



PLA-GNN: Computational inference of protein subcellular location alterations under drug treatments with deep graph neural networks

Ren-Hua Wang, Tao Luo, Han-Lin Zhang, Pu-Feng Du ^{*}

College of Intelligence and Computing, Tianjin University, Tianjin, 300350, China



ARTICLE INFO

Keywords:
PLA-GNN
TSA
Bortezomib
Tacrolimus
Mis-localized proteins

ABSTRACT

The aberrant protein sorting has been observed in many conditions, including complex diseases, drug treatments, and environmental stresses. It is important to systematically identify protein mis-localization events in a given condition. Experimental methods for finding mis-localized proteins are always costly and time consuming. Predicting protein subcellular localizations has been studied for many years. However, only a handful of existing works considered protein subcellular location alterations. We proposed a computational method for identifying alterations of protein subcellular locations under drug treatments. We took three drugs, including TSA (trichostain A), bortezomib and tacrolimus, as instances for this study. By introducing dynamic protein-protein interaction networks, graph neural network algorithms were applied to aggregate topological information under different conditions. We systematically reported potential protein mis-localization events under drug treatments. As far as we know, this is the first attempt to find protein mis-localization events computationally in drug treatment conditions. Literatures validated that a number of proteins, which are highly related to pharmacological mechanisms of these drugs, may undergo protein localization alterations. We name our method as PLA-GNN (Protein Localization Alteration by Graph Neural Networks). It can be extended to other drugs and other conditions. All datasets and codes of this study has been deposited in a GitHub repository (<https://github.com/quinlanW/PLA-GNN>).

1. Introduction

Proteins are sorted to appropriate subcellular compartments or secreted outside the cell after or along with the translation process [1,2]. The molecular function of a protein is highly correlated with its subcellular localization [3]. The aberrant translocation of a protein may affect its normal molecular function, and may involve it in an incorrect biological process [4,5]. Environmental stresses may alter protein sorting destinations [6], which is a response of a living cell to a changing environment. Protein mis-localization events are related to complex disorders, including Alzheimer's disease [7], amyotrophic lateral sclerosis [8] and acute myeloid leukemia [9]. Interfering protein sorting process by pharmaceutical substances is a kind of therapies to complex diseases [10,11]. Several practices have been performed [12].

Human protein subcellular localizations have been systematically mapped by experiments [13]. However, this mapping process is incredibly expensive and time consuming [14]. It is unlikely to determine every mis-localization event in a given cellular state by this way. The cellular state here means a cell in its normal living state or a disease

state or a disease state with drug perturbations. Therefore, computational estimations are considered as alternative approaches to determine protein mis-localization events [15–17].

In a fixed cellular state, predicting protein subcellular locations has been well studied [18–21]. There are many computational methods for predicting protein subcellular locations. These methods can predict protein subcellular location in a tissue-specific or a lineage-specific manner [20,22–24]. These computational approaches utilized protein sequences [18,19,25], structures [26,27] and interactions [16,28] to estimate protein subcellular locations. However, only a handful of studies tried to predict alterations of protein subcellular locations in different cellular states [17,29–31]. These studies generally fall into two categories, the image-based and the omics-based methods.

Image-based methods take immunohistochemical images [20] or immunofluorescence images [21] as input. They use image analysis algorithms along with machine learning models to identify protein subcellular locations in different cellular states. By comparing prediction results in different cellular states, these methods can report protein mis-localization events [20,21]. Omics-based methods take protein

* Corresponding author.

E-mail addresses: renhua_wang@tju.edu.cn (R.-H. Wang), luo_tao@tju.edu.cn (T. Luo), hanlin_zhang@tju.edu.cn (H.-L. Zhang), pdu@tju.edu.cn (P.-F. Du).

sequences and interactions as input. Systems biology methods are used to report mis-localization events. For example, Lee et al. integrated protein sequences, PPI (protein-protein interaction) networks and gene expression profiles to find mis-localized proteins in gliomas [31]. For another example, the PROLocalizer predictor used sequence mutations to detect protein mis-localizations in diseases [29,30].

Neither strategy can be applied as a common pipeline. Image-based methods face two challenges: the lack of fluorescence images and the limited resolution in immunohistochemical images [32]. Omics-based methods usually use the PPI networks in a normal state to mimic PPI networks in other cellular states, assuming the changes of PPIs can be ignored. This is due to the fact that PPI networks in different cellular states are usually not available [16]. However, this assumption has a paradox. Given that PPIs are usually physical interactions, if the sub-cellular location of a protein was changed, it would be less likely to interact with proteins in its original subcellular compartments. Its interacting proteins would be surely changed also. Therefore, assuming a universal PPI network in various cellular states just discarded the most informative changes. Although gene expressions may rescue this assumption to some extent, the prediction performances are inevitable affected [16].

Li et al. proposed the DPPN-SVM [17] method in accordance to the differential network biology concept [33]. They used gene expression profiles to estimate PPI networks in different cellular states. The PPI network in a given cellular state can be estimated by adding and removing certain interactions from the normal state network. By using this strategy, DPPN-SVM identified a serial of potentially mis-localized proteins in the breast cancer and validated them by other literatures.

Although attempts have been made in predicting mis-localized proteins in diseases, as far as we know, no existing study can computationally identify mis-localized proteins in drug therapies. In this work, we propose a new computational method for predicting mis-localized proteins in drug therapies. We estimated PPI networks under drug treatments. Graph neural network models were trained to aggregate high-order topological information of PPI networks, as it is reported that the high-order interaction information is more dominant in PPI networks [34,35]. We name our method as PLA-GNN (Protein Localization Alterations by Graph Neural Network).

We took TSA (trichostatin A), bortezomib, and tacrolimus as instances in our study. TSA, an antifungal biotic, is a potent and specific inhibitor of histone deacetylase (HDAC) activity [36]. Bortezomib is a dipeptide boronic acid derivative and a proteasome inhibitor. It is reported that bortezomib enhances Docetaxel-induced cell death level and has an inhibitory effect on cell migration in breast cancer [37]. Tacrolimus is a calcineurin inhibitor for preventing rejections in transplants, and for treating moderate to severe atopic dermatitis [38]. Our results indicated that, when administered, several proteins, which are highly related to pharmacological mechanisms of these drugs, may undergo protein localization alterations. This may provide useful information for pharmacological studies. Our method has the potential to become a common pipeline for predicting protein localization alterations in drug therapies.

2. Materials and methods

2.1. PPI network

We downloaded PPI records from the BioGRID database [39]. To construct a high-quality working dataset, we screened the raw PPI records strictly according to the following steps: (1) Only interactions between two human proteins were kept. (2) All interactions between two identical proteins were excluded. (3) Duplicate records were reduced. All redundant records were removed. (4) Non-physical interaction records were excluded. We kept only interactions with a type MI:0915 (physical association), MI:0407 (direct interaction) or MI:0403 (co-localization). All other types of interactions were excluded. After

above procedures, we obtained 1,376,072 interactions involving 24,041 proteins.

GO annotations were obtained from the UniProt database according to the UniProt/BioGRID ID mappings. We chose 12 GO terms for sub-cellular location annotations, including Cell cortex (GO:0005938), Cytosol (GO:0005829), Actin cytoskeleton (GO:0015629), Golgi apparatus (GO:0005794), Endoplasmic reticulum (GO: 0005783), Nucleolus (GO:0005730), Peroxisome (GO: 0005777), Mitochondrion (GO:0005739), Lysosome (GO:0005764), Centrosome (GO:0005813), Nucleus (GO:0005634) and Plasma membrane (GO:0005886). We chose only those GO terms with high confidence. The evidence codes: IDA (Inferred from Direct Assay), IEA (Inferred from Electronic Annotation), IPI (Inferred from Physical Interaction) and HDA (Inferred from High Throughput Direct Assay) were recognized as reliable evidences of the GO annotations. 8211 proteins were annotated with at least one of the above 12 subcellular locations.

Among the 8211 BioGRID proteins with annotations, 5249 proteins were with only one subcellular localization, 2134 proteins with two localizations, 629 proteins with three localizations, 163 proteins with four localizations, 30 proteins with five localizations, and six proteins with six localizations. Fig. 1(A) illustrated the breakdown of the dataset on localization multiplicity. The number of proteins in each subcellular location is presented in Fig. 1(B).

2.2. Co-expression network

We applied three GEO datasets in our study, including GSE74572, GSE30931, and GSE27182. The GSE74572 dataset presents gene expressions of A549 cell line under TSA treatment with controls. It contains 21 samples, including 18 cases and 3 controls, we took the TSA treatment only subset. The GSE30931 dataset studied breast cancer under bortezomib perturbations. It contains 12 samples, including 9 cases and 3 controls. We took the bortezomib perturbation only subset. The GSE27182 dataset provides expression profiles of the HK-2 cell line under tacrolimus perturbations with controls. It contains 12 samples, including 6 cases and 6 controls. We took the subset of “t12h”. The case and control samples in the dataset were grouped separately in each dataset. The hg133plus2.db and illuminaHumanv4.db annotation package was applied with Bioconductor to map gene expression values to proteins. In the case of many-to-one mapping, we used the average value as the final expression value of the gene.

Let $e_{k,u}$ be the expression value mapping to the u -th protein in the k -th sample, and c the number of samples in a group. We define the sample-wise expression vector \mathbf{e}_u as follows:

$$\mathbf{e}_u = [e_{1,u}, e_{2,u}, \dots, e_{c,u}]^T. \quad (1)$$

We now define the pair-wise PCC (Pearson Correlation Coefficient) of expression values between the u -th and the v -th protein as follows:

$$\rho_{u,v} = \frac{\mathbf{e}_u^T \mathbf{e}_v}{\sqrt{\mathbf{e}_u^T \mathbf{e}_u} \sqrt{\mathbf{e}_v^T \mathbf{e}_v}}, \quad (2)$$

where $\rho_{u,v}$ is the PCC between the u -th and the v -th proteins. Regardless of whether two proteins interact with each other, their PCCs are calculated as above.

2.3. Edge clustering coefficients

The Edge Clustering Coefficient (ECC) was originally proposed for detecting community structures in complex networks [40]. It has been applied in identifying essential proteins [41] and in predicting protein subcellular localizations [28]. ECC can be used to describe how close two interacting proteins are [41]. It can also be used as an indicator of how likely two proteins have a common subcellular location [28]. ECC is defined as follows:

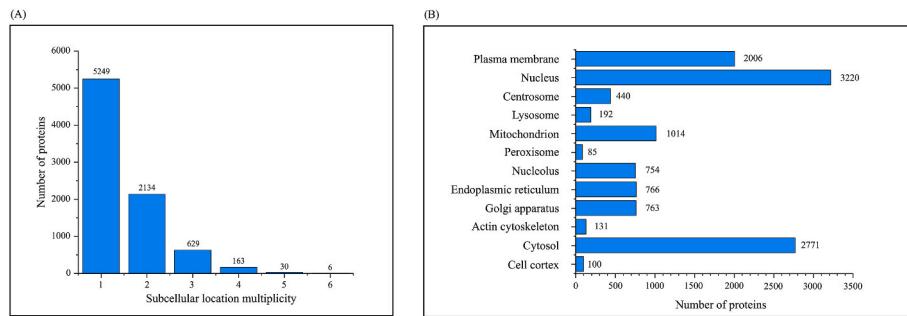


Fig. 1. The breakdown of the dataset. (A) The breakdown of the dataset for different localization multiplicity. (B) The number of proteins in each subcellular location.

$$\eta_{u,v} = \frac{z_{u,v}}{\min(d_u - 1, d_v - 1)}, \quad (3)$$

where $\eta_{u,v}$ is the ECC between the u -th and the v -th proteins, $z_{u,v}$ the number of triangles that involve the edge between the u -th and the v -th proteins, and d_u and d_v the degree of the u -th and the v -th proteins. The denominator represents the largest number of possible triangles that involve the u -th and the v -th proteins. We set $\eta_{u,v}$ to zero in the case that the denominator is zero.

2.4. Overall design of the PLA-GNN protocol

We introduce the dynamic PPI network for predicting alteration of protein subcellular locations. The topology structure of the PPI network in the drug perturbation state is estimated using the PPI network in the control state and differences of co-expression networks in different cellular states. We applied three layers GraphSAGE encoder to aggregate information from neighboring nodes in the network and a multilayer

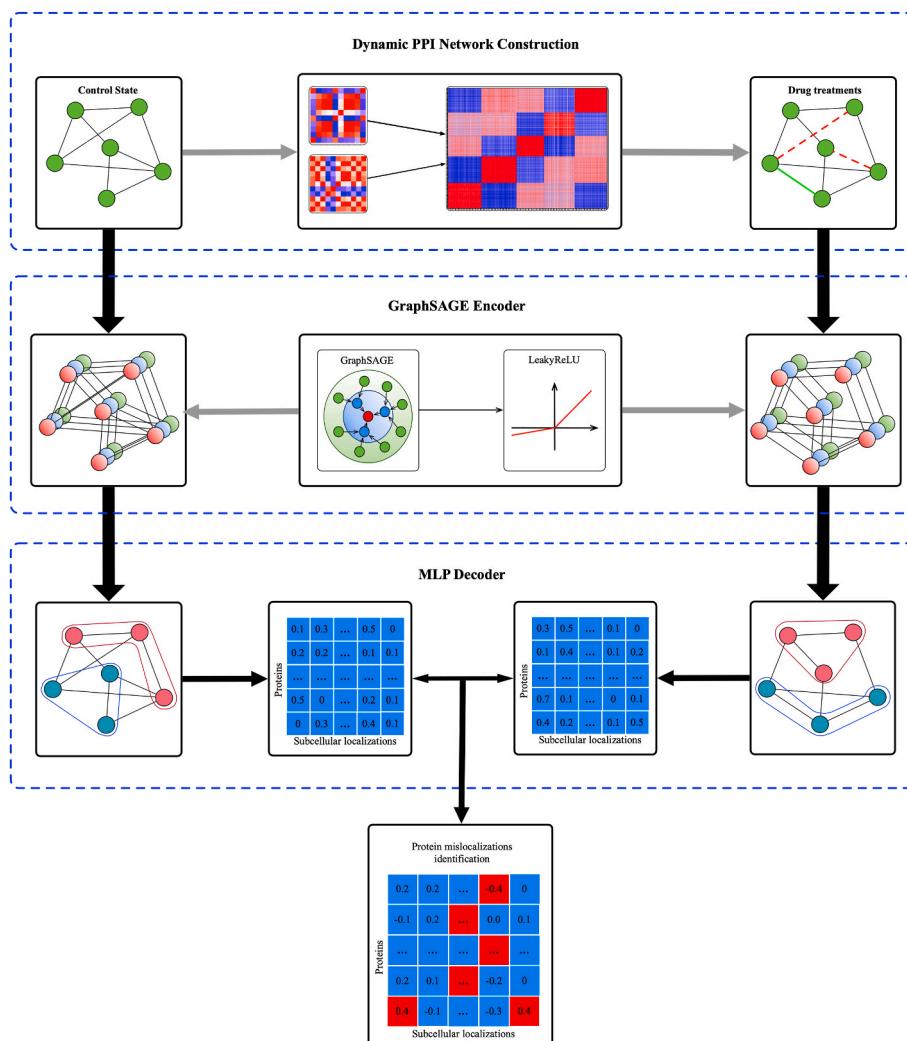


Fig. 2. - The flowchart of PLA-GNN. PLA-GNN contains three consecutive modules, which are the dynamic network construction module, the GraphSAGE feature aggregation module and the MLP decoding module. The PPI network in the drug treatment state is estimated from the control state and differential gene expression matrix. GraphSAGE aggregate features over the PPI network. MLP predictors produce the finale results.

perceptron as a decoder to quantitatively predict the localization of proteins. For each protein, a score was given to each possible subcellular localization to represent its potential in locating to that subcellular location. The alterations of protein subcellular locations are obtained by quantitatively comparing the localization scores between different states. The architecture of our predictor is illustrated as in Fig. 2.

2.5. Dynamic PPI network construction

In order to construct the PPI network in the drug perturbation state, we adjusted the topology of the PPI network according to the differences of PCC values under different conditions.

We term the drug perturbation state as θ_b , and the control state as θ_c . For the u -th and the v -th protein, we compute the PCC in θ_c and θ_b respectively, which can be noted as $\rho_{u,v}(\theta_c)$ and $\rho_{u,v}(\theta_b)$. We define the difference of PCC as follows:

$$\delta_{u,v} = \rho_{u,v}(\theta_b) - \rho_{u,v}(\theta_c), \quad (4)$$

and two threshold parameters:

$$t_+ = \delta + \kappa\sigma, \text{ and} \quad (5)$$

$$t_- = \delta - \kappa\sigma, \quad (6)$$

where δ is the average value of all $\delta_{u,v}$, σ the standard deviation of all $\delta_{u,v}$, and κ a parameter.

If the u -th protein and the v -th protein are two interacting proteins in the control state, the interaction would be removed under drug treatment if $\delta_{u,v} < t_-$ is satisfied. Similarly, if the u -th and the v -th proteins are two non-interacting proteins in the control state, the interaction between them should be established under drug treatment if $\delta_{u,v} > t_+$ is satisfied.

2.6. GraphSAGE encoder

We applied three layers GraphSAGE [42] encoder to aggregate the information in PPI networks in the control state and the drug perturbation state, respectively.

For the u -th protein in the PPI network, we defined the information vector ρ_u and η_u as follows:

$$\rho_u = [\rho_{u,1}, \rho_{u,2}, \dots, \rho_{u,m}]^T, \text{ and} \quad (7)$$

$$\eta_u = [\eta_{u,1}, \eta_{u,2}, \dots, \eta_{u,m}]^T, \quad (8)$$

where m is the total number of proteins in the network. To improve the GraphSAGE efficiency, PCA (Principle Component Analysis) was used to reduce the dimension of ρ_u and η_u to 250. We stitched the vector e_u , ρ_u and η_u together to get a universal feature vector f_u as follows:

$$f_u = [e_u \ \rho_u \ \eta_u]^T \quad (9)$$

The universal feature vectors were aggregated by the GraphSAGE encoder.

2.7. Multi-layer perceptron decoder

For the u -th protein, we note its score to the i -th subcellular localization in cellular state θ as $p_{u,i}(\theta)$ in the prediction results. If the u -th protein has subcellular locations annotations, we note this as a binary variable $r_{u,i}(\theta)$, which can be defined as follows:

$$r_{u,i}(\theta) = \begin{cases} 1 & \text{The } u\text{-th protein has the } i\text{-th subcellular location annotation,} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

We applied the multi-label loss function as follows:

$$L = \frac{1}{m} \sum_{i=1}^n \sum_{u=1}^m w_i r_{u,i}(\theta) s(p_{u,i}(\theta)) + (1 - r_{u,i}(\theta)) s(-p_{u,i}(\theta)), \quad (11)$$

where $s(\cdot)$ is a logit function, as follows:

$$s(x) = -\ln\left(\frac{1}{1 + \exp(-x)}\right), \quad (12)$$

w_i a location weight parameter, as follows:

$$w_i = \frac{m_0 - m_i}{m_i}, \quad (13)$$

m_0 the total number of proteins with subcellular location annotations, m_i the total number of proteins with the i -th subcellular location annotations, and n the total number of subcellular locations.

2.8. Subcellular location alteration scores

The localization score $p_{u,i}(\theta)$ varies in different ranges for different subcellular locations, we defined the standardized localization score as follows:

$$q_{u,i}(\theta) = \frac{\hat{p}_{u,i}(\theta)}{\sum_{i=1}^n \hat{p}_{u,i}(\theta)}, \quad (14)$$

where

$$\hat{p}_{u,i}(\theta) = \frac{p_{u,i}(\theta) - \min_p p_{u,i}(\theta)}{\max_p p_{u,i}(\theta) - \min_p p_{u,i}(\theta)}. \quad (15)$$

To quantify the likelihood of a protein mis-localization in the drug perturbation state, we defined the localization alteration score of the u -th protein in the i -th subcellular localization as follows:

$$\varphi_{u,i} = \frac{q_{u,i}(\theta_i) - q_{u,i}(\theta_c)}{q_{u,i}(\theta_c)}. \quad (16)$$

The $\varphi_{u,i}$ indicates the extent that the u -th protein would acquire or abandon the i -th subcellular localization. For each protein, we define the following two boundaries:

$$\sup(\varphi_u) = \max_i \varphi_{u,i}, \text{ and} \quad (17)$$

$$\inf(\varphi_u) = \min_i \varphi_{u,i}. \quad (18)$$

We sorted the proteins according to the $\sup(\varphi_u)$ and the $\inf(\varphi_u)$ in descending and ascending orders, respectively. The top-ranked proteins within a fixed proportion of the entire list are considered as proteins with localization alteration. Since protein mis-localization events are thought to be rare, this proportion is fixed as 5% in this work.

2.9. Performance evaluation methods

We performed 10 times 10-fold cross-validation to evaluate the prediction performance of our method in a single cellular state. We average the results of 10 times 10-fold cross-validation as the final result. It should be noted that the prediction performances in a single cellular state have no implication for the performance in predicting alteration of protein subcellular localizations between different states. For a single cellular state, the subcellular location of the u -th protein is determined by the standardized localization score $q_{u,i}(\theta)$.

We assigned the i -th subcellular localization to the u -th protein, if the following condition is satisfied:

$$q_{u,i}(\theta) \geq \max_i q_{u,i}(\theta) - \alpha \left(\max_i q_{u,i}(\theta) - \min_i q_{u,i}(\theta) \right), \quad (19)$$

where $\alpha \in [0,1]$ is a parameter. We use $S_u(\theta)$ to note the set of subcellular locations that are assigned to the i -th protein using the above condition.

Three statistics, including aiming (AIM), coverage (COV), and multi-label accuracy (mlACC) were applied to measure the prediction performances [43]. These statistics are defined as follows:

$$AIM = \frac{1}{m} \sum_{u=1}^m \frac{|S_u(\theta) \cap S_u|}{|S_u(\theta)|}, \quad (20)$$

$$COV = \frac{1}{m} \sum_{u=1}^m \frac{|S_u(\theta) \cap S_u|}{|S_u|}, \quad (21)$$

$$mlACC = \frac{1}{m} \sum_{u=1}^m \frac{|S_u(\theta) \cap S_u|}{|S_u(\theta) \cup S_u|}, \quad (22)$$

where S_u is the set of experimental protein subcellular localizations, and $|.|$ the cardinal operator in the set theory.

2.10. Parameter calibration

The parameters κ and α are calibrated in the following way. The parameter κ was chosen in [2.0, 3.0] manually to constrain the number of differential PCC values, which are out of the range $[t_-, t_+]$, to be just less than 0.3% of all values. When κ increases, the number of differential PCC values out of the range decreases. We set the κ values for TSA, bortezomib and tacrolimus treatment conditions to 2.91, 2.75 and 2.99, respectively. The parameter α was chosen in {0.1, 0.2, 0.3}. When choosing the value of α , we considered two factors. One is the prediction performances in the control state, while the other is the balance of the results among subcellular locations. We finally choose $\alpha = 0.1$. We manually adjusted the learning rate and epoch in the multi-layer perceptron. The learning rate is set to 0.00005, and epoch 200.

3. Results and discussions

3.1. Network topology adjustment

The PPI network has a total of 1,376,072 interactions in the control state. When creating the dynamic PPI network, a total of 577,969,681 differential PCC values of protein pairs are calculated for each of the three drugs. Topology adjustments were carried out according to these values. We finally obtained 2,202,772 interactions with the TSA treatment, 2,295,812 interactions with the bortezomib treatment, and 1,367,114 interactions with the tacrolimus treatment. Distributions of differential PCC values under drug treatments are presented in Fig. 3, where the distributions for all pairs, interacting pairs and non-interacting pairs are illustrated respectively.

We can see that the distribution of differential PCC values of interacting pairs and non-interacting pairs are almost the same for a given drug. However, for different drugs, the distributions have observable differences. After the network adjustment, 3,414, 3872 and 8958 interactions were removed in the TSA, bortezomib and tacrolimus treatment state. In the meantime, 830,114 and 923,612 interactions were established in the TSA and bortezomib perturbation states, respectively. Since the interacting pairs only occupy 0.23% of all protein pairs, our adjustment strategy surely results in more interaction established than removed. Interestingly, no interaction was established in the tacrolimus treatment state.

It seems that the adjustment procedure tends to add interactions rather than remove interactions. However, this is just a coincidence observation. We take the TSA treatment as an example. In the TSA treatment condition, the number of pairs with differential PCC values less than the lower boundary is 795012, while the number of pairs with differential PCC values larger than the upper boundary is 833714. The number of removed interaction is 3414, while the number of added interactions is 830114. These numbers leave the number of pairs with differential PCC less than the lower boundary and without interaction to be 791598. Based on these numbers, we validated that the adjustment is not biased on adding the interactions ($p = 0.82$, chi-square test).

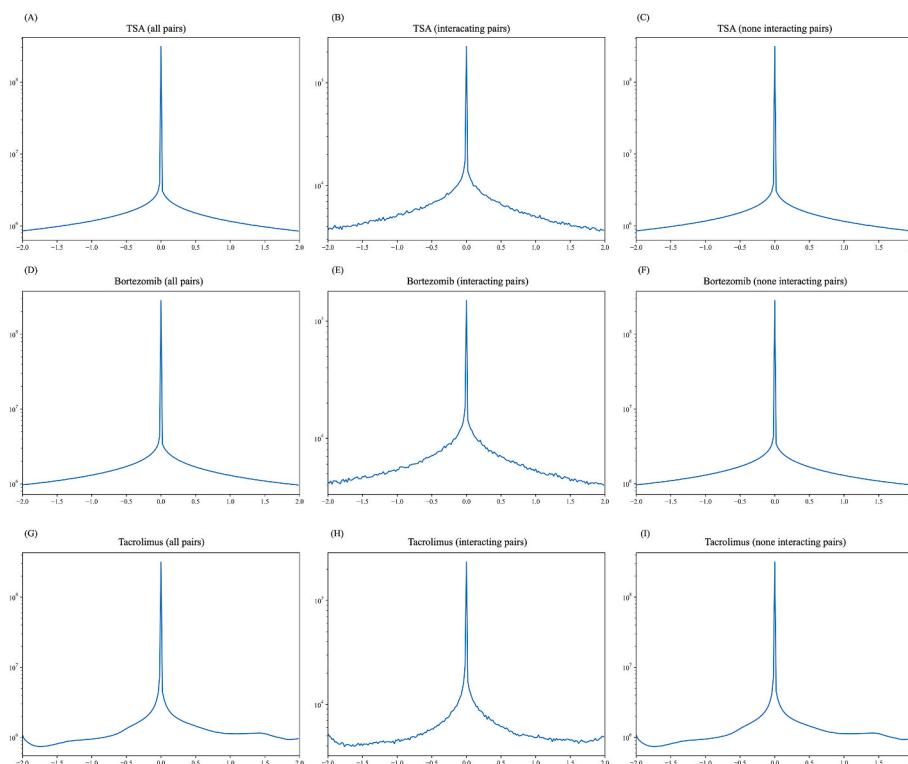


Fig. 3. Distributions of differential PCC values of protein pairs. (A) All protein pairs under TSA perturbation. (B) Only interacting pairs under TSA perturbation. (C) Only non-interacting pairs under TSA perturbation. (D) All protein pairs under bortezomib perturbation. (E) Only interacting pairs under bortezomib perturbation. (F) Only non-interacting pairs under bortezomib perturbation. (G) All protein pairs under tacrolimus perturbation. (H) Only interacting pairs under tacrolimus perturbation. (I) Only non-interacting pairs under tacrolimus perturbation.

With the TSA treatment, the average differential PCC value is -5.3×10^{-4} , with a standard deviation of 0.72. With the bortezomib treatment, the average differential PCC value is -6.2×10^{-4} , with a standard deviation of 0.68. For the tacrolimus treatment, the average differential PCC value is 0.035, with a standard deviation of 0.67. The average differential PCC values in the TSA and bortezomib treatment are essentially zero. But this value is definitely not zero for the tacrolimus treatment. From Fig. 3(A)–(F), the TSA and bortezomib treatment states both have symmetric distributions. However, this distribution is asymmetric for the tacrolimus treatment state (Fig. 3(G)–(I)). The positive half of the distribution is observably pulled up, which results in an average value larger than zero. This pushed the upper boundary for interaction adjustment beyond the maximal value of differential PCC values, resulting zero interaction established. Although this seems like some coincidence of numbers, we believe that there is a biological mechanism behind it. However, the interpretation of this phenomenon is beyond the scope of this work.

We have uploaded all adjusted interactions on Github (<https://github.com/quinlanW/PLA-GNN/blob/main/data/PPI.7z>).

3.2. Co-localization analysis of interacting proteins

We count the number of interactions, which have both proteins with subcellular location annotations, in the control state and in each of the three drug treatment states. In the control state, 572,240 interactions have both interacting proteins with annotations. Among these interactions, 319,828 pairs are within the same subcellular organelle, while 252,412 interactions are across different subcellular organelles (Fig. 4(A)). This observation is in line with literatures [44,45].

The number of interactions that are established and removed in the PPI network adjustment are also counted respectively for every drug. For the removed interactions, all three drug treatment states have

similar percentage values to the control state (Fig. 4(B)–(D)). Therefore, the removed interactions have no preference to whether the interacting pairs are within the same subcellular organelles or not (TSA, $p = 0.42$; bortezomib, $p = 0.60$; tacrolimus, $p = 0.79$; all chi-square test).

However, for the established interactions, the percentage values are obviously different to the control state. The established interactions have a strong preference to be across different subcellular organelles. For the TSA treatment state, about 65.5% established interactions are across different subcellular organelles ($p < 10^{-5}$, chi-square test). For the bortezomib treatment state, about 65.4% established interactions are across different subcellular organelles ($p < 10^{-5}$, chi-square test). Since there is no established interaction in the tacrolimus treatment state, we did not plot the corresponding pie chart.

As we have mentioned, if a protein has an altered subcellular location, its interacting proteins must be different. It would acquire new interacting partners in the new subcellular location. It would lose its original interacting peers in its original subcellular locations. Therefore, a reasonable hypothesis is that the removed interactions tend to be in the same subcellular organelle, while the established interactions are more likely to be across different subcellular organelles. The numbers in Fig. 4 validated this hypothesis.

3.3. Prediction performance analysis in the control state

To predict protein subcellular localization alteration, we established subcellular location predictors in both control and drug treatment states. Since there is no ground truth of subcellular locations in drug treatment states, we take the control state to validate the effectiveness of our method.

We performed 10 times 10-fold cross-validation to evaluate the prediction performances in the control state. We average the results of 10 times 10-fold cross-validation as the final result. The performance

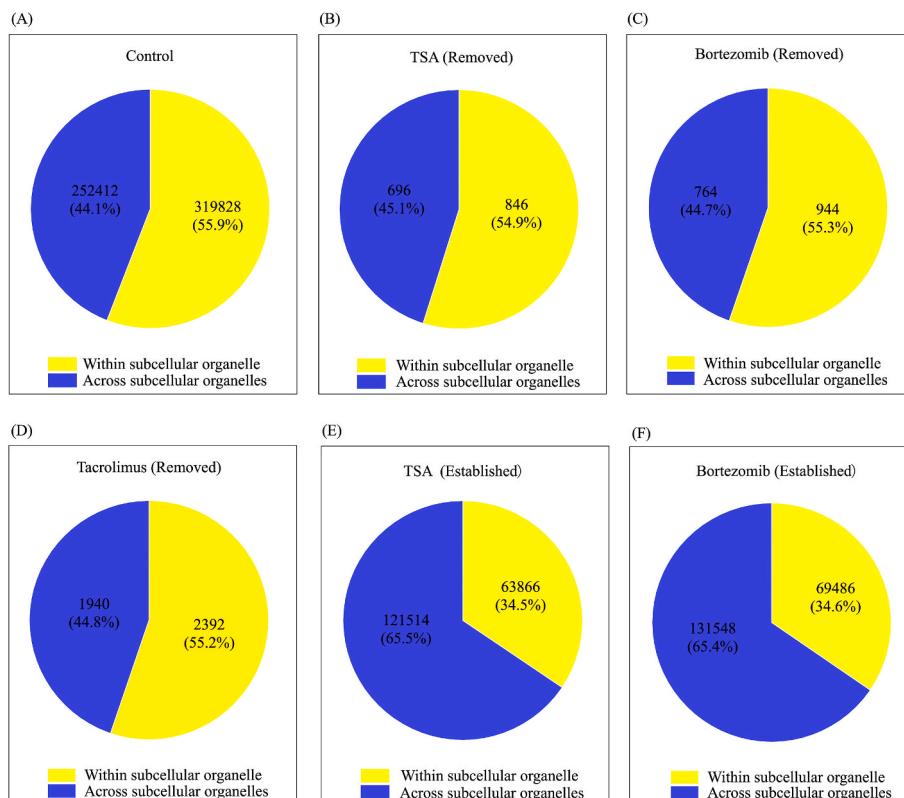


Fig. 4. - Number of interactions that are established or removed within the same subcellular organelle or across different subcellular organelles. (A) All interactions in the control state. (B) Removed interactions in the TSA treatment state. (C) Removed interactions in the Bortezomib treatment state. (D) Removed interactions in the Tacrolimus treatment state. (E) Established interactions in the TSA treatment state. (F) Established interactions in the bortezomib treatment state.

values are recorded in Fig. 5. Since gene expression profiles are different in the control state for different drugs, the prediction performances vary slightly. We used the performances of random guess trials as a baseline. Two random guess trials, limited and unlimited, were performed. The unlimited trial is to randomly assign random number of subcellular locations to a given protein. The limited trial is to randomly assign subcellular locations that have the same number to our prediction results. Though the performance values are not good enough, they are still in the same level of other multi-label protein subcellular location predictors [28]. Therefore, our method is effective in the control state.

We did not perform comparisons to other static protein subcellular location predictors for two reasons. One is that a fair comparison on our dataset is quite difficult, if not impossible. The other is that it is not necessary for us to make accurate predictions in a single state. Since we focused on the alteration of protein subcellular locations, errors in the control state and the drug treatment state will cancel each other to produce more accurate final results.

3.4. Effects of the parameter α

The cutoff value of the localization score, α , is an important parameter of the PLA-GNN method. Prediction performances in the control state may be affected by this parameter. In Table 1, we compared the prediction performances in the control state with different α . Different settings of the α parameter only affect the prediction performances in the control state slightly. However, the value of α also affects the distribution of predictions among subcellular locations. Fig. 6 compares the distribution of predictions among 12 subcellular locations under three drug perturbations with different settings of α . We measured the distribution difference of each perturbation state to the control state using the Jensen-Shannon distances [46]. Although the subcellular location distribution under drug perturbation may be different to the control state, we still try to minimize this difference, as the number of mis-localized proteins is only a small fraction. We finally choose $\alpha = 0.1$ as a universal parameter for simplicity and brevity.

Table 1

Prediction performances in the control state of each drug with different parameter values.

α	Treatment	AIM	COV	mlACC
0.1	TSA	0.512 ± 0.019	0.506 ± 0.014	0.406 ± 0.015
0.2	TSA	0.513 ± 0.020	0.507 ± 0.014	0.407 ± 0.015
0.3	TSA	0.513 ± 0.021	0.507 ± 0.015	0.407 ± 0.016
0.1	Bortezomib	0.523 ± 0.023	0.511 ± 0.017	0.414 ± 0.019
0.2	Bortezomib	0.524 ± 0.023	0.511 ± 0.017	0.414 ± 0.019
0.3	Bortezomib	0.523 ± 0.023	0.511 ± 0.018	0.414 ± 0.019
0.1	Tacrolimus	0.519 ± 0.022	0.511 ± 0.015	0.413 ± 0.016
0.2	Tacrolimus	0.520 ± 0.022	0.512 ± 0.015	0.413 ± 0.016
0.3	Tacrolimus	0.520 ± 0.020	0.512 ± 0.014	0.414 ± 0.015

3.5. Discovery of potentially mis-localized proteins in drug perturbation

We applied our method to three different drugs: TSA, tacrolimus, and bortezomib. For each drug, we list a set of representative proteins (Table 2). They have been predicted to acquire new subcellular locations or lose original subcellular locations under drug treatments. A comprehensive list of all proteins with mis-localization scores can be found in Table S1 in the supplementary material. A number of mis-localization events have been reported to be associated to pharmacological mechanisms of the corresponding drugs.

The first instance is the cytochrome C oxidase. Literatures have reported that TSA treatment induces cytochrome C releasing from mitochondria [36]. Since cytochrome C oxidase interact with the cytochrome C, it is expected that the cytochrome C oxidase would be increased in the cytoplasm. Our results indicated that under the TSA treatment, cytochrome C oxidase tends to lose the mitochondria localization, and acquire the cytoplasm localization.

Another example is the Smac protein. In our results, it is observed to have a significant increment in the localization of cell cortex under the bortezomib treatment. The literature [47] reported that bortezomib induces alterations in mitochondrial membrane potential releasing of mitochondrial protein Smac.

The third example is the protein ATP13A2. Literature shows that bortezomib treatment can redistribute ATP13A2 along with the

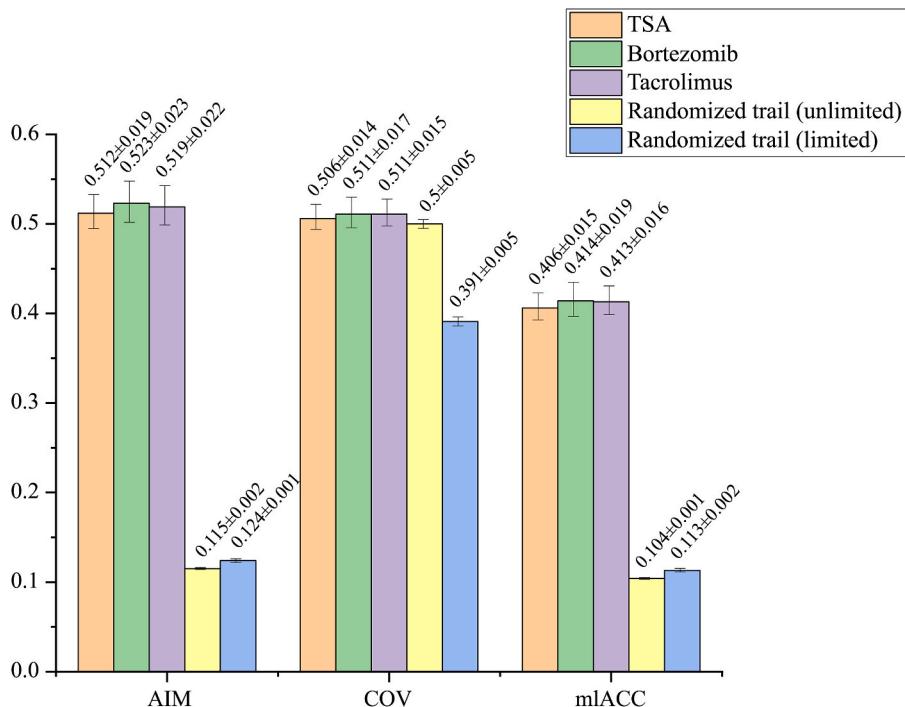


Fig. 5. 10-fold cross validation performances. We used three performance measures, including AIM, COV and mlACC. The randomized trail (unlimited) means the number of subcellular locations in random guesses are not restricted. The randomized trail (limited) mean the number of subcellular locations in random guesses are restricted to the real number of subcellular location for each protein. The performance values are presented in the form of average value \pm standard deviation. The average value and the standard deviation are obtained from 10 round of 10-fold cross-validations.

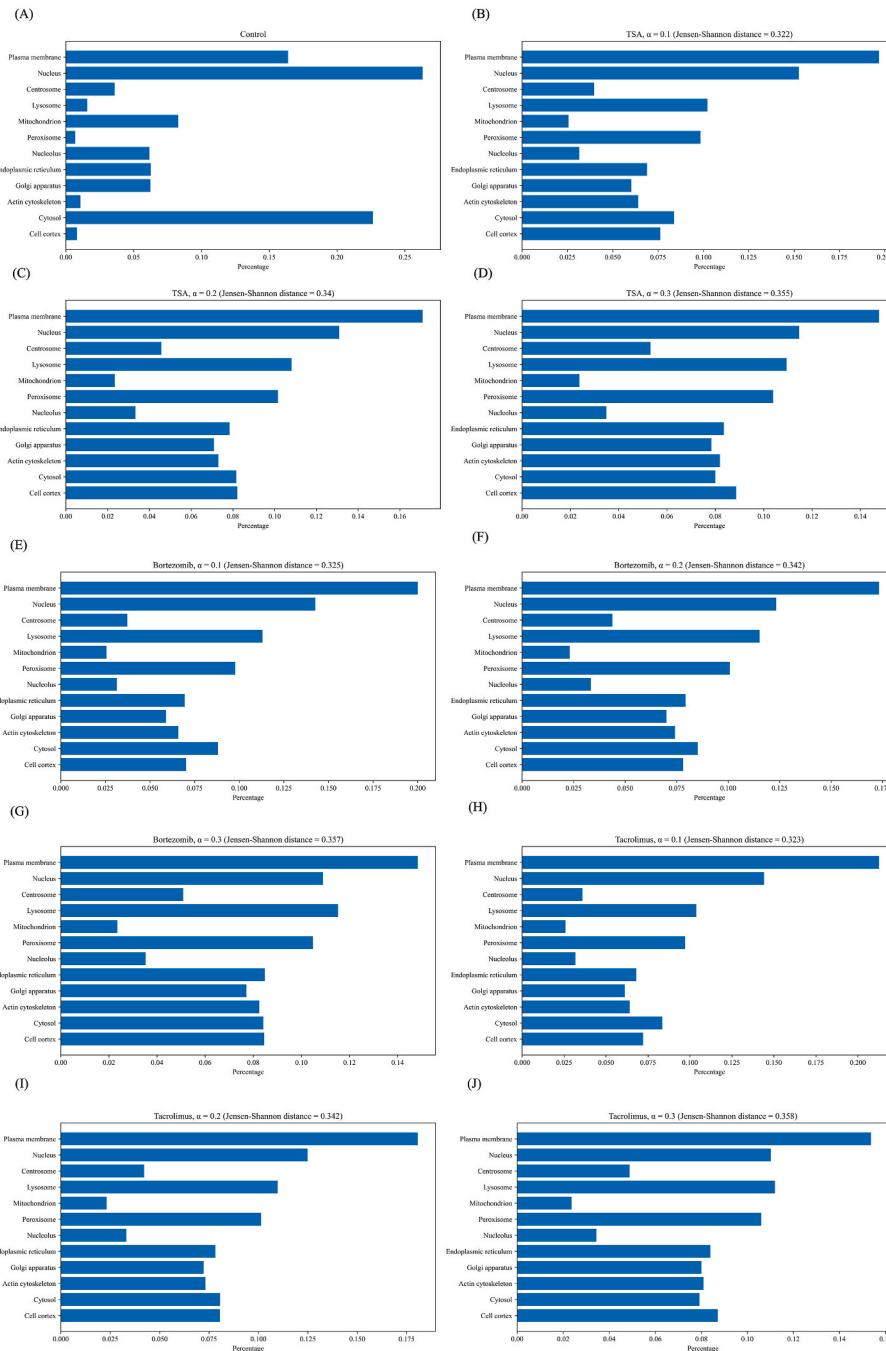


Fig. 6. Distributions of predicted subcellular locations with different settings of α . Jensen-Shannon Distances were calculated between the predicted results and the control state ground truth. By minimizing the Jensen-Shannon Distances in most cases, we choose $\alpha = 0.1$. (A) The distribution of control state ground truth. (B) The distribution of predicted subcellular locations in TSA treatment with $\alpha = 0.1$. (C) The distribution of predicted subcellular locations in TSA treatment with $\alpha = 0.2$. (D) The distribution of predicted subcellular locations in TSA treatment with $\alpha = 0.3$. (E) The distribution of predicted subcellular locations in bortezomib treatment with $\alpha = 0.1$. (F) The distribution of predicted subcellular locations in bortezomib treatment with $\alpha = 0.2$. (G) The distribution of predicted subcellular locations in bortezomib treatment with $\alpha = 0.3$. (H) The distribution of predicted subcellular locations in tacrolimus treatment with $\alpha = 0.1$. (I) The distribution of predicted subcellular locations in tacrolimus treatment with $\alpha = 0.2$. (J) The distribution of predicted subcellular locations in tacrolimus treatment with $\alpha = 0.3$.

peripheral pool of CD63+/LAMP+ [48]. Our results indicated that the ATP13A2 and CD63 were simultaneously acquiring the cell cortex localization after the bortezomib treatment.

The forth example is the MCP-1. It was reported that tacrolimus abolished the upregulation of MCP-1 by IL-1 β and TNF- α [38]. With our prediction results, both nuclear and cytoplasmic localization of MCP-1 had a noticeable decrement.

We can continue with these examples. Although not every instance in our list can be backed with literatures, we believe a major part of our results can be validated by existing literatures or future studies. It is possible that the protein mis-localization is not a rare phenomenon under drug treatments.

3.6. Computational mechanism analysis

The PLA-GNN method is designed to identify protein subcellular localization alterations. This is different to the traditional protein subcellular location predictions. In the traditional protein subcellular location prediction studies, the subcellular location annotations are recognized as static attributes of proteins. Protein features, like sequences, structures, and interactions, are also thought to be static information. This makes the traditional methods impossible to detect protein subcellular location alterations. predict the differences of protein subcellular locations between different states. However, this creates a dilemma in evaluating the prediction performances of the PLA-GNN. By reviewing similar existing studies [16,17], the accuracy of this kind of algorithm is usually evaluated by experiments or literature reports. This is why we use examples in the literatures to prove our predictions.

Table 2
Instances of subcellular location alteration predictions.

Drug	Protein	Location alteration ^a	Score	Rank
TSA	Cytochrome C oxidase	- Mitochondrion	-38.60%	-0.64%
		- Mitochondrion	-38.50%	-0.67%
		- Mitochondrion	-36.80%	-1.05%
		+ Cytosol	182.10%	0.38%
		+ Centrosome	251.70%	0.19%
		+ Centrosome	114.00%	1.06%
Bortezomib	Smac	+ Cell cortex	81.90%	1.90%
		+ Cell cortex	100.00%	1.30%
	ATP13A2	+ Cell cortex	92.30%	1.53%
	CD63	- Nucleolus	-43.20%	-0.12%
Tacrolimus	LAMP1	- Cytosol	-30.50%	-0.01%
	MCP-1	- Nucleus	-17.70%	-4.69%

^a The altered localization score is marked after the altered location in the percentage form. This score has no implication on any probability. The "+" prefix indicates an increment in the subcellular location. The "-" prefix indicates a decrement in the subcellular location. The ranks are quantiles in the protein list that is sorted using the boundary values in Eq. (17) and Eq. (18).

4. Conclusions

Computational prediction of protein subcellular localizations has been studied for over two decades. However, only a handful of studies considered protein subcellular location alterations in different cellular states. Notably, no existing study considered drug treatment states. We take the TSA, bortezomib, and tacrolimus as instances to develop PLA-GNN, which detects protein subcellular location alterations in drug perturbation states. We integrated gene expression profiles and PPIs to create a dynamic PPI network. The graph neural network algorithms were applied to aggregate high-order topology information. Although not directly, the prediction results can still be verified by existing literature. The algorithm of PLA-GNN can be extended to other kind of drugs and other conditions if the corresponding gene expression profiles can be obtained. For the future works, we believe a database containing manually curated mis-localization events will appear, which may serve as a reference resource for developing algorithms in predicting protein mis-localizations events.

Author contributions

RHW collected the data, constructed the model, implement the algorithm, performed experiments and partially wrote the manuscript. TL analyzed the results and partially wrote the manuscript. HLZ partially analyzed the results. PFD supervised the whole study, conceptualized the algorithm, analyzed the results and partially wrote the manuscript.

Funding

This work was supported by National Natural Science Foundation of China [NSFC 61872268].

Data availability statement

The code and data for reproducing the results of this paper is available in GitHub (<https://github.com/quinlanW/PLA-GNN>).

Declaration of competing interest

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2023.106775>.

References

- [1] Y. Nyathi, B.M. Wilkinson, M.R. Pool, Co-translational targeting and translocation of proteins to the endoplasmic reticulum, *Biochim. Biophys. Acta* 1833 (2013) 2392–2402, <https://doi.org/10.1016/j.bbamcr.2013.02.021>.
- [2] T.A. Rapoport, K.E. Matlack, K. Plath, B. Misselwitz, O. Staech, Posttranslational protein translocation across the membrane of the endoplasmic reticulum, *Biol. Chem.* 380 (1999) 1143–1150, <https://doi.org/10.1515/BC.1999.145>.
- [3] I. Mellman, W.J. Nelson, Coordinated protein sorting, targeting and distribution in polarized cells, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 833–845, <https://doi.org/10.1038/nrm2525>.
- [4] V. Schmidt, T.E. Willnow, Protein sorting gone wrong—VPS10P domain receptors in cardiovascular and metabolic diseases, *Atherosclerosis* 245 (2016) 194–199, <https://doi.org/10.1016/j.atherosclerosis.2015.11.027>.
- [5] Y. Guo, D.W. Sirkis, R. Schekman, Protein sorting at the trans-Golgi network, *Annu. Rev. Cell Dev. Biol.* 30 (2014) 169–206, <https://doi.org/10.1146/annurev-cellbio-100913-013012>.
- [6] L. Malinovska, S. Kroschwald, M.C. Munder, D. Richter, S. Alberti, Molecular chaperones and stress-inducible protein-sorting factors coordinate the spatiotemporal distribution of protein aggregates, *Mol. Biol. Cell* 23 (2012) 3041–3056, <https://doi.org/10.1091/mbc.E12-03-0194>.
- [7] C. Kontaxi, P. Piccardo, A.C. Gill, Lysine-directed post-translational modifications of Tau protein in Alzheimer's disease and related tauopathies, *Front. Mol. Biosci.* 4 (2017) 56, <https://doi.org/10.3389/fmbo.2017.00056>.
- [8] J.-E. Kim, Y.H. Hong, J.Y. Kim, G.S. Jeon, J.H. Jung, B.-N. Yoon, S.-Y. Son, K.-W. Lee, J.-I. Kim, J.-J. Sung, Altered nucleocytoplasmic proteome and transcriptome distributions in an in vitro model of amyotrophic lateral sclerosis, *PLoS One* 12 (2017), e0176462, <https://doi.org/10.1371/journal.pone.0176462>.
- [9] M. Prokocimer, A. Molchadsky, V. Rotter, Dysfunctional diversity of p53 proteins in adult acute myeloid leukemia: projections on diagnostic workup and therapy, *Blood* 130 (2017) 699–712, <https://doi.org/10.1182/blood-2017-02-763086>.
- [10] X. Wang, S. Li, Protein mislocalization: mechanisms, functions and clinical applications in cancer, *Biochim. Biophys. Acta Rev. Canc* (2014) 13–25, <https://doi.org/10.1016/j.bbcan.2014.03.006>, 1846.
- [11] R. Hill, B. Cautain, N. de Pedro, W. Link, Targeting nucleocytoplasmic transport in cancer therapy, *Oncotarget* 5 (2014) 11–28, <https://doi.org/10.18632/oncotarget.1457>.
- [12] M.-C. Hung, W. Link, Protein localization in disease and therapy, *J. Cell Sci.* 124 (2011) 3381–3392, <https://doi.org/10.1242/jcs.089110>.
- [13] P.J. Thul, L. Åkesson, M. Wiklund, D. Mahdessian, A. Geladaki, H.A. Blal, T. Alm, A. Asplund, L. Björk, L.M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Höber, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C.M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J.M. Schwenk, Å. Svertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K.S. Lilley, M. Uhlen, E. Lundberg, A subcellular map of the human proteome, *Science* (2017), <https://doi.org/10.1126/science.aal3321>.
- [14] R. Horwitz, G.T. Johnson, Whole cell maps chart a course for 21st-century cell biology, *Science* 356 (2017) 806–807, <https://doi.org/10.1126/science.aan5955>.
- [15] T. Ideker, N.J. Krogan, Differential network biology, *Mol. Syst. Biol.* 8 (2012) 565, <https://doi.org/10.1038/msb.2011.99>.
- [16] K. Lee, K. Byun, W. Hong, H.-Y. Chuang, C.-G. Pack, E. Bayarsaikhan, S.H. Paek, H. Kim, H.Y. Shin, T. Ideker, B. Lee, Proteome-wide discovery of mislocated proteins in cancer, *Genome Res.* 23 (2013) 1283–1294, <https://doi.org/10.1101/gr.155499.113>.
- [17] G.-P. Li, P.-F. Du, Z.-A. Shen, H.-Y. Liu, T. Luo, DPPN-SVM: computational identification of mis-localized proteins in cancers by integrating differential gene expressions with dynamic protein-protein interaction networks, *Front. Genet.* 11 (2020), 600454, <https://doi.org/10.3389/fgene.2020.600454>.
- [18] P. Du, C. Xu, Predicting multisite protein subcellular locations: progress and challenges, *Expert Rev. Proteomics* 10 (2013) 227–237, <https://doi.org/10.1586/epr.13.16>.
- [19] P. Du, T. Li, X. Wang, Recent progress in predicting protein sub-subcellular locations, *Expert Rev. Proteomics* 8 (2011) 391–404, <https://doi.org/10.1586/epr.11.20>.
- [20] Y.-Y. Xu, F. Yang, Y. Zhang, H.-B. Shen, An image-based multi-label human protein subcellular localization predictor (iLocator) reveals protein mislocalizations in cancer tissues, *Bioinformatics* 29 (2013) 2032–2040, <https://doi.org/10.1093/bioinformatics/btt320>.
- [21] L.P. Coelho, J.D. Kangas, A.W. Naik, E. Osuna-Highley, E. Glory-Afshar, M. Fuhrman, R. Simha, P.B. Berget, J.W. Jarvik, R.F. Murphy, Determining the subcellular location of new proteins from microscope images using local features, *Bioinformatics* 29 (2013) 2343–2349, <https://doi.org/10.1093/bioinformatics/btt392>.
- [22] K.-C. Chou, Z.-C. Wu, X. Xiao, iLoc-Euk: a multi-label classifier for predicting the subcellular localization of singleplex and multiplex eukaryotic proteins, *PLoS One* 6 (2011), e18258, <https://doi.org/10.1371/journal.pone.0018258>.
- [23] K.-C. Chou, Z.-C. Wu, X. Xiao, iLoc-Hum: using the accumulation-label scale to predict subcellular locations of human proteins with both single and multiple sites, *Mol. Biosyst.* 8 (2012) 629–641, <https://doi.org/10.1039/c1mb05420a>.
- [24] W.-Z. Lin, J.-A. Fang, X. Xiao, K.-C. Chou, iLoc-Animal: a multi-label learning classifier for predicting subcellular localization of animal proteins, *Mol. Biosyst.* (2013), <https://doi.org/10.1039/c3mb25466f>.
- [25] K.-C. Chou, H.-B. Shen, Recent progress in protein subcellular location prediction, *Anal. Biochem.* 370 (2007) 1–16, <https://doi.org/10.1016/j.ab.2007.07.006>.

- [26] X. Pan, H. Li, T. Zeng, Z. Li, L. Chen, T. Huang, Y.-D. Cai, Identification of protein subcellular localization with network and functional embeddings, *Front. Genet.* 11 (2020), 626500, <https://doi.org/10.3389/fgene.2020.626500>.
- [27] H. Zhou, Y. Yang, H.-B. Shen, Hum-mPLoc 3.0: prediction enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features, *Bioinformatics* 33 (2017) 843–853, <https://doi.org/10.1093/bioinformatics/btw723>.
- [28] P. Du, L. Wang, Predicting human protein subcellular locations by the ensemble of multiple predictors via protein-protein interaction network with edge clustering coefficients, *PLoS One* 9 (2014), e86879, <https://doi.org/10.1371/journal.pone.0086879>.
- [29] K. Laurila, M. Vihinen, PROlocalizer: integrated web service for protein subcellular localization prediction, *Amino Acids* 40 (2011) 975–980, <https://doi.org/10.1007/s00726-010-0724-y>.
- [30] K. Laurila, M. Vihinen, Prediction of disease-related mutations affecting protein localization, *BMC Genom.* 10 (2009) 122, <https://doi.org/10.1186/1471-2164-10-122>.
- [31] K. Lee, H.-Y. Chuang, A. Beyer, M.-K. Sung, W.-K. Huh, B. Lee, T. Ideker, Protein networks markedly improve prediction of subcellular localization in multiple eukaryotic species, *Nucleic Acids Res.* 36 (2008) e136, <https://doi.org/10.1093/nar/gkn619>.
- [32] A. Kumar, A. Rao, S. Bhavani, J.Y. Newberg, R.F. Murphy, Automated analysis of immunohistochemistry images identifies candidate location biomarkers for cancers, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 18249–18254, <https://doi.org/10.1073/pnas.1415120112>.
- [33] T. Ideker, N.J. Krogan, Differential network biology, *Mol. Syst. Biol.* 8 (2012) 565, <https://doi.org/10.1038/msb.2011.99>.
- [34] I.A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, M.A. Calderwood, M. Vidal, A.-L. Barabási, Network-based prediction of protein interactions, *Nat. Commun.* 10 (2019) 1240, <https://doi.org/10.1038/s41467-019-09177-y>.
- [35] O. Keskin, B. Ma, K. Rogale, K. Gunasekaran, R. Nussinov, Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach, *Phys. Biol.* 2 (2005) S24–S35, <https://doi.org/10.1088/1478-3975/2/2/S03>.
- [36] A.F. Taghiyev, N.V. Guseva, M.T. Sturm, O.W. Rokhlin, M.B. Cohen, Trichostatin A (TSA) sensitizes the human prostatic cancer cell line DU145 to death receptor ligands treatment, *Cancer Biol. Ther.* 4 (2005) 382–390, <https://doi.org/10.4161/cbt.4.4.1615>.
- [37] K. Mehdizadeh, F. Ataei, S. Hosseinkhani, Treating MCF7 breast cancer cell with proteasome inhibitor Bortezomib restores apoptotic factors and sensitizes cell to Docetaxel, *Med. Oncol.* 38 (2021) 64, <https://doi.org/10.1007/s12032-021-01509-7>.
- [38] S. Du, N. Hiramatsu, K. Hayakawa, A. Kasai, M. Okamura, T. Huang, J. Yao, M. Takeda, I. Araki, N. Sawada, A.W. Paton, J.C. Paton, M. Kitamura, Suppression of NF- κ B by cyclosporin A and tacrolimus (FK506) via induction of the C/EBP family: implication for unfolded protein Response1, *J. Immunol.* 182 (2009) 7201–7211, <https://doi.org/10.4049/jimmunol.0801772>.
- [39] R. Oughtred, C. Stark, B.-J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam, F. Zhang, S. Dolma, A. Willems, J. Coulombe-Huntington, A. Chatr-Aryamontri, K. Dolinski, M. Tyers, The BioGRID interaction database: 2019 update, *Nucleic Acids Res.* 47 (2019) D529, <https://doi.org/10.1093/nar/gky1079>.–D541.
- [40] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Defining and identifying communities in networks, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 2658–2663, <https://doi.org/10.1073/pnas.0400054101>.
- [41] J. Wang, M. Li, H. Wang, Y. Pan, Identification of essential proteins based on edge clustering coefficient, *IEEE ACM Trans. Comput. Biol. Bioinf.* 9 (2012) 1070–1080, <https://doi.org/10.1109/TCBB.2011.147>.
- [42] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017, in: <https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebca9-Abstract.html>. (Accessed 17 January 2023).
- [43] Y. Jiao, P. Du, Performance measures in evaluating machine learning based bioinformatics predictors for classifications, *Quant Biol* 4 (2016) 320–330, <https://doi.org/10.1007/s40484-016-0081-2>.
- [44] P.J. Thul, L. Åkesson, M. Wiklund, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L.M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hofer, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C.M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J.M. Schweln, Å. Siverstsson, E. Sjöstedt, M. Skogs, C. Städler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoğlu, F. Pontén, K. von Feilitzen, K.S. Lilley, M. Uhlen, E. Lundberg, A subcellular map of the human proteome, *Science* 356 (2017), <https://doi.org/10.1126/science.aal3321> eaal3321.
- [45] J.A. Christopher, C. Städler, C.E. Martin, M. Morgenstern, Y. Pan, C.N. Betsinger, D. G. Ratray, D. Mahdessian, A.-C. Gingras, B. Warscheid, J. Lehtiö, I.M. Cristea, L. J. Foster, A. Emili, K.S. Lilley, Subcellular proteomics, *Nat Rev Methods Primers* 1 (2021) 1–24, <https://doi.org/10.1038/s43586-021-00029-y>.
- [46] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, *IEEE Trans. Inf. Theor.* 49 (2003) 1858–1860, <https://doi.org/10.1109/TIT.2003.813506>.
- [47] D. Chauhan, G. Li, K. Podar, T. Hidemitsu, C. Mitsiades, R. Schlossman, N. Munshi, P. Richardson, F.E. Cotter, K.C. Anderson, Targeting mitochondria to overcome conventional and bortezomib/proteasome inhibitor PS-341 resistance in multiple myeloma (MM) cells, *Blood* 104 (2004) 2458–2466, <https://doi.org/10.1182/blood-2004-02-0547>.
- [48] S. Demirsoy, S. Martin, S. Motamedi, S. van Veen, T. Holemans, C. Van den Haute, A. Jordanova, V. Baekelandt, P. Vangheluwe, P. Agostinis, ATP13A2/PARK9 regulates endo-/lysosomal cargo sorting and proteostasis through a novel PI(3, 5) P2-mediated scaffolding function, *Hum. Mol. Genet.* 26 (2017) 1656–1669, <https://doi.org/10.1093/hmg/ddx070>.