Supplementary Material:

PhosIDN: an integrated deep neural network for improving protein phosphorylation site prediction by combining sequence and protein–protein interaction information

Hangyuan Yang 1, Ming-Hui Wang 1,2\*, Xia Liu1, Xing-Ming Zhao3,4,5 and Ao Li1,2,\*

1School of Information Science and Technology, University of Science and Technology of China, Hefei AH230027, China, 2Centers for Biomedical Engineering, University of Science and Technology of China, Hefei AH230027, China, 3Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China, 4MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence and Frontiers Center for Brain Science, China and 5Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China

**Table S1.** The number and residue type of sites in independent test data for each kinase group, family, subfamily and individual kinase

|  |  |  |  |
| --- | --- | --- | --- |
| Kinase | | Residue type | Number of sites |
| Group | AGC | S/T | 742 |
|  | Atypical | S/T | 236 |
|  | CAMK | S/T | 316 |
|  | CMGC | S/T | 1046 |
|  | TK | Y | 516 |
| Family | CDK | S/T | 412 |
| CK2 | S/T | 241 |
| MAPK | S/T | 301 |
| PKC | S/T | 386 |
| Src | Y | 283 |
| Subfamily | CDC2 | S/T | 199 |
| CDK2 | S/T | 156 |
| ERK1 | S/T | 180 |
| PKCa | S/T | 290 |
| SrcA | Y | 228 |
| Individual kinase | CDC2 | S/T | 199 |
| CK2a1 | S/T | 186 |
| PKACa | S/T | 180 |
| PKCa | S/T | 270 |
| SRC | Y | 185 |

**Table S2.** AUC values (%) of PhosIDN with sequence information for kinase-specific phosphorylation site prediction

|  |  |  |  |
| --- | --- | --- | --- |
| Kinase | | DCCNN | PhosIDNSeq |
| Subfamily | CDC2 | 94.3 | 95.8 |
|  | CDK2 | 94.3 | 96.1 |
|  | ERK1 | 94.2 | 97.6 |
|  | PKCa | 87.1 | 89.4 |
|  | SrcA | 83.0 | 85.9 |
| Individual kinase | PKCa | 90.8 | 94.5 |
| CDC2 | 95.2 | 95.8 |
| CK2a1 | 92.2 | 94.1 |
| PKACa | 95.6 | 96.9 |
| SRC | 82.6 | 85.4 |

**Table S3.** AUC values (%) of PhosIDN with both sequence and PPI information for kinase-specific phosphorylation site prediction

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kinase | | Baseline | IFENet\* | HFCNet\* | PhosIDN |
| Subfamily | CDC2 | 96.2 | 96.8 | 96.4 | 97.2 |
|  | CDK2 | 96.4 | 97.7 | 97.5 | 97.9 |
|  | ERK1 | 96.6 | 98.4 | 97.5 | 98.6 |
|  | PKCa | 89.4 | 90.1 | 89.9 | 92.1 |
|  | SrcA | 88.0 | 90.8 | 91.1 | 93.3 |
| Individual kinase | PKCa | 91.9 | 95.1 | 93.2 | 97.1 |
|  | CDC2 | 95.8 | 96.4 | 95.8 | 97.2 |
|  | CK2a1 | 94.3 | 96.2 | 95.8 | 96.7 |
|  | PKACa | 97.0 | 97.8 | 97.9 | 98.5 |
|  | SRC | 87.8 | 91.7 | 90.3 | 93.3 |

**Table S4**. The values (%) of Sn, Acc, MCC, Pre and F1 of PhosIDN for kinase-specific phosphorylation site prediction at medium and high stringency levels

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kinase | Method | Sp=90% | | | | | Sp=95% | | | | |
| Sn | Acc | Mcc | Pre | F1 | Sn | Acc | Mcc | Pre | F1 |
| Group | Baseline | 81.5 | 85.8 | 71.8 | 89.2 | 85.2 | 71.8 | 83.4 | 68.7 | 93.5 | 81.3 |
| CMGC | IFENet\* | 82.1 | 86.0 | 72.3 | 89.2 | 85.5 | 72.2 | 83.6 | 69.0 | 93.6 | 81.5 |
|  | HFCNet\* | 84.4 | 87.2 | 74.5 | 89.5 | 86.9 | 74.1 | 84.5 | 70.6 | 93.7 | 82.8 |
|  | PhosIDN | 87.4 | 88.7 | 77.5 | 89.8 | 88.6 | 77.7 | 86.3 | 73.8 | 94.0 | 85.1 |
| Group | Baseline | 67.0 | 78.1 | 58.3 | 87.7 | 76.0 | 46.4 | 70.0 | 47.2 | 91.2 | 61.5 |
| TK | IFENet\* | 76.8 | 83.1 | 67.1 | 89.1 | 82.5 | 62.5 | 78.3 | 60.6 | 93.3 | 74.9 |
|  | HFCNet\* | 70.8 | 80.0 | 61.6 | 88.3 | 78.6 | 49.4 | 71.5 | 49.7 | 91.7 | 64.2 |
|  | PhosIDN | 77.5 | 83.5 | 67.8 | 89.2 | 83.0 | 65.5 | 79.8 | 63.1 | 93.6 | 77.1 |
| Family  CDK | Baseline | 90.6 | 90.3 | 80.6 | 90.6 | 90.6 | 85.9 | 90.3 | 81.0 | 94.8 | 90.1 |
| IFENet\* | 92.5 | 91.3 | 82.5 | 90.8 | 91.6 | 89.7 | 92.2 | 84.6 | 95.0 | 92.3 |
| HFCNet\* | 93.9 | 92.0 | 84.0 | 90.9 | 92.4 | 89.7 | 92.2 | 84.6 | 95.0 | 92.3 |
| PhosIDN | 97.2 | 93.7 | 87.5 | 91.2 | 94.1 | 93.0 | 93.9 | 87.9 | 95.2 | 94.1 |
| Family  CK2 | Baseline | 85.6 | 87.6 | 75.2 | 89.9 | 87.7 | 78.4 | 86.3 | 73.9 | 94.2 | 85.6 |
| IFENet\* | 87.2 | 88.4 | 76.8 | 90.1 | 88.6 | 79.6 | 87.9 | 75.2 | 94.2 | 86.3 |
| HFCNet\* | 85.6 | 87.6 | 75.2 | 89.9 | 87.7 | 78.4 | 86.3 | 73.9 | 94.2 | 85.6 |
| PhosIDN | 89.6 | 89.6 | 79.2 | 90.3 | 90.0 | 84.0 | 89.2 | 79.0 | 94.6 | 89.0 |
| Family  MAPK | Baseline | 86.9 | 88.4 | 76.8 | 89.9 | 88.4 | 82.4 | 88.7 | 78.1 | 94.7 | 88.1 |
| IFENet\* | 88.9 | 89.4 | 78.7 | 90.1 | 89.5 | 83.7 | 89.4 | 79.3 | 94.8 | 88.9 |
| HFCNet\* | 90.2 | 90.0 | 80.1 | 90.2 | 90.2 | 85.0 | 90.0 | 80.6 | 94.9 | 89.7 |
| PhosIDN | 90.4 | 90.1 | 80.2 | 90.4 | 90.4 | 87.6 | 91.4 | 83.0 | 95.0 | 91.2 |
| Family  PKC | Baseline | 64.1 | 77.2 | 56.3 | 86.6 | 73.7 | 46.4 | 70.7 | 47.2 | 89.9 | 61.2 |
| IFENet\* | 71.9 | 81.1 | 63.2 | 87.9 | 79.1 | 53.1 | 74.1 | 52.8 | 91.1 | 67.1 |
| HFCNet\* | 64.3 | 77.2 | 56.3 | 86.9 | 73.9 | 53.1 | 74.1 | 52.8 | 91.1 | 67.1 |
| PhosIDN | 83.3 | 86.8 | 73.7 | 89.4 | 86.3 | 66.7 | 76.2 | 64.2 | 92.8 | 77.6 |
| Family  Src | Baseline | 65.7 | 78.1 | 57.8 | 86.8 | 74.8 | 45.0 | 70.3 | 46.5 | 90.0 | 60.0 |
| IFENet\* | 71.4 | 80.9 | 62.8 | 87.7 | 78.7 | 57.1 | 76.3 | 56.6 | 92.0 | 70.5 |
| HFCNet\* | 65.7 | 78.1 | 57.8 | 86.8 | 74.8 | 50.7 | 73.1 | 51.3 | 91.0 | 65.1 |
| PhosIDN | 72.6 | 83.2 | 64.3 | 89.5 | 80.2 | 59.4 | 79.5 | 57.9 | 92.1 | 72.2 |
| Subfamily  CDC2 | Baseline | 91.4 | 90.5 | 80.9 | 88.5 | 89.9 | 83.9 | 89.9 | 80.1 | 94.0 | 88.6 |
| IFENet\* | 92.5 | 91.0 | 81.9 | 88.7 | 90.5 | 84.9 | 90.5 | 81.0 | 94.0 | 89.3 |
| HFCNet\* | 91.4 | 90.5 | 80.9 | 88.5 | 89.9 | 87.1 | 91.5 | 83.0 | 94.2 | 90.5 |
| PhosIDN | 96.8 | 93.0 | 86.2 | 89.1 | 92.8 | 92.5 | 94.0 | 87.9 | 94.5 | 93.5 |
| Subfamily  CDK2 | Baseline | 84.6 | 87.2 | 74.5 | 89.2 | 86.8 | 61.5 | 78.2 | 59.8 | 92.3 | 73.8 |
| IFENet\* | 97.4 | 93.6 | 87.4 | 90.5 | 93.8 | 91.0 | 92.9 | 86.0 | 94.7 | 92.8 |
| HFCNet\* | 94.9 | 92.3 | 84.7 | 90.2 | 92.5 | 75.6 | 85.3 | 71.9 | 93.7 | 83.7 |
| PhosIDN | 98.7 | 94.2 | 88.8 | 90.6 | 94.5 | 91.0 | 92.9 | 86.0 | 94.7 | 92.8 |
| Subfamily  ERK1 | Baseline | 91.7 | 91.1 | 82.1 | 91.7 | 91.7 | 77.1 | 85.6 | 72.8 | 94.9 | 85.1 |
| IFENet\* | 96.9 | 93.9 | 87.8 | 92.1 | 94.4 | 90.6 | 92.8 | 85.7 | 95.6 | 93.0 |
| HFCNet\* | 95.8 | 93.3 | 86.7 | 92.0 | 93.9 | 90.6 | 92.8 | 85.7 | 95.6 | 93.0 |
| PhosIDN | 96.9 | 93.9 | 87.8 | 92.1 | 94.4 | 90.6 | 92.8 | 85.7 | 95.6 | 93.0 |
| Subfamily  PKCa | Baseline | 65.5 | 78.3 | 57.6 | 85.8 | 74.3 | 55.4 | 75.9 | 55.0 | 90.6 | 68.8 |
| IFENet\* | 65.5 | 78.3 | 57.6 | 85.8 | 74.3 | 59.7 | 77.9 | 58.6 | 91.2 | 72.2 |
| HFCNet\* | 66.2 | 78.6 | 58.2 | 86.0 | 74.8 | 56.2 | 76.8 | 56.2 | 91.2 | 69.5 |
| PhosIDN | 75.5 | 83.1 | 66.5 | 87.5 | 81.1 | 59.7 | 77.9 | 58.6 | 91.2 | 72.2 |
| Subfamily  SrcA | Baseline | 68.8 | 78.1 | 58.9 | 89.8 | 77.9 | 46.9 | 68.0 | 46.0 | 92.3 | 62.2 |
| IFENet\* | 75.8 | 82.0 | 65.4 | 90.7 | 82.6 | 63.3 | 77.2 | 59.7 | 94.2 | 75.7 |
| HFCNet\* | 76.6 | 82.5 | 66.2 | 90.7 | 83.1 | 53.9 | 71.9 | 51.8 | 93.2 | 68.3 |
| PhosIDN | 80.5 | 84.6 | 69.9 | 91.2 | 85.5 | 68.8 | 80.3 | 64.4 | 94.6 | 79.6 |
| Kinase  CDC2 | Baseline | 90.3 | 89.9 | 79.9 | 88.4 | 89.4 | 86.0 | 91.0 | 82.0 | 94.1 | 89.9 |
| IFENet\* | 90.3 | 89.9 | 79.9 | 88.4 | 89.4 | 89.2 | 92.5 | 84.9 | 94.3 | 91.7 |
| HFCNet\* | 91.4 | 90.5 | 80.9 | 88.5 | 89.9 | 88.2 | 92.0 | 83.9 | 94.3 | 91.1 |
| PhosIDN | 95.7 | 92.5 | 85.1 | 89.0 | 92.2 | 90.3 | 93.0 | 85.9 | 94.4 | 92.3 |
| Kinase  CK2a1 | Baseline | 79.8 | 84.9 | 70.0 | 87.7 | 83.5 | 74.2 | 84.9 | 71.0 | 93.0 | 82.5 |
| IFENet\* | 85.4 | 87.6 | 75.2 | 88.4 | 86.9 | 84.3 | 89.8 | 79.8 | 93.8 | 88.8 |
| HFCNet\* | 84.3 | 87.1 | 74.2 | 88.2 | 86.2 | 83.1 | 89.2 | 78.8 | 93.7 | 88.1 |
| PhosIDN | 91.0 | 90.3 | 80.6 | 89.0 | 90.0 | 88.8 | 91.9 | 83.9 | 94.0 | 91.3 |
| Kinase  PKACa | Baseline | 92.6 | 91.1 | 82.2 | 90.6 | 91.6 | 80.9 | 87.8 | 76.6 | 95.0 | 87.4 |
| IFENet\* | 95.7 | 92.8 | 85.6 | 90.9 | 93.3 | 80.9 | 87.8 | 76.6 | 95.0 | 87.4 |
| HFCNet\* | 94.7 | 92.2 | 84.5 | 90.8 | 92.7 | 85.1 | 90.0 | 80.6 | 95.2 | 89.9 |
| PhosIDN | 96.8 | 93.3 | 86.8 | 91.0 | 93.8 | 90.4 | 92.8 | 85.7 | 95.5 | 92.9 |
| Kinase  PKCa | Baseline | 71.6 | 80.4 | 62.3 | 88.6 | 79.2 | 58.9 | 76.3 | 57.6 | 93.3 | 72.2 |
| IFENet\* | 83.7 | 86.7 | 73.6 | 90.1 | 86.8 | 78.0 | 86.3 | 74.0 | 94.8 | 85.6 |
| HFCNet\* | 77.3 | 83.3 | 67.5 | 89.3 | 82.9 | 63.1 | 78.5 | 61.2 | 93.7 | 75.4 |
| PhosIDN | 92.9 | 91.5 | 82.9 | 91.0 | 91.9 | 85.1 | 90.0 | 80.6 | 95.2 | 89.9 |
| Kinase  SRC | Baseline | 66.3 | 77.3 | 57.0 | 87.8 | 75.6 | 41.8 | 67.0 | 43.3 | 91.1 | 57.3 |
| IFENet\* | 84.7 | 87.0 | 74.2 | 90.2 | 87.4 | 59.2 | 76.2 | 57.7 | 93.5 | 72.5 |
| HFCNet\* | 71.4 | 80.0 | 61.6 | 88.6 | 79.1 | 65.3 | 79.5 | 62.8 | 94.1 | 77.1 |
| PhosIDN | 84.7 | 87.0 | 74.2 | 90.2 | 87.4 | 67.3 | 80.5 | 64.6 | 94.3 | 78.6 |

**Table S5.** AUC values (%) of different methods for general phosphorylation site prediction

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sites | PPSP | Netphos3.0 | Musite | MusiteDeep | DeepPhos | DeepPhos-71 | PhosIDNSeq | PhosIDN |
| S,T | 74.2 | 63.2 | 57.9 | 77.6 | 80.4 | 80.2 | 81.1 | 82.2 |
| Y | 64.1 | 56.1 | 54.0 | 66.6 | 71.6 | 71.5 | 72.6 | 74.0 |

DeepPhos-71, DeepPhos with maxmal window size of 71

**Table S6.** The values (%) of Sn, Acc, MCC, Pre and F1 of different methods for general phosphorylation site prediction

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sites | Method | Sp=90% | | | | | Sp=95% | | | | |
| Sn | Acc | Mcc | Pre | F1 | Sn | Acc | Mcc | Pre | F1 |
| S,T | PPSP | 32.4 | 61.2 | 27.4 | 76.4 | 45.5 | 17.6 | 56.3 | 19.9 | 77.9 | 28.8 |
|  | Netphos3.0 | 24.2 | 57.1 | 18.8 | 70.7 | 36.0 | 13.9 | 54.5 | 15.3 | 73.7 | 23.4 |
|  | Musite | 20.1 | 57.9 | 24.3 | 82.7 | 32.3 | 20.1 | 57.9 | 24.3 | 82.7 | 32.3 |
|  | MusiteDeep | 44.4 | 67.2 | 38.7 | 81.6 | 57.5 | 31.0 | 63.0 | 33.9 | 86.1 | 45.6 |
|  | DeepPhos | 47.8 | 68.9 | 41.7 | 82.7 | 60.6 | 33.9 | 64.4 | 36.5 | 87.1 | 48.8 |
|  | DeepPhos-71 | 47.3 | 68.7 | 41.3 | 82.6 | 60.2 | 33.5 | 64.2 | 36.1 | 87.0 | 48.3 |
|  | PhosIDNSeq | 49.0 | 69.5 | 42.8 | 83.1 | 61.6 | 34.7 | 64.8 | 37.2 | 87.4 | 49.6 |
|  | PhosIDN | 52.1 | 71.1 | 45.5 | 83.9 | 64.3 | 37.9 | 66.4 | 40.1 | 88.3 | 53.0 |
| Y | PPSP | 23.2 | 56.6 | 17.7 | 69.9 | 34.8 | 12.5 | 53.7 | 13.2 | 71.3 | 21.2 |
| Netphos3.0 | 15.5 | 52.7 | 8.0 | 60.5 | 24.6 | 8.5 | 51.8 | 7.1 | 63.2 | 15.0 |
| Musite | 11.9 | 53.5 | 12.5 | 70.7 | 20.4 | 11.9 | 53.5 | 12.5 | 70.7 | 20.4 |
| MusiteDeep | 25.0 | 57.5 | 19.7 | 71.4 | 37.0 | 17.7 | 56.4 | 20.0 | 78.0 | 28.9 |
| DeepPhos | 30.1 | 60.1 | 25.1 | 75.1 | 43.0 | 18.6 | 56.8 | 21.1 | 78.8 | 30.1 |
| DeepPhos-71 | 30.3 | 60.1 | 25.3 | 75.2 | 43.2 | 18.4 | 56.7 | 20.8 | 78.6 | 29.8 |
| PhosIDNSeq | 31.5 | 60.8 | 26.5 | 75.9 | 44.5 | 19.6 | 57.3 | 22.2 | 79.6 | 31.5 |
| PhosIDN | 33.7 | 61.9 | 28.7 | 77.2 | 47.0 | 21.1 | 58.0 | 23.9 | 80.9 | 33.5 |

**Table S7.** AUC values (%) of different methods for kinase-specific phosphorylation site prediction

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kinase | | GPS | PPSP | MusiteDeep | DeepPhos | PhosphoPredict | PhosIDNSeq | PhosIDN |
| Subfamily | CDC2 | 91.1 | 82.1 | - | 93.6 | - | 95.8 | 97.2 |
|  | CDK2 | 92.4 | 81.6 | - | 95.9 | - | 96.1 | 97.9 |
|  | ERK1 | 93.3 | 84.4 | - | 95.9 | - | 97.6 | 98.6 |
|  | PKCa | 71.0 | 75.8 | - | 86.3 | - | 89.4 | 92.1 |
|  | SrcA | 76.8 | 66.3 | - | 82.4 | - | 85.9 | 93.3 |
| Individual kinase | PKCa | 70.7 | 70.7 | - | 90.7 | - | 94.5 | 97.1 |
| CDC2 | 91.1 | 79.8 | - | 96.4 | - | 95.8 | 97.2 |
| CK2a1 | 84.4 | 87.0 | - | 93.3 | - | 94.1 | 96.7 |
| PKACa | 88.7 | 83.5 | - | 96.2 | - | 96.9 | 98.5 |
| SRC | 75.7 | 64.3 | - | 82.7 | 88.0 | 85.4 | 93.3 |

**Table S8.** The values (%) of Sn, Acc, MCC, Pre and F1 of different methods for kinase-specific phosphorylation site prediction at high stringency level

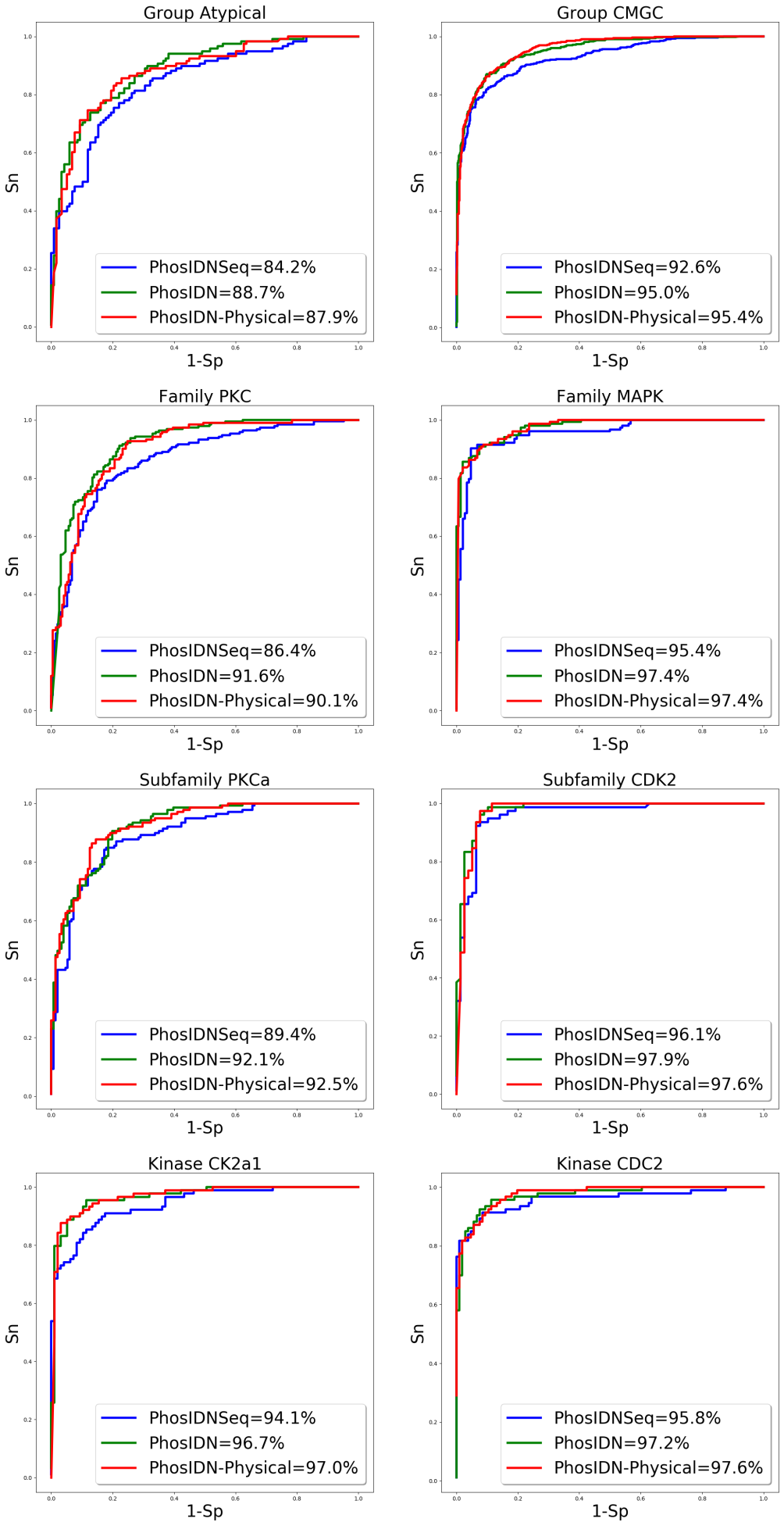
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Kinase | Method | Sn | Acc | Mcc | Pre | F1 |
| Group | GPS | 9.8 | 48.4 | 9.3 | 70.8 | 17.3 |
| CAMK | PPSP | 20.8 | 54.4 | 23.1 | 83.7 | 33.3 |
|  | DeepPhos | 57.8 | 74.7 | 55.7 | 93.5 | 71.4 |
|  | PhosIDNSeq | 68.2 | 80.4 | 64.5 | 94.4 | 79.2 |
|  | PhosIDN | 70.8 | 82.1 | 66.9 | 94.6 | 81.0 |
| Group | GPS | 50.1 | 72.5 | 50.4 | 91.0 | 64.6 |
| CMGC | PPSP | 29.9 | 62.3 | 32.8 | 85.8 | 44.4 |
|  | DeepPhos | 67.2 | 81.1 | 64.8 | 93.1 | 78.1 |
|  | PhosIDNSeq | 75.2 | 85.0 | 71.6 | 93.8 | 83.5 |
|  | PhosIDN | 77.7 | 86.3 | 73.8 | 94.0 | 85.1 |
| Group | GPS | 5.6 | 48.8 | 1.8 | 55.6 | 10.2 |
| TK | PPSP | 16.1 | 54.3 | 18.3 | 78.2 | 26.7 |
|  | DeepPhos | 40.1 | 66.7 | 41.8 | 89.9 | 55.4 |
|  | PhosIDNSeq | 40.5 | 66.9 | 42.9 | 90.0 | 55.8 |
|  | PhosIDN | 65.5 | 79.8 | 63.1 | 93.6 | 77.1 |
| Family  CDK | GPS | 64.3 | 79.1 | 61.8 | 93.2 | 76.1 |
| PPSP | 46.0 | 69.7 | 46.8 | 90.7 | 61.1 |
| MusiteDeep | 86.4 | 90.5 | 81.5 | 94.9 | 90.4 |
| DeepPhos | 87.3 | 91.0 | 82.3 | 94.9 | 90.9 |
| PhosIDNSeq | 89.7 | 91.2 | 84.6 | 95.0 | 92.3 |
| PhosphoPredict | 85.9 | 90.3 | 81.0 | 94.8 | 90.2 |
| PhosIDN | 93.0 | 93.9 | 87.9 | 95.2 | 94.1 |
| Family  CK2 | GPS | 44.8 | 68.9 | 45.3 | 90.3 | 59.9 |
| PPSP | 48.8 | 71.0 | 48.7 | 91.0 | 63.5 |
| MusiteDeep | 72.8 | 83.4 | 68.9 | 93.8 | 82.0 |
| DeepPhos | 76.0 | 85.1 | 71.7 | 94.1 | 84.1 |
| PhosphoPredict | 83.2 | 86.8 | 78.3 | 94.5 | 88.5 |
| PhosIDNSeq | 72.0 | 83.0 | 68.2 | 93.8 | 81.5 |
| PhosIDN | 84.0 | 89.2 | 79.0 | 94.6 | 89.0 |
| Family  MAPK | GPS | 79.1 | 87.0 | 75.2 | 94.5 | 86.1 |
| PPSP | 36.6 | 65.5 | 39.2 | 88.9 | 51.9 |
| MusiteDeep | 79.7 | 87.4 | 75.8 | 94.6 | 86.5 |
| DeepPhos | 83.7 | 89.4 | 79.3 | 94.8 | 88.9 |
| PhosphoPredict | 84.3 | 89.7 | 79.9 | 94.9 | 89.3 |
| PhosIDNSeq | 79.7 | 87.4 | 75.8 | 94.6 | 86.5 |
| PhosIDN | 87.6 | 91.4 | 83.0 | 95.0 | 91.2 |
| Family  PKC | GPS | 14.6 | 54.9 | 15.8 | 73.7 | 24.4 |
| PPSP | 20.3 | 57.8 | 22.8 | 79.6 | 32.4 |
| MusiteDeep | 33.3 | 64.3 | 35.8 | 86.5 | 48.1 |
| DeepPhos | 36.5 | 65.8 | 38.6 | 87.5 | 51.5 |
| PhosphoPredict | 66.7 | 76.2 | 64.2 | 92.8 | 77.6 |
| PhosIDNSeq | 36.0 | 65.5 | 38.2 | 87.3 | 50.9 |
| PhosIDN | 66.7 | 76.2 | 64.2 | 92.8 | 77.6 |
| Family  Src | GPS | 7.1 | 51.6 | 4.7 | 58.8 | 12.7 |
| PPSP | 8.6 | 52.3 | 7.3 | 63.2 | 15.1 |
| DeepPhos | 35.0 | 65.4 | 37.8 | 87.5 | 50.0 |
| PhosIDNSeq | 44.3 | 70.0 | 45.9 | 89.9 | 59.3 |
| PhosIDN | 59.4 | 79.5 | 57.9 | 92.1 | 72.2 |
| Subfamily  CDC2 | GPS | 80.7 | 88.4 | 77.3 | 93.8 | 86.7 |
| PPSP | 21.5 | 60.8 | 25.3 | 80.0 | 33.9 |
| DeepPhos | 86.0 | 90.9 | 82.0 | 94.1 | 89.9 |
| PhosIDNSeq | 89.3 | 92.5 | 84.9 | 94.3 | 91.7 |
| PhosIDN | 92.5 | 94.0 | 87.9 | 94.5 | 93.5 |
| Subfamily  CDK2 | GPS | 61.5 | 78.2 | 59.8 | 92.3 | 73.8 |
| PPSP | 30.8 | 62.8 | 33.4 | 85.7 | 45.3 |
| DeepPhos | 65.4 | 80.1 | 63.1 | 92.7 | 76.7 |
| PhosIDNSeq | 68.0 | 81.4 | 65.2 | 93.0 | 78.5 |
| PhosIDN | 91.0 | 92.9 | 86.0 | 94.7 | 92.8 |
| Subfamily  ERK1 | GPS | 68.8 | 81.1 | 65.5 | 94.3 | 79.5 |
| PPSP | 32.3 | 61.7 | 34.7 | 88.6 | 47.3 |
| DeepPhos | 80.2 | 87.2 | 75.6 | 95.1 | 87.0 |
| PhosIDNSeq | 87.5 | 91.1 | 82.5 | 95.5 | 91.3 |
| PhosIDN | 90.6 | 92.8 | 85.7 | 95.6 | 93.0 |
| Subfamily  PKCa | GPS | 23.7 | 60.7 | 26.4 | 80.5 | 36.7 |
| PPSP | 22.3 | 60.0 | 24.9 | 79.5 | 34.8 |
| DeepPhos | 41.0 | 68.9 | 42.8 | 87.7 | 55.9 |
| PhosIDNSeq | 43.9 | 70.3 | 45.3 | 88.4 | 58.7 |
| PhosIDN | 59.7 | 77.9 | 58.6 | 91.2 | 72.2 |
| Subfamily  SrcA | GPS | 11.7 | 48.3 | 11.8 | 75.0 | 20.3 |
| PPSP | 14.8 | 50.0 | 15.9 | 79.2 | 25.0 |
| DeepPhos | 39.8 | 64.0 | 40.1 | 91.0 | 55.4 |
| PhosIDNSeq | 47.7 | 68.4 | 46.7 | 92.4 | 62.9 |
| PhosIDN | 68.8 | 80.3 | 64.4 | 94.6 | 79.6 |
| Kinase  CDC2 | GPS | 80.7 | 88.4 | 77.3 | 93.8 | 86.7 |
| PPSP | 30.1 | 64.8 | 34.1 | 84.8 | 44.4 |
| DeepPhos | 88.1 | 91.9 | 83.9 | 94.2 | 91.1 |
| PhosIDNSeq | 83.9 | 90.0 | 80.1 | 94.0 | 88.6 |
| PhosIDN | 90.3 | 93.0 | 85.9 | 94.4 | 92.3 |
| Kinase  CK2a1 | GPS | 73.0 | 84.4 | 70.0 | 92.9 | 81.8 |
| PPSP | 43.8 | 70.4 | 45.5 | 88.6 | 58.7 |
| DeepPhos | 74.2 | 85.0 | 71.0 | 93.0 | 82.5 |
| PhosIDNSeq | 74.2 | 85.0 | 71.0 | 93.0 | 82.5 |
| PhosIDN | 88.8 | 91.9 | 83.9 | 94.0 | 91.3 |
| Kinase  PKACa | GPS | 75.5 | 85.0 | 71.8 | 94.7 | 84.0 |
| PPSP | 15.9 | 53.9 | 18.4 | 78.9 | 26.6 |
| DeepPhos | 80.9 | 87.7 | 76.6 | 95.0 | 87.4 |
| PhosIDNSeq | 81.9 | 88.3 | 77.6 | 95.1 | 88.0 |
| PhosIDN | 90.4 | 92.8 | 85.7 | 95.5 | 92.9 |
| Kinase  PKCa | GPS | 8.5 | 50.0 | 7.7 | 66.7 | 15.1 |
| PPSP | 14.2 | 52.9 | 16.1 | 76.9 | 23.9 |
| DeepPhos | 60.3 | 77.0 | 58.8 | 93.4 | 73.3 |
| PhosIDNSeq | 76.6 | 82.5 | 72.7 | 94.7 | 84.7 |
| PhosIDN | 85.1 | 90.0 | 80.6 | 95.2 | 89.9 |
| Kinase  SRC | GPS | 45.9 | 69.2 | 46.7 | 91.8 | 61.2 |
| PPSP | 3.1 | 46.5 | 4.0 | 42.9 | 5.7 |
| DeepPhos | 54.1 | 73.5 | 53.5 | 92.9 | 68.4 |
| PhosphoPredict | 41.8 | 67.0 | 43.3 | 91.1 | 57.3 |
| PhosIDNSeq | 42.9 | 67.6 | 44.2 | 91.3 | 58.3 |
| PhosIDN | 67.3 | 80.5 | 64.6 | 94.3 | 78.6 |

**Table S9.** AUC values of PhosIDN on well- and poorly-annotated kinase families

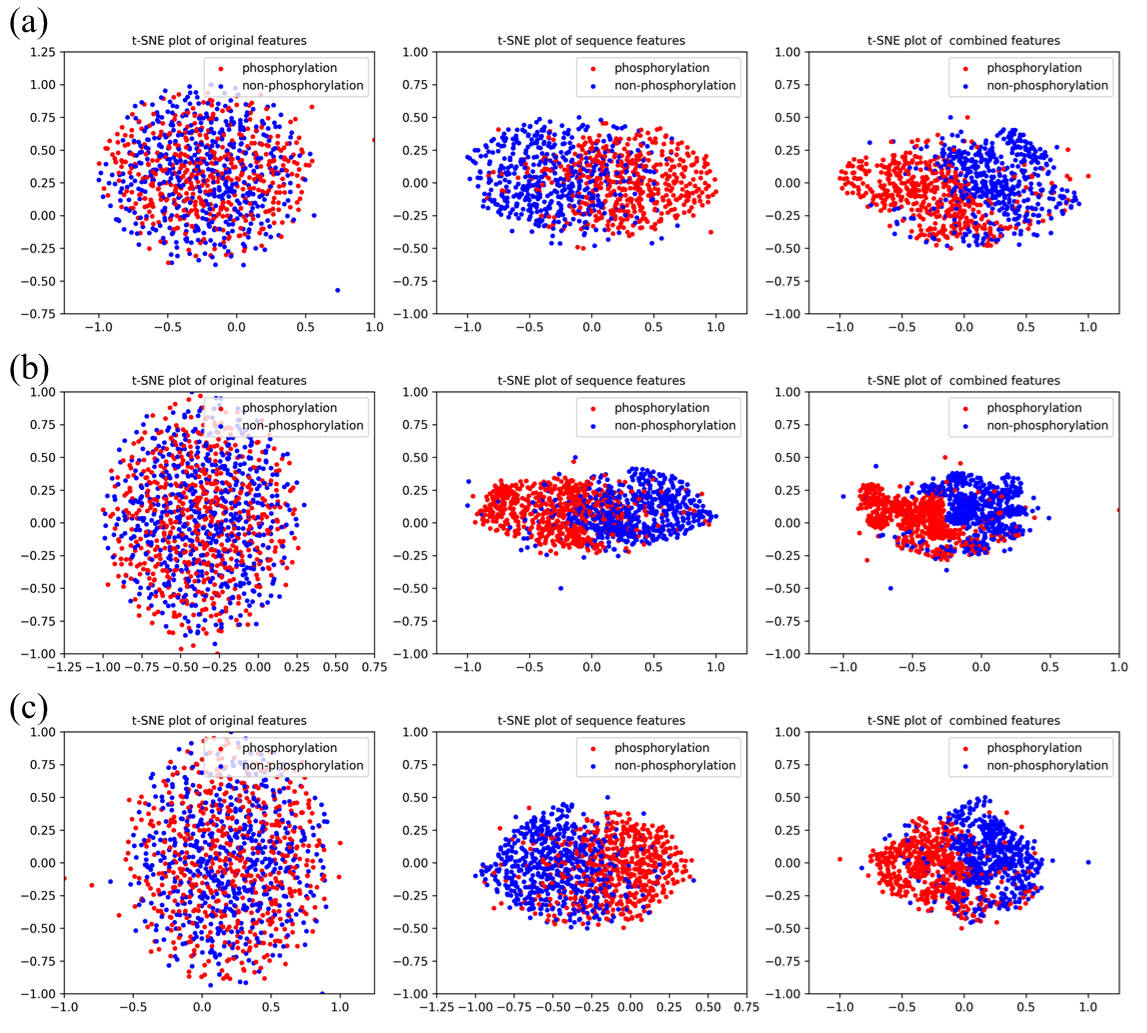
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Well-annotated (>200 sites) | | | Poorly-annotated (<100 sites) | | |
| Kinase family | Number of sites in independent test data | AUC(%) | Kinase family | Number of sites in independent test data | AUC(%) |
| CDK | 412 | 98.2 | Aur | 92 | 96.5 |
| CK2 | 241 | 97.0 | CAMK2 | 57 | 95.1 |
| MAPK | 301 | 97.4 | CK1 | 82 | 94.1 |
| PKC | 386 | 91.6 | GRK | 44 | 95.3 |
| Src | 283 | 90.3 | Abl | 62 | 90.1 |
| Average | 325 | 94.9 | Average | 67 | 94.2 |

**Table S10.** AUC values of PhosIDN on different kinases in a kinase family

|  |  |  |  |
| --- | --- | --- | --- |
| Family | Kinase | Number of sites in independent test data | AUC(%) |
| Src | SRC | 185 | 93.3 |
| HCK | 15 | 90.8 |
| FYN | 51 | 91.2 |
| LCK | 42 | 90.7 |
| PKC | PKCa | 270 | 97.1 |
| PKCg | 19 | 90.5 |
| PKCd | 60 | 91.2 |
| PKCe | 27 | 91.2 |
| PKCz | 28 | 95.8 |



**Figure S1.** ROC curves with different PPI information for kinase-specific phosphorylation site prediction. PhosIDN-Physical refers to PhosIDN using physical interactions from BioGRID database.



**Figure S2.** Visualization of original one-hot encoding features, sequence features extracted by PhosIDNSeq and combined features extracted by PhosIDN. The red dot represents the phosphorylation sites with kinase annotation belonging to (a) family CK2, (b) family MAPK or (c) family Src, the blue dot represents the non-phosphorylation sites