Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Evaluation of Pathology Foundation Models and Transcriptomic Integration for Glioblastoma Survival Prediction

Federica Rignanese

frignanese@fbk.eu

Fondazione Bruno Kessler

Gianmarco Sabbatini

aizoOn

Lisa Novello

Fondazione Bruno Kessler

Stefano Bovo

Fondazione Bruno Kessler

Flavio Ragni

Fondazione Bruno Kessler

Annarita Barone

Fondazione Bruno Kessler

Lorenzo Manganaro

aizoOn

Shahryar Noei

Fondazione Bruno Kessler

Matteo Pozzi

Orobìx

Marco Chierici

Fondazione Bruno Kessler

Paolo Falco

aizoOn

Giuseppe Jurman

Fondazione Bruno Kessler

Research Article

Keywords: Artificial Intelligence, Glioblastoma, Survival Analysis, Multimodal, Digital Pathology, Foundation Models, Transcriptomics

Posted Date: September 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10652791/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparative Evaluation of Pathology Foundation Models and Transcriptomic Integration for Glioblastoma Survival Prediction

Federica Rignanese[1,4*], Gianmarco Sabbatini[2], Lisa Novello[1], Stefano Bovo[1], Flavio Ragni[1], Annarita Barone[1,4], Lorenzo Manganaro[2], Shahryar Noei[1], Matteo Pozzi[1], Marco Chierici[1], Paolo Falco[2], Giuseppe Jurman[1,3]

[1*]Fondazione Bruno Kessler, Trento, Italy.
[2]aizoOn, Torino, TO, Italy.
[3]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, MI, Italy.
[4]Department of Mathematics, Università di Pavia, Pavia, PV, Italy.

*Corresponding author(s). E-mail(s): frignanese@fbk.eu;
Contributing authors: gianmarco.sabbatini@aizoongroup.com; lnovello@fbk.eu; sbovo@fbk.eu; fragni@fbk.eu; abarone@fbk.eu; lorenzo.manganaro@aizoongroup.com; snoei@fbk.eu; mat.poz96@gmail.com; chierici@fbk.eu; paolo.falco@aizoongroup.com; giuseppe.jurman@fbk.eu;

Abstract

Background Glioblastoma Multiforme (GBM) is a highly aggressive glioma difficult to stratify prognostically. Whole-slide histopathology images (WSIs) and RNA-seq profiles capture complementary phenotypic and molecular characteristics, yet their individual and combined value for GBM-specific survival prediction remains insufficiently characterized. Despite the promise of recent pathology foundation models to provide richer and more transferable image representations, their added prognostic value in GBM, either alone or in combination with transcriptomic data, has not yet been established. To address these gaps, this study evaluated deep survival models based on WSIs, RNA-seq data, and their integration for overall survival prediction in the TCGA-GBM cohort. WSI representations were derived by means of a general-domain ImageNet-pretrained ResNet50 as a baseline, and five recent histopathology-specific foundation-model

configurations. To disentangle modality effects from differences in sample size, models were evaluated in both full and size-matched cohorts.
Results Under size-matched conditions, multimodal integration consistently improved performance over WSI-only modelling across all six WSI encoders, despite the use of state-of-the-art pathology foundation models. Multimodal models achieved median C-indexes of 0.614–0.698, with a significant overall effect of modality. The CONCH-v1.5-based multimodal model achieved the highest median C-index of 0.698. RNA-seq provided the strongest prognostic signal, with held-out median C-indexes of 0.629 and 0.632 under full-budget and size-matched training, respectively. WSI-only models achieved median C-indexes of 0.543–0.587 under full-budget training and 0.512–0.549 under size-matched training. In this limited-size cohort and frozen-weight setting, WSI encoder choice did not significantly affect full-budget unimodal performance, and pathology foundation models did not systematically outperform the ResNet50 baseline. Overall, results suggest a positive but unconfirmed trend towards additional prognostic value in GBM from integrating histopathological and transcriptomic features.
Conclusions In this limited-size, frozen-encoder setting, integrating WSI and RNA-seq improved performance over WSI-only modelling across all encoders. The CONCH-v1.5-based multimodal model achieved the highest overall performance, while multimodal integration showed a positive but unconfirmed trend over RNA-seq alone. Pathology foundation models did not provide a consistent prognostic advantage over the ImageNet-pretrained ResNet50 baseline. Larger independent cohorts are needed to confirm the incremental value of multimodal integration.

Keywords: Artificial Intelligence, Glioblastoma, Survival Analysis, Multimodal, Digital Pathology, Foundation Models, Transcriptomics

1 Background

Glioblastoma (GBM) is the most aggressive and common adult diffuse glioma, a malignant primary tumor of the brain and central nervous system [1–3]. Despite advancements in treatments and in the biological understanding of the disease, its prognosis remains poor with a median survival of less than 15 months [1–6]. This reflects the infiltrative growth pattern, treatment resistance, and marked biological heterogeneity of GBM. Accurate survival prediction is therefore essential to support patient stratification, guide treatment planning, and inform follow-up strategies. However, due to its complex pathogenesis involving genetic, epigenetic, transcriptomic, histopathological, and microenvironmental alterations [4], prediction of survival remains challenging [7]. This complexity highlights the need to assess which data modalities best capture clinically meaningful prognostic information in GBM.

Diffuse gliomas have historically been classified and graded primarily on the basis of phenotypic characteristics observed on histopathological specimens, such as nuclear atypia, mitotic activity, necrosis, and microvascular proliferation [4, 8]. Recent WHO classifications have integrated molecular biomarkers into diagnostic criteria, recognizing genomic and molecular characterization as an essential modality that complements

traditional histopathology for the diagnosis, classification, and prognostic assessment of gliomas. This shift reflects a broader recognition that GBM prognosis cannot be fully understood from morphology alone, but requires integrating phenotypic and molecular information.

At the molecular level, GBM has been extensively characterized through bulk gene-expression profiling, single-cell analyses, and spatially informed studies, which identified clinically relevant transcriptional subtypes associated with distinct survival outcomes [8–10]. These findings validate the inter- and intra-tumor heterogeneity characterizing glioblastoma [2, 4, 11]. Recent transcriptomic analyses have further highlighted tumor microenvironment remodelling and transcriptional plasticity as key features of GBM heterogeneity, progression and treatment resistance [4, 6]. These findings support RNA-seq as a prognostically informative modality, capable of identifying differentially regulated genes and molecular programmes associated with survival [12, 13]. Several studies have therefore used gene-expression data for survival prediction using Cox-based models, penalized regression, pathway-guided neural networks, and deep survival architectures [14–16]. Nevertheless, RNA-seq is not always routinely available in clinical practice. This motivates interest in using routinely acquired and more easily accessible histopathological slides, which remain a fundamental source of information for characterizing GBM and cancer more broadly.

In pathology, Artificial Intelligence (AI) has improved the precision and throughput of histopathological image analysis, mitigating inter-observer variability and alleviating pathologist workloads through automated tumor detection and grading. Advancements in computer vision have enabled the analysis of gigapixel, multi-scale whole-slide images (WSIs) and the extraction of high-dimensional features from tissue morphology [17, 18]. WSIs represent high-resolution digital representations of the tumor and its surrounding microenvironment and encode morphological patterns reflecting underlying biological processes such as tumor aggressiveness, invasion, treatment resistance, and disease progression. Consequently, WSIs can provide information that goes beyond grading, supporting the stratification of patients according to survival risk. Studies across multiple cancer types have shown that WSI-based models can predict survival, supporting the hypothesis that tissue architecture encodes clinically meaningful prognostic signals [19–22]. In brain tumors, several studies have combined convolutional neural network (CNN)-derived histology representations with genomic or clinical variables, showing that tissue morphology contains prognostic information and can contribute to patient risk stratification [20–26].

More recently, the integration of transformer architectures and self-supervised learning has established the era of foundation models in cancer research [17, 27]. Pre-trained on massive, possibly multi-modal datasets, these models learn robust and generalizable visual representations that can be adapted to specific downstream tasks, often with minimal labelled data. Digital pathology foundation models, trained on extensive collections of histological images, have emerged as a promising alternative to conventional general-purpose feature extractors based on CNNs [28–31]. Being trained directly on pathology data, these models are expected to encode tissue morphology more effectively than encoders trained on natural images, potentially improving performance when labelled biomedical datasets are limited. However,

whether pathology-specific foundation models provide more prognostically informative representations than conventional ImageNet-pretrained CNNs for GBM survival prediction remains unclear.

While improved WSI encoders may enhance morphology-based prediction, histopathology captures only one layer of GBM biology. This has motivated increasing interest in multimodal models that integrate tissue morphology with molecular profiles. Because WSIs and RNA-seq capture different and potentially complementary views of the tumor, their integration may improve prognostic modelling by combining non-redundant biological information. Large-scale resources such as The Cancer Genome Atlas Glioblastoma cohort [32, 33] have created new opportunities to model GBM prognosis, providing matched histopathological and transcriptomic data and enabling systematic evaluation of predictive models.

Prior studies in glioma have suggested that integrating WSI-derived morphology with RNA-seq profiles can improve prognostic modelling [34–36]. However, much of this literature has relied on mixed TCGA-GBMLGG cohorts that include both lower-grade gliomas (LGG) and glioblastomas (GBM). These glioma subtypes have markedly different survival distributions, strongly influenced by grade, IDH status, and histological subtype. High predictive performance in this setting may therefore partly reflect the ability to differentiate lower-grade glioma from glioblastoma, rather than the ability to stratify survival within GBM itself. Thus, the prognostic value of histopathological, transcriptomic, and multimodal representations remains insufficiently characterized in GBM-specific cohorts.

GBM-specific multimodal studies integrating histopathology and transcriptomics remain limited. Moreover, its multimodal models incorporate demographic variables, whereas the present study focuses specifically on tumor-derived histopathological and transcriptomic information.

This study develops and evaluates deep survival models for overall survival prediction in patients with glioblastoma using WSIs and RNA-seq gene-expression profiles from the TCGA-GBM cohort. It addresses three research questions. First, it evaluates whether in the current experimental setup, WSI representations derived by means of pathology foundation models provide stronger prognostic information than embeddings derived from a conventional ImageNet-pretrained ResNet-50 baseline. Second, it assesses whether integrating RNA-seq features with WSI-based representations improves prognostic performance compared with WSI-only modelling under matched training conditions. Third, it compares histopathological, transcriptomic, and multimodal representations to determine which modelling strategy provides the strongest prognostic signal for GBM survival prediction, both overall and when controlling for differences in training-set size. To separate the effect of data modality from cohort size, models are evaluated under two experimental settings: a full-budget configuration setting, in which for each modality all available training patients are considered, and a controlled size-matched modality setting, in which WSI-only, RNA-seq-only, and multimodal models are trained on the same multimodal matched subset. Overall, this study provides a GBM-specific exploratory, controlled comparison of the prognostic value of transcriptomic, histopathological, and their integration under different

data-availability conditions, informing the development of future survival prediction pipelines for glioblastoma in larger and more clinically diverse cohorts.

2 Methods

This study was a retrospective observational secondary analysis of publicly available TCGA-GBM data, conducted for prognostic model development and validation.

2.1 Dataset and cohort selection

Table 1 Cohort characteristics by data modality.

	WSI	RNA-seq	WSI + RNA
Patients, n	356	137	80
Events, n (%)	301 (84.55%)	104 (75.9%)	64 (80.0%)
Censored, n (%)	55 (15.45%)	33 (24.1%)	16 (20.0%)
Age, median [IQR]	59 [50, 68]	61.5 [52, 70.25]	60 [50.75, 69]
Sex, n (%)			
Male	220 (61.8%)	87 (63.5%)	55 (68.8%)
Female	136 (38.2%)	50 (36.5%)	25 (31.3%)
OS time, median [IQR]	360 [185.5, 625.25]	331 [147.5, 532.75]	320 [173, 499.5]

Note: modality columns are not mutually exclusive.

This study used data from the public, multi-center Glioblastoma (GBM) cohort [32] of the Cancer Genome Atlas (TCGA). Data included histopathological hematoxylin and eosin (H&E)-stained diagnostic Whole-Slide Images (WSI), bulk RNA-seq gene expression profiles, and clinical survival information. The clinical endpoint was overall survival (OS), defined as time in days from diagnosis to death or last follow-up. Event status was encoded as 1 for deceased patients and 0 for right-censored patients, including patients alive or lost to follow-up. Diagnostic WSI images, normalised bulk RNA-seq gene expression profiles and survival information were retrieved from the NCI Genomics Data Common (GDC) [33].

Starting from 573 patients with available follow-up information, patients were excluded if they had: (i) missing or invalid OS and event-status information, (ii) missing or invalid clinical metadata, (iii) no available WSI or RNA-seq data or (iv) a histological classification other than primary untreated/de novo GBM. After these exclusions, 439 patients with valid clinical and survival data and at least one modality among WSI and RNA-seq were retained. No explicit outlier-detection procedure was applied beyond the exclusion of invalid survival records and non-eligible clinical or histological cases.

To ensure independent evaluation, a centre-based held-out test set was constructed by reserving all patients from Tissue Source Sites (TSSs) 27 and 41 having both modalities for final testing. These held-out centers included 26 patients and 42 WSIs. Final evaluation was restricted to the multimodal subset, comprising 17 patients with

both WSI and RNA-seq data and 27 WSIs, to allow direct comparison across RNA-seq, WSI, and multimodal models. The remaining 413 patients constituted the model-development cohort. Within this cohort, 137 patients had RNA-seq data and were used for RNA-seq model development, 356 patients had WSI data and were used for WSI model development, and 80 patients had both modalities and were used for multimodal model development (Table 1).

For patients with multiple diagnostic WSIs, all slides belonging to the same patient were assigned to the same data split to prevent patient-level leakage. RNA-seq profiles were treated as patient-level measurements, whereas WSI representations were available at slide level. In multimodal experiments, the same patient-level RNA-seq representation was paired with each available WSI from the corresponding patient.

WSIs were processed using the Trident Python framework [37]. Tissue regions were segmented and each slide was tiled into non-overlapping patches of 512 × 512 pixels at 20× magnification.

RNA-seq gene expression profiles were normalized using variance-stabilizing transformation (VST) and log2-transformed, resulting in a gene-level feature matrix comprising 60,660 features. All available annotated genes were retained for model development, and no additional low-expression filtering or feature selection was applied.

2.2 Model development

Feature extraction from WSI was performed using either pre-trained convolutional neural networks in a transfer learning paradigm or multiple pathology foundation models (Fig. 2). Slide-level embeddings were obtained with TITAN [30] and Prov-GigaPath [28], while patch-level embeddings were extracted using an ImageNet pre-trained ResNet-50 [38] that served as baseline, in addition to the patch-level encoder of Prov-GigaPath, UNI2-h [31], and CONCH v1.5 [29]. All feature extractors were kept frozen to exploit transferable representations and reduce overfitting, given the limited sample size. For patch-based approaches, slide-level representations were learned using an Attention-Based Multiple Instance Learning (ABMIL) [39] architecture, a well-suited approach in weakly supervised settings where only slide-level labels are available. Embedding dimensionalities for each WSI encoder are summarized in Table 2. The slide-level embeddings z^{WSI}, obtained either from ABMIL or from slide-level encoders, were projected into a fixed-dimensional latent space through a learned linear projection layer f_{proj} and provided as input to a survival prediction head. For patients with multiple diagnostic WSIs, each slide was treated as an independent sample and the performance metric was computed at the slide level.

A DeepHazard model [40], a neural-network-based extension of the Cox proportional hazards model [14, 40], was used both for RNA-seq training and for unimodal OS modeling (Fig. 2). The model was first trained on gene expression data to predict survival; the activations of the final hidden layer, of dimensionality d_{RNA} and treated as a hyperparameter, were then extracted as a fixed patient-level embedding z_{RNA} and used downstream for both unimodal survival prediction and multimodal fusion. The best performing dimensionality was selected by maximising the average C-index

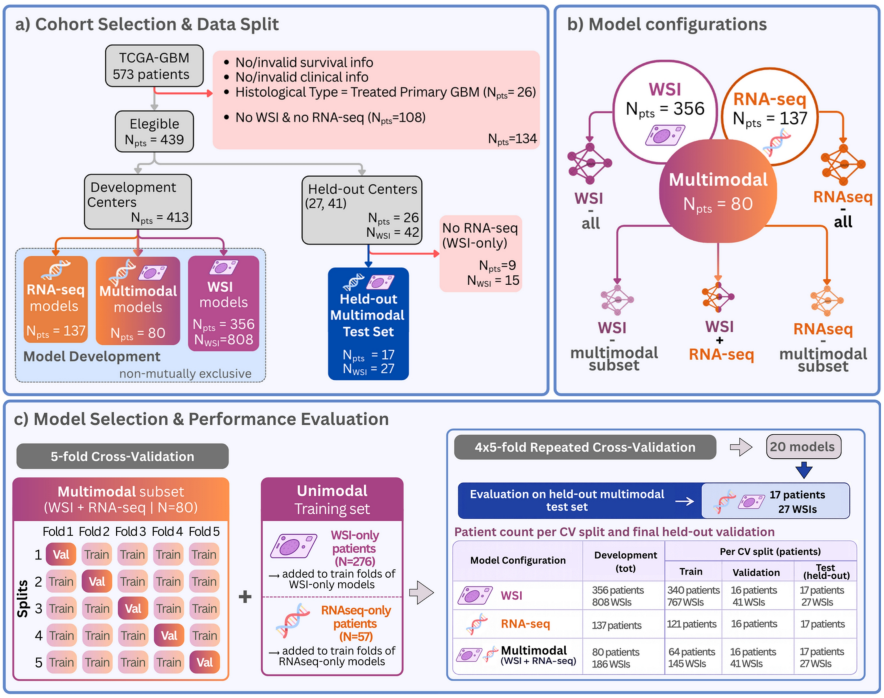


Fig. 1 Study design and evaluation workflow. a) TCGA-GBM cohort selection and centre-based data splitting strategy into development and held-out sets. The development cohort was used to define WSI-only, RNA-seq-only, and multimodal model-development subsets. b) Model configurations evaluated, including WSI-only, RNA-seq-only, WSI+RNA-seq multimodal, both in full- and matched-training-budget settings c) Model selection and performance evaluation strategy, with 5-fold cross-validation for hyperparameter selection and 4×5 repeated cross-validation followed by evaluation on the held-out multimodal test set.

over a 5-fold cross-validation on the training set, from the candidate set of {10, 100, 200, 500, 1000}.

Multimodal modelling employed an intermediate fusion strategy: the WSI-derived latent representation $z^{\mathrm{W}'_{\mathrm{proj}}}$ and the RNA-based embedding z^{G} were concatenated and used as input in a DeepHazard model for multimodal survival prediction. This design preserves modality-specific structure while enabling joint risk estimation.

2.3 Training details

The DeepHazard model [40] was adopted as the survival prediction architecture for both unimodal and multimodal experimental settings. All models were trained by minimizing the negative Cox partial log-likelihood with L1 and L2 regularization:

$$\mathcal{L}(\theta) = -\sum_{i:\delta_i=1}\left(h_i - \log\sum_{j:t_j\geq t_i}\exp(h_j)\right) + \lambda_1\|\theta\|_1 + \lambda_2\|\theta\|_2^2. \quad (1)$$

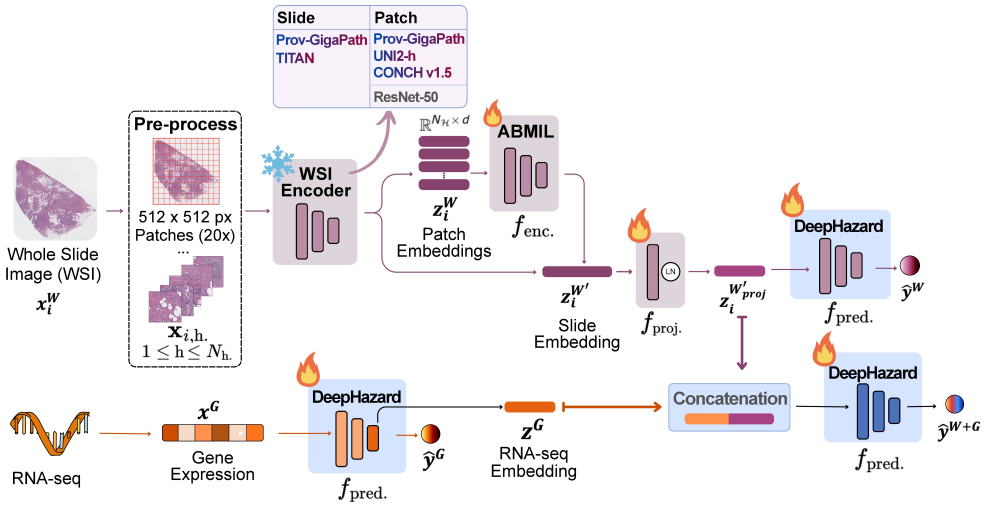


Fig. 2 Survival prediction framework. WSIs are tiled into 512 × 512 px patches at 20× magnification and encoded by a frozen WSI encoder. Patch embeddings $z_{i,h}^{W} \in \mathbb{R}^{N \times d}$ are aggregated via ABMIL, f_{enc}, into a slide-level representation $z_i^{\text{W}'}$, which is projected and layer-normalized by f_{proj} to obtain a fixed-dimensional WSI embedding $z_i^{W'_{proj}}$. RNA-seq gene-expression profiles x^G are encoded by a trained DeepHazard network, f_{pred}, to produce a genomic embedding z^G extracted from the last hidden layer. For multimodal prediction, $z_i^{W'_{proj}}$ and z^G are concatenated and passed to a joint DeepHazard predictor.

where h_i denotes the predicted log-hazard for patient i, t_i is the observed survival time, and $\delta_i \in \{0, 1\}$ is the right-censoring indicator.

Each model outputs a continuous risk score, where higher values indicate higher predicted mortality risk.

2.4 Experimental design and model configurations

In order to assess the contribution of each data modality to survival prediction performance, five different model configurations were evaluated (Fig. 1a). Two WSI models were trained, one on the full WSI development cohort (WSI-all, N=356) and one on the smaller multimodal subset (WSI-multimodal, N=80). Similarly, two RNA-seq models were developed, one trained on all patients having gene expression data (RNAseq-all; N=137) and one on the multimodal subset. The multimodal model integrating both WSI and RNA-seq data was trained exclusively on the 80 multimodal patients. Training unimodal models on the same multimodal patient subset allowed us to compare fairly the model performances and to assess the added value of multimodal fusion, without the confounding effect of training set size. Hyperparameters were tuned only once per modality, on the full-budget configuration (WSI-all and RNAseq-all).

2.5 Validation strategy and performance evaluation

A 5-fold Cross-Validation (CV) scheme (Fig. 1c) was employed for model development and hyperparameter tuning. The splits of the CV were defined exclusively on the subset of multimodal patients; patients with single-modality data were included exclusively in the training folds across all splits, effectively increasing the training set size without compromising evaluation integrity, that remained common across experiments. Fold assignments were kept aligned across all experimental conditions to ensure that all models were evaluated on identical patient splits, enabling direct statistical comparison. All splits were performed at the patient level to avoid data leakage. The optimal hyperparameter configuration was selected maximizing the average validation C-index [41] across the five folds.

Final performance was estimated using a 4×5 repeated cross-validation scheme applied to the development cohort, providing 20 separately trained models. Each model was evaluated on the same multimodal held-out test set (N=17) using the Concordance Index (C-index) [41]. For WSI-based models, each slide was treated as an independent sample, and the held-out test C-index was therefore computed at the slide level using 27 test slides. For RNA-seq models, predictions were generated at the patient level, and the held-out test C-index was computed across the 17 test patients. In multimodal models, RNA-seq profiles were patient-level representations and were propagated for each available slide from the same patient in order to enable slide-level fusion with WSI embeddings.

Model performance was reported using the median and interquartile range of held-out C-index values across repeated cross-validation folds.

2.6 Statistical assessment

Differences in prognostic performance across model configurations were assessed using the 20 held-out test C-index estimates obtained from the 4×5 repeated cross-validation. Repetition and fold assignments were kept aligned across configurations to enable paired comparisons. Global differences between conditions were assessed using repeated-measures ANOVA (rmANOVA). When a significant global effect was detected ($\alpha < 0.05$), post-hoc pairwise comparisons were performed using paired t-tests. To account for the dependency structure induced by cross-validation, the t-tests were adjusted using a custom implementation of the Nadeau and Bengio variance correction [42]. Because multiple slides from the same patient were not considered statistically independent, patient counts were used instead of slide counts when defining the effective training and test sample sizes for the Nadeau–Bengio correction. When comparing configurations trained on cohorts of different sizes, the smaller patient count was used as a conservative effective sample size. Post-hoc p-values were adjusted for multiple comparisons using the Holm–Bonferroni procedure [43]. Normality was assessed by means of the Shapiro–Wilk test [44]. Departure from sphericity was evaluated using the Greenhouse–Geisser epsilon, and Greenhouse–Geisser correction was applied where appropriate.

Three groups of statistical comparisons were performed. First, the effect of WSI encoder choice was assessed separately in the full-budget WSI-only and in the multimodal settings. For each analysis, a one-way repeated-measures ANOVA was fitted with WSI encoder as the within-subject factor. These analyses tested whether models based on the representations from the six WSI encoders differed in prognostic performance when trained using only WSI data, and whether encoder choice stayed influential after integration with transcriptomic features.

Second, the effect of RNA-seq integration was evaluated comparing WSI-only and WSI+RNA models under matched training conditions. A two-way repeated-measures ANOVA was fitted with modality setting and WSI encoder as within-subject factors to test the main effect of modality and encoder, and the modality-by-encoder interaction. For each WSI encoder, paired comparisons were performed between each size-matched WSI-only model and its corresponding WSI+RNA-seq model.

Third, the WSI-only and the WSI+RNA-seq model configurations with the highest median held-out C-index were selected for post-hoc exploratory comparison with RNA-seq models, performed by means of two complementary analyses. The configuration-level analysis compared the full-budget RNA-seq-only model with the best full-budget WSI-only and the best WSI+RNA-seq models, to identify the strongest configuration when each model was trained using all patients available for its required modality. The modality-level analysis compared the RNA-seq-only model with the best WSI-only and WSI+RNA-seq models trained on the same multimodal patient subset, thereby allowing modality effects to be assessed while controlling for training-set size. For each analysis, a one-way rmANOVA was performed using modality as within-subject factor.

2.7 Implementation details

Models were implemented in Python 3.10 using PyTorch 2.6.0 (CUDA 12.4 build). For WSI-only and multimodal models, all architectures were optimized using AdamW with a fixed weight decay of 1×10^{-4}. The ABMIL model was implemented through the MIL-Lab library (https://github.com/mahmoodlab/MIL-Lab) [45] and consisted of 1 fully-connected layer with an attention dimensionality of 384, a gating mechanism and dropout within the attention branch ($p = 0.25$). All linear layers were initialized with a Xavier normal initialization. Models were trained for a maximum of 50 epochs, with early stopping based on validation C-index (patience = 5 epochs) to avoid overfitting. Due to the multiple instance learning requirements, a batch size of 1 was used with gradient accumulation to simulate a batch size of 64. The Concordance index was computed using TorchSurv 0.1.4 [46].

Hyperparameters, including learning rate and network architecture, were selected via cross-validation using a grid search via Weights & Biases (v0.19.6); a total of 96 configurations (Supplementary Table S3) were evaluated for the WSI-only and multimodal settings. RNA-seq models were optimized using a restricted hyperparameter search. The evaluated hyperparameters included the network architecture, activation function and learning rate. The network consisted of two hidden layers, with the size of the last hidden layer evaluated among {10, 100, 200, 500, 1000} neurons and the preceding hidden layer fixed at twice that size. Two activation functions, ReLU and SELU, were evaluated, with SELU selected for the final models. Dropout was fixed

at $p = 0.0$, and the learning rate was evaluated between 1×10^{-3} and 1×10^{-4}, with 1×10^{-3} selected for the final models. Early stopping was not applied during training.

All experiments used a fixed random seed (42) to ensure reproducibility.

3 Results

3.1 RNA-seq embedding dimensionality

The prognostic performance of the RNA-based models showed an improving trend for increasing dimensionality of the RNA embedding, which plateaued at dimensionality 1000. The 500-dimensional embedding achieved the highest validation C-index while showing a slightly smaller train–validation gap than the 1000-dimensional configuration. Therefore, the 500-dimensional RNA-seq embedding was selected for all subsequent RNA-only and multimodal experiments.

3.2 Survival prediction performances

Figure 3 and Table 2 report the distributions of C-index values computed on the held-out test set across the 20 test folds of the repeated cross-validation, stratified by modality, WSI encoder architecture, and training budget.

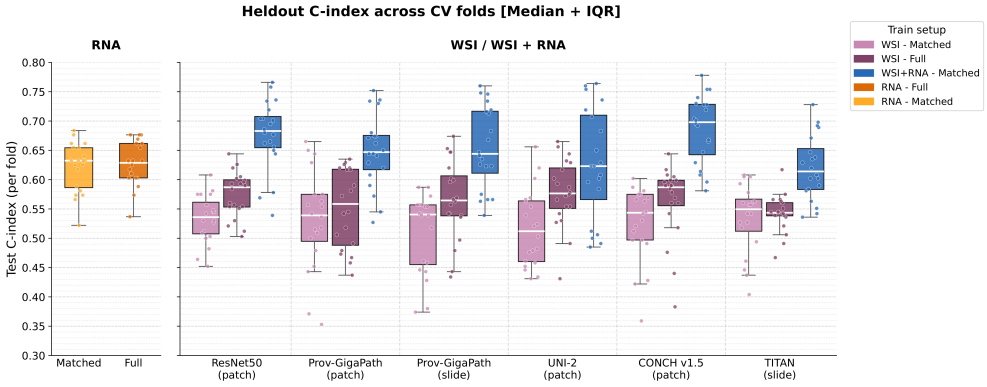


Fig. 3 Distribution of C-index computed on the held-out test set using models from a 4×5 CV, for a total of 20 estimates for each setting. Results are stratified by modality, training budget and encoder.

RNA-only models achieved similar held-out test performance under both training-budget conditions. When trained on the multimodal subset, the RNA-only model achieved a median C-index of 0.632 [0.586–0.654]. When trained using all available RNA-seq patients, performance remained comparable, with a median C-index of 0.629 [0.603–0.662].

WSI-only models showed more modest held-out performance, depending on both the encoder and the training-set size. Under the full-budget setting, median C-indexes ranged from 0.543 [0.538–0.561] for TITAN to 0.587 [0.553–0.600] for ResNet-50 and 0.587 [0.555–0.599] for CONCH v1.5.

Table 2 Held-out C-index performance stratified by modality, training-set setting, WSI encoder, and WSI embedding type. Performance is reported as median ± interquartile range (IQR) across repeated held-out estimates.

Modality	Train set	WSI encoder	Embed.	Median ± IQR
RNA-seq	Matched	–	–	**0.632 ± 0.068**
	Full	–	–	0.629 ± 0.059
WSI	Matched	TITAN	Slide	**0.549 ± 0.055**
		CONCH v1.5	Patch	0.543 ± 0.078
		GigaPath	Slide	0.540 ± 0.102
		GigaPath	Patch	0.539 ± 0.080
		ResNet50	Patch	0.536 ± 0.054
		UNI2-h	Patch	0.512 ± 0.103
	Full	CONCH v1.5	Patch	**0.587 ± 0.043**
		ResNet50	Patch	0.587 ± 0.046
		UNI2-h	Patch	0.576 ± 0.069
		GigaPath	Slide	0.564 ± 0.068
		GigaPath	Patch	0.558 ± 0.130
		TITAN	Slide	0.543 ± 0.022
WSI+RNA-seq	Matched	CONCH v1.5	Patch	**0.698 ± 0.086**
		ResNet50	Patch	0.683 ± 0.053
		GigaPath	Patch	0.647 ± 0.058
		GigaPath	Slide	0.644 ± 0.106
		UNI2-h	Patch	0.623 ± 0.144
		TITAN	Slide	0.614 ± 0.070

When training was restricted to the matched multimodal subset, WSI-only performance was generally lower and more variable with median C-indexes ranging from 0.512 [0.460–0.564] for UNI2-h to 0.549 [0.512–0.567] for TITAN (Table 2).

Increasing the training set size produced encoder-dependent gains (Table 3). The largest absolute median C-index gain was observed for UNI2-h (+0.064), followed by ResNet50 (+0.051) and CONCH v1.5 (+0.043).

Multimodal models achieved higher median C-indexes than their corresponding size-matched WSI-only models for all six encoders. The highest observed performance was obtained using CONCH v1.5, with a median C-index of 0.698 [0.642–0.728], followed by ResNet-50 at 0.683 [0.654–0.707]. Four out of six multimodal configurations exceeded the median C-index of both RNA-only models. With respect to the size-matched RNA model, multimodal models showed gains ranging from +0.018 for UNI2-h to +0.066 for CONCH v1.5.

3.3 Statistical assessment

Statistical testing was organized according to three groups of comparisons: (i) the effect of WSI encoder choice, (ii) the effect of multimodal integration over WSI-only modelling, and (iii) the comparison between the best-performing modality configurations.

First, the effect of the WSI encoder was evaluated separately in full-budget WSI-only models and in WSI+RNA multimodal models. In the full-budget WSI-only setting, the repeated-measures ANOVA did not reveal a significant encoder effect

Table 3 Median held-out C-index for WSI-only and multimodal WSI+RNA-seq models across WSI encoders. WSI-only full models were trained on the full WSI training set, whereas WSI-only matched and WSI/RNA matched models were trained on the same multimodal patient subset. Δ_f denotes the performance difference between WSI-only matched and WSI-only full; Δ_r denotes the performance difference between WSI-only matched and WSI+RNA-seq matched.

WSI enc.	**Embed.**	**Median C-index**			**Δ median C-index**	
		WSI full	**WSI match.**	**WSI+RNA match.**	$\Delta_{\text{WSI}_{\text{full}-\text{match.}}}$	$\Delta_{\text{multi}-\text{WSI}_{\text{match.}}}$
CONCH v1.5	Patch	0.587	0.543	0.698	0.043	0.154
ResNet50	Patch	0.587	0.536	0.683	0.051	0.147
UNI2-h	Patch	0.576	0.512	0.623	0.064	0.111
GigaPath	Patch	0.558	0.539	0.647	0.020	0.108
GigaPath	Slide	0.564	0.540	0.644	0.024	0.103
TITAN	Slide	0.543	0.549	0.614	−0.006	0.064

($F(5, 95) = 1.065$, $p > 0.05$). In contrast, the multimodal analysis detected a significant overall effect of the WSI encoder ($F(5, 95) = 5.413$, $p < 0.01$, $\epsilon_{\text{GG}} = 0.692$). However, none of the pairwise encoder-level comparisons remained significant after Holm adjustment.

Second, size-matched WSI-only and WSI+RNA models were compared using a two-way repeated-measures ANOVA to assess the added value of RNA-seq integration across WSI encoders. The analysis showed a strong main effect of modality, with higher held-out performance for WSI+RNA than for WSI-only models ($F(1, 19) = 123.736$, $p_{\text{GG}} < 0.01$). Neither the main effect of the WSI encoder ($F(5, 95) = 2.33$, $p_{\text{GG}} = 0.0768$) nor the modality-by-encoder interaction ($F(1, 19) = 123.736$, $p_{\text{GG}} = 0.0576$) reached significance after Greenhouse–Geisser correction. Mean differences favoured multimodal modelling for all six encoders and ranged from 0.086 to 0.160 C-index points. After Nadeau–Bengio and Holm correction, the improvement remained significant for ResNet50 ($\Delta_{\text{mean C-index}} = 0.143$, $p_{\text{corr}} < 0.01$), slide-level Prov-GigaPath ($\Delta_{\text{mean C-index}} = 0.145$, $p_{\text{corr}} < 0.01$), and CONCH v1.5 ($\Delta_{\text{mean C-index}} = 0.160$, $p_{\text{corr}} < 0.01$). The comparisons involving UNI2-h, patch-level Prov-GigaPath, and TITAN did not remain significant after adjustment.

Third, the best-performing configurations were compared with RNA-seq models. In the full-budget configuration analysis, the overall effect of model configuration was significant ($F(2, 38) = 29.992$, $p < 0.01$, $\epsilon_{\text{GG}} = 0.7474$). Post-hoc pairwise comparisons numerically favoured RNA-seq over WSI-only modelling and multimodal over both WSI-only and RNA-seq-only modelling. The multimodal versus RNA-seq contrast showed an unadjusted trend ($p_{\text{unc}} = 0.0569$), but none of the comparisons remained significant after Holm correction. In the size-matched modality analysis, a significant overall modality effect was observed ($F(2, 38) = 54.225$, $p < 0.01$, $\epsilon_{\text{GG}} = 0.922$). The CONCH v1.5 WSI+RNA model significantly outperformed the best size-matched WSI-only model, based on TITAN representations ($\Delta_{\text{mean}} = 0.152$, $p_{\text{corr}} < 0.01$).

Neither RNA-seq-only versus WSI-only nor WSI+RNA versus RNA-seq-only reached significance.

4 Discussion

This study evaluated the relative prognostic value of histopathology, transcriptomics, and their integrated representations for OS prediction in glioblastoma using the TCGA-GBM cohort. The analysis addressed whether pathology foundation models provided more informative WSI representations than an ImageNet-pretrained ResNet50 baseline, whether integrating RNA-seq improved prediction relative to WSI alone, and which modality or combination provided the strongest prognostic signal under both full and matched training budgets.

Three main patterns emerged. First, RNA-seq provided the numerically strongest unimodal prognostic signal. Second, integrating RNA-seq with WSI representations consistently improved performance relative to WSI-only modelling. Third, the best-performing multimodal configurations also exceeded RNA-seq-only modelling numerically, suggesting that histopathology may contribute complementary prognostic information when combined with a strong transcriptomic representation.

The clearest result was the improvement associated with adding RNA-seq to WSI representations under matched training conditions. The repeated-measures analysis identified a large main effect of modality, and the mean difference favoured WSI+RNA-seq for all six encoders. The direction of the effect was highly consistent, even though the strength of the statistical evidence varied by encoder. Numerically, the magnitude of the improvement varied across encoders, with CONCH v1.5 and ResNet50 showing the largest median multimodal gains. UNI2-h and TITAN showed smaller final multimodal performance, but both still improved relative to their corresponding size-matched WSI-only models. One possible explanation is that some encoders retain morphological information that is more complementary to transcriptomics, whereas others encode signals that overlap more strongly with molecular features. These findings indicate that the addition of transcriptomic information was beneficial across heterogeneous visual representations. The absence of a clear modality-by-encoder interaction is consistent with a broadly shared benefit of RNA-seq integration across encoders, although the limited sample size does not exclude encoder-specific differences in the magnitude of this benefit, as suggested by improvements in ResNet-, ProvGigaPath-, and CONCH-based models.

RNA-seq emerged as a stable unimodal predictor, with models achieving median held-out C-index values of approximately 0.63 under both training-budget conditions. Furthermore, the RNA-only model trained on the multimodal subset numerically exceeded the best full-budget WSI-only model, despite being trained on substantially fewer patients. The stability of RNA-seq performance across training budgets likely reflects the fact that the difference in the number of patients between successive training settings is negligible relative to the extremely high dimensionality of the feature space. As a result, the model might be exposed to essentially the same underlying molecular signal across all training budgets, suggesting that this dominant prognostic signal can be learned even from relatively small patient cohorts. Further improvements

would likely require larger cohorts to determine whether additional, weaker molecular prognostic patterns can be learned beyond this dominant signal. Although direct pairwise comparisons did not establish a definitive difference between the best RNA-seq and WSI-only models, the pattern was consistent across training budgets with transcriptomic profiles achieving a stronger and more stable unimodal prognostic signal than histopathological representations in the present cohort.

The comparison between multimodal and RNA-seq-only models was more nuanced but favourable to multimodal integration. Four of the six multimodal configurations achieved higher median C-indices than both RNA-seq-only models. CONCH v1.5 showed the largest difference relative to size-matched RNA-seq, improving from 0.632 to 0.698, while ResNet50 improved to 0.683. Both Prov-GigaPath configurations also exceeded RNA-seq alone, although by smaller margins. Conversely, the UNI2-h and TITAN multimodal models remained slightly below RNA-seq-only performance. The overall pattern therefore suggests that histopathological information may provide additional value when the WSI representation is sufficiently informative. The performance of CONCH v1.5 and ResNet50 is particularly relevant in this respect because both models produced numerical improvements over RNA-seq alone despite relying on substantially different pretraining strategies. The direct multimodal-versus-RNA comparisons did not survive correction for multiple testing; however, the observed median differences, consistent ranking of the strongest configurations, and limited size of the held-out cohort support interpreting these results as exploratory evidence that morphology may add complementary information beyond transcriptomics alone.

The observed results suggest that WSI and RNA-seq capture partially complementary but unequally informative prognostic signals, with RNA-seq providing the larger unimodal contribution and WSI adding a smaller complementary signal. This pattern is biologically plausible as bulk RNA-seq directly encodes molecular programmes associated with proliferation, immune activity, cellular state, treatment resistance, and microenvironmental remodelling, while histopathological images capture phenotypic manifestations which may only indirectly reflect the underlying molecular drivers. This interpretation is consistent with previous multimodal glioma studies suggesting that much of the predictive signal is modality-specific and additive, with relatively limited contribution from complex cross-modal interactions and better-performing architectures showing equivalent or lower levels of cross-modal interaction [49]. Under this interpretation, the use of a simple intermediate fusion approach by means of concatenation is a reasonable and conservative choice in such a limited-size cohort, where more expressive fusion architectures may increase overfitting risk without necessarily improving generalization.

Contrary to the expectation that pathology-specific foundation models would consistently outperform a generic natural-image encoder, the ImageNet-pretrained ResNet50 remained highly competitive, jointly ranking first in the full-budget WSI-only analysis together with CONCH v1.5 (median C-index = 0.587); in the multimodal setting, ResNet50 ranked second to CONCH v1.5, exceeding all the remaining pathology foundation models. No significant encoder effect was observed in the full-budget WSI-only analysis. Thus, within the present cohort, frozen-encoder configuration, and downstream survival architecture, pathology-specific pretraining did not produce a

consistent advantage over the ImageNet-pretrained baseline. A tentative explanation is that the histological features most associated with survival in GBM may be relatively coarse, therefore accessible to simpler encoders. Pathology foundation models are generally trained to capture broad morphological or image–text information across diverse tissues and tasks, whereas survival prediction may depend on relatively subtle and cohort-specific patterns. This result aligns with increasing evidence that the benefits of large, domain-specific pretrained models are context dependent [27] and that, depending on the task, cohort size, and downstream architecture, simpler or more generic pretrained models may remain competitive with larger, domain-specialized alternatives [47]. Nonetheless, the comparison was restricted to frozen representations, a limited GBM cohort, specific aggregation strategies and a single survival-prediction architecture. Foundation models may therefore still provide benefits under fine-tuning, larger labelled cohorts, different endpoints, or tasks more closely aligned with their pretraining objectives.

Increasing the WSI training set from the matched subset to the full WSI cohort improved median performance for five of the six encoders. The largest gains were observed for UNI2-h, ResNet50, and CONCH v1.5, whereas TITAN showed essentially unchanged performance. This suggests that additional training data can improve WSI-based survival modelling, but that the magnitude of the benefit depends on the representation and downstream architecture. Despite being trained on substantially more patients, the full-budget WSI-only models remained below RNA-seq-only and the strongest multimodal configurations. Thus, increasing the WSI-only training set was insufficient to close the observed performance gap with RNA-seq-only and the strongest multimodal configurations.

The GBM-specific design is important when interpreting these results. Many previous multimodal glioma survival studies used mixed TCGA-GBMLGG cohorts, where prediction may be strongly influenced by differences between lower-grade glioma and glioblastoma, including grade, IDH status, molecular subtype, and histological category. High performance in such settings may partly reflect discrimination between clinically and biologically distinct disease groups rather than prognostic stratification within GBM. By focusing exclusively on primary GBM, the present study addressed a more homogeneous but substantially more challenging prediction problem, in which survival-related variation had to be identified within a uniformly aggressive disease population.

The results are broadly comparable with PAGE-Net [24], one of the most closely related GBM-specific multimodal survival models. PAGE-Net integrated histopathology, gene expression, and demographic information and reported a C-index of 0.702, compared with their C-index value of 0.509 for histopathology alone and 0.640 for a gene-expression and age models. Although direct comparison is limited by differences in cohort definition, preprocessing, covariates, data availability, and validation design, the best model in the present study achieved a similar performance range without incorporating demographic variables. More importantly, the current work extends prior GBM studies by systematically comparing multiple WSI encoders and by separating modality effects from differences in training-set size.

This study indicates that RNA-seq provides the strongest and most stable unimodal prognostic signal in TCGA-GBM and captures survival-related information not fully represented by histopathology alone. Accordingly, integrating RNA-seq with WSI features consistently improved prognostic performance relative to WSI-only modelling. Within the present limited-sample, frozen-feature setting, however, features derived from pathology foundation models did not show a consistent advantage over those extracted with the conventional ImageNet-pretrained ResNet50 baseline, highlighting the need for controlled, task-specific benchmarking. Overall, these findings provide exploratory evidence that, in this cohort and experimental setting, modality composition appeared to have a larger influence on prognostic performance than increasing the size of the WSI-only training cohort.

This study has several limitations that should guide interpretation. First, the analysis was conducted on a relatively small GBM-specific cohort, particularly in the multimodal setting. Although a centre-held-out test set was used, it included only 17 patients and 27 WSIs, and no fully independent external validation cohort was available. Second, statistical comparisons were based on held-out C-index estimates from repeated cross-validation. This design increases the number of performance estimates and enables paired comparisons, but the estimates are not fully independent because training partitions overlap across folds and repetitions. The use of Nadeau–Bengio correction and Holm adjustment mitigates part of this issue, but the statistical results should still be interpreted with appropriate caution in light of these methodological choices. Third, multimodal modelling was limited to intermediate fusion by concatenation of WSI-derived and RNA-seq-derived embeddings. This strategy is reasonable for a limited-size multimodal cohort and provides a conservative baseline, but it does not establish whether more expressive fusion mechanisms could improve performance. Fourth, the study did not include formal interpretability analyses.

5 Conclusions

This study provides a controlled, glioblastoma-specific comparison of the value of histopathology, transcriptomic, and their integration for overall-survival prediction in different modality availability settings.

Across the evaluated configurations, RNA-seq provided the strongest and most stable unimodal prognostic signal, whereas WSI-only models achieved a more modest and encoder-dependent performance. Increasing the WSI training set generally improved WSI-only prediction but did not close the observed gap with transcriptomic or multimodal models. Under size-matched training conditions, integrating RNA-seq with WSI representations consistently improved performance over WSI-only modelling across all encoders. The CONCH v1.5 multimodal configuration achieved the highest median held-out C-index, and four of the six multimodal models numerically exceeded both RNA-seq-only configurations. These patterns suggest that RNA-seq provides the dominant observed unimodal contribution, while appropriately encoded histopathology may add complementary information. Pathology foundation models did not consistently outperform the ImageNet-pretrained ResNet50 baseline, indicating

that domain-specific pretraining alone does not guarantee superior survival representations in this setting. These findings provide exploratory support for the integration of RNA-seq and WSI features for survival prediction in the present cohort. Within the present experimental setting, modality composition appeared to have a larger and more consistent influence on prognostic performance than WSI encoder selection, supporting further evaluation of WSI–RNA integration in larger and independently validated GBM cohorts. Overall, this work provides a controlled GBM-specific assessment of the relative contributions of histopathology, transcriptomics, training-set size, and WSI encoder choice, thereby offering practical evidence to guide the design of future survival-prediction pipelines.

Declarations

Ethics approval and consent to participate. This study used publicly available, de-identified data from TCGA, TCIA, GDC, and UCSC Xena. No new patient recruitment, intervention, or direct contact with participants was performed.

During the preparation of this manuscript, the authors used several LLMs (ChatGPT) for grammar and spelling check, or paraphrasing/rewording in compliance with the editorial guidelines of the journal. The authors critically reviewed and edited the content as needed and take full responsibility for the content of the published article.

Consent for publication. Not applicable. This manuscript does not contain data, images, or details relating to any identifiable individual person.

Availability of data and materials. The datasets analysed during the current study are publicly available from The Cancer Genome Atlas (TCGA-GBM) project via the NCI Genomic Data Commons Data Portal, https://portal.gdc.cancer.gov/. Whole-slide images and RNA-seq expression data were downloaded from this repository.

Competing interests. The authors declare that they have no competing interests.

Funding. This work was supported by the Italian Ministry of Research, under the complementary actions to the NRRP "D34Health - Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care" Grant (# PNC0000001).

Authors' contributions. FR pre-processed the data, implemented and trained the histopathology-based survival models, and drafted the manuscript. AB downloaded the whole-slide image data and performed feature extraction. GS implemented and trained the survival network and the RNA-seq deep survival models. LN, SB, FR, LM, contributed to the design of the study and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

All authors read and approved the final manuscript.

Acknowledgements. The results shown here are in whole or part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga).

References

[1] Grochans S, Cybulska AM, Simińska D, Korbecki J, Kojder K, Chlubek D, et al. Epidemiology of Glioblastoma Multiforme–Literature Review. Cancers. 2022;14:2412. https://doi.org/10.3390/cancers14102412

[2] Verdugo E, Puerto I, Medina MÁ. An update on the molecular biology of glioblastoma, with clinical implications and progress in its treatment. Cancer Commun. Wiley Online Library; 2022;42:1083–111.

[3] Schaff LR, Mellinghoff IK. Glioblastoma and Other Primary Brain Malignancies in Adults: A Review. JAMA. 2023;329:574. https://doi.org/10.1001/jama.2023.0023

[4] Lan Z, Li X, Zhang X. Glioblastoma: An Update in Pathology, Molecular Mechanisms and Biomarkers. Int J Mol Sci. 2024;25:3040. https://doi.org/10.3390/ijms25053040

[5] Omuro A. Glioblastoma and Other Malignant Gliomas: A Clinical Review. JAMA. 2013;310:1842. https://doi.org/10.1001/jama.2013.280319

[6] Hoogstrate Y, Draaisma K, Ghisai SA, van Hijfte L, Barin N, de Heer I, et al. Transcriptome analysis reveals tumor microenvironment changes in glioblastoma. Cancer Cell. Elsevier; 2023;41:678-692. e7.

[7] Awuah WA, Ben-Jaafar A, Roy S, Nkrumah-Boateng PA, Tan JK, Abdul-Rahman T, et al. Predicting survival in malignant glioma using artificial intelligence. Eur J Med Res. 2025;30:61. https://doi.org/10.1186/s40001-025-02339-3

[8] Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell. 2006;9:157–73. https://doi.org/10.1016/j.ccr.2006.02.019

[9] Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17:98–110. https://doi.org/10.1016/j.ccr.2009.12.020

[10] Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, et al. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell. 2019;178:835-849.e21. https://doi.org/10.1016/j.cell.2019.06.024

[11] The Intratumoral Heterogeneity Reflects the Intertumoral Subtypes of Glioblastoma Multiforme: A Regional Immunohistochemistry Analysis [Internet]. https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2020.00494/full

[12] Yang J, Wang L, Xu Z, Wu L, Liu B, Wang J, et al. Integrated Analysis to Evaluate the Prognostic Value of Signature mRNAs in Glioblastoma Multiforme. Front Genet. 2020;11:253. https://doi.org/10.3389/fgene.2020.00253

[13] Yang Y, Yan R, Zhang L, Meng X, Sun W. Primary glioblastoma transcriptome data analysis for screening survival-related genes. J Cell Biochem. 2020;121:1901–10. https://doi.org/10.1002/jcb.29425

[14] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18:24. https://doi.org/10.1186/s12874-018-0482-1

[15] Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M. Cox-PASNet: Pathway-based Sparse Deep Neural Network for Survival Analysis. 2018 IEEE Int Conf Bioinforma Biomed BIBM [Internet]. Madrid, Spain: IEEE; 2018 [cited 2026 June 16]. p. 381–6. https://doi.org/10.1109/BIBM.2018.8621345

[16] Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. Markowetz F, editor. PLOS Comput Biol. 2018;14:e1006076. https://doi.org/10.1371/journal.pcbi.1006076

[17] Waqas A, Bui MM, Glassy EF, El Naqa I, Borkowski P, Borkowski AA, et al. Revolutionizing Digital Pathology With the Power of Generative Artificial Intelligence and Foundation Models. Lab Invest. 2023;103:100255. https://doi.org/10.1016/j.labinv.2023.100255

[18] Marra A, Morganti S, Pareja F, Campanella G, Bibeau F, Fuchs T, et al. Artificial intelligence entering the pathology arena in oncology: current applications and future perspectives. Ann Oncol. 2025;36:712–25. https://doi.org/10.1016/j.annonc.2025.03.006

[19] Yao J, Zhu X, Jonnagaddala J, Hawkins N, Huang J. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med Image Anal. 2020;65:101789. https://doi.org/10.1016/j.media.2020.101789

[20] Zhu X, Yao J, Zhu F, Huang J. WSISA: Making Survival Prediction from Whole Slide Histopathological Images. 2017 IEEE Conf Comput Vis Pattern Recognit CVPR [Internet]. Honolulu, HI: IEEE; 2017 [cited 2026 June 16]. p. 6855–63. https://doi.org/10.1109/CVPR.2017.725

[21] Luo C, Yang J, Liu Z, Jing D. Predicting the recurrence and overall survival of patients with glioma based on histopathological images using deep learning. Front Neurol. 2023;14:1100933. https://doi.org/10.3389/fneur.2023.1100933

[22] Chunduru P, Phillips JJ, Molinaro AM. Prognostic risk stratification of gliomas using deep learning in digital pathology images. Neuro-Oncol Adv. 2022;4:vdac111. https://doi.org/10.1093/noajnl/vdac111

[23] Shao Z, Chen Y, Bian H, Zhang J, Liu G, Zhang Y. HVTSurv: Hierarchical Vision Transformer for Patient-Level Survival Prediction from Whole Slide Image [Internet]. arXiv; 2023 [cited 2026 June 16]. https://doi.org/10.48550/ARXIV.2306.17373

[24] Hao J, Kosaraju SC, Tsaku NZ, Song DH, Kang M. PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data. Biocomput 2020 [Internet]. Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC; 2019 [cited 2026 June 1]. p. 355–66. https://doi.org/10.1142/9789811215636_0032

[25] Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci [Internet]. 2018 [cited 2025 Nov 6];115. https://doi.org/10.1073/pnas.1717139115

[26] Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. IEEE Trans Med Imaging. 2022;41:757–70. https://doi.org/10.1109/TMI.2020.3021387

[27] Neidlinger P, El Nahhas OSM, Muti HS, Lenz T, Hoffmeister M, Brenner H, et al. Benchmarking foundation models as feature extractors for weakly supervised computational pathology. Nat Biomed Eng [Internet]. 2025 [cited 2025 Oct 24]; https://doi.org/10.1038/s41551-025-01516-3

[28] Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024;630:181–8. https://doi.org/10.1038/s41586-024-07441-w

[29] Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, et al. A visual-language foundation model for computational pathology. Nat Med. 2024;30:863–74. https://doi.org/10.1038/s41591-024-02856-4

[30] Ding T, Wagner SJ, Song AH, Chen RJ, Lu MY, Zhang A, et al. A multimodal whole-slide foundation model for pathology. Nat Med. 2025;31:3749–61. https://doi.org/10.1038/s41591-025-03982-3

[31] Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024;30:850–62. https://doi.org/10.1038/s41591-024-02857-3

[32] Scarpace L, Mikkelsen T, Cha S, Rao S, Tekchandani S, Gutman D, et al. The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM) [Internet]. The Cancer Imaging Archive; 2016 [cited 2026 June 11]. https://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9

[33] Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, et al. The NCI Genomic Data Commons. Nat Genet. 2021;53:257–62. https://doi.org/10.1038/s41588-021-00791-5

[34] Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci [Internet]. 2018 [cited 2026 June 16];115. https://doi.org/10.1073/pnas.1717139115

[35] Steyaert S, Qiu YL, Zheng Y, Mukherjee P, Vogel H, Gevaert O. Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. Commun Med. 2023;3:44. https://doi.org/10.1038/s43856-023-00276-y

[36] Zhang Z, Zhao Y, Duan J, Liu Y, Zheng H, Liang D, et al. Pathology-genomic fusion via biologically informed cross-modality graph learning for survival analysis [Internet]. arXiv; 2024 [cited 2026 June 16]. https://doi.org/10.48550/arXiv.2404.08023

[37] Zhang A, Jaume G, Vaidya A, Ding T, Mahmood F. Accelerating Data Processing and Benchmarking of AI Models for Pathology [Internet]. arXiv; 2025 [cited 2026 June 11]. https://doi.org/10.48550/ARXIV.2502.06750

[38] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. arXiv; 2015 [cited 2026 June 11]. https://doi.org/10.48550/ARXIV.1512.03385

[39] Ilse M, Tomczak JM, Welling M. Attention-based Deep Multiple Instance Learning [Internet]. arXiv; 2018 [cited 2025 Mar 3]. https://doi.org/10.48550/arXiv.1802.04712

[40] Sabbatini G, Manganaro L. DeepHazard: A Tensorflow/Keras and PyTorch-Compatible Deep Learning Package for Survival Analysis. 2025;

[41] Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. Wiley Online Library; 1996;15:361–87.

[42] Nadeau C, Bengio Y. Inference for the generalization error. Adv Neural Inf Process Syst. 1999;12.

[43] Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. JSTOR; 1979;65–70.

[44] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. Oxford University Press; 1965;52:591–611.

[45] Shao D, Chen RJ, Song AH, Runevic J, Lu MY, Ding T, et al. Do Multiple Instance Learning Models Transfer? ArXiv Prepr ArXiv250609022 [Internet]. 2025; https://doi.org/https://doi.org/10.48550/arXiv.2506.09022

[46] Monod M, Krusche P, Cao Q, Sahiner B, Petrick N, Ohlssen D, et al. TorchSurv: A Lightweight Package for Deep Survival Analysis [Internet]. arXiv; 2024 [cited 2026 June 11]. https://doi.org/10.48550/ARXIV.2404.10761

[47] Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding Transfer Learning for Medical Imaging [Internet]. arXiv; 2019 [cited 2026 July 22]. https://doi.org/10.48550/ARXIV.1902.07208

[48] Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics. Int J Med Inf. 2021;153:104510. https://doi.org/10.1016/j.ijmedinf.2021.104510

[49] Swift I, Ye J, O'Reilly R. Quantifying Cross-Modal Interactions in Multimodal Glioma Survival Prediction via InterSHAP: Evidence for Additive Signal Integration [Internet]. arXiv; 2026 [cited 2026 May 18]. https://doi.org/10.48550/ARXIV.2603.29977

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.zip