

Supporting Information S5. The independent dataset S(5) for *S. cerevisiae*.
It is formed by 100 ψ -site-containing sequences and 100 false ψ -site-containing sequences. Each sample is 31-bp long with the uridine located at the center. None of the samples included here occurs in S(2) of the Supporting Information S2.

I. 100 ψ -site- containing sequences (positive samples)

```
>P1
CUAUCAUCGCUGAUCUCCCACUCCCUGAUCU
>P2
GAAGAGGUCAUCGGUUCGAUUCGGUUGCGU
>P3
GUAAGAUGCAAGAGUUCGAAUCUCUUAGCAA
>P4
GCGAAAGAUUAGAAAUCUUUUGGGCUUUGCC
>P5
GGUUAAGGCGAAAGAUAGAAAUCUUUUGGG
>P6
UUUAGGACCGAGCUUUAUGUGGAUGUCAUCA
>P7
GGACACUUCUGAUGUUCAAAAGAUAUUCCA
>P8
GGUACUGGACGAGAAUCGCAGAACAAUUUGA
>P9
CGUAGAUGUUUGUUGUCACCCACAACUGAA
>P10
GAGUUGUCGAGUUUUUGAGGUUAAGAAUGA
>P11
AAGGUCGAAAAGUUUCAGGCAGUUUCUCAG
>P12
CGUUGGGCCCCCGGUUCGAUUCGGGCUUGC
>P13
UGGUAAAAUCCAACGUUGCCAUCGUUGGGCC
>P14
UAAGCGCAAGUGGUUAGUGGUAAAAUCCAA
>P15
GGUUAAGGCGAAAGAUAGAAAUCUUUUGGG
>P16
GCGAAAGAUUAGAAAUCUUUUGGGCUUUGCC
>P17
GGCUUCAUUAACAUGACUUCAACUACGGAA
>P18
GUGGAGAUCAUCGGUCAAUCCGAUUGGAA
>P19
UUUGGUUUUCAAGUGAAUAGGCUACGUGAU
>P20
CAGUGGUUCAAGACGUCGCCUUUACACGGCG
>P21
UAGUGGUUAUCACUUUCGGUUUUGAUCCGGA
>P22
CACUUUCGGUUUUGAUCCGGACAACCCCGGU
>P23
```

AAUUGAUCUAUGUUGUAGCUGCGCUGGCGGC
 >P24
 AACUCCAGUUCUUUAUCUUCUUUCUCCGCUG
 >P25
 GCGUCUGACUUCUAAUCAGAAGAUUAUGGGU
 >P26
 UCUUCCGUGAUAGUUUAAUGGUCAGAAUGGG
 >P27
 CAGAAUGGGCGCUUGUCGCGUGCCAGAUCCG
 >P28
 GUGCCAGAUCCGGGUCAAUCCCCGUCGCG
 >P29
 AGAAAAAGCCAAUGAUGAGAUACAAGCCAUU
 >P30
 AUCGACAU AUGCUGGUACAUGGCAGUAGAA
 >P31
 GAAUAUACA UUCUAUAUCGAACCUGGCCAU
 >P32
 GAAACAAGAUUUCUGUAGCAUACUCGCUUCA
 >P33
 UACUUGUUUUCUUUUUGUGCCUUUGUUACG
 >P34
 UUGCUUUGUGGAAGUUCGAAACUCCAAAGUA
 >P35
 UGAGUGAUGGAAGUGAGUUAUCCGGAGAUC
 >P36
 AGGGUCAAAUCUUCGUGACCGUCAAUUACA
 >P37
 UGCAGCACAAAUUUGUAGACAGGCUGGUUUG
 >P38
 AGGAUUACUUGGACAUAACGGUUCUCCUAU
 >P39
 UCAAGACAAAAGUGUUCUUUCAUCUGCAGUG
 >P40
 UUGGCGUACAGAUUGAGUUGUGGCUGCUAC
 >P41
 CUUUUUUAAUGUCCGUUUCUAUGAUUGGGCU
 >P42
 AUUGUUCGAAGGUAAUGCCUUGAUCAGAAGA
 >P43
 CUGUUGGUCCUAGUUCGAUCCUGAGUGCGA
 >P44
 GCAGCAGAUUGCAAAUCUGUUGGUCCUAGU
 >P45
 UUAUCCGAUAUAGUGUACGGCUAUCACAUC
 >P46
 CGUGGAGACCGGGGUUCGACUCCCCGUAUCG
 >P47
 GGUAUGUUAUUUAUGUACGGGUAUGCGAAC
 >P48
 AUUCUUUUUUUGAUGAAUAGGAUAAGCUUG

>P49
CUGUUCUUUUCAGUGUAACAACUGAAAUGAC
>P50
UGUAGUAUCUGUUCUUUUCAGUGUAACAACU
>P51
GUGUAGUAUCUGUUCUUUUCAGUGUAACAAC
>P52
AAGUGUAGUAUCUGUUCUUUUCAGUGUAACA
>P53
CAAGUGUAGUAUCUGUUCUUUUCAGUGUAAC
>P54
AUCAAGUGUAGUAUCUGUUCUUUUCAGUGUA
>P55
UCAUUGUUCUUGGAUUUCA AUGGGUGCUGUC
>P56
UAAAUUUCGCCACUGUAGAUGAAGAAGACGA
>P57
AAA AUGAGAAGAGUGUAGAUGUAUUAUCCUU
>P58
CCAAGAUAGACUAUGUAAUGGUAAAGAACAU
>P59
AUGGCGGCGGGUGCCUUUGGAGCAGCAAUCG
>P60
AUGGUGUGGUCACUGUAAGAGAUUGGCCCCA
>P61
CCAUGGACGAGCCUGUAGUAUACAACGGUAA
>P62
ACAAAGGUCUCCUAUGAUUCCGGCGUUCGU
>P63
CUUUCUCAUACCCUGUAGACCAGACCUCUCU
>P64
AGAAUACUUUGAAGGUUAACCGAGGAAAUG
>P65
CGUGGAGACCGGGGUUCGACUCCCCGUAUCG
>P66
UUAUCCGAUUAUAGUGUAACGGCUAUCACAUC
>P67
GGACACUUCUGAUGUUCAAAAGAUAUUCCA
>P68
UAACUGUGGGAAUACUCAGGUAUUCGUAAGAU
>P69
GUAAGAUGCAAGAGUUCGAAUCUCUAGCAA
>P70
ACAAUUUUCACAGUUUAAGGCCAAGAACAAG
>P71
GCCCCGUUACACAUUUGAUACAACCGUAGA
>P72
CGGGAGGUCCCGGGUUCGAGUCCCGGCUCGC
>P73
GAUUCUCGCUUAGGGUGCGGGAGGUCCCGGG
>P74

GCGUGCGACUGUUAACGCAAGAUCGUGAGU
 >P75
 CGCAAGAUCGUGAGUACAACCCUCACUGGGG
 >P76
 CGUUGGGCCCCCGGUUGCAUUCGGGCUUGC
 >P77
 UGGUAAAUAUCCAACGUUGCCAUCGUUGGGCC
 >P78
 CUAGCGCAAGUGGUUAGUGGUAAAUCCAA
 >P79
 GGCUGUGUUCUUCUUUAUAAUUCUUAUCG
 >P80
 GGAAAAACCCGUUGCAGAAAGCGCAACUGGU
 >P81
 GAAAAAGUUCAGAAUUGCAGAAAGUGGUG
 >P82
 AGUGGUUCCUAGUGAUCAGCCACUAUCGG
 >P83
 CAUAAGGUUAGGGGUUGAGCCCCCUACAGG
 >P84
 GCAAUCGGUAGCGCGAUGACUCUUAUCAU
 >P85
 AAACAAAAGAAGCUGUCCAGAGAGCCCAAG
 >P86
 CCGGACAACCCCGGUUGGAAUCCGGGUAGGA
 >P87
 CACUUUCGGUUUUGAUCCGGACAACCCCGGU
 >P88
 GGUUAUCACUUUCGGUUUUGAUCCGGACAAC
 >P89
 UAGUGGUUAUCACUUUCGGUUUUGAUCCGGA
 >P90
 UGUUGCCGCUAAGUGAAGGAAGUCGGUAUC
 >P91
 CUGGUUAUUCUAUAUACUCACUUAUUACUU
 >P92
 UUCUGGUUAUUCUAUAUACUCACUUAUUAC
 >P93
 AAUGGCUCUUUUUGUAUUCGAAAGCUACA
 >P94
 UAAAAAGUUCGGCUAUCUCUUGGGCUCUGCC
 >P95
 UCUGCCCGCGCUGGUCAAUCCUGCUGGUG
 >P96
 CAUGGAUGAUUUUGUAGUAUGGCGGAAAC
 >P97
 GUGGAGAUCAUCGGUCAAUCCGAUUGGAA
 >P98
 AUACUAUUCAGUUUCAGAUUAAGGUUGCA
 >P99
 GCAAUUGGAAAAAUCUAUUAACCCAGAUGAA

>P100
CCAGUGCGUCUACUAUUACUCGGCCAAAUAU

II. 100 false ψ site containing sequences (negative samples)

>N1
UCGUAAUUUGAGAUCUCUGCAAAACAAUGCA
>N2
AAACAAUGCACCUCUCCUGGC AAAACAUCAAU
>N3
AAACAUCAAUGUCAAUUGUUUGAACGUCAAU
>N4
GAACGUCAAUUCUUGUCGUUGUCCGCAAGC
>N5
AAUUAUAUGGCUUGUAAUGGAAACAAGCAA
>N6
AAACAAGCAAGAUCUCCCAUACCGUUUCCC
>N7
ACCGUUUCCCCUGCAUGUAGAAUGCAACGAU
>N8
AUCAAUGUUUAAUCAUAAACAGAUCAAAGAGC
>N9
AUCAAAGAGCAGUGGUACUACAGAU GCGUCA
>N10
AGGUACGCAUAAGCGUGAACCCCGGUCGACG
>N11
CCGGUCGACGAUACAACAGAGCUGUUACAA
>N12
UAUAGCAAAGGACAGUAGAAACCUGAGUAAU
>N13
CCUGAGUAAUGGAUCUUUGAAUGAUUAUAAAC
>N14
UGAUAUUAACGAAAAUGAAGAGCUCCAAAAU
>N15
GCUCCAAAAUUUCCAAGAAAAAUCAGCGAA
>N16
UCCCCAAGGAAAAAAGCGAAACCAGAAAG
>N17
GUUAAUGCGGGAAGAUACAUUGCCUUGAAA
>N18
UGCCUUGAAACAACCCCAAGCUUGGGAGAU
>N19
GAGGAGAUUCGUCGUUAAGAACCAAGUCA
>N20
AACCAAGUCAUUCGGUACAAGUCCAAGAC
>N21
GUCCAAGACAUAACUGUCGAACCUCAAUCC
>N22
CGUAGGUAAAGUGUAUUUAGUGAGGGAACGC
>N23
CGCGACAAGUGAUCAUCCAUUUAUUGUGACA

>N24
UAUUGUGACACUGUAUCAUCCUUUCAAACC
>N25
AAACAUAUUACUGCAUCAAUCUGGUCAUGUC
>N26
UGGUCAUGUCAUGCUUUCUGACUUUGAUUUA
>N27
AGAUACAAAAUUUGUUCAGAUGGAUUCAGA
>N28
UGGAUUCAGAACUAAUCCUUUGUUGGUACU
>N29
UGGCUGUACUCCAUUUAAAGGAGAUAAUUCA
>N30
CGUCAAAUUUCCACAUGAUAAGGAAGUUUCG
>N31
GGAAGUUUCGAAGAAUGUAAAAGACCUGAUA
>N32
CUUCUUCAAAAAGUUCAGUGGUCGUUCUUA
>N33
CCCCCUCUAAUACCUGCAUUAUAUGAUAAC
>N34
UCCUUUUUAUAUUGUCUUGCAAUAAACACCCG
>N35
AAACGAUGAUGAAAUUGAUGAGGCUGAUCCA
>N36
GGCUGAUCCAUCCAUGAUUUUAAUUCUAUG
>N37
CUACUCUGAAAAUUAUACCUACGGAAAAAUU
>N38
AUCUUAUGAUAAAAUGUAAAAAUUAUUUA
>N39
AAACGAGAAAGUGAAUGAAAAUAUAUAUC
>N40
AUAUAUAUCAUUUAUGUCUGAUAUUGCUG
>N41
UACGUACCAUCCGCAUCAGUGGAUAUCCAAU
>N42
GAUAUCCAAUGAUAGUAAUUUCGCGAGUUUA
>N43
CGCGAGUUUAUCCGUUGCUGUUAUAUUAUCA
>N44
UAUAUUAUCACUUUUUAAUAUUCUUUCAA
>N45
GGAUCCUUCGCAAUCUUCUGAAAUACUG
>N46
CUCGCCAGUUUUUUGUCUUCACGUAAUCC
>N47
CCUUAUUAACGGAGAUUGAUUUCUCCAGC
>N48
ACCGAUUCGAGUGAGUACGUUUUCAAUAUG
>N49

UUCAAAUAUGCUUAAUCUGAUCUCUUCUGCG
 >N50
 GCCGAUCUGUGCCAUUAUAGUAAGCAGUGCC
 >N51
 AAGCAGUGCCACUUGCUAAUAUAAGAUGAU
 >N52
 GAUUUUACCGUUUUCUGGGGACAUCAUGAUA
 >N53
 CAUCAUGAUAUCAUUGGUACAUAAGAACA
 >N54
 GAUAAUUGGAUUUCUUGCAUUUUUUGCGAUU
 >N55
 UUUUGCGAUUAUGGCUGUUGACCAUUCACA
 >N56
 ACCAUUCACAAAAGUUGGUCUAACAUAUUUU
 >N57
 UAAGUCCUUGUAAUAUCUAGCUUUUGAGUC
 >N58
 UCUGGGAGUGGUAAAUCUACUGACCAUCUUC
 >N59
 UUUUAUCCAAUCAUCUGGCAAGUCCUAAUU
 >N60
 UUCAUCUCUAAAAUUAGAU AUGGACGUUUG
 >N61
 UGGACGUUUGAGAUUUUUCGUUUUCUGCC
 >N62
 UAUUUCUGCCAAUUCUCCUUUAACUGUGAC
 >N63
 UAACUGUGACCGUACUGAUUCGCUUCCCUU
 >N64
 GCUUUCUUUGAAUUAUUUAUUAUACCCUCU
 >N65
 CAUUACUUGCUUAUCUGAAUUUUUUUCCAUA
 >N66
 UUGUUUGCCUAUCCUCCAUCUGAUGACUUG
 >N67
 UGAUGACUUGAAAUGUCUGACAGGUAAGAU
 >N68
 UCUCAACAUUCUAAUCCAAACGAUGUCCUU
 >N69
 CUCCUGCUUGUGUAUUAAGGACAUGAAUA
 >N70
 UUUCGCUACAUGUAAUGGAAGAUCAUCUGUA
 >N71
 GAUUCGUUAUCUGUAGUUCUCAUCAGCAAGA
 >N72
 UCUUUCAAAAACGCUUGAUUUGCUGGCACCU
 >N73
 CUUAAUAGCGCUUGUUCUGCUUGCUCUACC
 >N74
 CUCUAGGUGAACGUUUAUCUGACAUCGGG

>N75
AAGUUUGAUGUGAAGUAUUCUGCUAACCGUU
>N76
CAGGUCCUUCACCUGUUGUCCAAGGGAGUA
>N77
AGGGGACUUUCUGGCUUUUUUUUUACGAAA
>N78
CUCUCCUCAUCAUCUCAGCCUCAACAUUU
>N79
UCCAACCGCAACUUCUGUUCUUGCUUAUGC
>N80
CCUGCUUAUUGUGGGUGUCCCGCCAUAUU
>N81
CGCCAUUAUUGUAAUAGAUUCAACAAAUA
>N82
UUUAUCAUUGAAAAUCACGUGAUCGCAUA
>N83
GAUCGCAUAUCCGUCAGGAGUGAUAAUA
>N84
UCGUCAUUGCACAAAUAGUUUAUUAUUCAC
>N85
UAUUAUUCACGACUCUUAACAACGACAAUUU
>N86
UAGACAGGUCGUCCGUGAUUUUACAUAUA
>N87
UACUACACAGACUACUUAUAGAAUUUGCGAA
>N88
AAUUUGCGAAGGAUUACCGAAGAAAAGCAC
>N89
AGAAAAGCACAGACCUUAUUGAGCUUUUGAA
>N90
UCAUAACCAGGAGUUCAAAAACAAACAGG
>N91
CACUUUUCAUUGAUCUUAUUUGAUAAAUCUGC
>N92
CACUAGAGUCCAAUCACGCGACUUAUUGCA
>N93
ACUUAUUGCAUUCUUGGAAGGUGAAAGUCU
>N94
UGCACGAUGGUCCAGUAUUGAUGAAUUUUU
>N95
AUUUGGCCAUUCAACUCAUAAGUGGUCGGU
>N96
AAGGUACCAGGAAAGUUCUGAAACCAUCCC
>N97
AAUCUUUACUCUACUUAUUAUCCAUUGCAUC
>N98
CCCCUGAAUAUCUUAUUUUAGCAUUAGUCA
>N99
CAUUAGUCAAGAAUGAAGCGGUUCGUUUU
>N100

GGUUCGUUUUAUUGAUAGAAAACAGGACAGU