Cross-Modal Prediction of DNA Methylation from Histology Images

Manahil Raza[1,2], Muhammad Dawood[3], Talha Qaiser[1], and Nasir Rajpoot[1,4]

[1] Tissue Image Analytics Centre, University of Warwick, Coventry, UK
manahil.raza@warwick.ac.uk
[2] Department of Engineering Science, University of Oxford, Oxford, UK
[3] Kennedy Institute of Rheumatology, NDORMS, University of Oxford, Oxford, UK
[4] Histofy Ltd, Coventry, UK

Abstract. DNA methylation is a key epigenetic regulator of gene expression and a well-established hallmark of cancer. However, clinical methylation profiling remains costly and time-consuming, limiting routine use, particularly in low-resource settings. Inferring DNA methylation from routinely acquired histology images may serve as a scalable and cost-effective surrogate approach, providing early indications of methylation patterns prior to molecular profiling. We propose GraphCpG, a graph neural network-based framework for predicting continuous DNA methylation beta values across 10,000 CpG sites from H&E-stained whole slide images (WSIs) of gliomas, a tumour class in which methylation plays a central diagnostic and prognostic role. We evaluate the inferred methylation profiles in clinically relevant downstream tasks, including survival stratification and prediction of mitotic and immune subtypes, which capture tumour proliferative activity and tumour microenvironment states, respectively. Across TCGA glioma cohorts, predicted methylation features support significant survival separation and improved subtype classification performance relative to clinical and histology-only baselines, with AUROC gains of $\approx 17\%$ and 3%, respectively. In some settings, methylation-based models achieve performance comparable to multi-modal approaches, and in certain cases yield improvements of up to $\approx 2\%$ AUROC, highlighting the predictive value of histology-derived methylation signals. Pathway enrichment of well-predicted CpG-associated genes identifies immune and inflammatory processes consistent with glioma biology, providing supportive evidence for the biological relevance of the inferred signals. To the best of our knowledge, this is the first study to demonstrate that immune and mitotic subtypes can be predicted from DNA methylation profiles inferred directly from WSIs, highlighting the potential of scalable, image-based molecular characterization in glioma.

Keywords: DNA Methylation · Computational Pathology · Graph Neural Networks.

1 Introduction

DNA methylation is an epigenetic modification involving the addition of a methyl group ($\text{-}CH_3$) to cytosine residues, typically at cytosine phosphate guanine

(CpG) sites, and plays a key role in gene regulation and genomic stability [18]. Aberrant DNA methylation is widely recognized as a hallmark of cancer and commonly occurs as either global hypomethylation, which promotes genomic instability or promoter hypermethylation, a well-characterized epigenetic alteration that silences tumour suppressor genes [6]. These changes serve as clinically valuable biomarkers for cancer classification and prognosis [3]. However, despite its clinical value, DNA methylation profiling remains costly, resource-intensive and time-consuming, limiting its routine use and accessibility in low-resource settings [8, 10]. In contrast, Hematoxylin and Eosin (H&E)-stained histology slides are routinely acquired and widely available. Inferring DNA methylation from histological images may provide a scalable and cost-effective surrogate approach, leveraging existing diagnostic data to offer early insights into tumour methylation patterns [27].

Recent work has shown that deep learning models can predict DNA methylation from whole slide images (WSIs). Early studies framed this as a binary classification task [29, 22], while others targeted specific alterations such as MGMT methylation [14]. More recent approaches have focused on directly predicting continuous methylation values from histology images, including beta-value inference for CNS tumour classification and multi-omics prediction using multilayer perceptrons (MLPs) [10, 9]. These approaches primarily relied on MLP-based or conventional machine learning (ML) models and did not explicitly model spatial relationships between histological regions for continuous methylation prediction. Furthermore, they did not assess whether predicted methylation profiles capture clinically relevant tumour phenotypes, such as mitotic activity and tumour microenvironment (TME) states.

Gliomas, including Glioblastoma Multiforme (GBM) and Lower-Grade Glioma (LGG), are a heterogeneous group of brain tumours comprising over 60% of primary brain cancers and approximately 25% of malignant cases [29]. GBM is highly aggressive with poor survival outcomes, whereas LGGs have a relatively better prognosis but remain difficult to treat and may progress to higher-grade disease over time [19]. The brain exhibits among the highest levels of DNA methylation across human tissues [18] and methylation plays an important role in glioma classification. Landmark methylation-based classifiers have defined clinically meaningful CNS tumour subtypes [2], leading to the inclusion of methylation profiling into the 2021 WHO classification of CNS tumours [17]. Moreover, DNA methylation patterns strongly correlate with molecular subtypes and survival, establishing them as promising biomarkers for glioma [3].

In this study, we introduce GraphCpG, a graph neural network (GNN)-based framework for predicting DNA methylation beta values directly from WSIs across thousands of CpG sites. These beta values indicate the degree of methylation at each CpG site, ranging from 0 (unmethylated) to 1 (fully methylated). Using these predicted methylation values, we perform survival stratification and clinically relevant downstream tasks, including the prediction of mitotic and immune subtypes. Mitotic subtypes reflect tumour proliferative activity [12], while immune subtypes describe the state of the TME [24]. DNA methylation is

linked to both processes: it records the history of cell division and regulates genes controlling mitosis, while in the immune system, methylation patterns reflects immune cell activity [30]. Together, these tasks highlight how methylation is associated with both tumour-intrinsic (mitosis) and tumour-extrinsic (immune) factors that drive cancer progression. Finally, we assess the biological relevance of the inferred methylation signals through pathway enrichment analysis.

2 Methodology

2.1 Data

We evaluated the proposed method on two cohorts from The Cancer Genome Atlas (TCGA) [26]: LGG and GBM, comprising 502 patients collectively, where both H&E-stained WSIs and methylation data are available. DNA methylation profiles were obtained from the Illumina HumanMethylation450 BeadChip (450K) array using the UCSC Xena database [7]. Model performance was assessed using patient-level stratified three-fold cross-validation.

2.2 Preprocessing of DNA Methylation values

Following prior work, we focused on CpG sites showing high variation in beta values across samples [10]. CpG sites with missing values were first removed. We then retained sites with a standard deviation $>= 0.15$. To exclude consistently hypermethylated or hypomethylated sites, we required that no more than 80% of samples had beta values below or above 0.5 at each site. Finally, we selected the $10,000$ most variable CpG sites for downstream analysis. This filtering emphasized CpG sites with substantial methylation variability across cancer patients.

2.3 Graph Construction from Whole Slide Images

For each WSI, tissue regions were first segmented using a U-Net based model [20] and then divided into non-overlapping 1024×1024 pixel patches at 0.50 microns-per-pixel (MPP), denoted as $X_i = \{x_1, x_2 \dots x_N\}$. Patches containing less than 35% viable tissue were excluded from further analysis. Patch-level features were extracted using a frozen CTransPath encoder [25] resulting in a 768-dimensional embedding $\mathbf{h}_i \in \mathbb{R}^{768}$ for each patch x_i at its corresponding spatial location $\mathbf{s}_i$. Using these representations, each WSI was modelled as a graph $G = (V, E)$, where nodes $V = \{\mathbf{n}_1, \mathbf{n}_2 \dots \mathbf{n}_N\}$ correspond to patches $\mathbf{n}_i = (\mathbf{h}_i, \mathbf{s}_i)$. Edges E were defined using Delaunay triangulation [4], connecting neighbouring nodes within 4000 pixels, to represent the underlying histological architecture.

2.4 Graph Neural Network for DNA Methylation Prediction

We developed a custom multi-output GNN that employs Transformer Conv (TConv) layers [23] to predict both patch-level and WSI-level DNA methylation beta values [5] for $c \in \{1 \cdots C\}$ CpG sites, as shown in Figure 1. The network comprises three TConv layers that iteratively update node representations

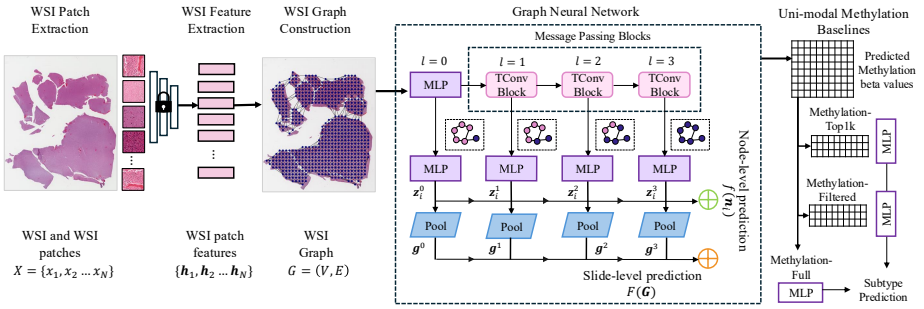

Fig. 1. GraphCpG framework and uni-modal methylation-based baselines: WSIs are tiled, encoded into a graph, and processed with a GNN to predict DNA methylation values, which are then used by MLP classifiers (Top1k, Filtered, Full) for downstream predictions.

through attention-weighted aggregation of spatial neighbours in the WSI graph. In each layer l, each node $\mathbf{n}_i$ computes a query vector $\mathbf{q}_i^{(l)} = \mathbf{W}_Q^{(l)}\mathbf{h}_i^{(l-1)}$, while each neighbour $\mathbf{n}_j$ in the neighbourhood $\mathcal{N}(i)$ computes a key $\mathbf{k}_j^{(l)} = \mathbf{W}_K^{(l)}\mathbf{h}_j^{(l-1)}$ and a value $\mathbf{v}_j^{(l)} = \mathbf{W}_V^{(l)}\mathbf{h}_j^{(l-1)}$. The updated representation of node i at layer l is given by:

$$\mathbf{h}_i^{(l)} = \mathbf{W}_{\text{root}}^{(l)}\mathbf{h}_i^{(l-1)} + \sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)}\,\mathbf{v}_j^{(l)} \tag{1}$$

where $\mathbf{W}_{\text{root}}^{(l)}\mathbf{h}_i^{(l-1)}$ is a root term used to preserve the node's own information, $\mathbf{W}_Q^{(l)}, \mathbf{W}_K^{(l)}, \mathbf{W}_V^{(l)}$ are learnable projection matrices and the coefficient $\alpha_{ij}^{(l)}$ represents the relative importance of neighbour j when updating node i at layer l. We use eight attention heads in each TConv layer and average the head outputs to obtain $\mathbf{h}_i^{(l)}$. At each layer, the node embedding is transformed by a small MLP $f_2(\cdot)$ and the final patch-level prediction for node $\mathbf{n}_i$ is obtained by aggregating the resulting layer-wise predictions:

$$\mathbf{z}_i^{(l)} = f_2^{(l)}(\mathbf{h}_i^{(l)}), \qquad f(\mathbf{n}_i) = \sum_{l=0}^{L} \mathbf{z}_i^{(l)} \tag{2}$$

where $\mathbf{z}_i^{(l)} \in \mathbb{R}^C$ denotes the node-level prediction for all CpG sites at layer l. After global mean pooling over nodes, $\mathbf{g}^{(l)} \in \mathbb{R}^C$ denotes the WSI-level prediction at layer l. The final WSI-level prediction $F(G) \in \mathbb{R}^C$ is obtained by summing the pooled outputs across all layers.

$$\mathbf{g}^{(l)} = \text{MeanPool}\left(\mathbf{z}^{(l)}\right), \qquad F(G) = \sum_{l=0}^{L} \mathbf{g}^{(l)} \tag{3}$$

We employ a pairwise ranking loss to train the GNN in an end-to-end manner:

$$\mathcal{L}_\theta = \sum_c \sum_{(i,j)\in P_c} \max\Big(0,\, 1 - (F_\theta^c(G_i) - F_\theta^c(G_j))\Big) \tag{4}$$

where $F_\theta^c(G)$ denotes the predicted methylation beta value for CpG site c and P_c is the set of patient pairs for which the ground-truth beta value at CpG site c for patient i is greater than that of patient j.

This loss formulation focuses on learning relative patient ordering instead of absolute predictions, making it well suited to DNA methylation data where absolute beta values are noisy and biologically meaningful information is often encoded in relative differences, particularly among the $10,000$ most variable CpG sites. WSIs were grouped at the patient level. Model selection used a 20% validation split, with the top 10 models ensembled for testing. Models were trained using the Adam optimizer (learning rate 0.001, weight decay 0.0001), with a batch size 16, trained up to 300 epochs, with early stopping (20-epoch patience).

2.5 Baselines for Downstream Prediction Tasks

To evaluate the utility of predicted methylation values, we compared GraphCpG with uni-modal, multi-modal, and clinical baselines for immune and mitotic subtype prediction. **Uni-Modal Baselines** included an image-only Attention-based Multiple Instance Learning (AMIL) model [11] trained on patch-level WSI features, and three methylation-only models trained on predicted beta values using MLPs (two layers, hidden size 512): (i) a full baseline using all $10,000$ CpG sites; (ii) a filtered baseline retaining CpG sites with prediction correlation > 0.5; and (iii) a ranked baseline using the top $1,000$ CpG sites based on prediction correlation, within the training folds. **Multi-Modal Baselines** combined histology and methylation features. The Dual-Branch model encoded mean-pooled histology features and top $1,000$ CpG features with 256-dimensional MLPs before concatenation. The Cross-Attention model projected both modalities into a shared 512-dimensional latent space and integrated them using cross-attention, enabling bidirectional interactions between image and methylation features. The resulting representations were concatenated and passed to an MLP for prediction. The **Clinical Baseline** used an MLP with patient age, sex, and histologic grade.

3 Results

3.1 Quantitative Analysis

Regression Results: To assess regression performance, we calculated the per-site Spearman correlation between predicted and observed beta values, reporting the average median correlation across folds. Inline with previous studies [9, 10], sites with a correlation > 0.4 were classified as "well-predicted". The model achieved an average median correlation of 0.42 with $7,252$ well-predicted sites, including $1,980$ sites with correlation > 0.5, as shown in Figure 2 (d).

Survival Analysis: We evaluated the prognostic value of histology-inferred methylation values using Cox regression to predict disease-specific survival (DSS), with survival times right-censored at ten years. For each training fold, we retained the top $1,000$ CpG sites based on regression correlation. From this subset, the ten most predictive sites were selected through univariate Cox analysis [9]. A multivariate Cox proportional hazards (CoxPh) model was subsequently trained using these sites and evaluated on the test sets. Performance was measured using the concordance index (C-index), reporting the average across folds. Patients were stratified into high-risk and low-risk groups based on the median predicted risk score. As shown in Figure 2 (a). the model achieved significant separation of survival outcomes, with a C-index of 0.70 ± 0.07 ($p < 0.001$). These findings suggest that the predicted methylation features contain prognostically relevant information that may be useful for stratifying patients into distinct risk groups.

Prediction of Mitotic Subtypes: Jahanifar *et al.* [12] stratified TCGA cases into Mitotic-Hot and Mitotic-Cold subgroups by quantifying mitotic activity in WSIs. Mitotic-Hot tumours demonstrate elevated mitotic activity and rapid proliferation compared to Mitotic-Cold tumours. We formulated a binary classification task (402 Cold and 95 Hot samples) to determine if the predicted methylation beta values could distinguish these subtypes. Table 1 summarizes the performance across clinical, uni-modal (UM) and multi-modal (MM) baselines. Methylation-based models consistently outperformed the clinical baseline and matched or exceeded histology-only performance, with multi-modal approaches yielding the strongest performance. The multi-modal Dual-Branch model achieved the highest overall performance (F1 = 0.724, AUROC = 0.815, BAcc = 0.748), indicating that combining predicted methylation signals with histological features may improve mitotic subtype classification compared with either modality alone.

Prediction of Immune Subtypes: We further evaluate the ability of the histology-inferred methylation values to distinguish between immune subtypes, specifically the Immunologically Quiet (C5: 337 samples) and Lymphocyte Depleted (C4: 165 samples) subtypes within the GBMLGG cohort [24, 7]. GBM samples were predominantly C4, whereas LGG included both C4 and C5 cases with a higher proportion of C5. C4 is associated with an immunosuppressed TME and poor prognosis, while C5 represents an immunologically quiet state with relatively better outcomes [24]. Table 1 reports performance across clinical, uni-modal (UM) and multi-modal (MM) baselines. Methylation-based models demonstrated competitive performance relative to clinical and histology-only baselines, suggesting that inferred DNA methylation profiles capture information relevant to immune subtype classification. However, because LGG and GBM exhibit distinct molecular characteristics and immune profiles, part of the predictive signal may be influenced by differences between tumor types rather than immune subtype characteristics alone. Notably, multi-modal fusion did not provide significant gains over methylation-only models, suggesting limited comple-

mentary value from histology features in this setting. The Methylation-Top1k baseline achieved the best overall performance (F1 = 0.721, AUROC = 0.767, BAcc = 0.710), demonstrating that selecting the most predictive CpG sites yields a more informative signal than relying on the entire feature space.

Table 1. Performance of clinical, uni-modal (UM) and multi-modal (MM) baselines for mitotic and immune subtype classification in TCGA-GBMLGG (mean ± standard deviation across folds), evaluated using F1-score, AUROC and balanced accuracy (BAcc).

	Model	F1	AUROC	BAcc
	Mitotic Subtype Classification			
	Clinical	0.562 ± 0.149	0.693 ± 0.078	0.607 ± 0.076
UM	Histology-AMIL	0.671 ± 0.043	0.790 ± 0.031	0.701 ± 0.045
UM	Methylation-Full	0.667 ± 0.015	0.781 ± 0.021	0.716 ± 0.028
UM	Methylation-Filtered	0.640 ± 0.056	0.779 ± 0.019	0.703 ± 0.028
UM	Methylation-Top1k	0.673 ± 0.035	0.781 ± 0.022	0.710 ± 0.028
MM	Dual-Branch	$\mathbf{0.724} \pm 0.074$	$\mathbf{0.815} \pm 0.031$	$\mathbf{0.748} \pm 0.044$
MM	Cross-Attention	0.717 ± 0.015	0.801 ± 0.020	0.746 ± 0.021
	Immune Subtype Classification			
	Clinical	0.599 ± 0.055	0.633 ± 0.178	0.610 ± 0.036
UM	Histology-AMIL	0.662 ± 0.059	0.723 ± 0.049	0.663 ± 0.051
UM	Methylation-Full	0.702 ± 0.054	0.748 ± 0.061	0.694 ± 0.051
UM	Methylation-Filtered	0.706 ± 0.036	0.765 ± 0.068	0.699 ± 0.034
UM	Methylation-Top1k	$\mathbf{0.721} \pm 0.047$	$\mathbf{0.767} \pm 0.065$	$\mathbf{0.710} \pm 0.043$
MM	Dual-Branch	0.709 ± 0.054	0.747 ± 0.075	0.698 ± 0.047
MM	Cross-Attention	0.700 ± 0.080	0.753 ± 0.069	0.687 ± 0.074

3.2 Qualitative Analysis

To investigate pathway-level associations, we mapped the 10,000 CpG sites to genes using the Illumina HM450 mask and manifest. For CpGs linked to multiple transcripts of the same gene, the closest transcription start site (TSS) was selected. We restricted the analysis to promoter-associated sites within ±1500 base pairs (bp) of a TSS [15]. Gene-level scores were defined as the maximum Spearman correlation across CpG sites associated with each gene, capturing the strongest gene-methylation association. These scores were used to rank genes for over-representation analysis (ORA) and gene set enrichment analysis (GSEA). Pathway enrichment was performed using the MSigDB hallmark 2020 library [16] with Enrichr [13]. ORA of the top 1,000 genes (adjusted $p < 0.05$) showed significant enrichment in Interferon Alpha Response, Inflammatory Response, and

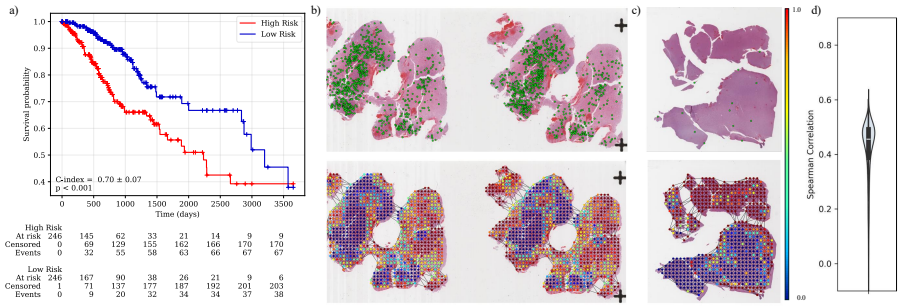

Fig. 2. (a) Kaplan–Meier curves stratifying patients into high-risk (red) and low-risk (blue) groups by the median predicted risk score; (b) Representative GBM (mitosis-hot, immune subtype C4) with detected mitotic figures (green dots) and corresponding predicted DNA methylation beta values (0–1); (c) Representative LGG (mitosis-cold, immune subtype C5) with detected mitotic figures and corresponding predicted beta values; (d) Spearman correlation between predicted and true methylation across the GBMLGG cohort.

Complement pathways. GSEA confirmed these results and additionally highlighted TNF-α/NF-κB signaling, Coagulation and Apoptosis (Normalized Enrichment Score (NES) > 0, False Discovery Rate (FDR) < 0.25). These pathways are important in glioma biology: NF-κB signaling drives proliferation, inflammation, immune evasion and treatment resistance [21], inflammation promotes tumour growth and an immunosuppressive TME [1], the complement pathway supports TME interactions and immune suppression [31] and Interferon alpha signalling reflects innate immune responses with potential antitumour effects [28]. Together, these results provide supportive evidence that the top-ranked CpG sites are associated with biologically relevant immune and inflammatory pathways implicated in glioma. The observed enrichment is consistent with the improved immune subtype classification performance of the Methylation-Top1k model relative to the clinical and uni-modal baselines (Table 1).

For visualization, we selected representative TCGA WSIs from cases with high prediction performance, including one GBM mitosis-hot, immune subtype C4 case and one LGG mitosis-cold, immune subtype C5 case. For each slide, a CpG site exhibiting substantial spatial heterogeneity was chosen. Figures 2 (b,c). show the WSIs with detected mitotic figures (green dots) and corresponding spatial maps of predicted DNA methylation beta values (0–1).

4 Conclusions

In this study, we introduced GraphCpG, a GNN-based framework for predicting DNA methylation beta values across $10,000$ CpG sites directly from H&E-stained WSIs in TCGA-GBMLGG. The inferred methylation profiles were informative for multiple clinically relevant downstream tasks, including survival anal-

ysis, mitotic subtype prediction, and immune subtype classification. Across tasks, models incorporating predicted methylation profiles from GraphCpG demonstrated competitive performance relative to clinical and histology-only baselines. For mitotic subtype prediction, the methylation-derived representations provided complementary predictive information, while for immune subtype classification the methylation-only model achieved the highest performance. Enrichment analyses of top-ranked CpG sites highlighted immune and inflammatory pathways, supporting the biological relevance of the predictions. Overall, our results suggest that routine histopathology images capture features reflective of the underlying epigenetic landscape, enabling the inference of DNA methylation patterns from standard clinical slides. These findings highlight the potential of inferred methylation profiles for molecular characterization and patient stratification in glioma, although further validation on independent, multi-institutional and EPIC-based cohorts is needed to establish generalizability.

Acknowledgments. The first author acknowledges the support of the University of Warwick, UK, where the research presented in this work was conducted during her PhD studies.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alorfi, N.M., Ashour, A.M., Alharbi, A.S., Alshehri, F.S.: Targeting inflammation in glioblastoma: An updated review from pathophysiology to novel therapeutic approaches. Medicine **103**(21), e38245 (2024)
2. Capper, D., Jones, D.T., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al.: Dna methylation-based classification of central nervous system tumours. Nature **555**(7697), 469–474 (2018)
3. Chen, X., Zhao, C., Zhao, Z., Wang, H., Fang, Z.: Specific glioma prognostic subtype distinctions based on dna methylation patterns. Frontiers in genetics **10**, 786 (2019)
4. Chew, L.P.: Constrained delaunay triangulations. In: Proceedings of the third annual symposium on Computational geometry. pp. 215–222 (1987)
5. Dawood, M., Vu, Q.D., Young, L.S., Branson, K., Jones, L., Rajpoot, N., Minhas, F.u.A.A.: Cancer drug sensitivity prediction from routine histology images. NPJ Precision Oncology **8**(1), 5 (2024)
6. Ehrlich, M.: Dna methylation in cancer: too much, but also too little. Oncogene **21**(35), 5400–5413 (2002)
7. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al.: Visualizing and interpreting cancer genomics data via the xena platform. Nature biotechnology **38**(6), 675–678 (2020)
8. Hauner, A., Onwuchekwa, C., Ariën, K.K.: Sample-to-result molecular diagnostic platforms and their suitability for infectious disease testing in low-and middle-income countries. Expert review of molecular diagnostics **24**(5), 423–438 (2024)

9. Hoang, D.T., Shulman, E.D., Dhruba, S.R., Nair, N.U., Barman, R.K., Lalchungnunga, H., Singh, O., Nasrallah, M.P., Stone, E.A., Aldape, K., et al.: Path2omics: Enhanced transcriptomic and methylation prediction accuracy from tumor histopathology. bioRxiv pp. 2025–02 (2025)
10. Hoang, D.T., Shulman, E.D., Turakulov, R., Abdullaev, Z., Singh, O., Campagnolo, E.M., Lalchungnunga, H., Stone, E.A., Nasrallah, M.P., Ruppin, E., et al.: Prediction of dna methylation-based tumor types from histopathology in central nervous system tumors with deep learning. Nature Medicine pp. 1–10 (2024)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
12. Jahanifar, M., Dawood, M., Zamanitajeddin, N., Shephard, A., Chohan, B.S., Bertram, C.A., Wahab, N., Eastwood, M., Aubreville, M., Raza, S.E.A., et al.: Pan-cancer profiling of mitotic topology & mitotic errors: Insights into prognosis, genomic alterations, and immune landscape. medRxiv pp. 2025–06 (2025)
13. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research **44**(W1), W90–W97 (2016)
14. Le, M.K., Kawai, M., Masui, K., Komori, T., Kawamata, T., Muragaki, Y., Inoue, T., Tahara, I., Kasai, K., Kondo, T.: Glioma image-level and slide-level gene predictor (glisp) for molecular diagnosis and predicting genetic events of adult diffuse glioma. Bioengineering **12**(1), 12 (2024)
15. Li, J.S., Riggins, K., Yang, L., Chen, C., Castro, P., Alfarkh, W., Zarrin-Khameh, N., Scheurer, M.E., Creighton, C.J., Musher, B., et al.: Dna methylation profiling at base-pair resolution reveals unique epigenetic features of early-onset colorectal cancer in underrepresented populations. Clinical Epigenetics **17**(1), 11 (2025)
16. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P.: The molecular signatures database hallmark gene set collection. Cell systems **1**(6), 417–425 (2015)
17. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H., Pfister, S.M., Reifenberger, G., et al.: The 2021 who classification of tumors of the central nervous system: a summary. Neuro-oncology **23**(8), 1231–1251 (2021)
18. Moore, L.D., Le, T., Fan, G.: Dna methylation and its basic function. Neuropsychopharmacology **38**(1), 23–38 (2013)
19. Ostrom, Q.T., Bauchet, L., Davis, F.G., Deltour, I., Fisher, J.L., Langer, C.E., Pekmezci, M., Schwartzbaum, J.A., Turner, M.C., Walsh, K.M., et al.: The epidemiology of glioma in adults: a "state of the science" review. Neuro-oncology **16**(7), 896–913 (2014)
20. Pocock, J., Graham, S., Vu, Q.D., Jahanifar, M., Deshpande, S., Hadjigeorghiou, G., Shephard, A., Bashir, R.M.S., Bilal, M., Lu, W., et al.: Tiatoolbox as an end-to-end library for advanced tissue image analytics. Communications medicine **2**(1), 120 (2022)
21. Ramar, V., Guo, S., Wang, G., Liu, M.: The pivotal role of nf-κb in glioblastoma: Mechanisms of activation and therapeutic implications. International Journal of Molecular Sciences **26**(16), 7883 (2025)
22. Raza, M., Dawood, M., Qaiser, T., Rajpoot, N.M.: A novel approach to linking histology images with dna methylation. In: European Conference on Computer Vision. pp. 335–349. Springer (2024)

23. Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y.: Masked label prediction: Unified message passing model for semi-supervised classification. arXiv preprint arXiv:2009.03509 (2020)
24. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al.: The immune landscape of cancer. Immunity **48**(4), 812–830 (2018)
25. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
26. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature genetics **45**(10), 1113–1120 (2013)
27. Yu, X., Zhao, H., Wang, R., Chen, Y., Ouyang, X., Li, W., Sun, Y., Peng, A.: Cancer epigenetics: from laboratory studies and clinical trials to precision medicine. Cell Death Discovery **10**(1), 28 (2024)
28. Zhang, X., Song, S., Zhou, D., Cui, H., Xie, J.: Identification of hub six interferon alpha response genes and immune cell infiltration characteristics in low-grade glioma. Scientific Reports **15**(1), 25210 (2025)
29. Zheng, H., Momeni, A., Cedoz, P.L., Vogel, H., Gevaert, O.: Whole slide images reflect dna methylation patterns of human tumors. NPJ genomic medicine **5**(1), 11 (2020)
30. Zhong, F., Lin, Y., Zhao, L., Yang, C., Ye, Y., Shen, Z.: Reshaping the tumour immune microenvironment in solid tumours via tumour cell and immune cell dna methylation: from mechanisms to therapeutics. British journal of cancer **129**(1), 24–37 (2023)
31. Zhu, H., Yu, X., Zhang, S., Shu, K.: Targeting the complement pathway in malignant glioma microenvironments. Frontiers in Cell and Developmental Biology **9**, 657472 (2021)