Network Biology: Exploring Methylation Features in Cancer through Machine Learning

Sicheng Jing[1,†], Yao Sun[2,*,†], Hongji Zhu[3,†], Zihan Wang[4,†]

{jingsicheng123@126.com[1], floaeam@gmail.com[2],
hongjizhu@outlook.com[3], zihanw626@outlook.com[4]}

Department of Biological Sciences, University of California San Diego, La Jolla, CA, 92092, USA[1]
Zhejiang University-University of Edinburgh Institute, International Campus, Zhejiang University, Haining, Zhejiang, 314400, China[2]
School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 19104, USA[3]
Zhejiang University-University of Edinburgh Institute, International Campus, Zhejiang University, Haining, Zhejiang, 314400, China[4]
*corresponding author
[†]All the authors contributed equally to this work and should be considered as co-first author.

Abstract. Despite the growing understanding of DNA methylation's role in cancer biology, many existing studies focus on analyzing individual methylation sites or predefined gene sets, leaving the interactions between co-methylated regions largely unexplored. In this study, we apply a network-based approach to explore methylation features across six cancer types using the *OhmNet* model. By constructing co-methylation networks and utilizing multi-layer network embedding, we identify significant co-methylation patterns associated with key genes implicated in tumorigenesis. Through pathway enrichment analysis, we discovered key pathways related to cell adhesion and axonogenesis, suggesting a novel link between DNA methylation and nerve-cancer crosstalk. Our work not only reveals unique insights into the methylation landscape of cancers but also introduces a scalable, label-free network-based approach for studying complex epigenetic regulation.

Keywords: Machine Learning, Cancer, Epigenetics

1 Introduction

Cancer research has made significant progress in recent decades, but the molecular mechanisms driving tumorigenesis remain a focal point for understanding and targeting the disease. DNA methylation is one of the crucial epigenetic mechanisms in regulating gene expression. Aberrations in DNA methylation patterns are now recognized as key features of various cancer types, contributing to the silencing of tumor suppressor genes and activation of oncogenes. These changes are pivotal for cancer initiation, progression, and metastasis. Analyzing these patterns allows researchers to understand cancer biology better and potentially identify biomarkers for early diagnosis, prognosis, and treatment stratification.

DNA methylation is an epigenetic modification involving adding a methyl group to the 5-carbon of cytosine in CpG dinucleotides, primarily facilitated by DNA methyltransferases (DNMTs) [1]. This modification can regulate gene expression, typically by silencing genes when it occurs

CONF-MLA 2024, November 21, Adana, Turkey

DOI 10.4108/eai.21-11-2024.2354642

in gene promoter regions. Aberrant DNA methylation has been implicated in various diseases, including cancer, where it is known to play a dual role. Tumor suppressor genes are often hypermethylated and silenced in cancer cells, while global hypomethylation can lead to genomic instability and the activation of oncogenes [2].

DNA methylation alterations are highly dynamic in the context of cancer, often exhibiting cancer-type-specific patterns. These methylation changes are not only markers of cancer but can also play active roles in driving tumorigenesis [1]. Understanding these patterns can inform targeted therapies, as epigenetic drugs that modify methylation status are increasingly being explored as therapeutic agents.

In recent years, machine learning and network analysis have become powerful tools in biology. The *OhmNet* model [3], a hierarchy-aware unsupervised node feature learning method, has demonstrated its potential in analyzing complex, multi-layered biological networks by focusing on the hierarchical organization of proteins across different tissues. In prior studies, *OhmNet* has been used to classify tissue-specific functions by integrating protein-protein interactions across different tissues. This model emphasizes multi-scale tissue organization, offering better predictions of cellular functions than flat network models.

Similar to protein-protein interaction (PPI) networks, the relationships among genes can be modeled as interactions that reflect co-regulation, co-methylation, or mutual influence in various biological contexts. By treating methylation profiles as a multi-layer network where each layer corresponds to a specific cancer type, we can explore the similarities and differences in methylation patterns across cancers.

One key advantage of applying *OhmNet* to methylation data is its ability to analyze large-scale molecular interactions without requiring prior labeling, much like its application in PPI networks. Methylation data, like protein-protein interactions, often lacks clearly defined categories or labels, especially when investigating epigenetic changes in cancer. *OhmNet*, as an unsupervised model, can identify patterns of co-methylation that might indicate functional groupings of genes or pathways involved in cancer development without needing predefined labels or categories [4].

Using a network-based approach allows for the identification of co-methylated gene modules, which could provide insights into common pathways affected by epigenetic alterations in different cancers. This modular view of methylation patterns has the potential to highlight key regulatory nodes that are disrupted across multiple cancers or that are unique to specific cancer types [5].

The main contributions and advantages of our study:

- **Comparative Analysis Across Cancer Types:** By constructing methylation networks for different cancer types, we aim to uncover both conserved and cancer-specific methylation signatures. This could reveal shared epigenetic mechanisms across cancers, suggesting common pathways of tumorigenesis.
- **Identifying Epigenetic Biomarkers:** Methylation signatures are increasingly being used as biomarkers for cancer diagnosis and prognosis.[2] Network analysis of methylation data could identify clusters of co-methylated genes that serve as robust biomarkers for specific cancer types.

2 Methods

2.1 Data Collection and Processing

We collected Illumina HumanMethylation450K data for six types of cancer from the TCGA and BioStudies databases: Breast invasive carcinoma, Ovarian serous cystadenocarcinoma, Uterine Corpus Endometrial Carcinoma, Stomach adenocarcinoma, Lung squamous cell carcinoma, and Liver hepatocellular carcinoma. The data were integrated and processed to remove any missing values (NA) to ensure the quality and completeness of the dataset for subsequent analyses.

2.2 Co-methylation Network Construction

We performed a correlation analysis on the DNA methylation levels of CpG sites to explore co-methylation relationships between different CpG loci. We conducted multiple random extractions of the beta values for 10,000 CpG sites (beta values reflect the methylation levels of CpG loci, ranging from 0 to 1). We then calculated the Pearson correlation coefficient for every pair of CpG sites to quantify their linear relationship.

Let **B** represent the matrix containing the beta values of all CpG sites across samples, where rows correspond to CpG loci and columns correspond to samples. For each pair of CpG sites *i* and *j*, their correlation is computed using the following formula.

$$r_{ij} = \frac{cov(B_i, B_j)}{\sigma_{B_i}\sigma_{B_j}} \tag{1}$$

In equation (1), where r_{ij} denotes the correlation between CpG sites *i* and *j*, cov($\mathbf{B}_i$, $\mathbf{B}_j$) represents the covariance between the beta value vectors of the two sites, and $\sigma_{\mathbf{B}}i$ and $\sigma_{\mathbf{B}}j$ are the standard deviations of the respective vectors.

To filter out noise and non-significant correlations, we applied a correlation threshold *T*. Only pairs of CpG sites with an absolute correlation value $|r_{ij}|$ greater than the threshold *T* were retained for subsequent analysis.

2.3 Feature Learning Using Ohmnet

Based on the CpG site correlation analysis, we constructed an undirected network, referred to as the co-methylation graph. Each node represents a CpG site, and edges between two nodes indicate that the Pearson correlation coefficient between the two CpG sites exceeds a predefined threshold *T*. Connections between CpG pairs that do not meet this threshold are considered insignificant and thus excluded from the network.

After inputting this network into the *Ohmnet* model, the model generated an embedding for each CpG site. *Ohmnet* uses a multilayer network embedding technique to map the original high-dimensional network structure into a lower-dimensional vector space, where each CpG site is represented as a vector in this space.

We applied dimensionality reduction algorithms, t-distributed Stochastic Neighbor Embedding (t-SNE), and Principal Component Analysis (PCA), to project the embeddings from the high-dimensional space into two dimensions. We labeled each CpG site based on the cancer type of the sample.

2.4 Pathway Enrichment Analysis

We divided the space into multiple equally sized grid cells based on the PCA results. The size of each grid cell was determined by the parameter *grid width*, and each data point was assigned to its corresponding grid cell based on its coordinates in the 2D space.

We recorded the number of CpG sites corresponding to different cancer types for each grid cell. To ensure that a grid cell contained a significant co-methylation pattern, we retained only those grid cells that included at least *min points* data points for each cancer type. CpG sites in these enriched cells were considered potential key regulatory sites within these cancers.

We then mapped the CpG sites from the enriched cells to their corresponding genes and performed Gene Ontology (GO) enrichment analysis on the resulting gene set. The GO analysis encompassed three domains: Molecular Function (MF), Cellular Component (CC), and Biological Process (BP). We employed hypergeometric testing, combined with the Benjamini-Hochberg multiple hypothesis testing correction, to identify significantly enriched GO terms.

3 Results

3.1 Study Design

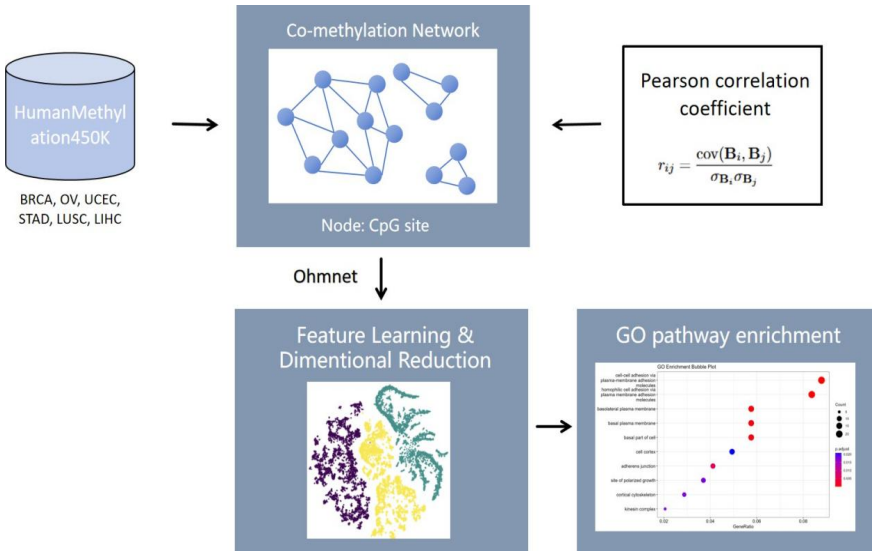

Fig. 1. Study Design

According to our study design (Fig. 1), we download methylation data for six types of cancer from the TCGA and BioStudies database. After standardized processing, we construct co-methylation networks for each type of cancer based on Pearson correlation coefficients. These networks are then input into the *Ohmnet* model for feature learning, followed by dimensionality reduction of the learned features. By analyzing the methylation sites in the enriched regions identified through dimensionality reduction, we identify the corresponding genes and further conduct gene enrichment analysis.

We collected methylation data for six types of cancer and constructed co-methylation networks for each. Using the *OhmNet* model, we learned the features and visualized the results. Based on methylation sites enriched in specific regions, we identified the corresponding genes and conducted further pathway enrichment analysis.

3.2 BRCA, OV, and UCEC

To evaluate whether *OhmNet* can identify shared CpG sites across different cancers, we first selected a few cancers known to exhibit similarities at the gene expression level. Specifically, we focused on breast invasive carcinoma (BRCA), ovarian serous cystadenocarcinoma (OV), and uterine corpus endometrial carcinoma (UCEC). These cancers share commonalities due to the involvement of several key genes, such as FGFR1OP2, BCL9L, and ACTB, which are known to play a role in their progression [6, 7, 8]. Our aim was to test whether our approach could accurately identify these genes. If successful, this would demonstrate the robustness of our method.

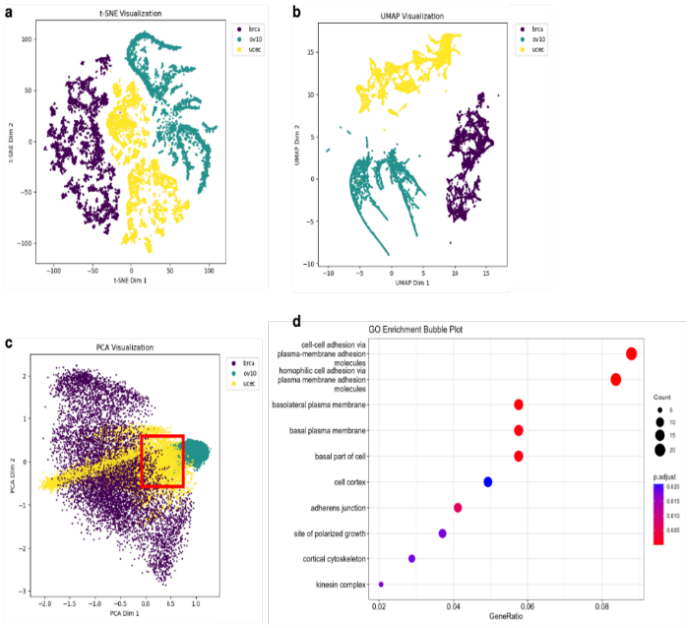

Fig. 2. Results of Breast Invasive Carcinoma (BRCA), Ovarian Serous cystadenocarcinoma (OV), and Uterine Corpus Endometrial Carcinoma (UCEC). **a.** The t-SNE visualization of *OhmNet* embedding results reveals distinct separation among the cancer clusters. **b.** The UMAP visualization illustrates well-separated cancer clusters from the *OhmNet* embedding results. **c.** The PCA visualization with the most densely overlapping region, is indicated by the red box. **d.** Results from the GO Enrichment analysis for genes within the overlapping region. The y-axis is the description of the corresponding pathways, while the x-axis shows the Gene Ratio, which represents the proportion of genes associated with a specific pathway relative to the total number of genes in the overlapping region. The size of the circles corresponds to the number of genes linked to each pathway, with larger circles indicating a higher gene count in the overlap. Additionally, the circle color reflects the p-adjust value; a smaller p-adjust signifies the greater significance of the results.

We began by obtaining methylation site data from the TCGA website, then constructed methylation networks (Methods) compatible with the *OhmNet* input format. As a result, *OhmNet* generated an output where each methylation site was represented in a 128-dimensional embedded matrix. To visualize the results, we applied three different dimensionality reduction techniques: t-SNE, UMAP, and PCA. Our goal was to visualize the embeddings while preserving both global and local structures in the data. Notably, the three cancers clustered distinctly in both t-SNE (Fig. 2a) and UMAP (Fig. 2b), highlighting the ability of *OhmNet* to effectively separate them. However, in the PCA plot (Fig. 2c), the clusters showed noticeable overlap. We believe this occurs because PCA emphasizes global variance, which helps retain broad patterns within the data. Unlike UMAP and t-SNE, which prioritize local structure and tend to separate clusters more distinctly, PCA maintains distance relationships across the entire dataset, which can result in overlapping patterns.

After confirming the embedding capabilities of *OhmNet*, we proceeded with downstream analysis. We began by selecting the most overlapped regions in the PCA plot (Fig. 2c). Our reasoning was that CpG sites (nodes) located near each other likely share similar 128-dimensional embeddings from *OhmNet*, which also suggests they have similar neighborhood structures or are closely connected in the constructed methylation site networks. We hypothesized that genes in this region are significant across all three cancers and may play crucial roles. Indeed, within this region, we successfully identified key genes contributing to Breast Cancer, Ovarian Cancer, and Endometrial Cancer: FGFR1OP2, BCL9L, and ACTB.

We also performed GO enrichment analysis to identify the pathways associated with these genes. As shown in the results (Fig. 2d), the most prominent pathway is cell-cell adhesion via plasma membrane adhesion molecules, which involves a relatively large number of genes and has a high gene ratio. The small p-adjust value further highlights its significance. The second-ranked pathway is homophilic cell adhesion via plasma membrane adhesion molecules. These findings are reasonable because these pathways are crucial in maintaining tissue structures, and disruption of these interactions facilitates cancer development [9, 10]. These results demonstrate that *OhmNet* can produce significant and meaningful outcomes, confirming that our study design effectively constructs a suitable input network for *OhmNet* to achieve these results.

3.3 STAD, BRCA, LUSC, and LIHC

After demonstrating *OhmNet* 's ability to identify CpG sites (or genes) associated with related cancers, we extended our approach to four cancers that lack obvious relationships. Our goal was to determine whether any shared CpG sites (or genes) exist across these cancers. If such genes are identified, their CpG sites could serve as potential therapeutic targets. Additionally, our study suggests that using *OhmNet* to uncover shared CpG sites may become a valuable way to discover key insights, potentially playing a broader role in future cancer research and treatment strategies. We obtained data on four cancer methylation sites from the TCGA website and constructed networks to serve as inputs for *OhmNet*. After conducting the analysis, we generated visualization plots by reducing dimensionality. The t-SNE plot (Fig. 3a) notably reveals that the Liver Cancer cluster is divided into two distinct parts, while the UMAP plot (Fig. 3b) demonstrates clear cluster separation. This phenomenon may be attributed to several factors. First, different liver cancer subtypes could exhibit distinct DNA methylation patterns. Second, t-SNE might amplify subtle differences, whereas UMAP provides a more global perspective. Lastly, the formation of two clusters could also be an artifact of how t-SNE manages the

complex structure of methylation data. In our PCA analysis, we repeated the previous pipeline to identify the most overlapping regions. However, we observed that the Stomach Cancer cluster did not overlap with any other cancer types (Fig. 3c). This finding suggests that the methylation site patterns in Stomach Cancer may differ significantly from those of other cancers. Given that over 50,000 methylation sites associated with these four cancers are available on the TCGA website, we repeated the process several times, randomly selecting 1,000 methylation sites for each iteration. Remarkably, all three types of visualization plots produced consistent results across these trials.

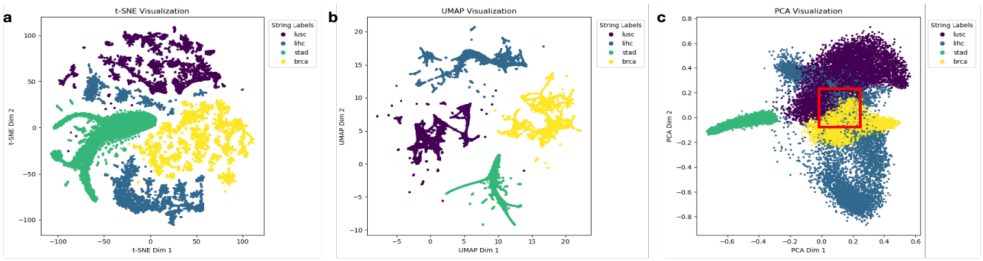

Fig. 3. Results of Stomach adenocarcinoma (STAD), Breast invasive carcinoma (BRCA), Lung squamous cell carcinoma (LUSC), and Liver hepatocellular carcinoma (LIHC). **a.** The t-SNE visualization of the *OhmNet* embedding results shows that the cancer clusters are roughly separated, though the Liver Cancer cluster appears to be truncated. **b.** The UMAP visualization showing that the cancer clusters are distinctly separated, highlighting clear delineations among the different cancer types. **c.** The PCA visualization identifies the most densely overlapping region, marked by the red box. Notably, the Stomach Cancer cluster does not overlap with any other clusters, indicating a unique expression profile.

From the GO analysis (Fig. 4), we focused on pathways that appeared in at least half of the total results. These include kidney development, pattern specification process, axonogenesis, central nervous system neuron differentiation, cell fate commitment, embryonic organ development, gland development, axon development, and renal system development. It's important to note that these pathways were generated solely from genes associated with BRCA, LUSC, and LIHC, as no overlapping regions were found for STAD. We anticipate that further exploration will uncover the roles these pathways may play in advancing cancer treatment.

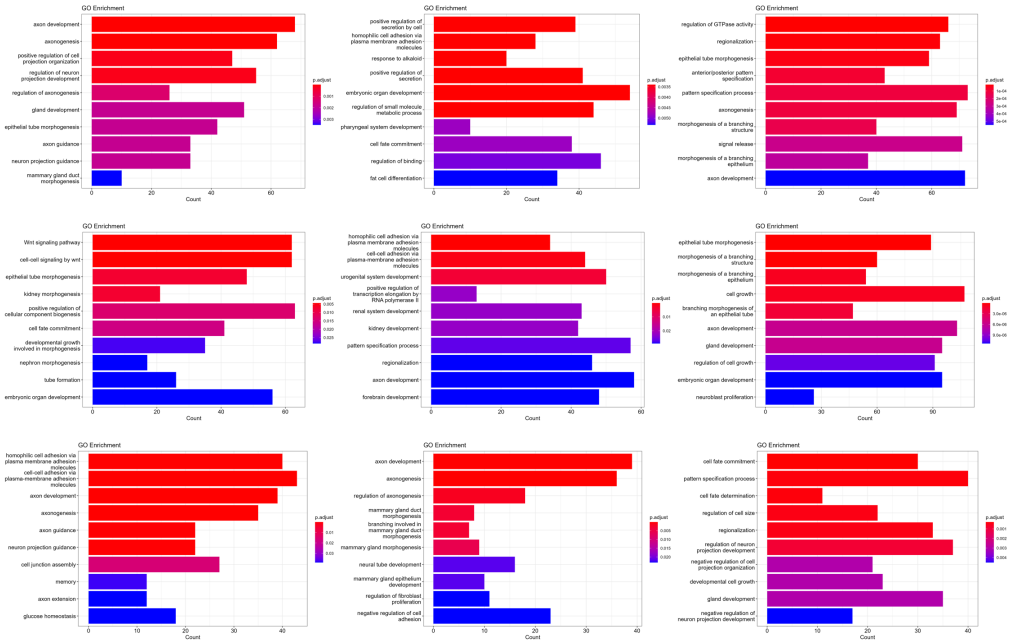

Fig. 4. GO Enrichment results of Stomach adenocarcinoma (STAD), Breast invasive carcinoma (BRCA), Lung squamous cell carcinoma (LUSC), and Liver hepatocellular carcinoma (LIHC). From the overall results, we selected nine key outcomes, each of which highlighted several important pathways. For each outcome, we randomly selected 1,000 CpG sites from the TCGA dataset.

4 Discussion

We generated networks of correlated CpG islands in tumors from online databases and applied the *Ohmnet* algorithm to these co-methylation networks for better feature learning. The GO enrichment result showed significant correlations of the common methylated CpG islands with gene function related to axon development and axonogenesis. This result suggested the involvement of neurogenesis in the tumor type we sampled (Breast invasive carcinoma, Lung squamous cell carcinoma, and Liver hepatocellular carcinoma), indicating the presence of nerve-cancer crosstalk. This phenomenon provides an important aspect in cancer research since there is a correlation between high intratumoral nerve density and high recurrence and poor prognosis in many types of solid tumors [11]. Our findings about functional enrichment in axonogenesis also indicate the important role of nerve-cancer crosstalk, which is often neglected in the study of tumor microenvironments. It was found that their interaction is a reciprocal process. Cancer cells promote innervation by secreting extracellular vesicles containing neurogenic molecules and axon guidance factors such as ephrin B1, while nerves can stimulate tumor growth by releasing certain neurotransmitters. The cancer-promoting microenvironment can be cultivated together by sympathetic and parasympathetic innervation [11]. Studies in mammary gland development have confirmed the role of axon guidance molecules in gland morphogenesis and epithelial homeostasis [12, 13]. Therefore, it is reasonable to assume that they also play a role in breast cancer. Studies confirmed that axon guidance molecules act as

tumor suppressors in the mammary tissue by inhibiting proliferation and metastasis [14]. Axon guidance molecules also play important roles in liver fibrosis, hepatocellular carcinoma, and multiple types of lung cancers [15, 16]. Our findings may re-emphasize the role of nerves and axon guidance molecules in cancer development from the methylation point of view, potentially contributing to the prosperous field of cancer neuroscience [17].

We used the *Ohmnet* algorithm [3] to automatically learn features of the co-methylation networks. *Ohmnet* is designed to optimize feature learning in multi-layer networks by utilizing the hierarchy information of layers based on the assumption that nodes in the nearby layers in the hierarchy share similar features. The embedding on a single layer is based on the *Node2Vec* algorithm [18], where it first uses random walk to embed nodes into low-dimension vector spaces, then defines an objective function to make node embedding that can predict the features of its neighborhood nodes. However, there are some shortcomings of this algorithm. *Node2Vec* belongs to shallow embedding approaches, which simply directly optimize an embedding vector for each node, lacking parameter sharing between different nodes [19]. Moreover, the logic behind the random walk method is to obtain an encoder network that represents the distances of input nodes. This might result in an over-emphasis on proximity information while compromising structural information [20]. Due to these limitations, the shallow embedding approaches are replaced by graph neural networks (GNNs)-based methods. For example, *Deep Graph Infomax* (DGI) replies on graph convolutional network architecture and trains encoder for maximizing mutual information between local and high-level graphs. It can better recognize the overall graph structure while preserving local information [20]. In the future, we could test the dataset on other graph representation learning methods and design more detailed downstream tasks to fully exploit the power of these algorithms. Due to the limitations of our computational resources and time, only a small percentage of available data were analyzed. The methylation data we used came from Infinium HumanMethylation450K, which is now replaced by the MethylationEPIC v2.0 with a significant increase of CpG islands covered [21, 22]. We could also sample more cancer types and possibly combine existing space and time information about the key transitions during cancer initiation and progression [23] to seek potential biomarkers or underlying mechanisms. Our work shows the potential of advanced computer science in analyzing biological data. The input can extend to networks other than DNA methylation, exploiting the power of state-of-the-art machine-learning technology in modern data-driven research.

5 Conclusion

Our study demonstrates the application of the *OhmNet* model in identifying co-methylation networks across different cancer types. By analyzing co-methylated gene clusters, we identified key regulatory nodes and pathways implicated in cancer progression, including those involved in cell adhesion and neurogenesis. Our findings support the emerging role of nerve-cancer crosstalk and axon guidance molecules in cancer biology. The results indicate that network-based approaches such as *OhmNet* hold promise for identifying shared epigenetic features and potential therapeutic targets across cancer types. Practically, these co-methylation patterns could possibly serve as epigenetic biomarkers for early diagnosis and prognosis, while cancer-specific patterns may improve personalized therapies targeting tumor-specific methylation changes.

Future work will aim to extend this analysis to additional data and explore advanced graph learning algorithms to further develop the cancer-specific methylation network.

Acknowledgements

Sicheng Jing, Yao Sun, Hongji Zhu, and Zihan Wang contributed equally to this work and should be considered co-first authors. We are very grateful to Professor Pietro Liò from the Department of Computer Science and Technology of the University of Cambridge for providing novel ideas and instructing us on basic knowledge in this field. We are thankful to Ash Hu for his generous help. Also, we thank all specimen donors and research groups involved in providing the data used in this study via TCGA and BioStudies.

References

[1] Karsakov, A. *et al.* Parenclitic network analysis of methylation data for cancer identification. *PloS one* 12, e0169661 (2017).

[2] Maros, M. E. *et al.* Machine learning workflows to estimate class probabilities for precision cancer diagnostics on dna methylation microarray data. *Nature protocols* 15, 479–512 (2020).

[3] Zitnik, M. & Leskovec, J. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* 33, i190–i198 (2017).

[4] Mo, Y., Peng, L., Xu, J., Shi, X. & Zhu, X. Simple unsupervised graph representation learning. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 7797–7805 (2022).

[5] Bartlett, T. E. *et al.* Corruption of the intra-gene dna methylation architecture is a hallmark of cancer. *Plos one* 8, e68285 (2013).

[6] Katoh, M. & Nakagama, H. Fgf receptors: Cancer biology and therapeutics. *Medicinal Research Reviews* 34, 280–300 (2014).

[7] Deka, J. *et al.* Bcl9/Bcl9l Are Critical for Wnt-Mediated Regulation of Stem Cell Traits in Colon Epithelium and Adenocarcinomas. *Cancer Research* 70, 6619–6628 (2010). URL https://doi.org/10.1158/0008-5472.CAN-10-0148. https://aacrjournals.org/ cancerres/article-pdf/70/16/6619/2636632/6619.pdf.

[8] Guo, C., Liu, S., Wang, J., Sun, M.-Z. & Greenaway, F. T. Actb in cancer. *Clinica Chimica Acta* 417, 39–44 (2013). URL https://www.sciencedirect.com/science/article/pii/ S0009898112005839.

[9] Okegawa, T., Pong, R.-C., Li, Y. & Hsieh, J.-T. The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochimica Polonica* 51, 445–457 (2004).

[10] Saiki, I. Cell adhesion molecules and cancer metastasis. *The Japanese Journal of Pharmacology* 75, 215–242 (1997).

[11] Silverman, D. A. *et al.* Cancer-associated neurogenesis and nerve-cancer cross-talk. *Cancer research* 81, 1431–1440 (2021).

[12] Brantley-Sieders, D. M. *et al.* The receptor tyrosine kinase epha2 promotes mammary adenocarcinoma tumorigenesis and metastatic progression in mice by amplifying erbb2 signaling. *The Journal of Clinical Investigation* 118, 64–78 (2008).

[13] Haldimann, M. *et al.* Deregulated ephrin-b2 expression in the mammary gland interferes with the development of both the glandular epithelium and vasculature and promotes metastasis formation. *International Journal of Oncology* 35, 525–536 (2009).

[14] Harburg, G. C. & Hinck, L. Navigating breast cancer: Axon guidance molecules as breast cancer tumor suppressors and oncogenes. *Journal of Mammary Gland Biology and Neoplasia* 16, 257–270 (2011).

[15] Nasarre, P., Potiron, V., Drabkin, H. & Roche, J. Guidance molecules in lung cancer. *Cell Adhesion and Migration* 4, 130–145 (2010).

[16] Basha, S., Jin-Smith, B., Sun, C. & Pi, L. The slit/robo pathway in liver fibrosis and cancer. *Biomolecules* 13, 785 (2023).

[17] Mancusi, R. & Monje, M. The neuroscience of cancer. *Nature* 618, 467–479 (2023).

[18] Grover, A. & Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 855–864 (Association for Computing Machinery, New York, NY, USA, 2016). URL https://doi.org/10.1145/2939672.2939754.

[19] Hamilton, W. L. *Graph Representation Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning (Springer International Publishing, Cham, 2020). URL https://link.springer.com/10.1007/978-3-031-01588-5.

[20] Veličković, P. *et al.* Deep graph infomax. *arXiv preprint arXiv:1809.10341* (2018).

[21] Pidsley, R. *et al.* Critical evaluation of the illumina methylationepic beadchip microarray for whole-genome dna methylation profiling. *Genome Biology* 17, 208 (2016).

[22] Solomon, O. *et al.* Comparison of dna methylation measured by illumina 450k and epic beadchips in blood of newborns and 14-year-old children. *Epigenetics* 13, 655–664 (2018).

[23] Rozenblatt-Rosen, O. *et al.* The human tumor atlas network: Charting tumor transitions across space and time at single-cell resolution. *Cell* 181, 236–249 (2020).