Recent Advances in the Prediction of Subcellular Localization of Proteins and Related Topics

Kenta Nakai[1]* and Leyi Wei[2]

[1]Institute of Medical Science, The University of Tokyo, Minato-Ku, Japan, [2]School of Software, Shandong University, Jinan, China

Prediction of subcellular localization of proteins from their amino acid sequences has a long history in bioinformatics and is still actively developing, incorporating the latest advances in machine learning and proteomics. Notably, deep learning-based methods for natural language processing have made great contributions. Here, we review recent advances in the field as well as its related fields, such as subcellular proteomics and the prediction/ recognition of subcellular localization from image data.

Keywords: subcellular localization, deep learning, subcellular proteomics, amino acid sequence analysis, image analysis

INTRODUCTION

In bioinformatics, prediction of subcellular localization sites of proteins from their amino acid sequences has remained to be an important field. Such studies are useful in understanding the mechanisms of their localization process, including the recognition of their sequence determinants (i.e., the sorting/targeting signals) coded in amino acid sequences, and in inferring the function of those proteins. The prediction problem has also been used as a benchmark place for introducing latest machine-learning algorithms. Many review articles including ours have been published in this field. In this review, we would like to introduce recent advances mainly published after our latest review (Imai & Nakai, 2020). We will also briefly review recent papers in related areas, such as the prediction/recognition of protein subcellular localization based on image data and the subcellular proteomics, which seems to give us hints for the future direction of this field.

GENERAL REVIEWS AND ASSESSMENT STUDIES

Several general review articles have been published recently (Nielsen, Konstantinos, et al., 2019; Kumar & Dhanda, 2020; Barberis et al., 2021; Jiang, Wang, Wang, et al., 2021; Pan et al., 2021). Of these, the review by Nielsen et al. is a retrospective of the field, emphasizing the prediction of signal peptides; Kumar and Dhanda give a full list of available tools and their classification; and the review by Jiang et al. emphasizes the mathematical foundation of various approaches, also giving a review of some benchmark datasets. Similarly, Shen et al. reports their critical evaluation of web-based tools for protein subcellular localization (Shen et al., 2020), though they seem to mainly focus on Gene Ontology-based predictors, which are basically out of the scope of this review. Although the prediction of subcellular localization of RNA molecules is also out of the scope, it is important to establish reliable data sources for training the prediction models (Cui et al., 2022) and how to integrate heterologous resources is an important issue (Savulescu et al., 2021).

As for the prediction of specific sorting signals, two assessment studies on the prediction of signal peptides based on experimental data of rather untypical organisms were recently published: these

organisms are phytoplasmas (Garcion et al., 2021) and a thermoacidophilic archaeon (Singhal et al., 2021). It seems that the current predictors are not highly reliable for the prediction of these organisms (probably due to the scarcity of their training data); among several versions of SignalP (see below), not the latest version was the most reliable. We believe that important future issues in the prediction of signal peptides are 1) the improvement of distinction between cleavable signal peptides and uncleaved N-terminal transmembrane segments and 2) the prediction of secretory proteins without (apparent) signal peptides (Lonsdale et al., 2016; Nielsen, Petsalaki, et al., 2019).

DEEP LEARNING AND LANGUAGE MODEL-BASED METHODS

As noted above, the prediction of protein subcellular localization has always been a playground where the latest machine learning algorithms are introduced. In recent years, deep learning-based methods have become quite popular and thus a number of papers have been published within a few years (Cong et al., 2020, 2022; Semwal & Varadwaj, 2020; Jiang, Wang, Yao, et al., 2021; Liao et al., 2021; Yuan et al., 2021). The architecture of deep learning models has made rapid progress and they have also been applied to bioinformatics, such as protein design (Ding et al., 2022). Convolutional neural networks (CNNs) are the standard model; Liao et al. introduced the PSSMs (position-specific scoring matrices) derived from PSI-BLAST (Altschul et al., 1997) for adding evolutionary information to input; Cong et al. used the ant-colony optimization for letting the prediction model self-evolving (Cong et al., 2020), which is a trend of deep learning. As another big trend, techniques that have been successfully used in natural language processing have been introduced. One of them is the use of (multi-head) self-attention mechanism, which was first introduced in Transformer (reviewed in Shreyashree et al., 2022). Both Jiang et al. and Cong et al. report the improvement of prediction performance with the use of the self-attention mechanism (Jiang, Wang, Yao, et al., 2021; Cong et al., 2022). Jiang et al. also claim that their method shows better performance in suborganellar prediction (see below).

Deep neural networks have also been used in the prediction of sorting signals, particularly signal peptides (J.-M. Wu et al., 2019). SignalP, a standard tool for signal-peptide prediction, has employed this technique since its version 5 (Almagro Armenteros et al., 2019). The same group also used a similar method, incorporating the self-attention mechanism, for TargetP, which can detect three kinds of targeting signals (Almagro Armenteros et al., 2019). One of the merits of using attention-based methods is that it enables us to see which parts of the input sequence are paid with greater "attention". Thus, Almagro Armenteros et al. found that the second amino acid, which appears after the first methionine, seems to be important for the recognition of targeting signals, such as the chloroplast transit peptides. Wu et al. used their Transformer-based model for a different purpose: they generated novel signal peptides from the model that learned a number of known signal peptides in many

organisms (Z. Wu et al., 2020). They confirmed experimentally that the generated peptides actually worked as functional signal peptides when appended to the N-terminus of cytosolic proteins in *Bacillus subtilis*, though these sequences were not similar to any of known ones.

After the appearance of Transformer, a kind of its successor model, BERT (Bidirectional Encoder Representations from Transformers), has become very popular in natural language processing. One of the characteristics of BERT is that the model has been pretrained with a very large unannotated (unlabled) training set and that users are usually expected to fine-tune the model with their own labeled data. This style, called the transfer learning, meaning that pre-trained results are transferred to specific topics, has become a new trend even in molecular biology, too. Heinzinger et al. used this approach and fine-tuned their model as an example, for the prediction of protein subcellular localization sites (Heinzinger et al., 2019). Jin et al. also applied their pre-trained model to the subcellular localization prediction (Jin & Yang, 2022). Nowadays, models pretrained with both DNA (Ji et al., 2021) and amino acid sequences, ProtTrans (Elnaggar et al., 2021), are available for end users. Indeed, SignalP ver. Six was constructed based on ProtTrans and a significant improvement of its performance is reported (Teufel et al., 2022).

The multi-head self-attention model with multi-scale (i.e., parallel for various scales) CNNs was also used for the prediction of subcellular localization sites of mRNAs (D. Wang et al., 2021). Their system not only predicts multiple localizations of mRNA isoforms but also is useful in interpreting the mechanisms/signals of isoform-specific localization, based on the analysis of attention weights.

MISCELLANEOUS ALGORITHMS

Of course, prediction methods which do not rely on deep learning but on other machine learning methods have also been published recently in this field. Most of them use general sequence features rather than hand-crafted features related to specific sorting signals and claim to be able to address the problem of proteins localized at multiple sites (i.e., multi-labeled proteins), though there still remains the problem that their training data do not seem to have been annotated with a uniform criterion (see below). Some of them proposed extensions of existing sequence features, such as the k-mer compositions (Li et al., 2019; Yao et al., 2019; Sahu et al., 2020), while some imported external information, such as Gene Ontology and protein-protein interactions (Chen et al., 2021; Liu et al., 2021; Zhang et al., 2021). One method employed an ensemble approach of multiple classifiers with voting, claiming that the approach is effective in addressing the problem of imbalanced sizes of training data between different localization sites (Wattanapornprom et al., 2021). Among these newly published reports, the approach reported by Alaa et al. might be a bit novel and have further room for improvements: they used Markov models to produce a feature vector which is based on the micro-similarities of the probability distributions between the input sequence and the

reference models (Alaa et al., 2019). Moreover, it might be notable that Wang et al. attempted to incorporate hand-crafted features derived from the three-dimensional structure information of input proteins, such as their substructure frequency (G. Wang et al., 2021). Although the improvement of performance may not look so striking in their analysis, this field would be promising because of the current significant improvement of the prediction accuracy in the three-dimensional structure of proteins through AlphaFold2 and related methods (Jumper et al., 2021). It seems that the 3D structure-based approach could be promising in understanding how some sorting signals that do not show apparent sequence similarity are specifically recognized. In addition, it has been known that there are some correlations between the pI value of proteins and their subcellular localization [reviewed in (Tokmakov et al., 2021)], which might be the reason why amino acid composition is an effective feature for the prediction. Comprehensive analyses using predicted structures would be useful in understanding the adaption of individual proteins to the environment provided by each localization site.

PREDICTION OF LOCALIZATION AT SPECIFIC ORGANELLES AND SUBORGANELLAR LOCALIZATION

Probably, partly because of a certain level of maturation in the prediction of protein subcellular localization and partly because of the progress of precise subcellular proteomics studies (see below), it seems that a boom in the prediction of suborganellar localization has come. Its most typical field is the prediction of submitochondrial localization [for a recent review, see (Martelli et al., 2021)]. In mitochondria, about 1000 proteins encoded in the nuclear genome are sorted into four places (the matrix, the inner membrane, the intermembrane space, and the outer membrane). Several predictors have been released in only a recent few years; most of them were based on (convolutional) deep learning. For example, Savojardo et al. developed DeepMito, which is based on CNN and the method was used to annotate the potential mitochondrial proteins of four species (Savojardo, Bruciaferri, et al., 2020; Savojardo, Martelli, et al., 2020). Wang et al. proposed another predictor (DeepPred-SubMito), which is also based on CNN and was taken care of the unbalanced sample sizes with the random over-sampling approach (X. Wang et al., 2020). Hou et al. developed another predictor (iDeepSubMito), which is based on CNN and the bidirectional LSTM (i.e., Long Short Term Memory) with the self-attention model (Hou et al., 2021). In contrast, Yu et al. developed a predictor (SubMito-XGBoost), which is based on eXtreme gradient boosting, a new method in traditional machine learning, combining gradient boosting and random forest ensemble learning (Yu et al., 2020). Unlike CNN-based methods, this method uses a variety of sequence features, such as the gapped dipeptide composition. It seems possible that such approaches will be useful in finding hidden sorting signals. More explicitly, Schneider et al. developed a specialized predictor (iMLP) for detecting internal matrix targeting-like sequences, which do not exist on the N-terminus, unlike well-known mitochondrial (matrix)

targeting peptides, using recurrent neural networks (RNNs) (Schneider et al., 2021).

Besides mitochondria, recent works on the prediction of suborganellar localization are not many. One for subnuclear localization (L. Wu et al., 2020), another for sub-peroxisomal localization (Anteghini et al., 2021), and another for distinguishing *cis*-Golgi and *trans*-Golgi proteins using deep learned 107 features (Lv et al., 2021). However, we believe that these works become pioneers for subsequent more elaborated works. In addition, there are a few works dealing with some specialized but biologically meaningful problems: Kaundal et al. proposed two-step predictions, where input proteins are classified into plastid or non-plastid proteins in the first step and the plastid proteins are further classified into one of the four types chloroplasts, chromoplasts, etioplasts, and amyloplasts) (Kaundal et al., 2013). Kaleel et al. presented a CNN-based predictor which simply detects endomembrane system and secretory pathway proteins from the others (Kaleel et al., 2020). It would be interesting to see how such an approach could complement traditional approaches based on the recognition of signal peptides.

At the end of this section, where a new trend in this field is introduced, we would like to add one paper, which may become a pioneering work in a new trend: prediction of tissue-specific subcellular localization of proteins. Zhu et al. attempted this based on the information of tissue-specific functional associations and protein-protein interactions (PPIs) (Zhu et al., 2019). They identified 1314 known differential localizations between nine types of tissues as well as 549 novel candidates, some of which were verified through literature survey. With the increase of more known examples, further approaches should be taken to improve the prediction accuracy.

LOCALIZATION OF BACTERIAL PROTEINS

As far as we notice, for the prediction of the subcellular localization of bacterial proteins, only a few new predictors have been released after the publication of our previous review (Imai & Nakai, 2020), where we already reviewed PSORTm for the prediction for metagenome data (Peabody et al., 2020) and PSORTdb 4.0, which contains both experimentally-verified and computationally-predicted subcellular localization information (Lau et al., 2021), for example. We have also mentioned PSO-LocBact, which gives a kind of consensus between various existing predictors using the particle swarm optimization (PSO) algorithm (Lertampaiporn et al., 2019). However, we noticed a few new works in more specific prediction problems: GP4 is a predictor specifically developed for the prediction of Gram-positive proteins (Grasso et al., 2021). Next, T3SEpp is a predictor for bacterial type III secreted effectors (T3SEs), which are used for the infection of Gram-negative bacteria, and thus the prediction has medical importance (Hui et al., 2020). Finally, BetAware-Deep is a web server for the discrimination and topology prediction of prokaryotic transmembrane beta-barrel proteins (Madeo et al., 2021): this discrimination is useful for the prediction of subcellular

localization because these beta-barrel proteins exist in the outer membrane of Gram-negative bacteria. In addition, the topology prediction could be also useful for further predictions because it is important to know that an additional signal faces to which side of the membrane (i.e., the periplasm and the outside of the cell).

SUBCELLULAR PROTEOMICS

For the development of reliable prediction methods, comprehensive lists of known (i.e., experimentally-verified) localization information are essential. Although such information is contained in standard databases, such as UniProt/Swiss-Prot (Bateman et al., 2021), it is desirable that the localization is determined in a uniform criterion because information that a protein is localized at a single site or multiple sites can be sometimes ambiguous, for example, and thus it must be decided in an objective manner. Recently, there have been notable advances in subcellular proteomics and thus we briefly introduce these advances in this section, hoping that these trends could be indicators for the future direction of novel prediction methods.

In the field of transcriptomics, obtaining the expression profile of individual cells (single cell RNA sequencing, scRNA-seq) has become rather popular. With such approaches, we can now explore what kind of cell types are contained in a certain tissue. Moreover, the techniques to clarify the spatial distribution of these cell types (spatial transcriptomics) is under rapid development. And the information of subcellular distribution of RNAs has begun to accumulate, too (Longo et al., 2021). This field will undoubtedly stimulate the further development of methods for the prediction of RNA subcellular localization.

In the field of proteomics, equivalent studies exist. For example, the importance of single-cell proteomics is increasing because the amount of mRNAs is not enough to know the amount of their protein products (Xie & Ding, 2022). However, it seems that currently the word, spatial proteomics, means subcellular proteomics, reflecting the current excitement in this field. Naturally, many review articles have been published in only a few years (Lundberg & Borner, 2019; Borner, 2020; Christopher et al., 2021, Christopher et al., 2022; Paul et al., 2021). Amongst them, Christopher et al. (2022) compares the methods used in subcellular transcriptomics and proteomics.

There are several approaches in subcellular proteomics and they can be classified in various ways. Perhaps, the most standard method is to directly observe the localization of a protein labeled by fluorescent antibody (or any other affinity reagents) through microscopic imaging *in situ*. The analysis of the obtained images requires computational procedures, where machine learning has made significant contributions (see the next section). The Subcellular section of the Human Protein Atlas (HPA) database contains the immunofluorescence images of 65% of the human protein-coding genes (Thul et al., 2017). Based on this collection, a competition for machine-learning-based image processing methods (the Kaggle competition for multi-label classification of cell organelles in proteome scale Human Protein Atlas data) was held (Ouyang et al., 2019).

Other direct methods include the physical separation of specific organelles and the biochemical fractionation of cells using centrifugation or detergents. Proteins contained/enriched in the separated organelle or the specific fraction are determined using tandem mass spectroscopy (MS). Since proteins that are co-localized at the same compartment should share the same distribution of their abundance between fractions, computational methods, such as the cluster analysis (known as correlation profiling), can be used to identify the subcellular localization of thousands of proteins simultaneously (Itzhak et al., 2019; Borner, 2020). If fractions are labeled with different isotope tags, like in the LOPIT (Localization of Organelle Proteins by Isotope Tagging) method, more accurate distinction can be made between organelles with similar densities (such as Golgi, plasma membrane and endoplasmic reticulum) (Elzek et al., 2021). Using an MS-based pipeline, Orre et al. identified the subcellular localization of about 12,000 proteins across five cell lines and the data are available as SubCellBarCode (Orre et al., 2019). It is interesting to see what kind of factors contribute to differential localization across cell types. Joshi et al. determined the three types of localization (cytosolic, nuclear, and membrane) of 6572 proteins in human T cells. They also monitored the time-course changes after the T cell receptor stimulation and identified about 200 potentially translocating proteins (Joshi et al., 2019). The data are released as TCellSubC. Huang et al. also constructed a database (PSL-LCCL) of protein subcellular localization (six organelles) in human cancer cell lines (Huang et al., 2022). Finally, a database of mitochondrial proteome (MitoCarta) was made mainly from an MS-based approach and its latest version contains the sub-organellar localization information (Rath et al., 2021). These resources could be useful in planning the future directions of subcellular localization predictions. It might be noteworthy that the advances in MS technology have enabled the peptidome and metabolome analyses in the single-cell level (Nemes, 2021).

The third systematic approach is to detect the protein-protein interaction network (Pino & Schilling, 2021). The detection can be made with co-immunoprecipitation or cross-linking. Recently, an approach called proximity labeling is increasingly used, where an expression plasmid containing the gene of a bait protein fused with an engineered protein, such as BioID (biotin ligase) or APEX (ascorbic acid peroxidase), is introduced into target cells; after the fused protein is expressed within the cells, the addition of biotin etc. causes labeling reactions, where nearby proteins are biotinylated; then the biotinylated proteins are collected and identified through LC-MS (Liquid Chromatography Mass Spectrometry). With this approach, no specific antibodies are required. Notably, a large BioID-based map of human cells (HEK293) was published recently (Go et al., 2021). In this work, the authors defined the intracellular locations of 4,145 proteins, using 192 subcellular markers. The data are provided at humancellmap.org. This approach seems to be superior in its sensitivity but it does not seem to be good at identifying proteins with multiple locations. Another group also identified the protein

proximity network in mitochondria, comprising 1,465 proteins (Antonicka et al., 2020).

IMAGE ANALYSIS

As compared to amino acid sequences, images that present proteins or subcellular locations with distinct patterns are more intuitive and interpretable. Benefitting from the advent of microscopic imaging techniques, there has been an increasing interest in protein subcellular localization based on the analyses of fluorescence microscope images and immunohistochemistry (IHC) images. The image-based protein subcellular localization methods can be roughly categorized into two groups, i.e., traditional machine learning methods, and deep learning methods. Traditional machine learning methods mainly rely on the design of hand-crafted image feature descriptors for predictive model construction. Typical image features in image processing field, such as Haralick features, Zernike features, and Local Binary Patterns (LBP), etc., are commonly used (Xu et al., 2018). Considering the microscope images are quite different from natural images, directly using the features are not sufficient to capture the key information in microscope images. Therefore, a number of methods based on essential properties of microscope images have been developed. Yang et al. proposed a frequency feature and intensity coding strategy to explore the local region information, improving feature representation of IHC images (Yang et al., 2019). Tahir et al. designed the threshold calculation LBP operator for feature extraction from fluorescence microscopy images (Tahir & Idris, 2020). To analyze the feature patterns from images of multi-location proteins, Xu et al. built the LDA models to extract various feature topics of IHC images and map them to subcellular locations (Xu et al., 2020). However, these hand-crafted image features are shallow and low-level, and cannot fully explore the specificity of different locations, due to the limited knowledge of imaging, thereby impacting the performance in protein subcellular localization.

Recently, deep learning has made breakthroughs in computer vision and has attracted considerable attention in biomedical image analysis, due to its excellent ability in learning high-latent image feature representations (Fuyong et al., 2018). Therefore, recent computational efforts in this field are more focused on deep learning methods. Convolutional neural networks are the very first deep networks that are introduced as the image feature extractor to capture the features for protein subcellular localization (Pärnamaa & Parts, 2017). The deep features learnt from deep neural networks are reportedly to be better than the hand-crafted features (Su et al., 2021; F.Wang & Wei, 2022). Considering the feature space learnt from CNNs may exist redundant information, Su et al. proposed a feature selection strategy to optimize the feature space, thereby improving the predictive performance and meanwhile reducing the computational complexity (Su et al., 2021). Similarly, Long et al. introduced the self-attention mechanism to capture the key features derived from the deep convolutional neural network (Long et al., 2020). Xue et al. combined multiple nonlinear decomposing algorithms to unmix effective feature patterns

from deep image feature representations (M.-Q. Xue et al., 2021). To further improve the predictive performance, Hu and colleagues employed a label-correlation relevancy strategy to enhance localization results (Hu et al., 2022). More recently, Wang et al. proposed a multi-scale feature representation learning framework and successfully learnt a set of comprehensive features from low-level to high-level (F. Wang & Wei, 2022). They demonstrate that the deep features from different scales are complementary and useful to capture the distinguishable information amongst different subcellular locations. In addition, some methods integrating deep features with traditional hand-crafted features are proposed. Xue et al. split images into representative patches as model inputs, and integrate feature engineering and deep learning methods (Z.-Z. Xue et al., 2020). UIIah et al. extracted different handcrafted and deep features learned from different viewpoints of the images (Ullah et al., 2021). However, some existing methods still suffer from data imbalance and insufficient data problems. To copy with the issues, Tu et al. proposed a self-supervised learning framework namely SIFLoc through introducing a hybrid data augmentation strategy and contrastive learning (Tu et al., 2022).

Furthermore, these protein subcellular location prediction methods have been used in location biomarker analysis (Fan et al., 2021; Long et al., 2020; F.; Wang & Wei, 2022; Xu et al., 2020; Z.-Z.; Xue et al., 2020). Proteins in the normal and cancerous images are likely to have different subcellular patterns. The methods based on images were expected to detect the differences. To assess the significance of location changes, an independent sample t-test is used to obtain the p values for the prediction results of the normal and cancerous images. Through the consistency evaluation analysis between HPA and Swiss-Prot, Xu et al. recently demonstrate that proteins having highly variable locations are more likely to be biomarkers of diseases (Xu et al., 2021). Thus, comprehensive analyses using microscopic images would be useful in speeding up the understanding of the mechanism of protein mislocalization and providing the accurate identification of cancer biomarkers.

FUTURE DIRECTIONS AND CONCLUDING REMARKS

As reviewed here, in only a few years, there have been many advances in this field. Deep learning-based methods will continue to play important roles in both image and sequence analyses. Moreover, since subcellular proteomics-based localization information has become increasingly popular, the need to predict the localization of unknown proteins computationally is becoming less important; rather, we believe that computational approaches should become even more important in the following aspects: 1) to assist the improvement of proteome-based experiments with the use of more sophisticated methods in their data analysis; 2) to contribute to understand specific molecular mechanisms of protein sorting, through the interpretation of learned features (such as the attentions), etc.; 3) to contribute to the understanding of "exceptional" mechanisms of protein sorting, such as the secretion without

N-terminal signal peptides (Nielsen, Petsalaki, et al., 2019); 4) to characterize the dynamic translocation processes upon external stimulation, disease, etc.; this includes the effort to interpret these phenomena through the change of their sequences (*via* alternative splicing) or environments. Lastly, 5) the contribution to synthetic biology, e.g., the design of novel targeting signals with desired characteristics, would become more important (Rajendran et al., 2010). The next few years will continue to be exciting for researchers in the field. Almagro

Armenteros et al., 2019,Jiang et al., 2021, Nielsen et al., 2019, Savojardo et al., 2020, Wang et al., 2021, Wu et al., 2020.

AUTHOR CONTRIBUTIONS

REFERENCES

Alaa, A., Eldeib, A. M., and Metwally, A. A. (2019). Protein Subcellular Localization Prediction Based on Internal Micro-similarities of Markov ChainsIEEE Engineering in Medicine and Biology Society. *Annu. Int. Conf. IEEE Eng. Med. Biol. SocAnnual Int. Conf.* 2019, 1355–1358. doi:10.1109/EMBC.2019. 8857598

Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., et al. (2019). Detecting Sequence Signals in Targeting Peptides Using Deep Learning. *Life Sci. Alliance* 2 (5). doi:10.26508/lsa. 201900429

Almagro, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* 37 (4), 420–423. doi:10.1038/s41587-019- 0036-z

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs. *Nucleic Acids Res.* 25 (17), 3389–3402. doi:10.1093/nar/25.17. 3389

Anteghini, M., Martins Dos Santos, V., and Saccenti, E. (2021). In-Pero: Exploiting Deep Learning Embeddings of Protein Sequences to Predict the Localisation of Peroxisomal Proteins. *Int. J. Mol. Sci.* 22 (12). doi:10.3390/ijms22126409

Antonicka, H., Lin, Z. Y., Janer, A., Aaltonen, M. J., Weraarpachai, W., Gingras, A. C., et al. (2020). A High-Density Human Mitochondrial Proximity Interaction Network. *Cell Metab.* 32 (3), 479–e9. e9. doi:10.1016/j.cmet.2020.07.017

Barberis, E., Marengo, E., and Manfredi, M. (2021). Protein Subcellular Localization Prediction. *Methods Mol. Biol. Clift. N.J.* 2361, 197–212. doi:10. 1007/978-1-0716-1641-3_12

Bateman, A., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., et al. (2021). UniProt: the Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100

Borner, G. H. H. (2020). Organellar Maps through Proteomic Profiling A Conceptual Guide. *Mol. Cell Proteomics* 19 (7), 1076–1087. doi:10.1074/ mcp.R120.001971

Chen, L., Li, Z., Zeng, T., Zhang, Y. H., Zhang, S., Huang, T., et al. (2021). Predicting Human Protein Subcellular Locations by Using a Combination of Network and Function Features. *Front. Genet.* 12, 783128. doi:10.3389/fgene. 2021.783128

Christopher, J. A., Stadler, C., Martin, C. E., Morgenstern, M., Pan, Y., Betsinger, C. N., et al. (2021). Subcellular Proteomics. *Nat. Rev. Methods Prim.* 1. doi:10. 1038/s43586-021-00029-y

Christopher, J. A., Geladaki, A., Dawson, C. S., Vennard, O. L., and Lilley, K. S. (2022). Subcellular Transcriptomics and Proteomics: A Comparative Methods Review. *Mol. Cell. Proteomics* 21 (2), 100186. doi:10.1016/j.mcpro.2021.100186

Cong, H., Liu, H., Chen, Y., and Cao, Y. (2020). Self-evolving Framework of Deep Convolutional Neural Network for Multilocus Protein Subcellular Localization. *Med. Biol. Eng. Comput.* 58 (12), 3017–3038. doi:10.1007/s11517-020-02275-w

Cong, H., Liu, H., Cao, Y., Chen, Y., and Liang, C. (2022). Multiple Protein Subcellular Locations Prediction Based on Deep Convolutional Neural Networks with Self-Attention Mechanism. *Interdiscip. Sci. Comput. Life Sci..* doi:10.1007/s12539-021-00496-7

Cui, T., Dou, Y., Tan, P., Ni, Z., Liu, T., Wang, D., et al. (2022). RNALocate v2.0: an Updated Resource for RNA Subcellular Localization with Increased Coverage

and Annotation. *Nucleic Acids Res.* 50 (D1), D333–D339. doi:10.1093/nar/ gkab825

Ding, W., Nakai, K., and Gong, H. (2022). Protein Design via Deep Learning. *Briefings Bioinforma.* 2022, bbac102. doi:10.1093/bib/bbac102

Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., et al. (2021). ProtTrans: Towards Cracking the Language of Lifes Code through Self-Supervised Deep Learning and High Performance Computing. *IEEE Trans. Pattern Anal. Mach. Intell.* 1, 1. –1. doi:10.1109/TPAMI.2021.3095381

Elzek, M. A. W., Christopher, J. A., Breckels, L. M., and Lilley, K. S. (2021). Localization of Organelle Proteins by Isotope Tagging: Current Status and Potential Applications in Drug Discovery Research. *Drug Discov. Today TechnolTechnologies* 39, 57–67. doi:10.1016/j.ddtec.2021.06.003

Fan, J., Liu, J., Xie, S., Zhou, C., and Wu, Y. (2021). Cervical Lesion Image Enhancement Based on Conditional Entropy Generative Adversarial Network Framework. *Methods.* doi:10.1016/J.YMETH.2021.11.004

Fuyong, X., Yuanpu, X., Hai, S., Fujun, L., and Lin, Y. (2018). Deep Learning in Microscopy Image Analysis: A Survey. *IEEE Trans. Neural Netw. Learn Syst.* 29 (10), 4550–4568. doi:10.1109/TNNLS.2017.2766168

Garcion, C., Béven, L., and Foissac, X. (2021). Comparison of Current Methods for Signal Peptide Prediction in Phytoplasmas. *Front. Microbiol.* 12, 661524. doi:10. 3389/fmicb.2021.661524

Go, C. D., Knight, J. D. R., Rajasekharan, A., Rathod, B., Hesketh, G. G., Abe, K. T., et al. (2021). A Proximity-dependent Biotinylation Map of a Human Cell. *Nature* 595 (7865), 120–124. doi:10.1038/s41586-021-03592-2

Grasso, S., van Rij, T., and van Dijl, J. M. (2021). GP4: an Integrated Gram-Positive Protein Prediction Pipeline for Subcellular Localization Mimicking Bacterial Sorting. *Brief. Bioinform* 22 (4). doi:10.1093/bib/bbaa302

Heinzinger, M., Elnaggar, A., Wang, Y., Dallago, C., Nechaev, D., Matthes, F., et al. (2019). Modeling Aspects of the Language of Life through Transfer-Learning Protein Sequences. *BMC Bioinforma.* 20 (1), 723. doi:10.1186/s12859-019- 3220-8

Hou, Z., Yang, Y., Li, H., Wong, K. C., and Li, X. (2021). iDeepSubMito: Identification of Protein Submitochondrial Localization with Deep Learning. *Brief. Bioinform* 22 (6). doi:10.1093/bib/bbab288

Hu, J. X., Yang, Y., Xu, Y. Y., and Shen, H. B. (2022). Incorporating Label Correlations into Deep Neural Networks to Classify Protein Subcellular Location Patterns in Immunohistochemistry Images. *Proteins* 90 (2), 493–503. doi:10.1002/prot.26244

Huang, F., Tang, X., Ye, B., Wu, S., and Ding, K. (2022). PSL-LCCL: a Resource for Subcellular Protein Localization in Liver Cancer Cell Line SK_HEP1. *Database J. Biol. Databases Curation* 2022, baab087. doi:10.1093/database/baab087

Hui, X., Chen, Z., Lin, M., Zhang, J., Hu, Y., Zeng, Y., et al. (2020). T3SEpp: an Integrated Prediction Pipeline for Bacterial Type III Secreted Effectors. *MSystems* 5 (4). doi:10.1128/mSystems.00288-20

Imai, K., and Nakai, K. (2020). Tools for the Recognition of Sorting Signals and the Prediction of Subcellular Localization of Proteins from Their Amino Acid Sequences. *Front. Genet.* 11, 607812. doi:10.3389/fgene.2020.607812

Itzhak, D. N., Schessner, J. P., and Borner, G. H. H. (2019). Dynamic Organellar Maps for Spatial Proteomics. *Curr. Protoc. Cell Biol.* 83 (1), e81. doi:10.1002/ cpcb.81

Ji, Y., Zhou, Z., Liu, H., and Davuluri, R. v. (2021). DNABERT: Pre-trained Bidirectional Encoder Representations from Transformers Model for DNA-Language in Genome. *Bioinformatics* 37 (15), 2112–2120. doi:10.1093/ bioinformatics/btab083

Jiang, Y., Wang, D., Wang, W., and Xu, D. (2021). Computational Methods for Protein Localization Prediction. *Comput. Struct. Biotechnol. J.* 19, 5834–5844. doi:10.1016/j.csbj.2021.10.023

Jiang, Y., Wang, D., Yao, Y., Eubel, H., Künzler, P., Møller, I. M., et al. (2021). MULocDeep: A Deep-Learning Framework for Protein Subcellular and Suborganellar Localization Prediction with Residue-Level Interpretation. *Comput. Struct. Biotechnol. J.* 19, 4825–4839. doi:10.1016/j.csbj.2021.08.027

Jin, Y., and Yang, Y. (2022). ProtPlat: an Efficient Pre-training Platform for Protein Classification Based on FastText. *BMC Bioinforma.* 23 (1), 66. doi:10.1186/s12859-022-04604-2

Joshi, R. N., Stadler, C., Lehmann, R., Lehtiö, J., Tegnér, J., Schmidt, A., et al. (2019). TcellSubC: An Atlas of the Subcellular Proteome of Human T Cells. *Front. Immunol.* 10, 2708. doi:10.3389/fimmu.2019.02708

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Kaleel, M., Zheng, Y., Chen, J., Feng, X., Simpson, J. C., Pollastri, G., et al. (2020). SCLpred-EMS: Subcellular Localization Prediction of Endomembrane System and Secretory Pathway Proteins by Deep N-To-1 Convolutional Neural Networks. *Bioinformatics* 36 (11), 3343–3349. doi:10.1093/bioinformatics/btaa156

Kaundal, R., Sahu, S. S., Verma, R., and Weirick, T. (2013). Identification and Characterization of Plastid-type Proteins from Sequence-Attributed Features Using Machine Learning. *BMC Bioinforma.* 14, S7. doi:10.1186/1471-2105-14-S14-S7

Kumar, R., and Dhanda, S. K. (2020). Bird Eye View of Protein Subcellular Localization Prediction. *Life* 10 (12), 347. doi:10.3390/life10120347

Lau, W. Y. V., Hoad, G. R., Jin, V., Winsor, G. L., Madyan, A., Gray, K. L., et al. (2021). PSORTdb 4.0: Expanded and Redesigned Bacterial and Archaeal Protein Subcellular Localization Database Incorporating New Secondary Localizations. *Nucleic Acids Res.* 49 (D1), D803–D808. doi:10.1093/nar/gkaa1095

Lertampaiporn, S., Nuannimnoi, S., Vorapreeda, T., Chokesajjawatee, N., Visessanguan, W., and Thammarongtham, C. (2019). PSO-LocBact: A Consensus Method for Optimizing Multiple Classifier Results for Predicting the Subcellular Localization of Bacterial Proteins. *Biomed. Res. Int.* 2019, 5617153. doi:10.1155/2019/5617153

Li, B., Cai, L., Liao, B., Fu, X., Bing, P., and Yang, J. (2019). Prediction of Protein Subcellular Localization Based on Fusion of Multi-View Features. *Molecules* 24 (5). doi:10.3390/molecules24050919

Liao, Z., Pan, G., Sun, C., and Tang, J. (2021). Predicting Subcellular Location of Protein with Evolution Information and Sequence-Based Deep Learning. *BMC Bioinforma.* 22 (Suppl. 10), 515. doi:10.1186/s12859-021-04404-0

Liu, Y., Jin, S., Gao, H., Wang, X., Wang, C., Zhou, W., et al. (2021). Predicting the Multi-Label Protein Subcellular Localization through Multi-Information Fusion and MLSI Dimensionality Reduction Based on MLFE Classifier, *Bioinformatics*, 38, 1223–1230. doi:10.1093/bioinformatics/btab811

Long, W., Yang, Y., and Shen, H. B. (2020). ImPLoc: a Multi-Instance Deep Learning Model for the Prediction of Protein Subcellular Localization Based on Immunohistochemistry Images. *Bioinformatics* 36 (7), 2244–2250. doi:10.1093/bioinformatics/btz909

Longo, S. K., Guo, M. G., Ji, A. L., and Khavari, P. A. (2021). Integrating Single-Cell and Spatial Transcriptomics to Elucidate Intercellular Tissue Dynamics. *Nat. Rev. Genet.* 22 (10), 627–644. doi:10.1038/s41576-021-00370-8

Lonsdale, A., Davis, M. J., Doblin, M. S., and Bacic, A. (2016). Better Than Nothing? Limitations of the Prediction Tool SecretomeP in the Search for Leaderless Secretory Proteins (LSPs) in Plants. *Front. Plant Sci.* 7, 1451. doi:10.3389/fpls.2016.01451

Lundberg, E., and Borner, G. H. H. (2019). Spatial Proteomics: a Powerful Discovery Tool for Cell Biology, *Nat. Rev. Mol. Cell Biol.* 20, 285–302. doi:10.1038/s41580-018-0094-y

Lv, Z., Wang, P., Zou, Q., and Jiang, Q. (2021). Identification of Sub-golgi Protein Localization by Use of Deep Representation Learning Features. *Bioinformatics* 36 (24), 5600–5609. doi:10.1093/bioinformatics/btaa1074

Madeo, G., Savojardo, C., Martelli, P. L., and Casadio, R. (2021). BetAware-Deep: An Accurate Web Server for Discrimination and Topology Prediction of Prokaryotic Transmembrane β-barrel Proteins. *J. Mol. Biol.* 433 (11), 166729. doi:10.1016/j.jmb.2020.166729

Martelli, P. L., Savojardo, C., Fariselli, P., Tartari, G., and Casadio, R. (2021). Computer-Aided Prediction of Protein Mitochondrial Localization. *Methods Mol. Biol. Clift. N.J.* 2275, 433–452. doi:10.1007/978-1-0716-1262-0_28

Nemes, P. (2021). Mass Spectrometry Comes of Age for Subcellular Organelles. *Nat. Methods* 18 (10), 1157–1158. doi:10.1038/S41592-021-01287-0

Nielsen, H., Petsalaki, E. I., Zhao, L., and Stühler, K. (2019). Predicting Eukaryotic Protein Secretion without Signals. *Biochim. Biophys. Acta Proteins Proteom* 1867 (12), 140174. doi:10.1016/j.bbapap.2018.11.011

Nielsen, H., Tsirigos, K. D., Brunak, S., von Heijne, G., and Gunnar Von Heijne, ·. (2019). A Brief History of Protein Sorting Prediction. *Protein J.* 38 (3), 200–216. doi:10.1007/s10930-019-09838-3

Orre, L. M., Vesterlund, M., Pan, Y., Arslan, T., Zhu, Y., Fernandez Woodbridge, A., et al. (2019). SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. *Mol. Cell* 73 (1), 166–e7. doi:10.1016/j.molcel.2018.11.035

Ouyang, W., Winsnes, C. F., Hjelmare, M., Cesnik, A. J., Åkesson, L., Xu, H., et al. (2019). Analysis of the Human Protein Atlas Image Classification Competition. *Nat. Methods* 16 (12), 1254–1261. doi:10.1038/s41592-019-0658-6

Pan, G., Sun, C., Liao, Z., and Tang, J. (2021). Machine and Deep Learning for Prediction of Subcellular Localization. *Methods Mol. Biol. Clift. N.J.* 2361, 249–261. doi:10.1007/978-1-0716-1641-3_15

Pärnamaa, T., and Parts, L. (2017). Accurate Classification of Protein Subcellular Localization from High-Throughput Microscopy Images Using Deep Learning. *G3 Genes|Genomes|Genetics* 7 (5), 1385–1392. doi:10.1534/g3.116.033654

Paul, I., White, C., Turcinovic, I., and Emili, A. (2021). Imaging the Future: the Emerging Era of Single-Cell Spatial Proteomics. *FEBS J.* 288 (24), 6990–7001. doi:10.1111/febs.15685

Peabody, M. A., Lau, W. Y. V., Hoad, G. R., Jia, B., Maguire, F., Gray, K. L., et al. (2020). PSORTm: a Bacterial and Archaeal Protein Subcellular Localization Prediction Tool for Metagenomics Data. *Bioinformatics* 36 (10), 3043–3048. doi:10.1093/bioinformatics/btaa136

Pino, L., and Schilling, B. (2021). Proximity Labeling and Other Novel Mass Spectrometric Approaches for Spatiotemporal Protein Dynamics. *Expert Rev. Proteomics* 18 (9), 757–765. doi:10.1080/14789450.2021.1976149

Rajendran, L., Knölker, H. J., and Simons, K. (2010). Subcellular Targeting Strategies for Drug Design and Delivery. *Nat. Rev. Drug Discov.* 9 (1), 29–42. doi:10.1038/nrd2897

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T. J., et al. (2021). MitoCarta3.0: an Updated Mitochondrial Proteome Now with Sub-organelle Localization and Pathway Annotations. *Nucleic Acids Res.* 49 (D1), D1541–D1547. doi:10.1093/nar/gkaa1011

Sahu, S. S., Loaiza, C. D., and Kaundal, R. (2020). Plant-mSubP: a Computational Framework for the Prediction of Single- and Multi-Target Protein Subcellular Localization Using Integrated Machine-Learning Approaches. *AoB PLANTS* 12 (3), plz068. doi:10.1093/aobpla/plz068

Savojardo, C., Bruciaferri, N., Tartari, G., Martelli, P. L., and Casadio, R. (2020). DeepMito: Accurate Prediction of Protein Sub-mitochondrial Localization Using Convolutional Neural Networks. *Bioinformatics* 36 (1), 56–64. doi:10.1093/bioinformatics/btz512

Savojardo, C., Martelli, P. L., Tartari, G., and Casadio, R. (2020). Large-scale Prediction and Analysis of Protein Sub-mitochondrial Localization with DeepMito. *BMC Bioinforma.* 21 (Suppl. 8), 266. doi:10.1186/s12859-020-03617-z

Savulescu, A. F., Bouilhol, E., Beaume, N., and Nikolski, M. (2021). Prediction of RNA Subcellular Localization: Learning from Heterogeneous Data Sources. *IScience* 24 (11), 103298. doi:10.1016/j.isci.2021.103298

Schneider, K., Zimmer, D., Nielsen, H., Herrmann, J. M., and Mühlhaus, T. (2021). iMLP, a Predictor for Internal Matrix Targeting-like Sequences in Mitochondrial Proteins. *Biol. Chem.* 402 (8), 937–943. doi:10.1515/hsz-2021-0185

Semwal, R., and Varadwaj, P. K. (2020). HumDLoc: Human Protein Subcellular Localization Prediction Using Deep Neural Network. *Curr. Genomics* 21 (7), 546–557. doi:10.2174/1389202921999200528160534

Shen, Y., Ding, Y., Tang, J., Zou, Q., and Guo, F. (2020). Critical Evaluation of Web-Based Prediction Tools for Human Protein Subcellular Localization. *Brief. Bioinform* 21 (5), 1628–1640. doi:10.1093/bib/bbz106

Shreyashree, S., Sunagar, P., Rajarajeswari, S., and Kanavalli, A. (2022). A Literature Review on Bidirectional Encoder Representations from Transformers, *Inventive*

ComputatioInventive Computation and Information Technologiesn and Information Technologies, 336, 305–320. doi:10.1007/978-981-16-6723-7_23

Singhal, N., Garg, A., Singh, N., Gulati, P., Kumar, M., and Goel, M. (2021). Efficacy of Signal Peptide Predictors in Identifying Signal Peptides in the Experimental Secretome of Picrophilous Torridus, a Thermoacidophilic Archaeon. *PloS One*, 16(8), e0255826. doi:10.1371/journal.pone.0255826

Su, R., He, L., Liu, T., Liu, X., and Wei, L. (2021). Protein Subcellular Localization Based on Deep Image Features and Criterion Learning Strategy. *Briefings Bioinforma.* 22 (4). doi:10.1093/bib/bbaa313

Tahir, M., and Idris, A. (2020). MD-LBP: An Efficient Computational Model for Protein Subcellular Localization from HeLa Cell Lines Using SVM. *Cbio* 15 (3), 204–211. doi:10.2174/1574893614666190723120716

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.*. doi:10.1038/s41587-021-01156-3

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., et al. (2017). A Subcellular Map of the Human Proteome. *Science* 356 (6340). doi:10.1126/science.aal3321

Tokmakov, A. A., Kurotani, A., and Sato, K. I. (2021). Protein pI and Intracellular Localization. *Front. Mol. Biosci.* 8, 775736. doi:10.3389/fmolb.2021.775736

Tu, Y., Lei, H., Shen, H.-B., and Yang, Y. (2022). SIFLoc: a Self-Supervised Pre-training Method for Enhancing the Recognition of Protein Subcellular Localization in Immunofluorescence Microscopic Images. *Briefings Bioinforma.* 23. doi:10.1093/bib/bbab605

Ullah, M., Han, K., Hadi, F., Xu, J., Song, J., and Yu, D. J. (2021). PScL-HDeep: Image-Based Prediction of Protein Subcellular Location in Human Tissue Using Ensemble Learning of Handcrafted and Deep Learned Features with Two-Layer Feature Selection. *Brief. Bioinform* 22 (6). doi:10.1093/bib/bbab278

Wang, D., Zhang, Z., Jiang, Y., Mao, Z., Wang, D., Lin, H., et al. (2021). DM3Loc: Multi-Label mRNA Subcellular Localization Prediction and Analysis Based on Multi-Head Self-Attention Mechanism. *Nucleic Acids Res.* 49 (8), e46. doi:10.1093/nar/gkab016

Wang, F., and Wei, L. (2022). Multi-scale Deep Learning for the Imbalanced Multi-Label Protein Subcellular Localization Prediction Based on Immunohistochemistry Images, *Bioinformatics* 11, btac123. doi:10.1093/bioinformatics/btac123

Wang, G., Zhai, Y. J., Xue, Z. Z., and Xu, Y. Y. (2021). Improving Protein Subcellular Location Classification by Incorporating Three-Dimensional Structure Information. *Biomolecules* 11 (11). doi:10.3390/biom11111607

Wang, X., Jin, Y., and Zhang, Q. (2020). DeepPred-SubMito: A Novel Submitochondrial Localization Predictor Based on Multi-Channel Convolutional Neural Network and Dataset Balancing Treatment. *Int. J. Mol. Sci.* 21 (16). doi:10.3390/ijms21165710

Wattanapornprom, W., Thammarongtham, C., Hongsthong, A., and Lertampaiporn, S. (2021). Ensemble of Multiple Classifiers for Multilabel Classification of Plant Protein Subcellular Localization. *Life* 11 (4), 293. doi:10.3390/life11040293

Wu, J. M., Liu, Y. C., and Chang, D. T. (2019). SigUNet: Signal Peptide Recognition Based on Semantic Segmentation. *BMC Bioinforma.* 20 (24), 677. doi:10.1186/s12859-019-3245-z

Wu, L., Huang, S., Wu, F., Jiang, Q., Yao, S., and Jin, X. (2020). Protein Subnuclear Localization Based on Radius-SMOTE and Kernel Linear Discriminant Analysis Combined with Random Forest. *Electronics* 9 (10), 1566. doi:10.3390/electronics9101566

Wu, Z., Yang, K. K., Liszka, M. J., Lee, A., Batzilla, A., Wernick, D., et al. (2020). Signal Peptides Generated by Attention-Based Neural Networks. *ACS Synth. Biol.* 9 (8), 2154–2161. doi:10.1021/acssynbio.0c00219

Xie, H., and Ding, X. (2022). The Intriguing Landscape of Single-Cell Protein Analysis, *Adv. Sci.* 9, 2105932. doi:10.1002/advs.202105932

Xu, Y.-Y., Yao, L.-X., and Shen, H.-B. (2018). Bioimage-based Protein Subcellular Location Prediction: a Comprehensive Review. *Front. Comput. Sci.* 12 (1), 26–39. doi:10.1007/s11704-016-6309-5

Xu, Y. Y., Shen, H. B., and Murphy, R. F. (2020). Learning Complex Subcellular Distribution Patterns of Proteins via Analysis of Immunohistochemistry Images. *Bioinformatics* 36 (6), 1908–1914. doi:10.1093/bioinformatics/btz844

Xu, Y. Y., Zhou, H., Murphy, R. F., and Shen, H. B. (2021). Consistency and Variation of Protein Subcellular Location Annotations. *Proteins* 89 (2), 242–250. doi:10.1002/prot.26010

Xue, M.-Q., Zhu, X.-L., Wang, G., and Xu, Y.-Y. (2021). DULoc: Quantitatively Unmixing Protein Subcellular Location Patterns in Immunofluorescence Images Based on Deep Learning Features. *Bioinformatics* 38, 827–833. doi:10.1093/bioinformatics/btab730

Xue, Z. Z., Wu, Y., Gao, Q. Z., Zhao, L., and Xu, Y. Y. (2020). Automated Classification of Protein Subcellular Localization in Immunohistochemistry Images to Reveal Biomarkers in Colon Cancer. *BMC Bioinforma.* 21 (1), 398. doi:10.1186/s12859-020-03731-y

Yang, F., Liu, Y., Wang, Y., Yin, Z., and Yang, Z. (2019). MIC_Locator: a Novel Image-Based Protein Subcellular Location Multi-Label Prediction Model Based on Multi-Scale Monogenic Signal Representation and Intensity Encoding Strategy. *BMC Bioinforma.* 20 (1), 522. doi:10.1186/s12859-019-3136-3

Yao, Y. H., Lv, Y. P., Li, L., Xu, H. M., Ji, B. B., Chen, J., et al. (2019). Protein Sequence Information Extraction and Subcellular Localization Prediction with Gapped K-Mer Method. *BMC Bioinforma.* 20 (22), 719. doi:10.1186/s12859-019-3232-4

Yu, B., Qiu, W., Chen, C., Ma, A., Jiang, J., Zhou, H., et al. (2020). SubMito-XGBoost: Predicting Protein Submitochondrial Localization by Fusing Multiple Feature Information and eXtreme Gradient Boosting. *Bioinformatics* 36 (4), 1074–1081. doi:10.1093/bioinformatics/btz734

Yuan, X., Pang, E., Lin, K., and Hu, J. (2021). Deep Protein Subcellular Localization Predictor Enhanced with Transfer Learning of GO Annotation. *IEEJ Trans. Elec Engng* 16 (4), 559–567. doi:10.1002/tee.23330

Zhang, Q., Li, S., Zhang, Q., Zhang, Y., Han, Y., Chen, R., et al. (2021). MpsLDA-ProSVM: Predicting Multi-Label Protein Subcellular Localization by wMLDAe Dimensionality Reduction and ProSVM Classifier. *Chemom. Intelligent Laboratory Syst.* 208, 104216. doi:10.1016/j.chemolab.2020.104216

Zhu, L., Hofestadt, R., and Ester, M. (2019). Tissue-Specific Subcellular Localization Prediction Using Multi-Label Markov Random Fields. *IEEE/ACM Trans. Comput. Biol. Bioinform* 16 (5), 1471–1482. doi:10.1109/TCBB.2019.2897683