Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(5): 6410–6429.
DOI: 10.3934/mbe.2021318
Received: 17 May 2021
Accepted: 06 July 2021
Published: 23 July 2021

*Research article*

# TYLER, a fast method that accurately predicts cyclin-dependent proteins by using computation-based motifs and sequence-derived features

**Jian Zhang[1,*], Xingchen Liang[1], Feng Zhou[1], Bo Li[2] and Yanling Li[1]**

[1] School of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China
[2] College of Electronic Science and Engineering, Jilin University, Changchun 130012, China

* **Correspondence:** Email: jianzhang@xynu.edu.cn.

## Supplementary

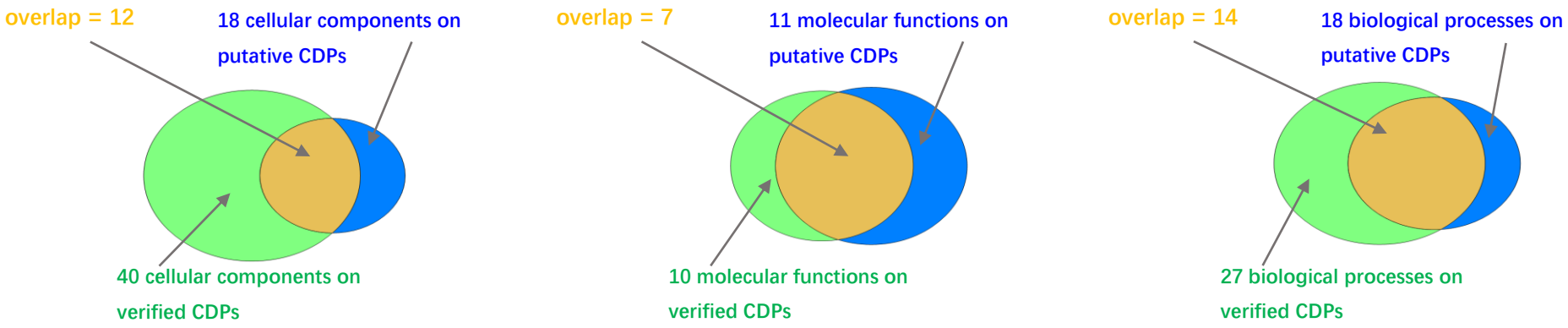


**Figure S1**. Each set and the overlap of significantly enriched subcellular locations and functions in verified and putative CDPs.

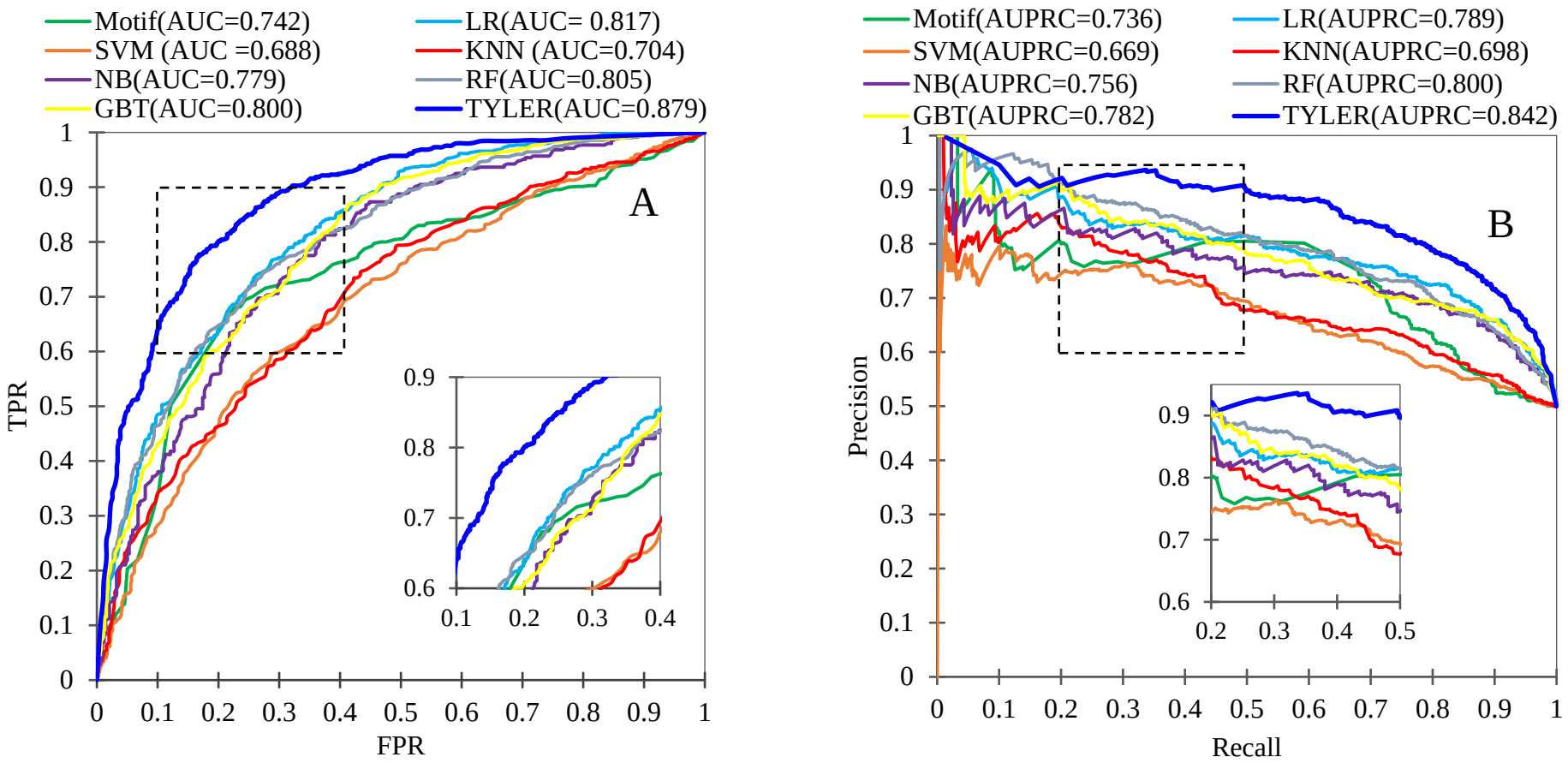


**Figure S2.** ROC curves and PRC curves for different methods on the TRAINING dataset.

**Table S1.** Description of features that are utilized for construction of the machine learning model.

| Feature type | Description | Number of features | Number of features per feature type |
|---|---|---|---|
| Evolutionary profile | Normalized position-specific substitution scores of hydroxylic residues that substituted by same group of hydroxylic residues | 11 | 121 |
| | Normalized position-specific substitution scores of tiny residues that substituted by same group of tiny residues | 11 | |
| | Normalized position-specific substitution scores of small residues that substituted by same group of small residues | 11 | |
| | Normalized position-specific substitution scores of acidic residues that substituted by same group of acidic residues | 11 | |
| | Normalized position-specific substitution scores of positive residues that substituted by same group of positive residues | 11 | |
| | Normalized position-specific substitution scores of polar residues that substituted by same group of polar residues | 11 | |
| | Normalized position-specific substitution scores of charged residues that substituted by same group of charged residues | 11 | |
| | Normalized position-specific substitution scores of hydrophobic residues that substituted by same group of hydrophobic residues | 11 | |
| | Normalized position-specific substitution scores of aromatic residues that substituted by same group of aromatic residues | 11 | |
| | Normalized position-specific substitution scores of sulphur residues that substituted by same group of sulphur residues | 11 | |
| | Normalized position-specific substitution scores of aliphatic residues that substituted by same group of aliphatic residues | 11 | |
| | Normalized position-specific substitution scores of hydrophobic residues that substituted by same group of hydrophobic residues | 11 | |
| Secondary structure | Fraction of residues located on the helix/sheet/coil of the protein | 3 | 30 |
| | Fraction of helix/sheet/coil segments within the protein | 3 | |
| | Fraction of residues in a given secondary structure motif type in the sequence (XHE, XHC, XEH, XEC, XCH, XCE, HEX, HCX, EHX, ECX, EHX, ECX, CHC, CHE, EHC, EHE, HCH, ECH, HCE, ECE, CEC, HEC, CEH, HEH) | 24 | |
| Physicochemical Properties | Fraction of {hydroxylic, tiny, small, acidic, positive, polar, charged, hydrophobic, aromatic, sulphur, aliphatic} residues within the protein | 11 | 33 |
| | Fraction of {hydroxylic, tiny, small, acidic, positive, polar, charged, hydrophobic, aromatic, sulphur, aliphatic} residues within longest secondary structure segment | 11 | |
| | Fraction of {hydroxylic, tiny, small, acidic, positive, polar, charged, hydrophobic, aromatic, sulphur, aliphatic} residues within longest secondary structure motif | 11 | |
| TOTAL | | | 184 |

**Table S2.** Selective enriched motifs in CDPs. The motifs are ranked by descending order of IGR.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | ADFGLAR | 16 | S-H\|-[T\|Q\|R] S[E\|Q\|N] | 31 | LE-CA | 46 | GC-I-F |
| 2 | YR-PELLL | 17 | S-FK-S | 32 | L [I\| M] [D\|G] [W\|H] | 47 | FESP |
| 3 | Y-[D-\|K-C] | 18 | RP[S\|P\|R\|Y] [P\|T\|S\|N\|R] [P\|T\|S\|R] | 33 | L[A\|D] [C\|R] [K\|F] L | 48 | EY-R-H |
| 4 | W[L-\|LI-] E-V | 19 | RLNL | 34 | L-W-Y-R | 49 | ES-Q-S |
| 5 | W-Y-R-P | 20 | QSS-P | 35 | L-KPQ | 50 | EME-D |
| 6 | V\|-V-V-T-WY\|-R\|-P | 21 | QL-LC | 36 | K[K\|L] [I\|C] [K\|G] | 51 | E-S[DP\|PT] |
| 7 | TR-SC | 22 | Q[K\|T] [V\|P] [Q\|K] | 37 | K-S[PP\|TR] | 52 | E-[E\|I] W [R\|P] |
| 8 | T-\|LWY | 23 | Q-Q-H-H | 38 | K-R[RI\|SN] | 53 | DR[F\|Y]LS |
| 9 | SSH-[S\|T] | 24 | P[G\|L\|Q] [P\|K\|H] [E\|P\|Q] | 39 | K-I-ADFG-AR | 54 | DI-TN |
| 10 | SRRR | 25 | P-SPK | 40 | ISP-K | 55 | DFGL-A |
| 11 | SPV-K | 26 | NNEN | 41 | IQN-S | 56 | DEDEE |
| 12 | SP-F\|-[P\|N] K | 27 | N[I\|Y]-D[T\|R] | 42 | I-D-M-W | 57 | D-M-W-[S\|VGC] |
| 13 | SN-QN | 28 | MHRD | 43 | HS-KR | 58 | CIF-[A\|E] |
| 14 | SLSS-S | 29 | LK-PQ | 44 | H-R-D-L-K | 59 | C[C\|L]-[L\|H]C |
| 15 | S-L-SS-S | 30 | LH-R-D | 45 | GDDD | 60 | C-E-FS |