Pacific Symposium on Biocomputing 27:373-384(2022)

Multi-Omic Graph Transformers for Cancer Classification and Interpretation

Emily Kaczmarek[1,†], Amoon Jamzad[1], Tashifa Imtiaz[2], Jina Nanayakkara[2], Neil Renwick[2], and Parvin Mousavi[1]

[1]*Medical Informatics Laboratory, School of Computing, Queen's University, Kingston, K7L 3N6, Canada*
[2]*Laboratory of Translational RNA Biology, Department of Pathology and Molecular Medicine, Queen's University, Kingston, K7L 3N6, Canada*
[†]*E-mail: emily.kaczmarek@queensu.ca*

Next-generation sequencing has provided rapid collection and quantification of 'big' biological data. In particular, multi-omics and integration of different molecular data such as miRNA and mRNA can provide important insights to disease classification and processes. There is a need for computational methods that can correctly model and interpret these relationships, and handle the difficulties of large-scale data. In this study, we develop a novel method of representing miRNA-mRNA interactions to classify cancer. Specifically, graphs are designed to account for the interactions and biological communication between miRNAs and mRNAs, using message-passing and attention mechanisms. Patient-matched miRNA and mRNA expression data is obtained from The Cancer Genome Atlas for 12 cancers, and targeting information is incorporated from TargetScan. A Graph Transformer Network (GTN) is selected to provide high interpretability of classification through self-attention mechanisms. The GTN is able to classify the 12 different cancers with an accuracy of 93.56% and is compared to a Graph Convolutional Network, Random Forest, Support Vector Machine, and Multilayer Perceptron. While the GTN does not outperform all of the other classifiers in terms of accuracy, it allows high interpretation of results. Multi-omics models are compared and generally outperform their respective single-omics performance. Extensive analysis of attention identifies important targeting pathways and molecular biomarkers based on integrated miRNA and mRNA expression.

Keywords: Graph Neural Networks; miRNA; mRNA; Cancer Classification

1. Introduction

Cancer detection and classification is currently based on histopathological analysis and additional ancillary testing. However, histopathologic classification can be subject to interobserver variability, and cancer classification is regularly revised by the World Health Organization and other panels of international experts based on novel technologies and findings.[1,2] In particular, integrated morphology-based and molecular analyses are becoming increasingly important for classification.[2] Next-generation sequencing has allowed the rapid acquisition of 'big' biological data, which may improve the current subjectivity of cancer classification and further understanding of cancer including interactions between biological molecules. There is, therefore, a need for improved development and application of computational methods that are capable

of handling large-scale data. Machine learning approaches, and more recently deep learning architectures have been developed to manage the requirements of big data.[3] However, the majority of these techniques do not directly encode discrete relationships between input data, and there remains a lack of interpretability on how predictions are made. Graph neural networks are a field of deep learning in which inputs are represented structurally as a set of nodes (objects) and edges (relationships between nodes), and numerically through data corresponding to nodes or edges (features). For example, biological molecules and their expression from next-generation sequencing can be modelled as nodes and node features respectively, with interactions between them included as edges. Graph neural networks allow modelling and extraction of features from interacting biological molecules. In particular, graph neural networks use message-passing between connected nodes to update node features based on neighbouring information. A novel graph neural network architecture, Graph Transformer Network (GTN),[4] was recently introduced that in addition to modeling interactions between data, provides interpretability through self-attention mechanisms which identify important parts of the graphs for prediction. They are, therefore, particularly well-suited for analysis and understanding of large biological networks.

microRNAs (miRNAs) are small (∼22 nt), non-coding RNA molecules that negatively regulate protein-coding messenger RNAs (mRNAs).[5] The dysregulation of both miRNA and mRNA expression has previously been associated with the presence and/or type of cancer. These molecules are particularly well-suited as cancer biomarkers since their expression can be specific to cell-type and disease-stage, and they can act as oncomiRs/oncogenes (promoting cancer growth) or tumour suppressors.[6–8] miRNAs target mRNAs through a 2-8 nt seed sequence, with the complementary pairing on the mRNA.[5] Targeting pathways are particularly complex due to the high number of targets per miRNA and the shared targets between miRNAs. miRNA targeting is additive, meaning the combined power of multiple miRNAs with the same target can repress the expression of a single mRNA. In addition, the expression of mRNAs is largely reliant on miRNAs. For example, the upregulation of a miRNA can lead to the increased targeting of a mRNA, resulting in downregulated mRNA expression. This creates a large network of molecules with dependent expression, which suggests that biomarkers should be considered as targeting pathways instead of single molecules. Although the relationship between miRNAs and mRNAs is well-suited for analysis with graph neural networks, there is currently a lack of studies in this area.

Single-omics analysis has been previously studied and shown to yield high classification performance and insight into cancer biomarkers. For example, Ramirez *et al.* (2020) classified data from 33 cancer types and a 'normal' group using graph convolutional neural networks, with an accuracy of 94.7%.[9] They use data from the The Cancer Genome Atlas (TCGA)[10] where each patient is modelled as a graph, with nodes representing mRNAs and edges either protein-protein interactions or co-expression of molecules. Multi-omics methods in the literature have used *early-*,[11,12] *late-*,[13,14] or *intermediate-fusion*[15,16] of data, referring to immediate concatenation of data types, individual data analysis followed by combining of results, or concatenation of information at some point between raw data and results, respectively. A particularly relevant study by Wang *et al.* (2021) uses multi-omic data from TCGA for clas-

sification tasks, including identification of breast cancer subtypes and grading of gliomas.[16] Three individual graph convolutional networks are used to analyze miRNA, mRNA, and DNA methylation data followed by integration of results for final classification. While a few studies include constraints based on targeting to influence feature selection or model training, the majority do not use targeting information or molecular interactions explicitly in computational model creation or analysis. In addition to machine learning approaches, multi-omic interactions have also been studied through disease networks. Wang *et al.* (2015) created a network of interacting miRNAs, genes, phenotypes, and other clinical data which can be used to identify molecules associated with disease.[17] However, the majority of disease networks do not incorporate molecular measurements at individual patient level but capture population trends to find associations or to perform classification. There exists a lack of studies that computationally encode targeting pathways and patient-specific molecular expressions for multi-omic cancer classification and interpretation.

In this paper, we present a new approach to directly incorporate the interaction and dependent-expression between miRNAs and mRNAs to perform cancer classification. We develop a novel way of forming graphs that model the biological interactions and targeting between miRNAs and mRNAs. Specifically, the biological communication between miRNAs and mRNAs can be modelled with our design through direct message-passing in graph neural networks. Data from TCGA[10] are used to create graphs based on miRNA-mRNA targeting pathways from TargetScan.[18] Graph Transformer Networks (GTNs)[4] are specifically selected for their highly interpretable self-attention mechanisms, used to analyze miRNA-mRNA interactions of our designed graphs. We contrast our approach with several baseline methods including Graph Convolutional Network (GCN), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). Single-omic models are also implemented for comparison to multi-omic performance. In addition to classification, we perform extensive interpretation of the network attention to both edges and nodes of the GTN. Pathways (edges) between miRNAs and mRNAs reveal important targeting relationships, while multi-omic biomarkers (nodes) are discovered using the relationship between interacting molecules.

2. Materials and Methods

The schematics of our proposed approach is shown in Fig. 1. Briefly, we obtain patient-matched miRNA and mRNA data for 12 cancer classes from TCGA. Data is preprocessed and each patient sample (of both miRNA and mRNA expressions) is mapped as a bipartite graph using targeting relationships obtained from TargetScan. Since the proposed graph creation approach incorporates the expressions of miRNA and mRNA as node features, it is important to distinguish and identify one node from another. This is performed through positional encoding, where nodes are assigned additional features to describe their relative position in the graph and differentiate nodes.[4] Our architecture includes an initial linear layer that increases the dimensionality of node features, followed by two Graph Transformer Layers (GTLs) for node update, and a Multilayer Perceptron (MLP) network for classification. Each GTL consists of self-attention mechanisms that update node features (miRNA/mRNA expressions) based on expressions of neighbouring nodes, thus providing contextual information from connected

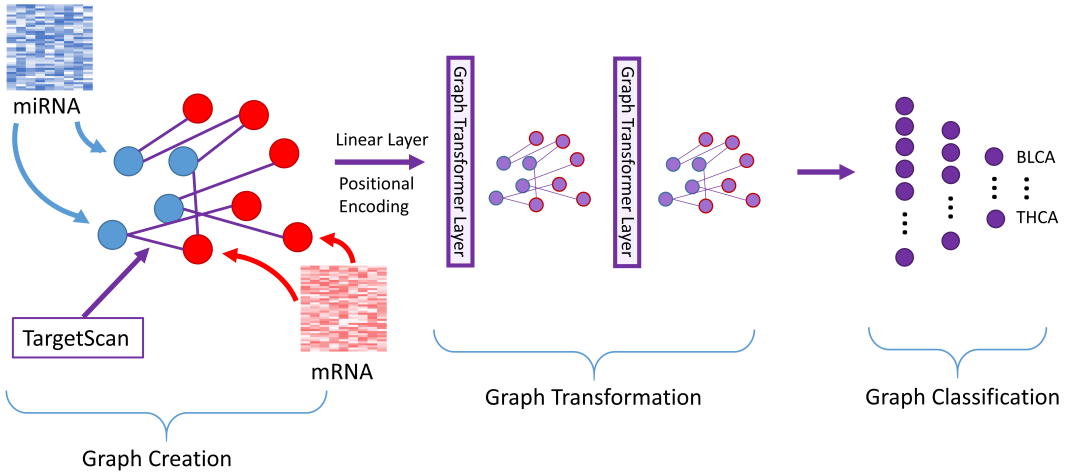

Fig. 1: Overview of our proposed multi-omic Graph Transformer Network. Preprocessed patient-matched miRNA and mRNA data from TCGA are formatted into bipartite graphs (one per patient) where each node represents a molecule, edges are based on miRNA-mRNA targeting, and node features are the expression of the respective molecule in that patient. The network consists of a linear layer, followed by two Graph Transformer Layers for message-passing between neighbouring node features, and a three-layer MLP for graph (patient) classification.

(interacting) nodes. This occurs through message-passing, where information from one node is used to update another. Message-passing of node features through graph edges can be used to *(i)* classify cancer using relationships between interacting molecules with dependent expression, and *(ii)* identify important targeting pathways based on this dependent expression. The features from the output of the last GTL are averaged as inputs to a 3-layer MLP and finally, graph classification is performed to categorize each patient to a cancer class. The attention of the trained GTN is then used to identify significant molecules and pathways for classification of different cancers.

2.1. *Data and preprocessing*

We use miRNA and mRNA TCGA data of 12 cancer types from the Broad Institute GDAC Firehose.[10] miRNA -5p and -3p expression counts are extracted from the miRNA isoform expression files, while mRNA data are from the RSEM normalized files (upper quartile normalized). miRNA and mRNA data are preprocessed separately through an established pipeline.[19] Figure 2 shows the distribution of samples for each cancer type after preprocessing (average number of samples per cancer is 475 (range

Fig. 2: Final data distribution. The number of samples per class ranges from 301 (CESC) to 750 (BRCA). Abbreviations used throughout the paper are carcinomas of: Bladder urothelial (BLCA); Breast invasive (BRCA); Cervical/endocervical (CESC); Head and neck squamous cell (HNSC); Pan-kidney (KIPAN); Brain lower grade glioma (LGG); Liver hepatocellular (LIHC); Lung adenocarcinoma (LUAD); Lung squamous cell (LUSC); Prostate (PRAD); Stomach and esophageal (STES); Thyroid (THCA).

301-750)). Final data from all cancers are aggregated and yield a total of 1,256 miRNAs and 20,502 mRNAs. miRNA and mRNA data are patient-matched based on the first 15 digits of the TCGA barcode, corresponding to the patient ID and sample ID. Only one sample per patient is included in the study. This yields 5,743 patient samples, 50% of which are used for training, 10% for validation, and 40% as test. Data is stratified by cancer type and patient (*i.e.*, training, validation, and test sets are mutually exclusive and do not contain any overlapping patients while they have similar composition of cancer samples). miRNA and mRNA filtering is then performed based on samples in the training and validation sets. First molecules with expression of zero in over 75% of samples are removed. Subsequently, miRNA and mRNA are filtered such that only the top 25% of molecules with the largest variance across samples are kept.

2.2. *Graph creation*

We create graphs based on the inherent targeting relationship between miRNAs and mRNAs. Each patient sample is represented by one graph, where each miRNA and mRNA represent a node, and node features are the respective expression of the molecule in that patient. Edges are formed between nodes based on miRNA-mRNA targeting. This forms a bipartite graph with edges only between different molecules, as shown in Fig. 1 (Graph Creation). Edges are bidirectional between miRNAs and mRNAs and pass information in both directions (i.e., a highly expressed mRNA influences the attention on the miRNA it's targeted by and vice versa). miRNA targets and context scores are acquired from TargetScan.[18] We keep the size of the graph limited by only selecting the top 10 targets per miRNA, using the highest TargetScan context scores. All graphs have identical structures in terms of the number of nodes and edges; the only differences are the node features. Following graph creation, nodes with one edge are pruned iteratively until all nodes have a minimum of two edges. The final graphs feature 168 miRNAs and 379 mRNAs for 5,743 patients (samples). Once final molecules are identified through graph pruning, miRNA and mRNA data are separately rank normalized to a similar scale for interaction in the GTN. Since each node represents a specific molecule and its corresponding expression, it is important to add the information about the location of nodes with respect to the whole graph to differentiate between them. To do so, the position of nodes can be embedded as additional node features that are specific to graph structure. For this study, Laplacian positional encoding is applied to all graphs before classification.[4]

2.3. *Network structure*

The network structure consists of a node-level linear layer, two Graph Transformer Layers, and a three-layer MLP. The dimensionality of node features (molecular and positional) is first increased in the node-level linear layer. The GTL layers preserve the structure of the graph while updating individual node features through message-passing. The update is based on an attention mechanism that incorporates information from neighbouring nodes that are connected by edges as presented in Eq. 1 and 2 below.

$$w_{ij}^{k,l} = \text{softmax}_j\left(\frac{Q^{k,l}h_i^l \cdot K^{k,l}h_j^l}{\sqrt{d_k}}\right). \tag{1}$$

$$\hat{h}_i^{l+1} = O_h^l \Big\|_{k=l}^{H} \Big(\sum_{j \epsilon \mathcal{N}_i} w_{ij}^{k,l} V^{k,l} h_j^l\Big). \quad (2)$$

Here, node features are represented by h_i^l and h_j^l, where i indicates the current node, j indicates the neighbouring nodes of i, and l represents the current layer. Each GTL uses H=8 parallel attention mechanisms or "heads" with different initializations to increase the stability and generalizability of the training process. The k^{th} attention $w_{ij}^{k,l}$ that determines the importance of the neighbour node j to node i, is calculated through the linear transformation of node features i.e. dot-product of learnable weights (Q, K), normalized by dimensionality of node features. Attentions of neighbour nodes are then used as weights for their corresponding node features (transformed with learnable weight matrix V). The weighted summation over the whole neighborhood is calculated for each attention head, and concatenated over all H heads (Eq. 2). The result is passed through an additional linear transformation (O), and later through a two-layer feedforward network with residual layers to generate the updated node features. After the last GTL, the graph is converted to a single feature vector by dismissing the edges and averaging all the nodes, and passed through a three-layer MLP for final classification.

2.4. *Implementation and ablation studies*

All implementation is in Google Colaboratory[20] using the Deep Graph Library.[21] The parameters of the network include training for 400 iterations with a batch size of 16, batch and layer normalization after each GTL, and dynamic learning rate of 0.001 reduced by a scale of 0.8 and a patience of 12. The model with the lowest validation loss is kept for final evaluation.

Extensive ablation studies are performed to choose the optimal hyperparameters for cancer classification. Specifically, the number of GTLs (1-5), attention heads (2-20, in steps of 2) and hidden features per attention head (1-15), type of normalization after each GTL (layer, batch, layer+batch), and the effect of positional encoding (implemented or not) are tested. For each parameter, five trials with different seeds are run, and the mean and standard deviation of the validation set performance are compared. The parameters that result in the highest performance are used in all subsequent analysis. The network with chosen parameters is run for 30 iterations with different seeds, and final performance is averaged across all models.

2.5. *Evaluation*

Comparison to baseline methods: The performance of the GTN is compared to GCN, RF, SVM, and MLP. The GCN is implemented using the Deep Graph Library with the same input graphs and parameters as the GTN, with the exception of an initial learning rate of 0.005. The RF, SVM, and MLP are implemented through Scikit-Learn[22] in Python with base parameters. The miRNA and mRNA expressions used in the GTN are directly used as inputs to the models. The GCN, RF and MLP run for thirty iterations using the same seeds as the GTN. The SVM is only run once (as performance does not change with different seeds).

Comparison to single-omics: We compare the performance of our approach for each model (GTN, GCN, RF, SVM, MLP) using single-omics of either miRNA or mRNA molecules. For the GTN and GCN, miRNA graphs are created with each node representing one miRNA,

and edges between nodes based on miRNAs targeting the same mRNA. Similarly, for mRNA graphs, all nodes represent a mRNA with edges based on being targeted by the same miRNA. These graphs use the same miRNA or mRNA nodes as the multi-omic graphs, but have a higher number of edges. The miRNA graphs feature 168 nodes and 4,162 edges, and the mRNA graphs have 379 nodes and 8,992 edges. All parameters are identical to the multi-omic networks. The single-omic RF, SVM, and MLP use miRNAs or mRNAs as inputs.

Attention: To interpret the attention of the GTN, the weights w_{ij} (Eq. 1) of the final GTL are used as the significance of each edge in the graph. After training, the attention of test samples is obtained and combined for each of the 30 models to yield a final attention value for all edges in the network. The attention to individual nodes is determined by the sum of the attention of all input edges to a node, normalized by the number of edges. All attention (both to edge and node) for a specific cancer class is summed and averaged by the number of samples in that class. The edge attention of the network identifies important miRNA-mRNA targeting pathways, while node attention determines significant miRNAs or mRNAs.

3. Results

3.1. *Ablation studies*

The final parameters found to optimize the classification performance are two GTLs, eight attention heads, six hidden features, layer and batch normalization, and implementing positional encoding. In addition to resulting in the highest performance, two GTLs will cause each node to incorporate features from other nodes a maximum of two edges away. Due to the bipartite graph structure, this means miRNAs will receive information from their immediate mRNA targets, and other miRNAs with the same targets (or vice versa). This mimics the true interaction of biological molecules; any connections past this point are not as important, biologically. Increasing the number of attention heads and hidden features improves performance until the chosen parameters above; after this point, the performance decreases. The normalization approach, and positional encoding do not influence the performance noticeably.

3.2. *Classification accuracy*

Overall, the GTN achieves a classification accuracy of 93.56%. The confusion matrix depicting the classification accuracy per class is shown as a heat map in Fig. 3. The diagonal of the confusion matrix is dark, indicating a high proportion of samples are correctly classified as their respective cancer. The two lung cancers (LUAD, LUSC, Fig. 3 red square) have a number of misclassifications between them, as well as squamous cell carcinomas from different anatomic sites (CESC, HNSC, and LUSC, Fig. 3 marked by *).

The performance of the GTN is next compared to four baseline models: a GCN, RF, SVM, and MLP. The mean accuracy of 30 trials is displayed in Table 1, with the standard deviation shown in brackets. The GCN has the lowest perofrmance of all models. The RF, SVM, and MLP all achieve comparable, slightly higher accuracy than the GTN. However, the GTN offers in-depth interpretability, which is not possible at the same level of depth with machine learning models. The performance of the multi-omic GTN and machine learning classifiers are

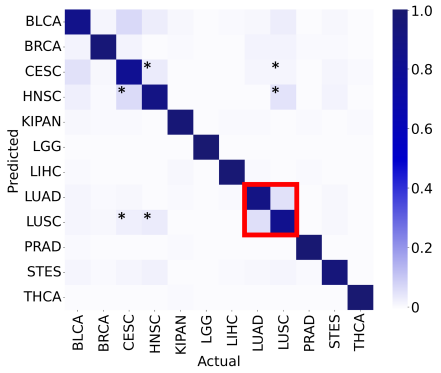

Table 1: Classification accuracy using Graph Transformer Network, Graph Convolutional Network, Random Forest, Support Vector Machine, and Multilayer Perceptron for multi-omic and single-omic studies.

Data	miRNA+mRNA	miRNA	mRNA
GTN	93.56 (0.91)	92.66 (0.82)	91.96 (1.09)
GCN	77.98(12.97)	79.61(12.51)	71.15 (8.17)
RF	96.29 (0.21)	94.29 (0.23)	94.69 (0.21)
SVM	96.52 (0.00)	94.17 (0.00)	95.60 (0.00)
MLP	97.97 (0.15)	95.90 (0.21)	96.89 (0.12)

Fig. 3: Left: Confusion matrix showing the classification accuracy as a heat map. The dark diagonal indicates each cancer is classified with high accuracy. Misclassifications are between two lung classes (LUAD and LUSC, red square), and squamous cell carcinomas (CESC, HNSC, and LUSC, *).

then compared against single-omic models. All multi-omic models outperform respective single-omic analyses apart from the GCN, shown in Table 1.

3.3. *GTN attention*

Pathway attention: The attention of the GTN is analyzed following training to determine important miRNA-mRNA biomarkers and targeting pathways. First, the edge attention (Eq. 1) is studied to identify important targeting pathways between miRNAs and mRNAs. We choose to analyze the targeting attention using breast cancer as a case study. Fig. 4 displays five of the top 15 targeting pathways (edges) identified by the GTN. Image (a) shows the ability of the attention to detect single edges, while image (b) finds a larger pathway consisting of multiple interacting miRNAs and mRNAs.

Molecule attention: We next analyze attention by individual molecules. Node attention is found by summing the attention of each edge leading into a destination node and normalizing by the number of edges (*i.e.*, a node with more edges will not have higher attention automatically). To simplify, five cancers are chosen to compare (BRCA, LGG, KIPAN, PRAD, and THCA); these cancers are commonly studied and have a high number of test samples in the dataset, preventing reliance on outliers. Node attention is averaged per chromosome, shown in Fig. 5 for miRNAs (a, blue) and mRNAs (b, red).

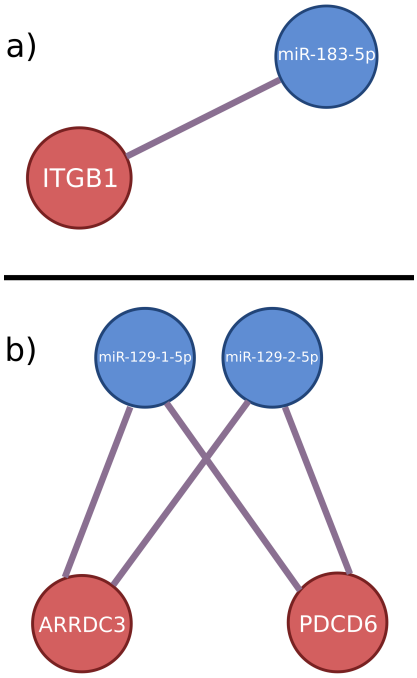

Fig. 4: Depiction of five of the top 15 edges identified by the attention of the GTN for breast cancer. Image (a) depicts a known targeting pathway, while (b) displays a number of molecules previously associated with breast cancer.

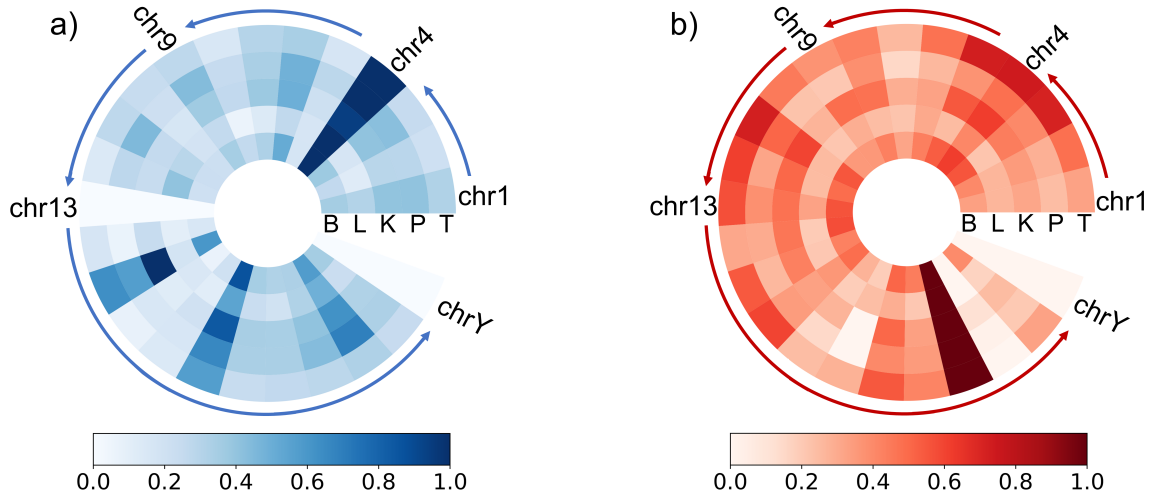

Fig. 5: Node attention of the GTN summarized for five cancers across chromosomes for miRNAs (a, blue) and mRNAs (b, red). Each layer of the sunburst represents one class. Dark squares indicate high attention to the miRNAs or mRNAs originating from that chromosome. The colour of each square (chromosome attention) is normalized against all other squares for that cancer. In both images, none of the final miRNAs or mRNAs originate from chromosome Y. Abbreviations: BRCA (B); LGG (L); KIPAN (K); PRAD (P); THCA (T).

High attention (darker squares) is seen in all five classes for miRNAs on chromosome 4 and 18 (Fig. 5 (a)). In Fig. 5 (b), mRNAs from chromosome 21 have high attention for all five classes. No molecules in this study originate from chromosome Y.

4. Discussion

In this study, we develop a Graph Transformer Network capable of discovering important miRNA-mRNA cancer pathways. The GTN classifies the 12 cancer types with an accuracy of 93.56%. Classification is compared to a GCN, RF, SVM, and MLP. The GTN does not perform as highly as some of the baseline models, but has the advantage of high interpretability. All multi-omic models outperform each of their respective single-omic miRNA and mRNA models, aside from the GCN. Extensive interpretation of network attention is performed to identify important miRNA-mRNA targeting pathways and potential biomarkers for cancer. Ablation studies are used to determine optimal hyperparameters of the network. In addition, design parameters are carefully considered to avoid overfitting to training data due to large input graphs and high number of classes. We constrain graph size by initially filtering a large number of molecules, only selecting 10 targets per miRNA, and pruning nodes with one edge. Only 12 classes from TCGA with a high number of samples are used. The design of the GTN (*e.g.*, number of layers), in addition to choice of external parameters such as graph size and number of classes, are chosen to optimize performance and generalizability.

The multi-omic GTN is able to classify the 12 cancer types with an accuracy of 93.56%. Figure 3 displays the confusion matrix showing high accuracy for each class. There are a number of misclassifications between LUAD and LUSC, which are two distinct lung cancer subtypes. The misclassifications may be due to the similar tissue properties (*i.e.*, overlapping

gene expressions) between the two cancers. It has also been shown that lung cancers can have mixed components of adenocarcinoma and squamous cell carcinoma,[23] which may be an additional factor in these misclassifications and important considerations for treatment selection. In addition, a number of misclassifications are also seen between CESC, HNSC, and LUSC. These three classes are (largely) squamous cell carcinoma samples from different anatomic sites; these likely share similar gene and miRNA expression profiles.

The performance of the multi-omic GTN is compared to a GCN, RF, SVM, and MLP (Table 1). The MLP outperforms all other models with an accuracy of 97.97%, followed by the SVM, RF, GTN, and GCN, respectively. The GCN and GTN have the lowest accuracy, likely due to the larger inputs to the network with both nodes and edges. In addition, both models take an average of each node feature across all nodes to use as input to the MLP. This summation may cause a large amount of information to be lost due to the high number of nodes. However, the GTN still achieves high performance, and has the additional benefit of excellent interpretability not possible with the other models. Specifically, the attention of the GTN can be analyzed to determine potential biomarkers and targeting pathways, which is not easily achieved (or possible) with the other models.

All multi-omic models are compared with and outperform their respective single-omic models using both miRNA and mRNA, aside from the GCN (Table 1). Most single-omics performance is similar to their respective multi-omics performance; it is possible the multi-omics based selection of individual miRNAs and mRNAs during preprocessing may have positively influenced the single-omics performance. For the graph networks, single miRNA or mRNA graphs require a substantial number of edges between miRNA-miRNA nodes (4,162) or mRNA-mRNA nodes (8,992) to retain the same amount of targeting information as in the miRNA-mRNA graphs. The high number of edges likely increases both the difficulty in determining important edges and the chance of overfitting to training data.

The edge attention of the GTN is used to determine important miRNA-mRNA pathways for each cancer. Figure 4 shows five of the top 15 edges identified by attention for breast cancer. Image (a) displays a validated targeting pathway identified in the literature, previously associated with melanoma and cervical cancer.[24] Due to the upregulated expression of miR-183 in breast cancer, it is possible this pathway is disrupted, and should be further explored.[24] In image (b), all molecules presented in this diagram have been previously associated as potential biomarkers for breast cancer. While these edges have not been identified as targeting pathways in cancer, they may be underexplored, or could indicate important relationships between the expression of these molecules (*e.g.*, co-expressed). Further research into specific expression should be performed for verification. The attention of GTNs can therefore be used for multiple pathway analyses (*i.e.*, locating pathways with large inverse targeting relationships, or identifying related biomarkers with high expression).

To find the attention to specific molecules, edge attention is aggregated for all edges at a destination node, and averaged for molecules from the same chromosome. In Fig. 5 (a), the miRNAs from chromosome 4 and 18 have high attention for all cancers. In this study, only miR-1269a originates from chromosome 4, which is highly associated with numerous cancers. Four miRNAs originate from chromosome 18: miR-133a, miR-1, miR-187, and miR-122. While

these have previously been identified as oncomiRs and tumour suppressors, miR-133a and miR-1 are also specific to muscle tissue.[25] It is possible the differing levels of muscle within the cancer types in this study are used for decision-making by the GTN; tissue specificity should be considered when identifying potential biomarkers. Figure 5 also identifies unique signatures of cancer. For example, in (a), the second inner ring (LGG) has noticeably different attention for chromosomes 6, 12, and 15. These can be analyzed further to identify specific biomarkers of LGG. The same analysis is performed for mRNAs (Fig. 5 (b)). All five cancers have high attention for chromosome 21, which consists of TFF2 and KCNJ15. These genes have previously been shown to affect cell apoptosis and migration in many cancers.

There are a number of limitations to this study. Healthy samples are not included due to the low number per tissue type in TCGA. It is possible some cancers may have improved classification based on tissue signature instead of neoplasticity signature. In addition, extensive pruning of the graph is performed. This includes high initial molecule filtering, and pruning nodes with one edge. While this is necessary to avoid overfitting (due to the high number of both nodes and edges), important pathways may have been eliminated from analysis. Lastly, additional interactions between mRNA-mRNA, and mRNA-RNA binding proteins are not included, which are also involved in mRNA regulation. Nonetheless, this initial study identifies the strong potential of graph neural networks for classification, data integration, and biomarker identification. It can be extended for specific applications, *e.g.*, using all targets of a few miRNAs for specific pathway analysis.

5. Conclusion

In this study, we propose a novel method of modelling miRNA-mRNA interactions through graph neural networks to classify 12 cancers. Specifically, graphs are formed using patient-specific miRNA and mRNA targeting relationships as edges, and expressions as node features, to model the direct biological communication between molecules. We use a highly interpretable GTN to classify 12 cancers from TCGA with an accuracy of 93.56%. While the GTN does not outperform all baseline models in terms of accuracy, its importance lies in high interpretability of targeting pathways and individual molecules. The performance of each model is also compared to single-omic miRNA and mRNA models. Four of the five multi-omic models outperform their respective single-omic model, showing the importance of integrating multiple data types. In addition, we perform extensive evaluation of attention to both edges (representing targeting pathways), and nodes (representing individual molecules), identifying important molecular biomarkers for cancer. Future work should examine datasets with balanced neoplastic and non-neoplastic samples, and incorporating TargetScan context scores into edge weights of graph inputs.

Acknowledgements

This work was supported by Ontario Graduate Scholarship; Vector Institute; National Sciences and Engineering Research Council of Canada; SEAMO New Clinician Scientist Award; AHSC AFP Innovation Fund; OMPRN Cancer Pathology Translational Research Award; and New Frontiers Research Fund.

References

1. M. Van Bockstal, M. Berlière, F. Duhoux and C. Galant, Interobserver variability in ductal carcinoma in situ of the breast, *American Journal of Clinical Pathology* **154** (2020).
2. G. Rindi *et al.*, A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal, *Modern Pathology* **31** (2018).
3. Y. Lecun, Y. Bengio and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
4. V. P. Dwivedi and X. Bresson, A generalization of transformer networks to graphs, in *AAAI Workshop on Deep Learning on Graphs: Methods and Applications*, 2021.
5. D. Bartel, Metazoan microRNAs, *Cell* **173**, 20 (2018).
6. P. Landgraf *et al.*, A mammalian microRNA expression atlas based on small RNA library sequencing, *Cell* **129**, 1401 (2007).
7. N. Renwick *et al.*, Multicolor microRNA FISH effectively differentiates tumor types, *J Clin Invest* **123**, 2694 (2013).
8. E. Y. H. P. Lee and W. J. Muller, Oncogenes and tumor suppressor genes, *Cold Spring Harb Perspect Biol* **2**, p. a003236 (2010).
9. R. Ramirez *et al.*, Classification of cancer types using graph convolutional neural networks, *Frontiers in Physics* **8** (2020).
10. B. I. of MIT Harvard, Firehose broad GDAC (2016).
11. Y. Cun and H. Fröhlich, Network and data integration for biomarker signature discovery via network smoothed t-statistics, *PLOS One* **8**, p. e73074 (2013).
12. Gade *et al.*, Graph based fusion of miRNA and mRNA expression data improves clinical outcome prediction in prostate cancer, *BMC Bioinformatics* **12** (2011).
13. O. P. Günther *et al.*, A computational pipeline for the development of multi-marker bio-signature panels and ensemble classifiers, *BMC Bioinformatics* **13** (2012).
14. F. Rohart, B. Gautier, A. Singh and K. A. Lê Cao, mixOmics: An R package for 'omics feature selection and multiple data integration, *PLOS Computational Biology* **13**, p. e1005752 (2017).
15. D. Kim *et al.*, Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction, *Journal of the American Medical Informatics Association* **22**, 109 (2015).
16. T. Wang *et al.*, MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification, *Nature Communications* **12** (2021).
17. L. Wang, D. Himmelstein, A. Santaniello, P. Mousavi and S. Baranzini, iCTNet2: integrating heterogeneous biological interactions to understand complex traits, *F1000Research* **5** (2015).
18. V. Agarwal, G. W. Bell, J.-W. Nam and D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, *eLife* **4**, p. e05005 (2015).
19. J. Nanayakkara *et al.*, Characterizing and classifying neuroendocrine neoplasms through microRNA sequencing and data mining, *NAR Cancer* **2**, p. zcaa009 (2020).
20. Google, Google. what is colaboratory? (n.d.), Available at: `https://colab.research.google.com/notebooks/intro.ipynb`.
21. M. Wang *et al.*, Deep graph library: Towards efficient and scalable deep learning on graphs.
22. F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python.
23. C. Li and H. Lu, Adenosquamous carcinoma of the lung, *OncoTargets and Therapy* **11**, p. 4829–4835 (2018).
24. D. Cao, M. Di, J. Liang, S. Shi, Q. Tan and Z. Wang, MicroRNA-183 in cancer progression, *Journal of Cancer* **11** (2020).
25. J.-F. Chen *et al.*, The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation, *Nature Genetics* **38** (2006).