User manual for ProteinMatrix

ProteinMatrix1.2, released on Sep 30, 2017

The ProteinMatrix program was designed for the prediction of protein properties, such as protein residue depth and secondary structures.

Please address comments/bug-reports to Dr. Yan (yanrenxiang@fzu.edu.cn).

There are two steps to run the program for protein zinc-binding site prediction.
First, PSSM profile generated by PSI-BLAST as
blastpgp -b 1000 -j 3 -h 0.001 -d nr -i seq.fasta -C psitmp.chk -Q pssm.txt

Second, input the PSSM file to ProteinMatrix and the prediction will be generated as
java -jar ProteinMatrix.jar pssm.txt zinc
Here, pssm.txt is an input PSSM file and zinc is the flag of the running program.

Only those residues annotated as * in the output file were predicted, Label:1/Zn-binding site Label:0/non-zn-binding site. One output file is as

#	AA	Label	CScore	
1	E	1	0.999	*
2	H	1	0.991	*
3	Q	0	0.000	
4	G	0	0.000	
5	A	0	0.000	
6	L	0	0.000	
7	V	0	0.000	
8	C	0	0.792	*
9	L	0	0.000	
10	A	0	0.000	
11	E	0	0.355	*
12	S	0	0.000	
13	C	0	0.265	*
14	L	0	0.000	
15	T	0	0.000	
16	L	0	0.000	
17	E	0	1.000	*
18	G	0	0.000	
19	P	0	0.000	
.	.	.	.	
.	.	.	.	

where CScore is the confident score for the prediction. In the test datasets, the prediction score is reliable if CScore is higher than 0.5.

There are two steps to run the program for protein feature prediction.
First, PSSM profile generated by PSI-BLAST as
blastpgp -b 1000 -j 3 -h 0.001 -d nr -i seq.fasta -C psitmp.chk -Q pssm.txt

Second, input the PSSM file to ProteinMatrix and the prediction will be generated as
java -jar ProteinMatrix.jar A2T3Q0.pssm pssm2features
Here, pssm.txt is an input PSSM file and pssm2features is the flag of the running program.

The predictions are
AA SE RE JSD SS SS-H SS-E SS-C SA Rdepth
AA: amino acids
SE: Shannon entropy-based conservation measure
RE: Relative entropy-based conservation measure
JSD: Jensen-Shannon divergence conservation score
SS: secondary structure. H: Helix E: strand C: coil
SS-H: probability for alpha-helix
SS-E: probability for beta-strand
SS-C: probability for coil
SA: relative solvent accessibility
nonTM: probability for non-TM region
TM: probability for TM region
Rdepth: relative residue depth

References

1. Renxiang Yan, Xiaofeng Wang, Lanqing Huang, Yarong Tian and Weiwen Cai (2017) Transmembrane region prediction by using sequence-derived features and machine learning methods，***RSC Advances***, 29200-29211
2. Renxiang Yan, Xiaofeng Wang, Weiming Xu, Weiwen Cai, Juan Lin, Jian Li, Jiangning Song (2016), A neural network learning approach for improving the prediction of residue depth based on sequence-derived features, ***RSC Advances***, 67729-67738
3. Renxiang Yan, Xiaofeng Wang, Lanqing Huang, Feidi Yan, Xiaoyu Xue, and Weiwen Cai (2015) Prediction of structural features and application to outer membrane protein identification. ***Scientific Reports***, 5, 11586.
4.Renxiang Yan, Jiangning Song, Weiwen Cai and Ziding Zhang (2015) A short review on protein secondary structure prediction methods. ***Pattern Recognition in Computational Molecular Biology: Techniques and Approaches*** (Wiley Series in Bioinformatics). Chapter 6. DOI: 11.1002/9781119078845.ch6, ISBN: 978-1-118-89368-5, 99-112.
5. Renxiang Yan, Xiaofeng Wang, Lanqing Huang, Jun Lin, Weiwen Cai and Ziding Zhang (2014) GPCRserver: an accurate and novel G protein-coupled receptor predictor. ***Mol Biosyst***, 10:2495-2504.