Special Protein Molecules Computational Identification

Quan Zou * ⬤ and Wenying He

School of Computer Science and Technology, Tianjin University, Tianjin 300354, China; hwying1234@tju.edu.cn
* Correspondence: zouquan@tju.edu.cn

Abstract: Computational identification of special protein molecules is a key issue in understanding protein function. It can guide molecular experiments and help to save costs. I assessed 18 papers published in the special issue of *Int. J. Mol. Sci.*, and also discussed the related works. The computational methods employed in this special issue focused on machine learning, network analysis, and molecular docking. New methods and new topics were also proposed. There were in addition several wet experiments, with proven results showing promise. I hope our special issue will help in protein molecules identification researches.

Keywords: bioinformatics; machine learning; feature selection; protein classification; network analysis; molecular docking

1. Introduction

With the development of next generation sequencing technologies, the size of biological databases have increased dramatically in terms of the number of samples. It is fast and cheap to obtain biological sequences but relatively slow and expensive to extract function information because of limitations of traditional biological experimental technologies. Protein, as the product of gene expression and the important material basis of life activity, participates in almost all life activities and biological processes. For some special protein molecules, the detection of new ones is time-consuming and costly. Some special proteins are present, such as cytokines, enzymes, cell-penetrating peptides, anticancer peptides, cancerlectins, and G protein-coupled receptors. In order to save the wet experimental costs, researches first select some candidates through computer programs. The "computer program" is the key step in selecting candidates. High false positive software would lead to high spending on the validation process.

In this special issue, these "computer program" approaches and algorithms are discussed. Numerous sequence-based "golden features" have been proposed for these problems, such as Chou's PseAAC. Ever since the concept of PseAAC was proposed, it has penetrated into nearly all fields of protein identification. However, it is suggested that special features and classification methods should be proposed for special protein molecular. "Golden features" could hardly apply to all kinds of proteins. In this special issue, submissions focused on a kind of special protein molecules, collected related data sets, got better prediction performance (especially low false positive), and developed friendly software tools or web servers.

We received 36 submissions. After rigorous reviewing process, 18 papers were published. They come from different countries, including China, Russia, Canada, Australia, USA, Poland, etc. These papers could be categorized into three subtopics. As shown in Figure 1.
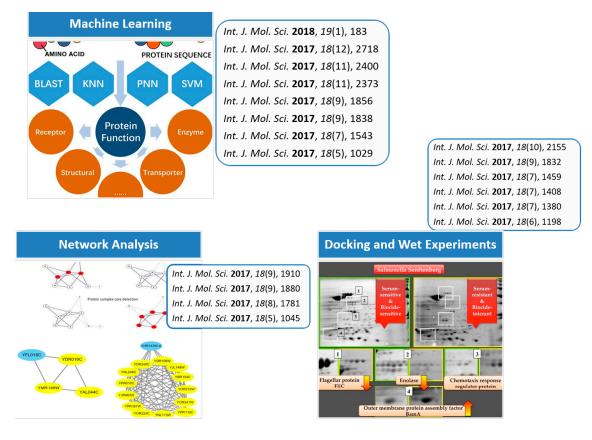
Figure 1. Subtopics of our special issue [38,52,63].

2. Machine Learning Related Researches

2.1. Protein–Protein Interaction Prediction

The first subtopic is to identify or predict protein function with machine learning methods. Two papers focused on protein–protein interaction prediction. Protein–protein interactions (PPIs) play crucial roles in almost all cellular processes. Correctly predicting protein–protein interactions contributes to precise protein function prediction [1,2]. Most of them focus on the PPIs predictions from various data types, including 3D structural information, gene ontology and annotations, and gene fusion. Wang et al. [3] proposed a sequence-based approach (DNN-LCTD) combining deep neural networks (DNN) and Local Conjoint Triad Description (LCTD) feature representation. Experimental results showed that DNN-LCTD is very promising for predicting PPIs. Wang et al. [4] using the Zernike moments (ZM) descriptor on the PSSM combined with Probabilistic Classification Vector Machines (PCVM) classifier developed the PCVMZM predictor for predicting the PPIs from protein amino acids sequences. It was proved to be a robust, powerful and feasible PPI prediction method. Ding et al. [5] developed a random forest algorithm based predictor using a multivariate mutual information feature representation scheme and normalized Moreau-Broto Autocorrelation information from protein sequence. Another work [6] is a novel matrix-based protein sequence representation approach to identify PPIs, using an ensemble learning method for classification. The matrix of Amino Acid Contact (AAC) was constructed based on the statistical analysis of residue-pairing frequencies in a data-set of 6323 protein–protein complexes. The feature vector was extracted by applying algorithms of Histogram of Oriented Gradient (HOG) and Singular Value Decomposition (SVD) on the Substitution Matrix Representation (SMR) matrix of protein sequence.

Drug-target interaction is a special PPI. Because the experimental prediction of drug-target interaction (DTIs) is time-consuming and expensive, computational technology with high accuracy plays a crucial rule in the large-scale rapid prediction of DTIs. Shen et al. [7] proposed DAWN a kind

of Drug-target interactions predictor combining discrete Wavelet transform and Network features. Most importantly, DAWN as a kind of machine learning approach of feature vector-based method, has the desired effect under the condition of without network information. In the same year, they also developed the second tool [8] using molecular substructure fingerprints, Multivariate Mutual Information (MMI) of proteins, and network topology.

Hotspot has important significance in the determination of protein–protein interactions [9]. Many methods have been developed for the hotspot predictions [10,11] and even protein binding site predictions [12]. Most of the works focused on the hotspot predictions from a curated small partial dataset of the whole protein sequences [13]. In Jiang's work [14], the issue of hotspot determination was approached from whole natural protein sequences, and a random projection ensemble system based on k nearest neighbor algorithm to identify hotspot residues by sequence information alone was developed. Experimental results showed that although this method did not perform well enough in the real applications of hotspots, it was very promising in the determination of hotspot residues from whole sequences.

2.2. Special Proteins Identification

Besides protein–protein interaction, DNA binding proteins, ion channel proteins, and amyloids have also attracted researchers' attentions. DNA binding protein is a kind of special protein molecule, whose identification is one of the most important tasks in studying the function of proteins. In this regard, many computational predictors have been proposed [15–21]. In a special issue, Zhang et al. [22] proposed a new approach to extract evolutionary information from the Position Specific Frequency Matrix (PSFM) and incorporate the evolutionary information, and a computational predictor was proposed for DNA binding protein identification. Experimental results showed that this predictor outperformed some existing state-of-the-art approaches in this field. DNA-protein interactions play a key role in a variety of biological processes, especially in cellular metabolism. Endowed with a ditto multi-scale idea in essence, Shen et al. [23] addressed a kind of competitive method called Multi-scale Local Average Blocks (MLAB) algorithm. Different from the structure-based route, MLAB exploited a strategy that not only extracted local evolutionary information from primary sequence, but also used predicted solvent accessibility. Moreover, the construction on the predictor of DNA-protein binding sites wields an ensemble weighted sparse representation model with random under-sampling.

Ion channels are membrane proteins which are widely distributed in all cells. They have been shown to be extensively involved in various physiological and pathological processes, including regulating neuronal and cardiac excitability, muscle contraction, hormone secretion, fluid movement, and immune cell activation. Different ion channels play their unique roles in different biological processes. With the rapid development of next-generation sequencing technologies, the accumulation of proteomic data provides uswith a platform to systematically investigate and predict ion channels and their types. Several studies have focused on the prediction of ion channels and their types [24–26]. The paper published in the special issue [27] proposed a new prediction model to quickly predict ion channels and their types. An improved feature extraction method combining dipeptide composition with the physicochemical property correlation between two residues was developed to formulate protein samples. Subsequently, the analysis of variance (ANOVA) combined with the incremental feature selection (IFS) was employed to find out the optimal features which can produce the maximum accuracy. As a result, authors achieved the overall accuracies of 87.8% for discriminating ion channels from non-ion channels, 94.0% for distinguishing between voltage-gated ion channel and ligand-gated ion channels and 92.6% for four types of voltage-gated ion channels, respectively. Based on the proposed models, a web server called IonchanPred 2.0 (http://lin.uestc.edu.cn/server/IonchanPredv2.0) was established. The free predictor will be most useful to most wet-experimental scholars. A few groups have focused on the outer membrane protein recently, Wang et al. introduced the predicted topology structure as a mainly structure-specific

feature to this classical type of ion channel protein, improved the precision of outer membrane identification [28], inter-barrel contact prediction [29] and fold recognition [30].

In this special issue, Antonets et al. [31] detected amyloidogenic proteins in the proteomes of plants. Amyloids are protein fibrils with characteristic spatial structure. The main computational method for them was phylogenetic analysis together with machine learning techniques. This kind of protein also includes DNA and RNA binding ones, which showed that different kinds of proteins have comment characters. To summarize, effective protein features and machine learning techniques are still essential and challenging in the future.

2.3. Protein Subcellular Localization and Function Analysis

Besides PPI, special proteins, protein subcellular localization, and function prediction are traditional challenges and attract researchers. In general, only when the protein is located in the correct subcellular location, can the protein function normally. Therefore, prediction of protein subcellular localization is an important component of proteomics, and it can aid the identification of drug targets. Due to the technical limitation and high cost of time and money in traditional experimental methods, research on protein subcellular location annotation with the machine learning technique has become a focused research problem in bioinformatics. When we use machine learning technologies to predict protein subcellular location, we need to extract the features of protein sequences, and then use the classifier to realize the protein classification. Thus feature extraction and dimension reduction are important techniques for analyzing the complex and high dimensional biological data in protein subcellular location. In order to improve the prediction accuracy of protein subcellular location, an appropriate algorithm for reducing data dimension should be used before classification. Wang et al. [32] proposed two feature fusion expressions and then used the linear discriminant analysis (LDA) method for dimension reduction. Considering the general nonlinear property in protein sequence data, they [33] introduced the nonlinear kernel discriminant analysis (KDA) method to reduce the high dimensionality in some feature data in this special issue. In this paper, an improved Gauss kernel parameter selection algorithm was proposed to predict subcellular location. It was proposed by maximizing the differences of reconstruction errors between edge normal samples and internal normal samples. The proposed method did not only show the same effect as traditional methods, but also reduced the computational time and improved the efficiency. It should be noted that LDA and KDA methods cannot only reduce the data dimensionality, but also take use of some classification information in the data, resulting in an ideal classification effect. Besides, there have been some new dimensional reduction algorithms which have been tried in other pattern recognition fields, such as face recognition [34].

Knowledge of protein function is the key to the understanding of the biological process and disease development and to the discovery of new therapeutic targets [35]. Various in-silico methods have been developed for protein function prediction [36], which complement one another due to their distinct underlying theory [37]. A comprehensive comparison of the performances between those popular prediction algorithms was conducted based on the information from 93 functional protein families [38], which observed a substantially higher sensitivity of BLAST and a significantly reduced false discovery rate of machine learning.

Since machine learning is a key issue in protein research, it is essential to extract numerical features from the protein primary sequence. Some recent studies showed that evolutionary information and the sequence-order effects are very important for extracting the features of proteins [39,40]. In their special issue, Du et al. [41] developed the UltraPse program to convert biological sequences into digital features. Unlike the PseAAC-Builder [42] or PseAAC-General [43], the UltraPse program can be used on DNA/RNA sequences as well as protein sequences. The program is a good starting point in predicting special protein functional characters, especially the exact subcellular localization of proteins [44].

3. Network Techniques Related Researches

Network analysis is also an important technique for protein identification and function research. Identification of disease genes is very important in medicine. For a disease, extracting its disease genes as completely as possible is helpful in understanding its pathogenesis, thereby designing effective treatments. To date, several network methods have been proposed to identify genes related to different diseases, such as the guilt by association (GBA) based method [45], the shortest path algorithm based method [46–48], the flow propagation algorithm [49], and the random walk with restart (RWR) algorithm based method [50]. In view of the fact that the RWR algorithm can make full use of the whole network, a RWR algorithm based method was proposed by Lu et al. [51] to identify disease genes of uveitis, a serious eye disease that may cause blindness in both young and middle-aged people. The method first applied the RWR algorithm on a protein–protein interaction (PPI) network using validated uveitis-related genes as seed nodes. Second, the obtained genes were filtered by a permutation test that can exclude false positive genes produced by the PPI network. Finally, they extracted important genes from the remaining genes by evaluating their associations to validate genes. Several putative genes were accessed and some have been determined to be important for the pathogenesis of uveitis.

Li et al. [52] employed the advanced network clustering algorithm for protein complex identification. Their method could detect the overlapping complex from the PPI network. Cluster analysis of biological networks is an important topic in systems biology. Up to now, a number of computational methods and tools have been proposed for analyzing biological networks and identifying protein complexes [53]. Various plugins based on cytoscape, such as CytoNCA [54], ClusterViz [55], DyNetViewer [56], CytoCtrlAnalyser [57], were developed to analyze biological networks from different perspectives. CytoCluster [58] in our special issue is a popular clustering tool which integrates six clustering algorithms and BinGO function. Since it was established in July 2013, CytoCluster has been downloaded more than 11,200 times from the Cytoscape App Store and has been applied to different biological networks analyses.

4. Docking and Wet Experiments Researches

Docking is still an interesting and hot topic in protein structure and function analysis, especially in the drug design process. Adenosine monophosphate-activated protein kinase (AMPK) plays a critical role in the regulation of energy metabolism. Huang et al. [59] employed molecular docking to get potential β1-selective AMPK activators. Finally, 12 novel compounds were selected as potential starting points for the design of direct β1-selective AMPK activators. Hou et al. [60] investigated the relationship between scopoletin structure and *TcPMCA1*(a gene name)-inhibiting activity of scopoletin and other 30 coumarin derivatives by employing docking and three-dimensional quantitative structure-activity relationships (3D-QSAR). This work offers additional insights into the mechanism underlying the interaction of scopoletin with *TcPMCA1* gene. Together with this work, the other three works in this special issue also carried out wet experiments. Besides wet experiments, Ding et al. [61] completed bioinformatics analysis and molecular dynamics simulation on glucose 1-dehydrogenase (GDH). Chandler et al. [62] extracted insulin-binding protein and insulin-like peptides in the Eastern spiny lobster, *Sagmariasus verreauxi*. Molecular modelling, including docking, showed various interaction and regulation. Futoma-Koloch et al. [63] laid special stress on analyzing the relationship between triamine-biocide tolerance of *Salmonella enterica* serovar Senftenberg with antimicrobial susceptibility, serum resistance, and outer membrane proteins.

To conclude, papers in this special issue cover several emerging topics of computational identification and bioinformatics analysis of special protein molecules. We fervently hope that this particular issue will attract considerable interest in the relevant fields. We are grateful to *Int. J. Mol. Sci.* for providing the chance to organize this special issue. We also thank the reviewers for their efforts in guaranteeing the high quality of this special issue. Finally, we thank all those who contributed to this special issue. *Int. J. Mol. Sci.* has promised to continue with the same topic as a new special issue in

2018. Besides special protein molecules, nucleic acids with special modifications identification (such as RNA m6A [64], protein phosphorylation [65] and methylation, etc.) will also be welcomed in the 2018 special issue. I hope more authors and readers will contribute, especially to the follow-up works from this special issue.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zeng, J.; Zou, Q.; Wu, Y.; Li, D.; Liu, X. An Empirical Study of Features Fusion Techniques for Protein–protein Interaction Prediction. *Curr. Bioinform.* **2016**, *11*, 4–12. [CrossRef]
2. Garzón, J.I.; Deng, L.; Murray, D.; Shapira, S.; Petrey, D.; Honig, B. A computational interactome and functional annotation for the human proteome. *eLife* **2016**, *5*, e18715. [CrossRef] [PubMed]
3. Wang, J.; Zhang, L.; Jia, L.; Ren, Y.; Yu, G. Protein–protein Interactions Prediction Using a Novel Local Conjoint Triad Descriptor of Amino Acid Sequences. *Int. J. Mol. Sci.* **2017**, *18*, 2373. [CrossRef] [PubMed]
4. Wang, Y.; You, Z.; Li, X.; Chen, X.; Jiang, T.; Zhang, J. PCVMZM: Using the Probabilistic Classification Vector Machines Model Combined with a Zernike Moments Descriptor to Predict Protein–Protein Interactions from Protein Sequences. *Int. J. Mol. Sci.* **2017**, *18*, 1029. [CrossRef] [PubMed]
5. Ding, Y.; Tang, J.; Guo, F. Predicting protein–protein interactions via multivariate mutual information of protein sequences. *BMC Bioinform.* **2016**, *17*, 398. [CrossRef] [PubMed]
6. Ding, Y.; Tang, J.; Guo, F. Identification of Protein–Protein Interactions via a Novel Matrix-Based Sequence Representation Model with Amino Acid Contact Information. *Int. J. Mol. Sci.* **2016**, *17*, 1623. [CrossRef] [PubMed]
7. Shen, C.; Ding, Y.; Tang, J.; Xu, X.; Guo, F. An Ameliorated Prediction of Drug-Target Interactions Based on Multi-Scale Discrete Wavelet Transform and Network Features. *Int. J. Mol. Sci.* **2017**, *18*, 1781. [CrossRef] [PubMed]
8. Ding, Y.; Tang, J.; Guo, F. Identification of Drug-Target Interactions via Multiple Information Integration. *Inf. Sci.* **2017**, *418–419*, 546–560. [CrossRef]
9. Chen, P.; Li, J. Sequence-based identification of interface residues by an integrative profile combining hydrophobic and evolutionary information. *BMC Bioinform.* **2010**, *11*, 402. [CrossRef] [PubMed]
10. Chen, P.; Li, J.; Wong, L.; Kuwahara, H.; Huang, J.Z.; Gao, X. Accurate prediction of hot spot residues through physicochemical characteristics of amino acid sequences. *Proteins-Struct. Funct. Bioinform.* **2013**, *81*, 1351–1362. [CrossRef] [PubMed]
11. Pan, Y.; Wang, Z.; Zhan, W.; Deng, L. Computational identification of binding energy hot spots in protein-RNA complexes using an ensemble approach. *Bioinformatics* **2017**. [CrossRef] [PubMed]
12. Chen, P.; Hu, S.; Zhang, J.; Gao, X.; Li, J.; Xia, J.; Wang, B. A sequence-based dynamic ensemble learning system for protein ligand-binding site prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *13*, 901–912. [CrossRef] [PubMed]
13. Hu, S.S.; Peng, C.; Bing, W.; Li, J. Protein binding hot spots prediction from sequence only by a new ensemble learning method. *Amino Acids* **2017**, *49*, 1773–1785. [CrossRef] [PubMed]
14. Jiang, J.; Wang, N.; Chen, P.; Zheng, C.; Wang, B. Prediction of Protein Hotspots from Whole Protein Sequences by a Random Projection Ensemble System. *Int. J. Mol. Sci.* **2017**, *18*, 1543. [CrossRef] [PubMed]
15. Liu, B.; Wang, S.; Dong, Q.; Li, S.; Liu, X. Identification of DNA-binding proteins by combining auto-cross covariance transformation and ensemble learning. *IEEE Trans. NanoBiosci.* **2016**, *15*, 328–334. [CrossRef] [PubMed]
16. Liu, B.; Xu, J.; Fan, S.; Xu, R.; Zhou, J.; Wang, X. PseDNA-Pro: DNA-Binding Protein Identification by Combining Chou's PseAAC and Physicochemical Distance Transformation. *Mol. Inform.* **2015**, *34*, 8–17. [CrossRef] [PubMed]
17. Liu, B.; Wang, S.; Wang, X. DNA binding protein identification by combining pseudo amino acid composition and profile-based protein representation. *Sci. Rep.* **2015**, *5*, 15479. [CrossRef] [PubMed]
18. Liu, B.; Xu, J.; Lan, X.; Xu, R.; Zhou, J.; Wang, X.; Chou, K.-C. iDNA-Prot|dis: Identifying DNA-Binding Proteins by Incorporating Amino Acid Distance-Pairs and Reduced Alphabet Profile into the General Pseudo Amino Acid Composition. *PLoS ONE* **2014**, *9*, e106691. [CrossRef] [PubMed]

19. Song, L.; Li, D.; Zeng, X.; Wu, Y.; Guo, L.; Zou, Q. nDNA-Prot: Identification of DNA-binding Proteins Based on Unbalanced Classification. *BMC Bioinform.* **2014**, *15*, 298. [CrossRef] [PubMed]

20. Wei, L.; Tang, J.; Zou, Q. Local-DPP: An Improved DNA-binding Protein Prediction Method by Exploring Local Evolutionary Information. *Inf. Sci.* **2017**, *384*, 135–144. [CrossRef]

21. Qu, K.; Han, K.; Wu, S.; Wang, G.; Wei, L. Identification of DNA-Binding Proteins Using Mixed Feature Representation Methods. *Molecules* **2017**, *22*, 1602. [CrossRef] [PubMed]

22. Zhang, J.; Liu, B. PSFM-DBT: Identifying DNA-Binding Proteins by Combing Position Specific Frequency Matrix and Distance-Bigram Transformation. *Int. J. Mol. Sci.* **2017**, *18*, 1856. [CrossRef] [PubMed]

23. Shen, C.; Ding, Y.; Tang, J.; Song, J.; Guo, F. Identification of DNA-protein Binding Sites through Multi-Scale Local Average Blocks on Sequence Information. *Molecules* **2017**, *22*, 2079. [CrossRef] [PubMed]

24. Liu, W.X.; Deng, E.Z.; Chen, W.; Lin, H. Identifying the subfamilies of voltage-gated potassium channels using feature selection technique. *Int. J. Mol. Sci.* **2014**, *15*, 12940–12951. [CrossRef] [PubMed]

25. Lin, H.; Ding, H. Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition. *J. Theor. Biol.* **2011**, *269*, 64–69. [CrossRef] [PubMed]

26. Chen, W.; Lin, H. Identification of voltage-gated potassium channel subfamilies from sequence information using support vector machine. *Comput. Biol. Med.* **2012**, *42*, 504–507. [CrossRef] [PubMed]

27. Zhao, Y.W.; Su, Z.D.; Yang, W.; Lin, H.; Chen, W.; Tang, H. IonchanPred 2.0: A Tool to Predict Ion Channels and Their Types. *Int. J. Mol. Sci.* **2017**, *18*, 1838. [CrossRef] [PubMed]

28. Wang, H.; Liu, B.; Sun, P.P.; Ma, Z.Q. A Topology Structure Based Outer Membrane Proteins Segment Alignment Method. *Math. Probl. Eng.* **2013**, *2013*, 541359. [CrossRef]

29. Zhang, L.; Wang, H.; Yan, L.; Su, L.; Xu, D. OMPcontact: An Outer Membrane Protein Inter-Barrel Residue Contact Prediction Method. *J. Comput. Biol.* **2017**, *24*, 217–228. [CrossRef] [PubMed]

30. Wang, H.; He, Z.; Zhang, C.; Zhang, L.; Xu, D. Transmembrane protein alignment and fold recognition based on predicted topology. *PLoS ONE* **2013**, *8*, e69744. [CrossRef] [PubMed]

31. Antonets, K.S.; Nizhnikov, A.A. Predicting Amyloidogenic Proteins in the Proteomes of Plants. *Int. J. Mol. Sci.* **2017**, *18*, 2155. [CrossRef] [PubMed]

32. Wang, S.; Liu, S. Protein Sub-Nuclear Localization Based on Effective Fusion Representations and Dimension Reduction Algorithm LDA. *Int. J. Mol. Sci.* **2015**, *16*, 30343–30361. [CrossRef] [PubMed]

33. Wang, S.; Nie, B.; Yue, K.; Fei, Y.; Li, W.; Xu, D. Protein Subcellular Localization with Gaussian Kernel Discriminant Analysis and Its Kernel Parameter Selection. *Int. J. Mol. Sci.* **2017**, *18*, 2718. [CrossRef] [PubMed]

34. Wang, S.; Liu, P. A New Feature Extraction Method Based on the Information Fusion of Entropy Matrix and Covariance Matrix and Its Application in Face Recognition. *Entropy* **2015**, *17*, 4664–4683. [CrossRef]

35. Li, Y.H.; Yu, C.Y.; Li, X.X.; Zhang, P.; Tang, J.; Yang, Q.; Fu, T.; Zhang, X.; Cui, X.; Tu, G. Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* **2017**, *46*, D1121–D1127.

36. Li, B.; Tang, J.; Yang, Q.; Li, S.; Cui, X.; Li, Y.; Chen, Y.; Xue, W.; Li, X.; Zhu, F. NOREVA: Normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* **2017**, *45*, W162–W170. [CrossRef] [PubMed]

37. Li, Y.H.; Xu, J.Y.; Tao, L.; Li, X.F.; Li, S.; Zeng, X.; Chen, S.Y.; Zhang, P.; Qin, C.; Zhang, C. SVM-Prot 2016: A Web-Server for Machine Learning Prediction of Protein Functional Families from Sequence Irrespective of Similarity. *PLoS ONE* **2016**, *11*, e0155290. [CrossRef] [PubMed]

38. Yu, C.; Li, X.; Yang, H.; Li, Y.H.; Xue, W.; Chen, Y.; Tao, L.; Zhu, F. Assessing the Performances of Protein Function Prediction Algorithms from the Perspectives of Identification Accuracy and False Discovery Rate. *Int. J. Mol. Sci.* **2018**, *19*, 183. [CrossRef] [PubMed]

39. Liu, B. BioSeq-Analysis: A platform for DNA, RNA, and protein sequence analysis based on machine learning approaches. *Brief. Bioinform.* **2018**. [CrossRef] [PubMed]

40. Liu, B.; Liu, F.; Wang, X.; Chen, J.; Fang, L.; Chou, K.-C. Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res.* **2015**, *43*, W65–W71. [CrossRef] [PubMed]

41. Du, P.F.; Zhao, W.; Miao, Y.Y.; Wei, L.Y.; Wang, L. UltraPse: A Universal and Extensible Software Platform for Representing Biological Sequences. *Int. J. Mol. Sci.* **2017**, *18*, 2400. [CrossRef] [PubMed]

42. Du, P.; Wang, X.; Xu, C.; Gao, Y. PseAAC-Builder: A cross-platform stand-alone program for generating various special Chou's pseudo-amino acid compositions. *Analyt. Biochem.* **2012**, *425*, 117–119. [CrossRef] [PubMed]

43. Du, P.; Gu, S.; Jiao, Y. PseAAC-General: Fast Building Various Modes of General Form of Chou's Pseudo-Amino Acid Composition for Large-Scale Protein Datasets. *Int. J. Mol. Sci.* **2014**, *15*, 3495–3506. [CrossRef] [PubMed]

44. Jiao, Y.S.; Du, P.F. Predicting Golgi-resident protein types using pseudo amino acid compositions: Approaches with positional specific physicochemical properties. *J. Theor. Biol.* **2015**, *391*, 35–42. [CrossRef] [PubMed]

45. Oti, M.; Snel, B.; Huynen, M.A.; Brunner, H.G. Predicting disease genes using protein–protein interactions. *J. Med. Genet.* **2006**, *43*, 691. [CrossRef] [PubMed]

46. Chen, L.; Hao Xing, Z.; Huang, T.; Shu, Y.; Huang, G.; Li, H.-P. Application of the Shortest Path Algorithm for the Discovery of Breast Cancer-Related Genes. *Curr. Bioinform.* **2016**, *11*, 51–58. [CrossRef]

47. Zhang, J.; Yang, J.; Huang, T.; Shu, Y.; Chen, L. Identification of novel proliferative diabetic retinopathy related genes on protein–protein interaction network. *Neurocomputing* **2016**, *217*, 63–72. [CrossRef]

48. Chen, L.; Yang, J.; Xing, Z.; Yuan, F.; Shu, Y.; Zhang, Y.; Kong, X.; Huang, T.; Li, H.; Cai, Y.D. An integrated method for the identification of novel genes related to oral cancer. *PLoS ONE* **2017**, *12*, e0175185. [CrossRef] [PubMed]

49. Zhang, J.; Zhang, Z.; Chen, Z.; Lei, D. Integrating Multiple Heterogeneous Networks for Novel LncRNA-disease Association Inference. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**. [CrossRef] [PubMed]

50. Zhu, L.; Su, F.; Xu, Y.; Zou, Q. Network-based method for mining novel HPV infection related genes using random walk with restart algorithm. *BBA-Mol. Basis Dis.* **2018**. [CrossRef] [PubMed]

51. Lu, S.; Yan, Y.; Li, Z.; Chen, L.; Yang, J.; Zhang, Y.; Wang, S.; Liu, L. Determination of Genes Related to Uveitis by Utilization of the Random Walk with Restart Algorithm on a Protein–Protein Interaction Network. *Int. J. Mol. Sci.* **2017**, *18*, 1045. [CrossRef] [PubMed]

52. Li, B.; Liao, B. Protein Complexes Prediction Method Based on Core-Attachment Structure and Functional Annotations. *Int. J. Mol. Sci.* **2017**, *18*, 1910. [CrossRef] [PubMed]

53. Li, M.; Meng, X.; Zheng, R.; Wu, F.X.; Li, Y.; Pan, Y.; Wang, J. Identification of protein complexes by using a spatial and temporal active protein interaction network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**. [CrossRef] [PubMed]

54. Tang, Y.; Li, M.; Wang, J.; Pan, Y.; Wu, F.X. CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *BioSystems* **2015**, *127*, 67–72. [CrossRef] [PubMed]

55. Wang, J.; Chen, G.; Chen, G.; Li, M.; Wu, F.X.; Pan, Y. ClusterViz: A cytoscape APP for cluster analysis of biological network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *12*, 815–822. [CrossRef] [PubMed]

56. Li, M.; Yang, J.; Wu, F.X.; Pan, Y.; Wang, J. DyNetViewer: A Cytoscape app for dynamic network construction, analysis and visualization. *Bioinformatics* **2018**. [CrossRef] [PubMed]

57. Wu, L.; Min, L.; Wang, J.; Wu, F.X. CytoCtrlAnalyser: A Cytoscape app for biomolecular network controllability analysis. *Bioinformatics* **2017**. [CrossRef] [PubMed]

58. Li, M.; Li, D.; Tang, Y.; Wu, F.; Wang, J. CytoCluster: A Cytoscape Plugin for Cluster Analysis and Visualization of Biological Networks. *Int. J. Mol. Sci.* **2017**, *18*, 1880. [CrossRef] [PubMed]

59. Huang, T.; Sun, J.; Zhou, S.; Gao, J.; Liu, Y. Identification of Direct Activator of Adenosine Monophosphate-Activated Protein Kinase (AMPK) by Structure-Based Virtual Screening and Molecular Docking Approach. *Int. J. Mol. Sci.* **2017**, *18*, 1408. [CrossRef] [PubMed]

60. Hou, Q.L.; Luo, J.X.; Zhang, B.C.; Jiang, G.F.; Ding, W.; Zhang, Y.Q. 3D-QSAR and Molecular Docking Studies on the TcPMCA1-Mediated Detoxification of Scopoletin and Coumarin Derivatives. *Int. J. Mol. Sci.* **2017**, *18*, 1380. [CrossRef] [PubMed]

61. Ding, H.; Gao, F.; Yu, Y.; Chen, B. Biochemical and Computational Insights on a Novel Acid-Resistant and Thermal-Stable Glucose 1-Dehydrogenase. *Int. J. Mol. Sci.* **2017**, *18*, 1198. [CrossRef] [PubMed]

62. Chandler, J.C.; Gandhi, N.S.; Mancera, R.L.; Smith, G.; Elizur, A.; Ventura, T. Understanding Insulin Endocrinology in Decapod *Crustacea*: Molecular Modelling Characterization of an Insulin-Binding Protein and Insulin-Like Peptides in the Eastern Spiny Lobster, *Sagmariasus verreauxi*. *Int. J. Mol. Sci.* **2017**, *18*, 1832. [CrossRef] [PubMed]

63. FutomaKołoch, B.; Dudek, B.; Kapczyńska, K.; Krzyżewska, E.; Wańczyk, M.; Korzekwa, K.; Rybka, J.; Klausa, E. Relationship of Triamine-Biocide Tolerance of Salmonella enterica Serovar Senftenberg to Antimicrobial Susceptibility, Serum Resistance and Outer Membrane Proteins. *Int. J. Mol. Sci.* **2017**, *18*, 1459. [CrossRef] [PubMed]

64. Chen, W.; Xing, P.; Zou, Q. Detecting N6-methyladenosine sites from RNA transcriptomes using ensemble Support Vector Machines. *Sci. Rep.* **2017**, *7*, 40242. [CrossRef] [PubMed]
65. Wei, L.; Xing, P.; Tang, J.; Zou, Q. PhosPred-RF: A novel sequence-based predictor for phosphorylation sites using sequential information only. *IEEE Trans. NanoBiosci.* **2017**, *16*, 240–247. [CrossRef] [PubMed]