

PhosphoPredict: A bioinformatics tool for prediction of human kinase-specific phosphorylation substrates and sites by integrating heterogeneous feature selection

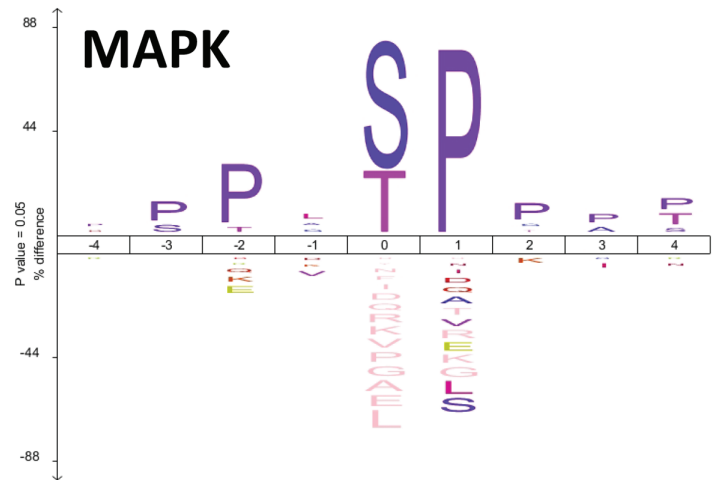
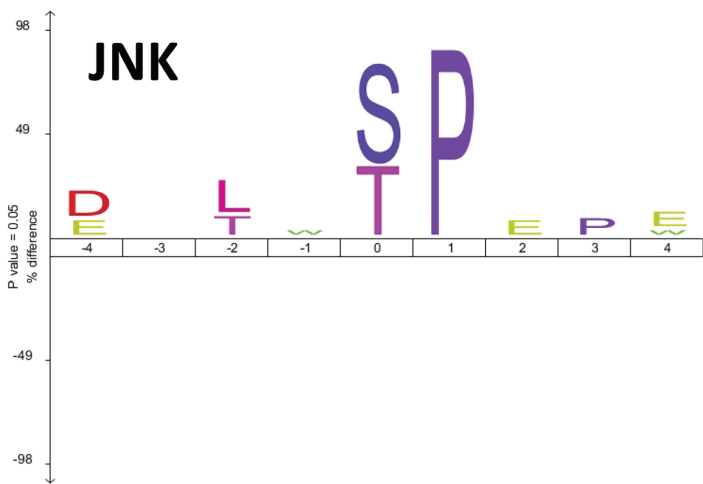
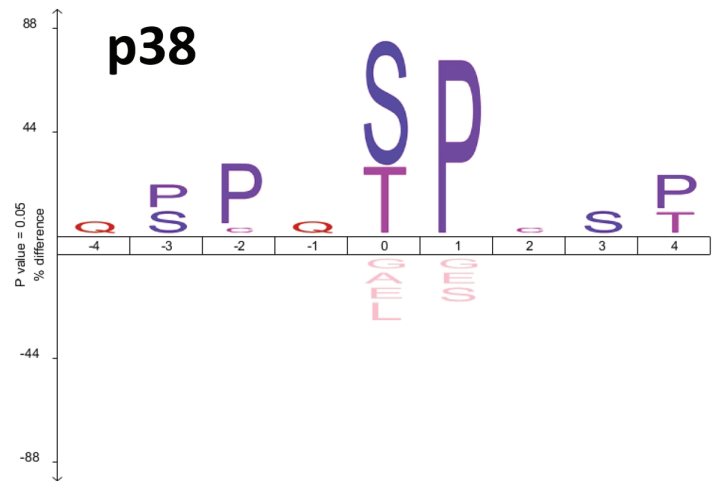
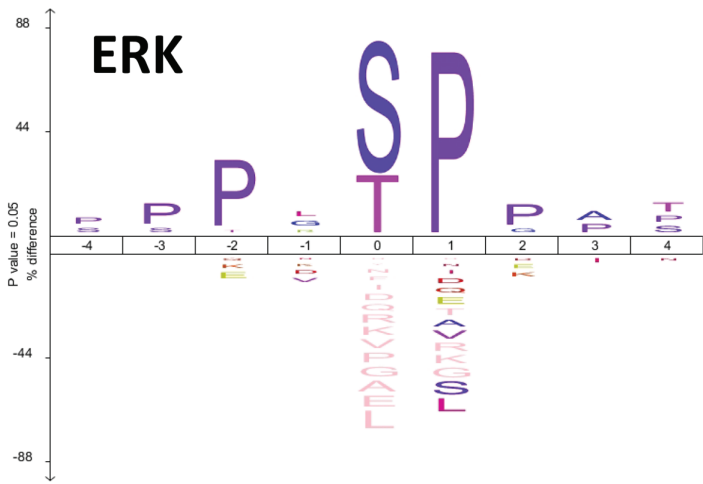
Jiangning Song, Huilin Wang, Jiawei Wang, Andre Leier, Tatiana Marquez-Lago, Bingjiao Yang, Ziding Zhang, Tatsuya Akutsu, Geoffrey I. Webb and Roger J. Daly

Supplementary Information

Supplementary Figure S1. Sequence logos generated using pLogo for ERK, p38 and JNK MAPKs, and the overall consensus motif for the MAPK family.

Supplementary Figure S2. Example of output webpage from the online web server of PhosphoPredict.

Supplementary Table S1. Prediction performance of different sequence encoding schemes based on the 5-fold cross-validation tests.



Supplementary Figure S1. Sequence logos generated using pLogo for ERK, p38 and JNK MAPKs, and the overall consensus motif for the MAPK family.

PhosphoPredict Results:

The prediction output results of kinase-specific phosphorylation sites for the submitted sequences are provided in the following Table. In addition, intermediate files (e.g. predicted native disorder, secondary structure and solvent accessibility) generated during the prediction process are also provided for users' reference.

Model File: PKB Threshold: All

[Export as txt](#) [Export as excel](#)

input-1 *

Rank	Position	Surrounding Sequence	Regulatory Kinase	Probability Score
1	280	PRVT S GGVS	PKB	0.246

Original Sequence: ●

MAVPFVEDWDLVQLGEGAYGEVQLAVNRVTEEAVALVKIVDMKRAVDCPENIKKEICINKMLNHENVVFKYHRRREGNIQYLFLEYCSGGELFDRIEPDIGMPEPDAQRFHQLMAGVYVHLHGIGITHRDIKPENLLLLDERDNLKISDFGLATVFRYNNRERLLNKMCGTLPYAPPELLKRREFHAEPV
 DVWSCGIVLTAMLAGELPWDQPSDSCQEYSDWKEKKTYNLPWKKIDSAPLALLHLKLVENPSARITIPDIKKDRWYNKPLKGAKRPRVTSGGVSESPSGFSKHQSNLDFSPVNSASSEENVYSSSQPEPRTGLSLWDTSPSYIDKLQGISFSQPTCPDHMLLNSQLLGTGSSQNPWQRLVK
 RMTRFFTKLDADKSYQCLKETCEKLGQYQWKKSCMNQVTISTDRRNNKLFKVNLLMDDKILVDFRLSKDGLFKRHFLKIKGKLLDIVSSQ

Native Disorder: ●

Solvent Accessibility

This is the solvent accessibility information for the sequence, which provides a score from 0 to 6, representing the extent of solvent accessibility from buried to exposed.

Secondary
 CCCCCC EEEEECHHHHHHHHHHHHHHHHHHHHHCCCEEEEEEECCCEEEEEEECCCHHHHHHHCCGCCHHHHHHHHHHHHHHHHHHHHCCCEEECCGCCCEEECCCEEECCCHHHHHCCGCCCEEECCCHHHHC
 CCCCCC CHHHHHHHHHCCCEEECCCHHHHHHHHHHHHHHHHHHHHHCCCEEECCCHHHHHCCCEEECCCHHHHHCCCEEECCCHHHHHCCCEEECCCHHHHHCCCEEECCCHHHHHCCCEEECCCHHHHC
 CCCCCC HHHHHHHHHCCCEEECCCEEEEEEECCCEEEEEEECCCEEEEEEECCCHHHHHHHHHHHHHHHHHHHHHCHHEEEEC

Solvent Accessibility: ●

44243134030021013011010110223525420001002253025035303210200310412100201211445310000000041120022015635153430130010002001000311001110101100013632010010000001334443410422111000001004443131120000000
 00000001010223431031023135552411214302430130023002321521120440362410333155334334244442444344334414451533324344454424444544434443232133334314412441212222335312224324433333333542152211010334044004302
 4005506150445433213133455340201010100201330002022342121110340045025303300122

Please note that submitted sequences with the sequence name labeled as "*" mean:
 The sequence you entered DOES NOT provide a valid Uniprot ID and thus protein functional features (functional annotations retrieved from KEGG, GO, PPI and Pfam etc) will not be extracted and used for making the prediction.

Supplementary Figure S2. Example of output webpage from the online web server of PhosphoPredict.