PAPER

SPIDER: Constructing cell-type-specific protein-protein interaction networks

Yael Kupershmidt,[1] Simon Kasif[2] and Roded Sharan[1],*

[1]Blavatnik School of Computer Science, Tel Aviv University, 69978 Tel Aviv, Israel and [2]Department of Biomedical Engineering, Boston University, Boston MA, 02215, USA

*Corresponding author. roded@tauex.tau.ac.il

Abstract

Protein-protein interactions play essential roles in the buildup of cellular machinery and provide the skeleton for cellular signaling. However, these biochemical roles are context dependent and interactions may change across cell type, time and space. In contrast, protein-protein interactions detection assays are run in a single condition that may not even be an endogeneous condition of the organism, resulting in static networks that do not reflect full cellular complexity. Thus, there is a need for computational methods to predict cell-type-specific interactions. Here we present SPIDER, a graph attention based model for predicting cell-type-specific protein-protein interaction networks. In contrast to previous attempts at this problem that were unsupervised in nature, our model's training is guided by experimentally measured cell-type-specific networks which benefits its performance. We evaluate our method against experimental data of cell-type-specific networks from both human and mouse, and show that it outperforms current approaches by a large margin. We further demonstrate the ability of our method to generalize the predictions to datasets of tissues lacking prior protein-protein interactions experimental data. We leverage the networks predicted by the model to facilitate the identification of tissue-specific disease genes. Our code and data are available at https://github.com/Kuper994/SPIDER.

Key words: Protein-protein interaction, cell-type-specific network, graph attention, deep learning

Introduction

Protein-protein interaction (PPI) networks are the workhorse behind molecular processes in an organism's cell. Understanding these networks can assist in the prediction of protein function, elucidation of protein complexes, identification of disease pathways and drug target prioritization [1, 2, 3]. However, cellular conditions greatly affect protein function and abundance and, ultimately, protein interaction. Thus, protein-protein interactions are context-specific and depend on the tissue or cell type in question [4, 5]. This context-dependence is not reflected in public databases as most experimental data to date comes from artificial systems that measure the interactions in some arbitrary conditions. Indeed, profiling interactions in a specific cell type could be challenging due to the high number of proteins, the variability of their biochemical properties, the high rate of transient interactions and the low abundance of some proteins [6]. Due to these difficulties, very scarce cell-type specific data is available, calling for computational inference of such interactions.

While there is a growing number of works that leverage deep learning to predict protein-protein interactions (see, e.g., [7, 8, 9]), only few works aim at contextualizing them. Early work in this domain has primarily focused on creating biological context networks [10, 11]. To this end, for a given context, the general (non-context-specific) PPI network is filtered to include only the proteins associated with that context, typically determined through gene ontology (GO) annotations. Subsequently, various methods have been developed to construct tissue-specific networks. One such method, Node Removal (NR) (cf. [12]), removes from the network nodes that are not expressed in a given tissue (see, e.g. [13]). A more refined, Edge Reweighting (ERW) method, scores each interaction by the product of the probabilities that its nodes are expressed in the tissue in question [12]. Another approach introduced in [14] infers tissue-specific associations in an unsupervised manner based on protein co-abundance information. To calibrate the association scores the authors rely on protein complex annotations from the CORUM collection [15]. Less systematic approaches are based on text mining but as a result are affected by literature biases (see, e.g. [16]).

A related body of work relies on tissue-specific networks to conduct downstream prediction tasks. Ohmnet [17] predicts the cellular function of a protein using a multi-layer network that reflects the known tissue hierarchy. Each layer in the network represents a tissue-specific network that is built using an NR-like strategy of maintaining only nodes that represent tissue-specific or ubiquitous genes. GIANT [18] utilizes NR-like tissue-specific networks and data from genome-wide association studies to construct tissue-specific functional networks. The

recent PINNACLE model uses tissue-specific networks to learn tissue-specific protein representations [19]. Those networks are again built by maintaining genes with higher than average expression, compared to other cell types or tissues.

Recent experimental advances have allowed for the first time the direct measurement of tissue-specific interactions. Specifically, the BioPlex network project [6] measures two human cell-type-specific networks in two cell lines: HEK293T and HCT116. Additionally, seven mouse tissue networks were measured by Skinnider et al. [20]. The availability of experimentally curated tissue- and cell-type-specific networks provides a solid foundation for the development of condition-specific network predictors, facilitating the potential curation of new contexts.

In this work, we introduce SPIDER (Supervised Protein Interaction DEtectoR), a graph attention-based framework for constructing condition-specific PPI networks. SPIDER utilizes diverse protein and interaction features to extract interaction patterns within the dataset that are smoothed using the graph attention component. We showcase the model's proficiency in capturing interaction characteristics and predicting tissue-specific interactions among a set of proteins based on their features. We demonstrate the superiority of these networks over earlier works in both human and mouse. We further demonstrate the model's capability to generalize predictions to datasets describing cell types or tissues the model has not been trained on. Finally, We show the application of SPIDER to construct tissue-specific networks and cancer-type-specific networks across a host of human tissues and their utility in identifying disease genes. Overall, the SPIDER model is shown to be a robust and efficient method for computationally inferring context-specific networks with applications to downstream classification tasks.

Materials and methods

We developed a deep learning framework for inferring a protein-protein interaction (PPI) network in a condition of interest. Our model receives as input two types of information: (i) general information that includes a set of potential interactions, derived from experimental measurements in arbitrary conditions, their detection techniques and their proteins' cellular location; and (ii) condition-specific information on gene expression and protein abundance. These gene-level and interaction-level features are fed into a neural network for their integration and subsequent use for prediction of condition-specific interactions. The model is trained using real experimental data of protein-protein interactions that were directly measured in the specified condition. The model's architecture is sketched in Figure 1. Further details regarding the data and methodology are outlined below.

Protein-protein interaction network data

We used PPI information from both mouse and human. For mouse, seven tissue-specific networks were taken from [20]. These tissue-specific networks varied in size, spanning a range of 20,000-35,000 interactions and involving 1300-2000 proteins. A general (non-context specific) network was produced by taking the union of these networks and the mouse BioGRID network [21], resulting in 202,443 interactions among 11,795 proteins.

For human, a general network was sourced from BioGRID [21], spanning 1,030,456 physical interactions

involving 19,609 proteins. Cell-type specific networks for HEK293T and HCT116 were taken from the Bioplex project [22]. The HEK293T network spans 118,162 interactions between 13,957 proteins; the HCT116 network includes 70,966 interactions among 10,115 proteins. The general network covers about 89% of both HEK293T and HCT116 interactions. Additionally, a text mining-based network for HeLa was taken from [16] and includes 253,506 interactions (87% of which are covered by the general network) among 16,423 proteins. As acknowledged by the authors of [16], their text mining approach may extract interactions that do not occur under the specified conditions, thus we expect such interactions to have lower reliability than experimentally-based ones. For this reason, we used only the HeLa network which was the largest network in [16].

Henceforth, we focus on interactions that are included in the general BioGRID network. As PPI datasets are known to suffer from high false positive rates [23], information regarding the experimental methods used for interaction discovery was obtained from BioGRID and used for reliability estimation. For every interaction, we recorded the assays by which it was discovered and the number of times it was discovered in each. The technique features included Affinity chromatography technology, Anti-bait coimmunoprecipitation, Anti-tag coimmunoprecipitation, Two-hybrid, Pull down, Transcriptional complementation assay, Affinity technology, Biochemical, Enzymatic study, Proximity labeling technology, and a final category grouping all other experimental interaction detection methods. These features were used to characterize human interactions only, as a large portion of the mouse interactions were only reported in [20].

Gene and protein expression data

Human gene expression data, measured in each of the three cell types, were sourced from the Cancer Cell Line Encyclopedia (CCLE) [24]. The data includes $\log_2(\text{TPM} + 1)$ values, which were normalized to a range between 0 and 1. For the mouse networks, the gene expression data were extracted from [25]. These expression values underwent log transformation to align with the human gene expression features and were subsequently normalized to values between 0 and 1.

Protein abundance data for HEK293T and HCT116 cells were taken from the Bioplex project [6]. Each cell line had five samples, with the protein abundance values computed by the median across those samples. Protein abundance data for HeLa cells were sourced from [26]. Each protein was measured across 30 samples and the protein value was calculated as the median of the sample-specific values. Protein abundance information for the mouse tissues was taken from [20]. Protein abundance values for both human and mouse were normalized to be between 0 and 1. Co-abundance values were computed by Pearson correlation of the sample-specific values.

Cellular localization

Cellular location information was derived from the gene ontology annotation. For each protein, a vector was constructed, describing its probability of being localized in a predefined set of primary compartments within the cell (63 for human and 62 for mouse). These compartments were delineated based on terms employed in the LOCATE database [27]. The probability of a protein being localized to a compartment was considered to be uniformly distributed across all GO cellular compartments it is associated with from the specified set. The

co-localization of protein pairs was computed by the Pearson correlation between their localization vectors.

Neural network architecture

The SPIDER architecture consists of several layers (Figure 1). The first layer projects the various input features into a joint space. This is accomplished by passing each feature, such as gene expression, protein abundance and co-abundance, through a feedforward neural network (FFNN) component. Next, a graph attention layer aggregates information on each protein from its network neighbors, as described in [28]. Subsequently, for each interaction, we utilize a deep set function to merge the embeddings of its two proteins using an order-invariant function. As shown in [29], such a function can be approximated by a neural network architecture where each element of the pair is passed through a fully connected layer ϕ, the outputs are summed, and the result passes through another fully connected layer ρ. The resulting embedding is concatenated with data pertaining to the interaction itself, including the co-localization of the proteins, their co-abundance and information regarding the assays used to detect the interaction. These features are then processed through a feed-forward component, ultimately producing the probability of the interaction being a part of the cell-type-specific network.

Model training and hyperparameter tuning

The network models were trained for 750 epochs using an Adam optimizer with a weight decay of 0.01 and a learning rate of 0.001 (see hyperparameter tuning process below). The loss function employed was binary cross-entropy, computed as the difference between the predictions of each interaction and its true label determined by the gold standard context-specific network. For the training and testing process, the nodes of the general network were categorized into three groups: train, validation and test, of sizes 70%, 15% and 15%, respectively. The graphs induced by those nodes sets defined the corresponding edge train, validation and test sets.

The model's hyperparameters include the embedding size for the representation of proteins, the learning rate, the dropout probability of the feed-forward components and the number of training iterations. The dimensions of the feed-forward layers were determined based on the proteins' embedding size. A range of hyperparameter values was explored, including embedding sizes of $[32, 64, 128]$, learning rates of $[1e-3, 5e-4, 1e-4]$, dropout probabilities of $[0.2, 0.3, 0.4]$, and iteration numbers of $[500, 750, 1000]$. Given the data imbalance, the combination yielding the highest AUPRC score on the validation set was selected. The tuning was conducted using the HEK293T dataset as the most extensively researched and experimentally curated network out of the ten human and mouse datasets. This optimal configuration consisted of an embedding size of 64, a learning rate of 0.001, a dropout probability of 0.3, and 750 epochs.

Baseline methods

We compared our classification results against three baseline methods: node removal (NR), edge reweighting (ERW) (cf. [12]) and Co-abundance [14]. In NR, the condition-specific network is constructed by projecting gene expression onto the network nodes. Only nodes representing expressed genes are retained in the network, while nodes corresponding to non-expressed genes are removed along with the edges connected to them. The cutoff for a gene to be considered expressed can vary. In ERW, the weights of the edges are modified according to the expression of the genes comprising it. The algorithm sets a penalty factor rw, and modifies each edge weight $w_{i,j}$ by multiplying it by rw^n where n ranges from 0 to 2 and represents the number of lowly or unexpressed genes among the interacting genes. Following [12], we used a penalty factor of $rw = 0.001$. The
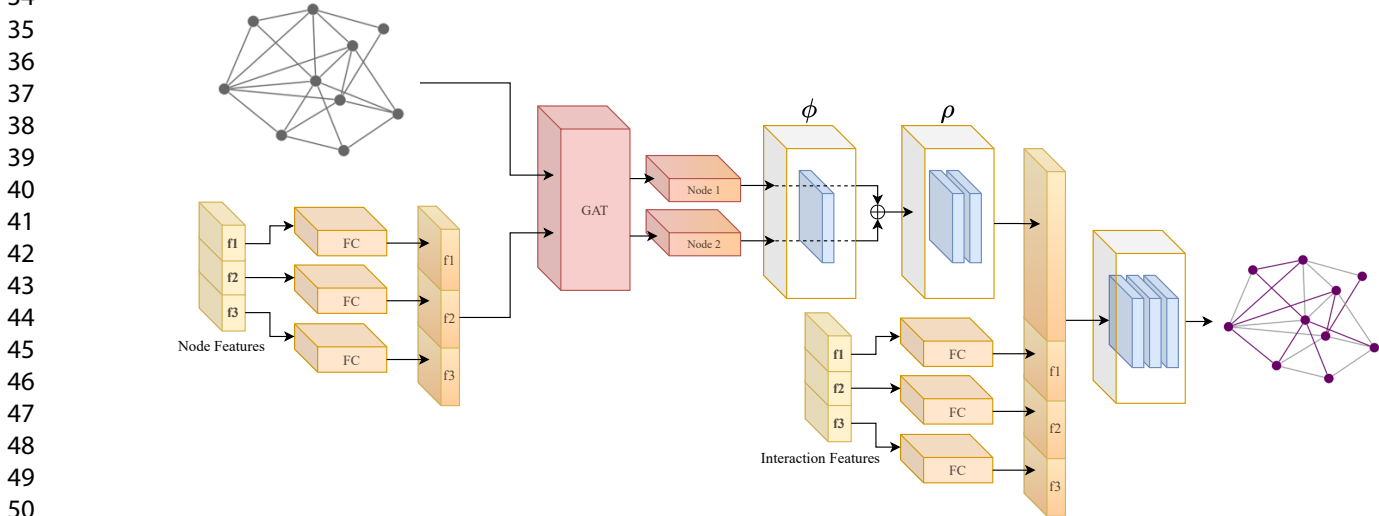

Fig. 1. SPIDER's neural network architecture. Node features, including gene expression, protein expression and protein localization, as well as interaction features, including protein co-expression, protein co-localization and interaction discovery techniques, are projected into a joint space using a feedforward neural network (FFNN) component with two fully connected (FC) layers. Node features are further processed with a graph attention (GAT) layer and a deepest component which represents a ϕ function with one FC layer and a ρ function with two FC layers. The final FFNN component consists of three FC layers and outputs the interaction's probability.
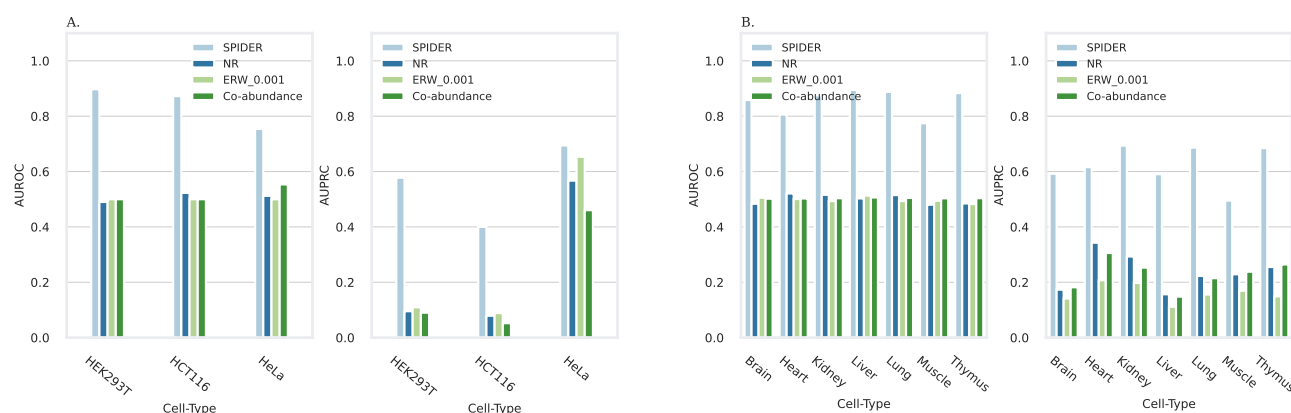
Fig. 2. AUROC and AUPRC scores for the human (A) and mouse (B) models and comparison to previous approaches.

AUROC and AUPRC scores for both methods were computed by varying the gene expression threshold. In Co-abundance, the association between a pair of proteins is computed based on the Pearson correlation between their expression values across samples. These scores are further refined using a logistic model that is based on protein complex information from CORUM as explained in [14].

Results

We developed SPIDER, a graph attention based model for inferring cell type specific protein-protein interactions. The predictions employ both protein-level features, such as protein expression and location, and interaction-level features, such as detection assays. In the following we describe the performance of the model, its comparison to previous work and its application to predicting tissue-specific networks and disease genes.

Performance evaluation

We trained ten cell-type specific models using gold standard networks of various cell types, including three human cell types (HEK293T, HCT116, and HeLa) and seven mouse tissues (Brain, Heart, Kidney, Liver, Lung, Muscle and Thymus). We assessed the performance of the models in cross-validation by measuring the areas under the ROC and precision-recall curves (Figure 2). Each classifier was trained using both context-specific datasets including gene expression and protein abundance information, as well as, non-context specific ones including protein location and interaction detection information.

For the human data, the HEK293T and HCT116 models received AUROCs of 0.9 and 0.86, respectively, while the HeLa model achieved an AUROC of 0.77. The relatively lower value for HeLa may reflect the fact that the HeLa ground truth network was not experimentally measured and its text-mining construction may be biased. Furthermore, the three models performed well w.r.t. the AUPRC measure, yielding a score of 0.63 for the HEK293T model, 0.34 for the HCT116 model and 0.71 for the HeLa model, substantially higher than the rate of positive interactions in those datasets (12%, 6% and 30%, respectively). Similarly good results were obtained for the mouse data with an average AUROC of 0.85 and AUPRC of 0.62.

We compared our performance to three unsupervised baselines: node removal (NR), edge reweighting (ERW) and co-abundance (see Methods for details). One potential advantage of SPIDER over previous unsupervised methods w.r.t. the human dataset is the reliance on information regarding the experimental evidence for each interaction, which allows it to focus on more reliable interactions. To remove this bias, we performed the comparison when restricting attention to a subset of reliable interactions. To this end, we applied the pipeline in [30] to score interactions and for each method set a threshold that maximized the AUROC score obtained. The results are portrayed in Figure 2 and show the superiority of SPIDER over the three baselines across all cell types.

Generalization and transfer learning

Next, we evaluated SPIDER's capacity to generalize predictions to new cell types or tissues that the model has not been trained on. To this end, we applied each of the models trained on a specific cell type to predict the other cell types (within the same species). The generalization results are presented in Table 1.

For human, both HEK293T and HCT116 models demonstrate the ability to predict one another network nearly as effectively as their original models. However, as shown in Table 1, neither the HEK293T model nor the HCT116 model generalize well to HeLa. This may reflect the fact that the latter network was deduced from text mining, exposing it to potential literature biases, while the other two networks were systematically explored through experimentation. For mouse, the results reveal that each mouse tissue-specific network can be predicted with high accuracy by any of the other tissue-specific models, with an average AUROC of 0.81. In some cases, such as the predictions of the Kidney model on the Liver dataset, the generalized model performs as well as the original model. In the case of the Brain model predictions on the Muscle dataset, the generalized model outperforms the original Muscle-trained model. This suggests promising generalization capabilities of the models across different cellular contexts within the same species.

In an effort to enhance the human network predictions, we pursued the fine-tuning of the HEK293T model, which had the most comprehensive experimental dataset, in order to represent other cell types, leveraging transfer learning techniques. Specifically, after the original training of the HEK293T model, the weights of the initial layers of the model

Dataset / Model	HEK293T	HCT116	HeLa
HEK293T	(-)	0.85	0.54
HCT116	0.85	(-)	0.56
HeLa	0.69	0.57	(-)

(a)

Dataset / Model	Brain	Heart	Kidney	Liver	Lung	Muscle	Thymus
Brain	(-)	0.72	0.82	0.87	0.85	0.77	0.86
Heart	0.78	(-)	0.83	0.87	0.85	0.76	0.81
Kidney	0.81	0.75	(-)	0.89	0.87	0.75	0.86
Liver	0.83	0.76	0.86	(-)	0.88	0.76	0.86
Lung	0.81	0.75	0.85	0.88	(-)	0.76	0.88
Muscle	0.74	0.67	0.76	0.79	0.78	(-)	0.82
Thymus	0.82	0.76	0.85	0.89	0.88	0.76	(-)

(b)

Table 1. Generalization AUROC scores of the different human (a) and mouse (b) models.

that process the protein-level features were frozen. Thereafter, the model underwent additional training on datasets of different sizes from either HCT116 or HeLa cell types. To prevent overfitting on relatively small training sets, the fine-tuning was executed with a low learning rate of $5e - 4$ and limited to only 100 training epochs. Figure 3 illustrates the score improvement achieved through the fine-tuning of the HEK293T model with growing fractions of data from the target cell type. The improvement in both AUROC and AUPRC is evident already after introducing the HEK293T model with 10% of the dataset. This is particularly notable given the small fraction of edges that the HEK293T network covers in the two other networks (31% of HCT116 and 20% of HeLa, see Figures 3.B and 3.C), underscoring the model's generalization power.

Building tissue-specific networks

As a next step, we constructed tissue-specific networks using transcriptomic and proteomic data curated by the Genotype-Tissue Expression (GTEx) project [31]. This dataset encompasses gene and protein expression from 32 human tissues with multiple samples per tissue.

In order to apply SPIDER to these data, we trained a human predictive model using both the HEK293T and HCT116 datasets. The HeLa dataset was excluded due to its non-experimental assembly. This model underwent training and validation using the combined train and validation sets of the two cell lines, respectively. Its performance was evaluated on the test sets of the two cell lines, achieving an AUROC score of 0.89 and an AUPRC of 0.57 on the HEK293T test set, and an AUROC score of 0.87 and an AUPRC of 0.33 on the HCT116 test sets, closely resembling the scores of the original cell lines models.

Next, we applied the combined model to the GTEx data to construct the tissue-specific networks. Since each of the GTEx tissues had only a few samples (2-10) and their gene expression values displayed low variation (mean coefficient of variation per gene in the range $0.2 - 0.49$), we constructed one network per tissue by utilizing the median gene- and protein-expression values to compute interaction classification features. The median expression values underwent the same normalization process as the cell line data. We also created an unweighted version of the networks by setting a threshold on the

probabilities of the edges output by the model. This threshold was set to 0.39, as the one that maximizes the combined model's F1-score on the HEK293T and HCT116 test sets.

In order to validate the constructed networks, we sought to compare their pairwise similarities to those implied by the curated Brenda tissue ontology [32]. Following [19], we computed a pairwise tissue ontology distance by the sum of path lengths between the two tissue nodes to their lowest common ancestor. We correlated these distances with the corresponding distances between the (weighted) tissue-specific networks. The distance between a pair of networks was computed using Pearson distance (defined as $1 - r$, where r denotes the Pearson correlation). As a benchmark, we also computed this correlation for tissue-specific networks constructed according to the NR method, where the gene expression threshold was chosen as the value maximizing the F1-score on the combined HEK293T and HCT116 dataset. The correlation obtained by the SPIDER networks was 0.23 ($p < 1.52e - 5$), while the other method led to a close-to-zero negative correlation of -0.02.

As another validation of the tissue-specific networks, we checked if the interactions within them are enriched with proteins that are known to be specifically expressed in the corresponding tissues. To this end, we utilized the tissue-enriched gene sets defined by The Human Protein Atlas [33], focusing on tissues with at least 25 annotated proteins. For each protein we computed its relative degree (i.e., degree percentile) change between any tissue-specific network (unweighted version) and the general network. We then compared the distribution of these values in a given tissue between the tissue-specific proteins and all other proteins using a Wilcoxon rank-sum test. The results are portrayed in Figure 4 and, reassuringly, all tissues display significant differences.

To further validate the tissue-specific networks, we tested their utility at pinpointing tissue-specific disease genes. To this end, we used a dataset reported in [34] of 1, 117 tissue-associated disease genes affecting 23 tissues.In order to predict tissue-specific genes, we used network propagation [35] in a 3-fold cross-validation setting. That is, each fold was predicted based on propagating from a seed composed of the genes from the other two folds. For the propagation process, we used the network's adjacency matrix normalized by vertex

degrees (stochastic normalization) and a default 0.8 value for the α parameter that describes the tradeoff between network smoothing and prior information. To remove genes for which no interaction prediction could be made, each tissue-specific gene group was filtered to include only genes measured by GTEx in the respective tissue. Gene sets with less than 25 genes were discarded. Figure 5 presents the resulting AUROC scores, comparing them to predictions made using the general network.

Evidently, the tissue-specific networks exhibit a notable advantage in predicting tissue-specific disease genes.
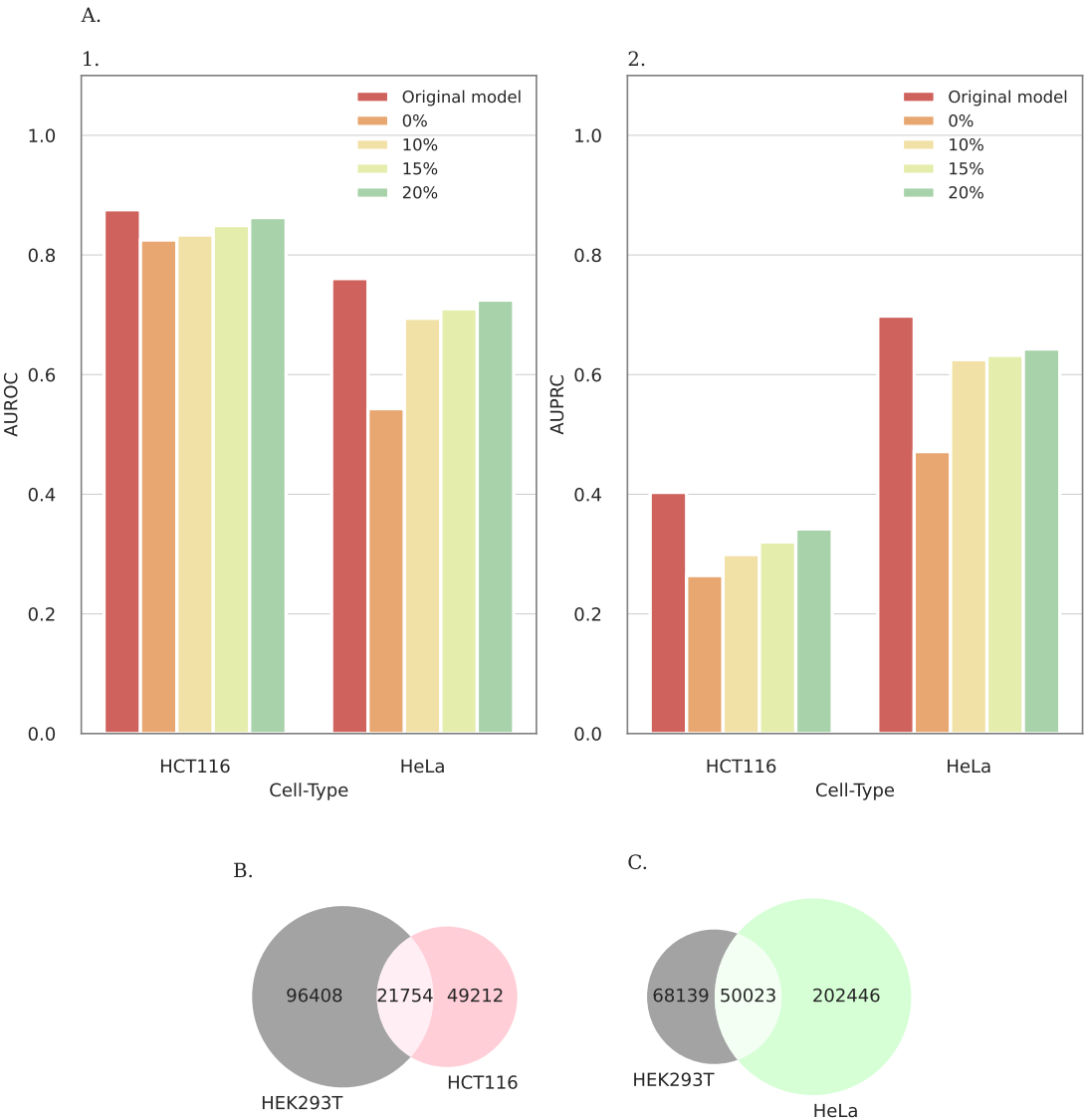

Fig. 3. A: AUROC (A.1) and AUPRC (A.2) scores for the transfer learning from the HEK293T model using varying portions of training data from a target cell (HCT116 or HeLa). The left bar in each panel corresponds to the original model that was trained on the target cell. The 0% bar represents the performance of the HEK293T model w.r.t. other cell types without further training. **B,C:** Venn diagrams displaying the intersection of the HEK293T network with the two other human networks.
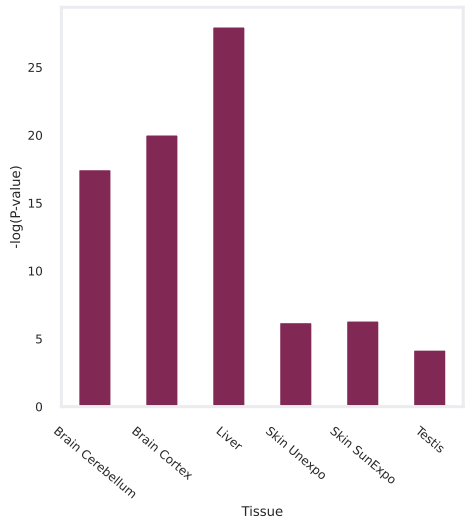
Fig. 4. Significance values (-log(*p*-value)) of a Wilcoxon rank-sum test between degree differences of tissue-specific proteins and all other proteins, where (relative) degree differences are computed between the general network degree and a tissue-specific network degree.

Identifying cancer driver genes

Last, we tested the application of the model in constructing a cancer-type-specific network and utilizing it for the identification of cancer driver genes. To this end, three datasets from the Cancer Genome Atlas (TCGA) were incorporated (Breast Cancer, Colon Adenocarcinoma, and Ovarian Cancer). The selection of these datasets was based on the availability of protein expression for at least 30% of the proteins in the network.

Unlike the GTEx data, TCGA datasets included hundreds of samples with high gene expression variation (mean coefficient of variation of 1.6-2.8). Thus, we generated an individual network for each sample within a cancer type, employing the
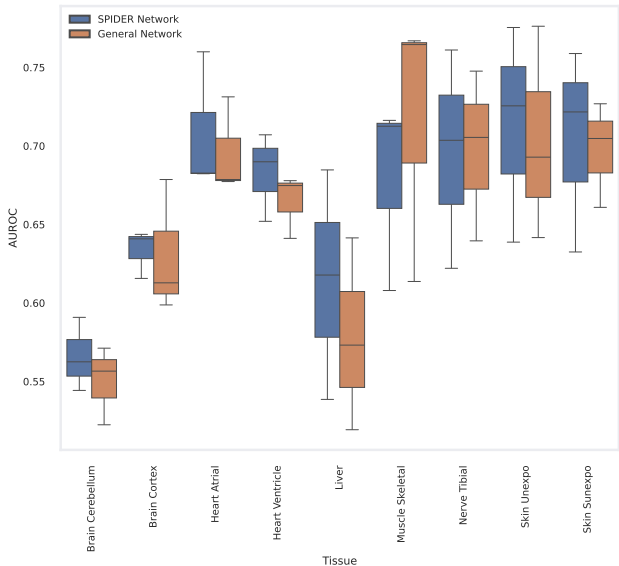
HEK293T and HCT116 combined model and the sample's features. As there are no multiple measurements per sample, the co-abundance feature was computed using the sample's protein abundance and the protein abundance measurements of the 5 samples that are most similar to it (measured via the cosine similarity of the samples' protein expression vectors). Then, we aggregated the sample-specific networks to form a *consensus network* for each cancer type, where the weight of each edge in the consensus network was determined as the mean edge weight across all sample-specific networks.

Finally, we used the consensus network to predict cancer driver genes. For each sample, we propagated its mutated gene set in the corresponding consensus network to score all other genes. The mutated gene set included genes that were reported to be mutated or to exhibit copy number alterations. These predictions were then averaged over all samples of a specific cancer type. A gold standard set of driver genes was defined by the COSMIC Cancer Gene Census (CGC) [36]. Figure 6 presents the enrichment of the CGC genes within the top-ranked genes. For any K, the Top-K enrichment was calculated according to:

$$\frac{n/K}{|CGC|/N} \qquad (1)$$

Where n is the number of CGC genes within the K top-ranked genes, and N denotes number of proteins in the consensus network. Evidently, the predictions made by the sample-specific networks dominate those made with a general network across all three cancer types.

Conclusions

Protein-protein interaction networks are the workhorse of the cell but can vary substantially between cell types and tissues. In this study, we have successfully developed and trained a predictive model for cell-type-specific protein-protein interactions. Our model integrates gene expression information, protein abundance data, protein location information and interaction detection assay data, and demonstrates improvements on multiple evaluation fronts and benchmarks as compared to previous methods in this space.

One of the advantages of using advanced machine learning models of this type is their ability to generalize to novel cell types where relevant experimental data has been collected without doing full scale protein-protein interaction studies which might be cost or labor prohibitive. Indeed, we show that our model can generalize to yet-unseen tissues or cell types. Additionally, we implemented a transfer learning strategy to fine-tune such predictions and demonstrated that even with a small training set the model can be refined effectively to predict the target network with sufficient accuracy. Finally, we showcased the utility of the framework in predicting tissue-specific disease genes and cancer-driver genes. SPIDER can thus serve as a useful resource for downstream tissue-specific classification tasks.

While SPIDER has shown good performance, several of its limitations should be acknowledged. First, condition-specific features could be enhanced depending on the availability of such data. Second, training and validation are limited by the relatively small number of condition-specific networks, particularly in human. With the accumulation of condition-specific data, future work could investigate how networks vary between different conditions, providing insights into the
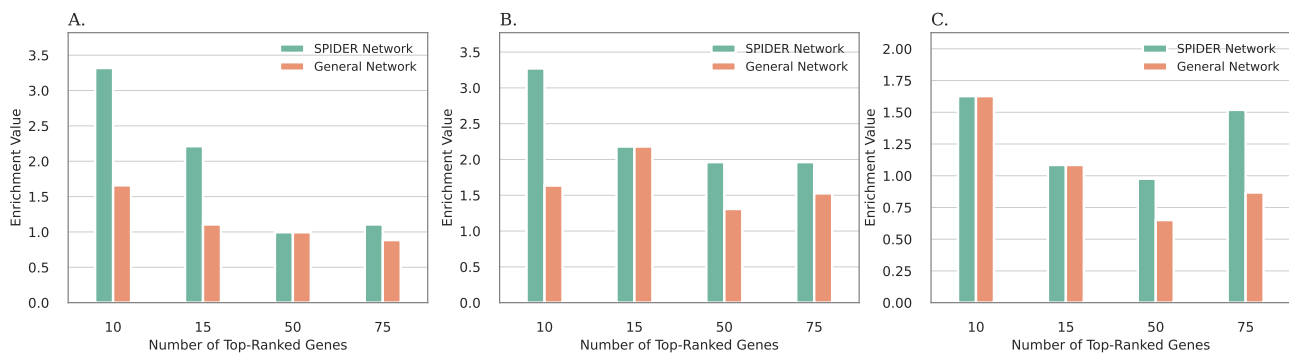

Fig. 5. AUROC box plots for tissue-associated disease gene prediction.

Fig. 6. Enrichment values w.r.t. cancer-associated genes of top K predictions in Breast Cancer (A), Colon Adenocarcinoma (B), and Ovarian Cancer (C) cancer-specific networks.

dynamic nature of PPIs and their roles in different biological contexts.

Competing interests

No competing interest is declared.

Funding

References

1. Roded Sharan, Igor Ulitsky, and Ron Shamir. Network-based prediction of protein function. *Molecular systems biology*, 3(1):88, 2007.

2. David E Gordon, Gwendolyn M Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M White, Matthew J O'Meara, Veronica V Rezelj, Jeffrey Z Guo, Danielle L Swaney, et al. A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583(7816):459–468, 2020.

3. Nahid Safari-Alighiarloo, Mohammad Taghizadeh, Mostafa Rezaei-Tavirani, Bahram Goliaei, and Ali Asghar Peyvandi. Protein-protein interaction networks (ppi) and complex diseases. *Gastroenterology and Hepatology from bed to bench*, 7(1):17, 2014.

4. Alice Bossi and Ben Lehner. Tissue specificity and the human protein interaction network. *Molecular systems biology*, 5(1):260, 2009.

5. Eran Segal, Nir Friedman, Daphne Koller, and Aviv Regev. A module map showing conditional activity of expression modules in cancer. *Nature genetics*, 36(10):1090–1098, 2004.

6. Edward L Huttlin, Lily Ting, Raphael J Bruckner, Fana Gebreab, Melanie P Gygi, John Szpyt, Stanley Tam, Gabriela Zarraga, Greg Colby, Kurt Baltier, et al. The bioplex network: a systematic exploration of the human interactome. *Cell*, 162(2):425–440, 2015.

7. Joseph Szymborski and Amin Emad. Rapppid: towards generalizable protein interaction prediction with awd-lstm twin networks. *Bioinformatics*, 38(16):3958–3967, 2022.

8. Raulia Syrlybaeva and Eva-Maria Strauch. Deep learning of protein sequence design of protein–protein interactions. *Bioinformatics*, 39(1):btac733, 2023.

9. Shubhrangshu Ghosh and Pralay Mitra. Matpip: A deep-learning architecture with explainable ai for sequence-driven, feature mixed protein-protein interaction prediction. *Computer Methods and Programs in Biomedicine*, 244:107955, 2024.

10. John Rachlin, Dikla Dotan Cohen, Charles Cantor, and Simon Kasif. Biological context networks: a mosaic view of the interactome. *Molecular Systems Biology*, 2(1):66, 2006.

11. Martin H Schaefer, Tiago JS Lopes, Nancy Mah, Jason E Shoemaker, Yukiko Matsuoka, Jean-Fred Fontaine, Caroline Louis-Jeune, Amie J Eisfeld, Gabriele Neumann, Carol Perez-Iratxeta, et al. Adding protein context to the human protein-protein interaction network to reveal meaningful interactions. *PLoS computational biology*, 9(1):e1002860, 2013.

12. Oded Magger, Yedael Y Waldman, Eytan Ruppin, and Roded Sharan. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS computational biology*, 8(9):e1002690, 2012.

13. Maya Ziv, Gil Gruber, Moran Sharon, Ekaterina Vinogradov, and Esti Yeger-Lotem. The tissuenet v.3 database: Protein-protein interactions in adult and embryonic human tissue contexts. *Journal of Molecular Biology*, 434(11):167532, 2022. Computation Resources for Molecular Biology.

14. Diederik S Laman Trip, Marc van Oostrum, Danish Memon, Fabian Frommelt, Delora Baptista, Kalpana Panneerselvam, Glyn Bradley, Luana Licata, Henning Hermjakob, Sandra Orchard, et al. An atlas of protein-protein associations of human tissues prioritizes candidate disease genes. *bioRxiv*, pages 2024–05, 2024.

15. Andreas Ruepp, Brigitte Waegele, Martin Lechner, Barbara Brauner, Irmtraud Dunger-Kaltenbach, Gisela Fobo, Goar Frishman, Corinna Montrone, and H-Werner Mewes. Corum: the comprehensive resource of mammalian protein complexes—2009. *Nucleic acids research*, 38(suppl_1):D497–D501, 2010.

16. Anthony Federico and Stefano Monti. Contextualized protein-protein interactions. *Patterns*, 2(1):100153, 2021.

17. Marinka Zitnik and Jure Leskovec. Predicting multicellular function through multi-layer tissue networks.

Bioinformatics, 33(14):i190–i198, 2017.

18. Casey S Greene, Arjun Krishnan, Aaron K Wong, Emanuela Ricciotti, Rene A Zelaya, Daniel S Himmelstein, Ran Zhang, Boris M Hartmann, Elena Zaslavsky, Stuart C Sealfon, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics*, 47(6):569–576, 2015.

19. Michelle M Li, Yepeng Huang, Marissa Sumathipala, Man Qing Liang, Alberto Valdeolivas, Ashwin N Ananthakrishnan, Katherine Liao, Daniel Marbach, and Marinka Zitnik. Contextualizing protein representations using deep learning on protein networks and single-cell data. *bioRxiv*, 2023.

20. Michael A Skinnider, Nichollas E Scott, Anna Prudova, Craig H Kerr, Nikolay Stoynov, R Greg Stacey, Queenie WT Chan, David Rattray, Jörg Gsponer, and Leonard J Foster. An atlas of protein-protein interactions across mouse tissues. *Cell*, 184(15):4073–4089, 2021.

21. Rose Oughtred, Chris Stark, Bobby-Joe Breitkreutz, Jennifer Rust, Lorrie Boucher, Christie Chang, Nadine Kolas, Lara O'Donnell, Genie Leung, Rochelle McAdam, Frederick Zhang, Sonam Dolma, Andrew Willems, Jasmin Coulombe-Huntington, Andrew Chatr-aryamontri, Kara Dolinski, and Mike Tyers. The BioGRID interaction database: 2019 update. *Nucleic Acids Research*, 47(D1):D529–D541, 11 2018.

22. Edward L Huttlin, Raphael J Bruckner, Jose Navarrete-Perea, Joe R Cannon, Kurt Baltier, Fana Gebreab, Melanie P Gygi, Alexandra Thornock, Gabriela Zarraga, Stanley Tam, et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, 184(11):3022–3040, 2021.

23. Xiaoqing Peng, Jianxin Wang, Wei Peng, Fang-Xiang Wu, and Yi Pan. Protein–protein interactions: detection, reliability assessment and applications. *Briefings in bioinformatics*, 18(5):798–819, 2017.

24. Mahmoud Ghandi, Franklin W Huang, Judit Jané-Valbuena, Gregory V Kryukov, Christopher C Lo, E Robert McDonald, Jordi Barretina, Ellen T Gelfand, Craig M Bielski, Haoxin Li, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature*, 569(7757):503–508, 2019.

25. Shuhei Noguchi, Takahiro Arakawa, Shiro Fukuda, Masaaki Furuno, Akira Hasegawa, Fumi Hori, Sachi Ishikawa-Kato, Kaoru Kaida, Ai Kaiho, Mutsumi Kanamori-Katayama, et al. Fantom5 cage profiles of human and mouse samples. *Scientific data*, 4(1):1–10, 2017.

26. Daniel N Itzhak, Colin Davies, Stefka Tyanova, Archana Mishra, James Williamson, Robin Antrobus, Jürgen Cox, Michael P Weekes, and Georg HH Borner. A mass spectrometry-based approach for mapping protein subcellular localization reveals the spatial proteome of mouse primary neurons. *Cell reports*, 20(11):2706–2718, 2017.

27. J Lynn Fink, Rajith N Aturaliya, Melissa J Davis, Fasheng Zhang, Kelly Hanson, Melvena S Teasdale, Chikatoshi Kai, Jun Kawai, Piero Carninci, Yoshihide Hayashizaki, et al. Locate: a mouse protein subcellular localization database. *Nucleic Acids Research*, 34(suppl_1):D213–D217, 2006.

28. Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

29. Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

30. LF Signorini, T Almozlino, and Roded Sharan. Anat 3.0: a framework for elucidating functional protein subnetworks using graph-theoretic and machine learning approaches. *BMC bioinformatics*, 22:1–6, 2021.

31. Lihua Jiang, Meng Wang, Shin Lin, Ruiqi Jian, Xiao Li, Joanne Chan, Guanlan Dong, Huaying Fang, Aaron E Robinson, François Aguet, et al. A quantitative proteome map of the human body. *Cell*, 183(1):269–283, 2020.

32. Marion Gremse, Antje Chang, Ida Schomburg, Andreas Grote, Maurice Scheer, Christian Ebeling, and Dietmar Schomburg. The brenda tissue ontology (bto): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic acids research*, 39(suppl_1):D507–D513, 2010.

33. Mathias Uhlen, Per Oksvold, Linn Fagerberg, Emma Lundberg, Kalle Jonasson, Mattias Forsberg, Martin Zwahlen, Caroline Kampf, Kenneth Wester, Sophia Hober, et al. Towards a knowledge-based human protein atlas. *Nature biotechnology*, 28(12):1248–1250, 2010.

34. Eyal Simonovsky, Moran Sharon, Maya Ziv, Omry Mauer, Idan Hekselman, Juman Jubran, Ekaterina Vinogradov, Chanan M Argov, Omer Basha, Lior Kerber, et al. Predicting molecular mechanisms of hereditary diseases by using their tissue-selective manifestation. *Molecular Systems Biology*, page e11407, 2023.

35. Lenore Cowen, Trey Ideker, Benjamin J Raphael, and Roded Sharan. Network propagation: a universal amplifier of genetic associations. *Nature Reviews Genetics*, 18(9):551–562, 2017.

36. Zbyslaw Sondka, Sally Bamford, Charlotte G Cole, Sari A Ward, Ian Dunham, and Simon A Forbes. The cosmic cancer gene census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer*, 18(11):696–705, 2018.