Table S2. The performance of profile HMMs trained with MDDLogo-identified substrate motifs based on five-fold cross-validation

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MDDLogo-clustered subgroup** | **Entropy plot** | **Number of positive data** | **Number of negative data** | **Sn** | **Sp** | **Acc** | **MCC** | **AUC** |
| All carbonylated K sites |  | 256 | 512 | 0.72 | 0.71 | 0.71 | 0.41 | 0.81 |
| CarbK\_1 |  | 54 | 108 | 0.81 | 0.81 | 0.81 | 0.61 | 0.91 |
| CarbK\_2 |  | 52 | 104 | 0.74 | 0.74 | 0.74 | 0.46 | 0.84 |
| CarbK\_3 |  | 35 | 70 | 0.71 | 0.71 | 0.71 | 0.41 | 0.81 |
| CarbK\_4 |  | 33 | 66 | 0.77 | 0.82 | 0.80 | 0.59 | 0.89 |
| CarbK\_5 |  | 26 | 52 | 0.69 | 0.75 | 0.73 | 0.43 | 0.82 |
| CarbK\_6 |  | 33 | 66 | 0.76 | 0.76 | 0.76 | 0.49 | 0.85 |
| CarbK\_7 |  | 23 | 46 | 0.74 | 0.74 | 0.74 | 0.45 | 0.84 |
| All carbonylated R sites |  | 115 | 230 | 0.68 | 0.66 | 0.67 | 0.33 | 0.73 |
| CarbR\_1 |  | 56 | 92 | 0.76 | 0.74 | 0.75 | 0.47 | 0.85 |
| CarbR\_2 |  | 33 | 86 | 0.70 | 0.68 | 0.69 | 0.35 | 0.75 |
| All carbonylated T sites |  | 109 | 218 | 0.70 | 0.69 | 0.69 | 0.37 | 0.77 |
| CarbT\_1 |  | 46 | 92 | 0.74 | 0.74 | 0.74 | 0.45 | 0.84 |
| CarbT\_2 |  | 38 | 76 | 0.75 | 0.75 | 0.75 | 0.48 | 0.85 |
| All carbonylated P sites |  | 109 | 218 | 0.72 | 0.72 | 0.72 | 0.42 | 0.82 |
| CarbP\_1 |  | 43 | 86 | 0.81 | 0.81 | 0.81 | 0.61 | 0.90 |
| CarbP\_2 |  | 22 | 44 | 0.73 | 0.73 | 0.73 | 0.43 | 0.82 |
| CarbP\_3 |  | 21 | 42 | 0.70 | 0.68 | 0.69 | 0.34 | 0.73 |