) for *H. sapiens*.** It is formed by 100  $\psi$ -site-containing sequences and 100 false  $\psi$ -site-containing sequences. Each sample is 21-bp long with uridine located at the center. None of the samples included here occurs in S(1) of the Supporting Information S1.

### **I. 100 $\psi$ -site-containing sequences (positive samples)**

```

>P1
GCUAAACAGGUACUGCUGGGC
>P2
UUAUUGAGUGUCUACUGUGUG
>P3
GAUAAACUGUACGCAUAUAU
>P4
UUGUCGGUGUAAACAAAUGG
>P5
UCGGGCCUAGUCAAACCUUU
>P6
UUUUUAAGUAACAGGGGUCU
>P7
GGCCGGUCUGAGCGGAUCAC
>P8
UAGCUAUCGCUUCUCGGCCUU
>P9
UGAAAGUAACUUGCCCGAGC
>P10
ACUAUUCUGUAAAAUCAGGA
>P11
GCAGCUGCCUUCCAACAGCC
>P12
CAAAAUGACUUCGUUCUUCU
>P13
UUCAGAUACUACAUAUUUUU
>P14
CCGAAUCAACUUGCCGUGUU
>P15
GACUCAAAGUACUCUCCUUC
>P16
CUACCCACCUUCCCAGAAGU
>P17
GGACAAUAUAUAAAUGGAUU
>P18
GAGGACAAUAUAUAAAUGGA
>P19
GUGUAGUAUCGUUCUUAUCA
>P20
AAGUGUAGUAUCUGUUCUUAU
>P21
UCAAGUGUAGUAUCUGUUCUU
>P22
AGAUCAAGUGAGUAUCUGUU
>P23

```

CUUCUCGGCCUUUUGGCUAAG  
>P24  
AGUAAUCGCUUCUCGGCCUUU  
>P25  
GAGUAAUCGCUUCUCGGCCUU  
>P26  
GAUGUAUUGUUGCACUCUUC  
>P27  
AUGAUUCUAUUAUAGUAUUCU  
>P28  
UGUUUUUGUAUGUUGCUCCU  
>P29  
UUCUUUUUUUUGGCCUUUCUC  
>P30  
GCUAAACAGGUACUGCUGGGC  
>P31  
CCAUAUCGCUUCUCGGCCUU  
>P32  
CAUAUCGCUUCUCGGCCUUU  
>P33  
UGUAAUAUUUAUCCUGGAC  
>P34  
UAGUAUCUGUCUUAUCAGUU  
>P35  
GUAGUAUCUGUCUUAUCAGU  
>P36  
GUGUAGUAUCGUUCUUAUCA  
>P37  
AAGUGUAGUACUGUUCUUAU  
>P38  
UCAAGUGUAGUAUCUGUUCUU  
>P39  
AGAUCAAGUGUAGUAUCUGUU  
>P40  
GUAUUGAGUGUCUACUGUGUG  
>P41  
UUUUCAUCACUAUGGCUUAGC  
>P42  
GCAUCAAAACUCACUUUUUG  
>P43  
AUUGGUGGUAAGUGGUGAGC  
>P44  
GAUAAAAGGCUAAUAUCCAGA  
>P45  
GGUCCCUGGUUGGAUCCCGGG  
>P46  
AGACUGAAGAUCAAAGGUCC  
>P47  
GGGAGAGCGUAGACUGAAGA  
>P48  
UGGGAGAGCGUAGACUGAAG

>P49  
AAAUCCUUUCUAAAUUGCAUG  
>P50  
CAUAAAAAGUUUUUUCUUCAG  
>P51  
AGAGUAUGGAUCCGAUAUGA  
>P52  
AAGACAUGAAUAAGAACUGAU  
>P53  
GACUUUCAAUUAUCUGUGUGA  
>P54  
GCCUUUUCUUUGUUUGUAACU  
>P55  
AGCCAUCAGGUAAGCCAAGAU  
>P56  
CUUCUCGGCCUUUUGGCUAAG  
>P57  
AUCCAUCGCUUCUCGGCCUUU  
>P58  
GGGCCCAGGGUGCUGUGGAGA  
>P59  
AUUGUCCUCCUUCUGAAGCCC  
>P60  
CCUCCUUUUCUGAGGAAGGUG  
>P61  
AUUGGAACGAUACAGAGAAGA  
>P62  
AGACUAUACUUCAGGGAUCA  
>P63  
GCGCCCCAAUAUUAUGACUG  
>P64  
UAAGUUAUUUUGCUCUCACUG  
>P65  
GCAAUUUGGUCCACCACAUC  
>P66  
ACUCAAUACUACCUGGCAGG  
>P67  
CACUCAAUACUACCUGGCAG  
>P68  
CUGGCUGCUGAGGUCUUUUC  
>P69  
AUUGUUGAUAUUGCCCAGCA  
>P70  
GGGCCUCAGUAGCUCUCAAG  
>P71  
UCCCAUGGUGAAUGGUUAGC  
>P72  
GUUAGCACUCUGGACUUUGAA  
>P73  
GGACUUUGAAUCCAGCGAUCC  
>P74

GCGAUCCGAGUUCAAAUCUCG  
>P75  
CGAUCCGAGUUCAAAUCUCGG  
>P76  
UCAUUUAUGUAUAUUUAUCA  
>P77  
CCUUUCCAGUACUCCUUAUA  
>P78  
UAAGUUAUUUGCUCUCACUG  
>P79  
UCAAGUGUAGU AUCUGUUCUU  
>P80  
GGUAGGUGAGUUUAAAGUCUU  
>P81  
CUCUUACCUGUAAAAUCAGG  
>P82  
GCAACAGAGUCAACUAUCUC  
>P83  
CAUUUGCUGUACUCUGGAGA  
>P84  
UCAAGUGUAGU AUCUGUUCUU  
>P85  
GUAAAAGGGUACUCUCAUAC  
>P86  
UUUUUAUUUUUGGAUGAAUAU  
>P87  
CUUCUCGGCCUUUUGGCUAAG  
>P88  
AACUAUCGCUUCUCGGCCUUU  
>P89  
AAACUAUCGCUUCUCGGCCUU  
>P90  
CCCUGGAGGUCCAAUCCUGC  
>P91  
UUCUCCAUGAUUCGUGCAUCU  
>P92  
CUAAUUAUGCUGGACUGUUUU  
>P93  
AUUGGAACGAUACAGAGAAGA  
>P94  
AUAUUUCUCAUUCUUUUAGU  
>P95  
UAUACUAAAAUGGAACGAUA  
>P96  
AUUGGAACGAUACAGAGAAGA  
>P97  
ACACGCAAUUCGUGAAGCGU  
>P98  
AAGUGUAGUAUCUGUUCUUAU  
>P99  
UCAAGUGUAGU AUCUGUUCUU

>P100  
UCAAGUGUAGUAUCUGUUCUU

## II. 100 false $\psi$ -site-containing sequences (negative samples)

>N1  
GUGAUUAUAACUCAGUGGCAGA  
>N2  
GGCCUUGGAUUUCAUCCCCAG  
>N3  
GGAGAGGGAGUGGGAACAGGA  
>N4  
UUUGCAAGACUCCUAGUACCU  
>N5  
UGUGUAGCAAUGGUGUCCAGG  
>N6  
AGUAACAAGUUCAGGUUCACC  
>N7  
GCAAAGUCACUCUAUUCUGAU  
>N8  
CCCAAAGGUUACUUA AUGUU  
>N9  
UAGGUUCCUGUGCUUGCCAU  
>N10  
CUAAGAGGUUUGUUGUCCUAU  
>N11  
UGGAAGUCUUUCCUUUAAAG  
>N12  
UCUCUUAGCAUCAGACACUUA  
>N13  
AGAGAGAGAAUGAGAAUCAUC  
>N14  
GUGGAAUGAAUAGACUUAACU  
>N15  
GUCAGGAGGCUGUCUUACGUA  
>N16  
CACAAUUGCAUGUGGAAGCUG  
>N17  
CAAUAACUCAUCCUACAGCC  
>N18  
CCACAAACGGUUUAAGCUUGA  
>N19  
GUCACAAUAAUCAUCAUUUCA  
>N20  
UCCCUUCAAUA AAAAAAAAAAU  
>N21  
CAUUUCUGAAUCAGAUGUAU  
>N22  
CUAUCAUAGUUGGGUUUAAGA  
>N23  
AUCAGAACAUUGGGUAUAUUC

>N24  
CACCAUGGUGUCUGGGAGCAC  
>N25  
ACAUUACCCCUCUUUCCCGC  
>N26  
ACCAACGAUCUGCUUGUGAAC  
>N27  
AGAGCUUUAGUCCAGAGCAAG  
>N28  
CCCCCGCCUUUUUUUCUGUUG  
>N29  
UAAAUUUUGUAUGCAAUUA  
>N30  
UUUAGAGGAAUAGGGAAAGUG  
>N31  
GACGUGUCUGUGUUUCUCAA  
>N32  
GGGUCCGGACUGUUUGACACU  
>N33  
GAUGAAUGCUUCUCAAAAGU  
>N34  
UUAACAGUUUCAUUUGGAAG  
>N35  
UAGGGUCGCCUAAGUCAACA  
>N36  
UCACAGAUGCUCAGCAGGCA  
>N37  
ACCAUAUGUUAGAAAUAAA  
>N38  
CCAGCCGCGGUGCCAGCAAAG  
>N39  
AACAGACACAUACUUGAACU  
>N40  
UGUUCUGAGUUCUACUGUCUU  
>N41  
ACCCAAUUGCUGGAAACUCU  
>N42  
CUUAUGACUGGACUUCAGAA  
>N43  
AAAGAAGGAUCCAAAGACAA  
>N44  
ACUCAAAUUCUAGAUGACCA  
>N45  
AGGCAGACAGAGGAAGAGUA  
>N46  
AUGGAAUCCUUUUGUUUUGU  
>N47  
UGUUCUGUUGUGUCAAGUGC  
>N48  
AAAAUAUAAUUGUUGAAUA  
>N49

UGUGUGCUUCUGUCCUACUAC  
>N50  
AUUUCUUCCAUUUUUAAUUA  
>N51  
AAAGUAGAGCUGAGGCCACU  
>N52  
CUUGUUCUGUACUCACUGUA  
>N53  
GGACCCACCUAAAAGUAUAA  
>N54  
UCCUGAGAGUUCACGCUGAGC  
>N55  
CUUUUCUCUCUCUCCUGAAA  
>N56  
ACUGAAGUGUCCCAAAGCUA  
>N57  
UGUGUAAAGGUUGGUUCUCA  
>N58  
UCUCUCUCUCUCUCUCUCUCU  
>N59  
UGUAGGUGGGUAGUAGGUGAG  
>N60  
CAGCUGGGAGUAAAUACUCU  
>N61  
GUGGAACCUCUCUAGUUAAAA  
>N62  
GUAACCAGUCUGUGGGAAGUA  
>N63  
AAAGCAACAUCCCUGCUGGA  
>N64  
GGCUCCAGGAUCCUAAGGGAC  
>N65  
GUCUGUACUCUAAGGGGACAU  
>N66  
UUAAAUGCAUCUCCCUCAUU  
>N67  
AAAUGAUGACUGAUGCUACUA  
>N68  
UGUUUAAACAUGGAUUUAAAC  
>N69  
GUUUUUUCAUGUUUUUAUU  
>N70  
UCACUGUGGGUCUGGGCUUUA  
>N71  
AGACCCUCAUUUAGCUGCCU  
>N72  
AGCCUUCAGAUGAAGGGGGG  
>N73  
UCUCUGCUAAUAUACAUCUG  
>N74  
GAGUUCAGCCUCAGAACUUG

>N75  
UCAGCCACCCUACCCUACUUG  
>N76  
GACCAUGUCUUUAAAAGACAA  
>N77  
GUGGUUGACUUUGGGUUUCUU  
>N78  
AUGUGUUUGUUUGUUUGUUUG  
>N79  
UUUGUUUUGCUCCUGACACCA  
>N80  
CCACCCUCUUUAAGUAGAUU  
>N81  
GUGACCAGAAUAGUAACUAAA  
>N82  
AUGUUGAAUUUAUUUGCUUAA  
>N83  
CAAAUGUGGCUUAAAUUUUUA  
>N84  
AGGAUCAUUAUGAAAGAUGAA  
>N85  
UAGCUCCCCUUCUCUGCUUG  
>N86  
UGAACACGUAUGCCAAUGGAC  
>N87  
UCUGCUCCCGUGUUACAGUGU  
>N88  
GACCUAACUUUGGAUACUUUU  
>N89  
UCCUCUAUAGUAACCACAUU  
>N90  
AAUUUCAAAAUGCAGAGAAA  
>N91  
UGGAUCACUUUGCAUCAGUAG  
>N92  
GGCUGGUAAAUGAAAUACUG  
>N93  
GACCAUCACAUUUUCCUGGU  
>N94  
GCUUCUUUGUUUAUUCAUUUG  
>N95  
GCUAUUCCAUGUUCUGUAC  
>N96  
CAUCAAUCUUUCUCAGUUUGU  
>N97  
GAACAUGAGCUUUGAGAUUC  
>N98  
AUUCAGGAGGUCUCAGAACAC  
>N99  
UAAGGCUUUAUGUCUCCUAA  
>N100

UCUUAACUCUUGGGGCUGGUA