EBioMedicine 61 (2020) 103093

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Incremental prognostic value and underlying biological pathways of radiomics patterns in medulloblastoma

Jing Yan[a,1], Shenghai Zhang[b,1], Kay Ka-Wai Li[c,1], Weiwei Wang[d], Ke Li[e], Wenchao Duan[e], Binke Yuan[f], Li Wang[d], Lei Liu[b], Yunbo Zhan[e], Dongling Pei[e], Haibiao Zhao[e], Tao Sun[e], Chen Sun[e], Wenqing Wang[e], Zhen Liu[e], Xuanke Hong[e], Xiangxiang Wang[e], Yu Guo[e], Wencai Li[d], Jingliang Cheng[a], Xianzhi Liu[e], Ho-Keung Ng[c], Zhicheng Li[b,*], Zhenyu Zhang[e,*]

[a] Department of MRI, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China
[b] Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China
[d] Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China
[e] Department of Neurosurgery, The First Affiliated Hospital of Zhengzhou University, Jian she Dong Road 1, Zhengzhou, Henan 450052, China
[f] Institute for Brain Research and Rehabilitation, South China Normal University, China

ARTICLE INFO

Article History:
Received 14 July 2020
Revised 9 October 2020
Accepted 9 October 2020
Available online xxx

Keywords:
Medulloblastoma
Molecular
Radiomics
Prognosis
Pathway

ABSTRACT

Background: To develop a radiomics signature for predicting overall survival (OS)/progression-free survival (PFS) in patients with medulloblastoma (MB), and to investigate the incremental prognostic value and biological pathways of the radiomics patterns.

Methods: A radiomics signature was constructed based on magnetic resonance imaging (MRI) from a training cohort (n = 83), and evaluated on a testing cohort (n = 83). Key pathways associated with the signature were identified by RNA-seq (GSE151519). Prognostic value of pathway genes was assessed in a public GSE85218 cohort.

Findings: The radiomics-clinicomolecular signature predicted OS (C-index 0.762) and PFS (C-index 0.697) better than either the radiomics signature (C-index: OS: 0.649; PFS: 0.593) or the clinicomolecular signature (C-index: OS: 0.725; PFS: 0.691) alone, with a better calibration and classification accuracy (net reclassification improvement: OS: 0.298, P = 0.022; PFS: 0.252, P = 0.026). Nine pathways were significantly correlated with the radiomics signature. Average expression value of pathway genes achieved significant risk stratification in GSE85218 cohort (log-rank P = 0.016).

Interpretation: This study demonstrated radiomics signature, which associated with dysregulated pathways, was an independent parameter conferring incremental value over clinicomolecular factors in survival predictions for MB patients.

Funding: A full list of funding bodies that contributed to this study can be found in the Acknowledgements section.

1. Introduction

Medulloblastoma (MB) is the most common pediatric cancer of the central nervous system (CNS), with a 5-year overall survival (OS) rate of 54% [1,2]. Remarkable progress on MB has revealed this tumor to consist of at least four molecular subgroups: wingless (WNT), sonic hedgehog (SHH), Group 3 and Group 4 [3-4]. Notable findings have been discovered in the aspects of different tumor origins [5], various genetic profiles [6], differential phenotypes [3,4], diverse clinical traits [4], distinct prognosis [3,4] and potential therapeutic responses [2] among the four subgroups. Moreover, the distinction among the four subgroups has prompted an international consensus on MB

***Abbreviations*: MB**, Medulloblastoma; CNS, Central nervous system; SHH, Sonic hedgehog subgroup; WNT, Wingless subgroup; FFPE, Formalin-fixed paraffin-embedded; MRI, Magnetic resonance imaging; T1, Precontrast T1-weighted imaging; T1c, Contrast-enhanced T1-weighted imaging; T2, T2-weighted imaging; FLAIR, Fluid-attenuated inversion recovery; ADC, Apparent diffusion coefficient; DWI, diffusion-weighted imaging; VOI, volume of interest; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix; GLSZM, gray level size zone matrix; GLDM, gray level dependence matrix; NGTDM, neighborhood gray-tone difference matrix; LoG, Laplacian of Gaussian; ICC, intraclass correlation coefficient

* Corresponding authors.
E-mail addresses: zc.li@siat.ac.cn (Z. Li), fcczhangzy1@zzu.edu.cn (Z. Zhang).
[1] Jing Yan, Shenghai Zhang and Kay Ka-Wai Li contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2020.103093

Research in context

Evidence before this study

The 2016 WHO Classification of Central Nervous System (CNS) tumors has included the four molecular subgroups, namely sonic hedgehog (SHH), wingless (WNT), Grade 3 and Group 4 for classification of medulloblastoma (MB). However, a radiomics signature associated with the survivals of MB patients has not yet been developed, and whether radiomics signature is able to offer incremental prognostic value beyond molecular subgroups and clinical risk factors has not been investigated. We searched published literatures on PubMed and Web of Science with the following terms: "(radiomics OR radiomic) AND magnetic resonance AND (survival OR prognosis) AND medulloblastoma", without date restriction or limitation to English language publications. This search did not identify any previous publications investigating the prognostic values of radiomics signatures based on magnetic resonance imaging (MRI) on MB.

Added value of this study

In the current study, a radiomics signature was constructed based on magnetic resonance imaging (MRI) from a training cohort (n = 83), and evaluated on a testing cohort (n = 83). By incorporating radiomics signature and clinicomolecular factors, the resulted radiomics-clinicomolecular model predicted OS and PFS better than either the radiomics signature or the clinicomolecular model alone. Furthermore, a radiogenomics analysis using paired MRI and RNA-seq in a GSE151519 cohort revealed nine pathways significantly associated with the radiomics signature. Average expression value of pathway genes was demonstrated to have significant prognostic value in an external GSE85218 cohort.

Implications of all the available evidence

This study demonstrated radiomics signature, which associated with dysregulated pathways, was an independent parameter conferring incremental value over clinicomolecular factors in survival predictions for MB patients.

categorization [3], which is included in the World Health Organization (WHO) classification of CNS tumors [7].

Recent advances in quantitative image analysis enabled better characterization of the imaging phenotypes of the tumors, which may offer complementary information for prognostication of MB patients. A rapidly emerging field called radiomics converts digitally encrypted medical images into minable quantitative data [8], and has been utilized to increase the accuracy in survival prediction of cancers. For example, in patients with glioblastoma, several studies have shown radiomics models improved the prognostic performance over the established biomarker and clinical risk factors [9]. As for MB, although two studies attempted to use multi-parametric MRI-based nomograms or radiomics for predicting molecular subgroups of MB [10,11], to our best knowledge, a radiomics signature associated with the survivals of MB patients has not yet been developed, and whether radiomics signature is able to offer incremental prognostic value beyond molecular subgroups and clinical risk factors has not been investigated. Early evidence has revealed the association between radiomic patterns and underlying signaling pathways in several cancers [12,13,14]. However, the biological pathways associated with the radiomics features that stratify MB patients for prognosis remains elusive, which becomes one of the barriers preventing radiomics analysis into clinical translation.

This study leverages radiomics to unravel the potential of preoperative multi-parametric MRI, which includes precontrast T1-weighted imaging (T1), contrast-enhanced T1-weighted imaging (T1c), T2-weighted imaging (T2), T2-weighted fluid-attenuated inversion recovery (FLAIR) imaging, and apparent diffusion coefficient (ADC) maps, for prediction of both OS and progression-free survival (PFS) in MB patients. Furthermore, we explored the biological basis of the prognostic radiomics features by revealing key pathways associated with the radiomics features that confer prognostic significance in MB patients.

2. Methods

2.1. Study design

The overall design of our study was illustrated in Fig. 1, including two parts: radiomics profiling and radiogenomics analysis. Based on a radiomics cohort with MRI data, we first identified the most prognostic radiomics features and combined them into a radiomics signature for survival prediction, and then assessed the incremental prognostic value of the signature beyond existing clinicomolecular risk factors. Based on a radiogenomics cohort (GSE151519) with both MRI and RNA-seq, we identified key biological pathways underlying the radiomics features, where the prognostic value of the genes contained in the pathways was validated in public database on MB (GSE85218, including RNA-seq and survival data) published by Cavalli FMG et al. [15].

2.2. Patients

This study was approved by the Human Scientific Ethics Committee of the First Affiliated Hospital of Zhengzhou University (No. 2019-KY-176). Informed consents were obtained from all patients who provided fresh tumor specimens. For other patients, informed consents were waived by the Committee. During January 2009 to July 2019, 248 pathologically diagnosed primary MB patients were surgically treated in the Department of Neurosurgery, the First Affiliated Hospital of Zhengzhou University. Then, 166 of 248 cases were further selected by the following criteria: (1) availability of survival data, (2) availability of preoperative MR images, (3) availability of multi-parametric MRI, including axial T1, T1c, T2, FLAIR, and ADC maps, (4) availability of sufficient image quality without significant artifacts, determined by neuroradiologists and neurosurgeons. The selection procedure was depicted in **Supplementary Figure S1**. Patients were randomly divided to a training data set (n = 83) and a testing data set (n = 83). Clinical factors, including gender, age, preoperative Karnofsky performance status (KPS) scale, extent of resection (complete resection or incomplete resection) and adjuvant therapies (postoperative radiation therapy and chemotherapy) were collected from the medical record system. Extent of resection was evaluated based on neurosurgeons' reports and postoperative MRI. We define complete resection as no residual tumor (gross total resection), and incomplete resection as any residual tumor reported (near-total resection or sub-total resection). Follow-up data were acquired by medical records, telephone follow-up, out-patient clinic, or visiting. OS was calculated as the time interval between the date of surgery and the date of death or the end of follow-up. PFS was calculated as the time interval between the date of surgery and the date of recurrence, progression or the end of follow-up. Moreover, in the current cohort, fresh frozen tissues of 17 MB patients with corresponding preoperative multi-parametric MRI were acquired for RNA-seq. The RNA-seq data has been deposited into NCBI Gene Expression Omnibus under accession code GSE151519. A radiogenomics analysis were conducted based on the data of radiomics and transcriptomes of these patients.

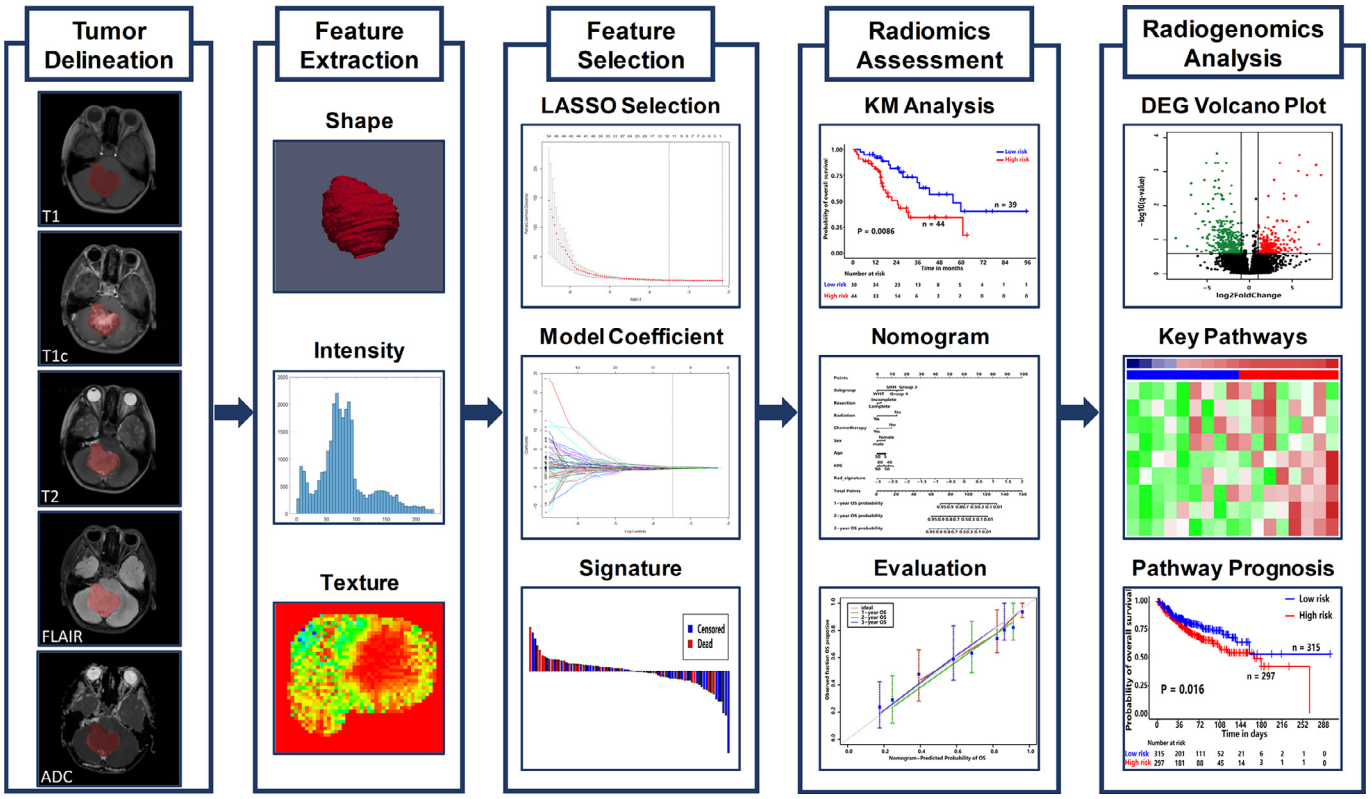

Fig. 1. The overview of the study design, including tumor delineation from multi-parametric MRI, radiomics feature extraction from delineated tumor regions, feature selection for building an optimal signature, statistical analysis for radiomics model assessment, and radiogenomics analysis for identifying the key pathways.

2.3. Molecular subgroups affiliation

Formalin-fixed paraffin embedded (FFPE) tissues were available in all the 166 MB cases and used for detection of molecular subgroups (WNT, SHH, Group 3, and Group 4) based on a nanoString assay according to the protocols previously described by Northcott PA et al. [16]. WNT, SHH, Group 3 and Group 4 subgroups were categorized by the CodeSet, which consists of 22 genes with 5–6 signature genes in each subgroup: WNT (DKK2, EMX2, GAD1, TNC, WIF1), SHH (ATOH1, EYA1, HHIP, PDLIM3, SFRP1), Group 3 (EGFL11, GABRA5, IMPG2, MAB21L2, NPR3, NRL), Group 4 (EOMES, KCNA1, KHDRBS2, OAS1, RBM24, UNC5D), and three housekeeping genes (ACTB, GAPDH, LDHA). The heatmap showing the expression of 22 MB signature genes was depicted in **Supplementary Figure S2**.

2.4. MRI acquisition

All MR images were acquired during routine clinical work-up on 3.0 T clinical scanners (Siemens Magnetom Skyra/Trio TIM; GE Discovery MR750; Philips Ingenia). The brain imaging protocol at our institution included the following sequences: (a) axial and sagittal T1; (b) axial T2; (c) axial FLAIR; (d) diffusion-weighted imaging (DWI) and the corresponding ADC maps generated with the software incorporated into the MRI unit; (e) axial, sagittal, and coronal T1c obtained after intravenous administration of a 0.1 mmol/kg dose of gadolinium-based contrast agent (Gadolinium-diethylenetriamine pentaacetic acid [Gd-DTPA], Bayer Healthcare, Leverkusen, Germany, or Gadoteric Acid Meglumine Salt Injection, Hengrui Healthcare, Jiangsu, China). Details on MRI acquisition parameters are available in **Supplementary material**.

2.5. RNA sequencing

Fresh frozen tissues of 17 MB patients with MRI were used for RNA-seq. All RNA-seq data were normalized using the fragments per kilobase transcriptome per million reads (FPKM) method [17]. Details of the RNA samples preparation and sequencing can be found in **Supplementary material**.

2.6. Tumor delineation

First, preprocessing was performed for image standardization. The preprocessing began with N4ITK-based bias field distortion correction. Then, all voxels were isotropically resampled into $1 \times 1 \times 1$ mm^3 using trilinear interpolation. Rigid registration was performed with a 3D Slicer software on the multi-parameter MR images for each patient using axial resampled T1c as a template with mutual information similarity metric, generating the registered images namely rT1, rT1c, rT2, rFLAIR, and rADC. Histogram matching was performed to normalize the intensity distribution. Three-dimensional volume of interest (VOI) of tumor contours were manually delineated slice-by-slice using the ITK-SNAP software by a neuroradiologist (J.Y. with 11 years' experience) in the axial plane primarily from rFLAIR images, where rT2 and rT1c were used to cross-check the extension of the tumor and fine tune the tumor contour. The VOIs were delineated as the tumor region, including the contrast-enhancing area, the non-enhancing area, and the necrotic portions of the tumor. To select robust features against intra-rater and inter-rater variations, the VOI delineation process was repeated on 30 patients by the same radiologist (J.Y.) and by another neurosurgeon (Z.Y.Z. with 11 years' experience), yielding an intra-rater test data set and an inter-rater test data set respectively. Within the VOI, fixed bin number discretization with bin number of 150 was performed.

2.7. Radiomics features extraction

Radiomics features were calculated using the PyRadiomics tool version 3.0, an open-source python package for radiomics feature extraction from medical images [18]. Based on the VOI, three groups

of features were extracted, including shape features, first-order intensity features, and higher-order texture features. The shape features were calculated from the delineated VOI. The intensity features and texture features were extracted from both the original images and the transformed images obtained by applying wavelet transform or Laplacian of Gaussian (LoG) filtering onto the original images. All extracted features were summarized in **Supplementary Table S1**. The detailed calculations of the radiomics features used can be found in the literature reported by Zwanenburg A et al. [19]. We extracted 5929 features from all five MR sequences in total.

2.8. Radiomics signature construction

A three-stage feature selection approach was performed. First, features with low repeatability were excluded. The intra-rater and inter-rater repeatability for each feature was quantified by intraclass correlation coefficient (ICC) calculated on the intra-rater test data set and inter-rater test data set respectively. Any feature with ICC < 0.85 were discarded. Univariate concordance index (C-index) for each remaining feature was calculated to assess their association with OS. The features with univariate C-index ≥ 0.60 (positive association) or ≤ 0.40 (negative association) were selected as better prognostic variables for further analysis. Next, the least absolute shrinkage and selection operator (LASSO) penalized Cox proportional hazards regression [20] was used on the training data set to select the optimal feature subset and build the radiomics signature. LASSO shrunk all feature coefficients towards zero and set the coefficients of many irrelevant features exactly to zero. To optimize the model, 10-fold cross validation with minimum criteria was used, so the final model gave minimum cross validation error. The remaining non-zero features were used to construct the final regression model and were combined into a radiomics signature. The R package glmnet was used for LASSO modeling.

2.9. Statistical analysis

This study adhered to IBSI guidelines for reporting all necessary details in **Supplementary Table S2** to ensure reproducibility.

Validation of radiomics signature: Statistical analysis was performed using R version 3.5.3. *P*-value < 0.05 was considered significant. The differences in clinical, pathological, treatment and survival were assessed by a Wilcoxon test or Chi-square test. The potential association of the established radiomics signature with OS and PFS was first assessed in the training data set and then tested in the testing data set by using Kaplan-Meier analysis. According to a radiomics signature-based cutoff value determined by using an optimal cutoff analysis with X-tile software [21], patients were stratified into low-risk and high-risk subgroups. The cutoff value was estimated on the training data set and validated on the testing data set. A weighted log-rank test (the G-rho rank test, *rho* = 1) was applied to test the significant differences in the survival between the high-risk and low-risk subgroups [22]. The assessment of the radiomics signature as an independent prognostic factor was performed by integrating clinicomolecular risk factors such as age, sex (female or male), KPS, molecular subgroups (WNT, SHH, Group 3 or Group 4), extent of resection (complete or incomplete), radiation therapy (yes or no) and chemotherapy (yes or no) into the multivariate Cox proportional hazard model. Stratified analysis was performed to assess the potential association of the radiomics signature with the OS and PFS in subgroups stratified by clinicomolecular risk factors.

Incremental value of the radiomics signature: To demonstrate the incremental value of the radiomics signature over the clinicomolecular risk factors for individualized assessment of OS and PFS, both a clinicomolecular nomogram and a radiomics-clinicomolecular nomogram was constructed in the training data set. The clinicomolecular nomogram consisted of independent clinicomolecular risk factors including age, sex, KPS, molecular subgroups, extent of resection, radiation therapy and chemotherapy. Based on the clinicomolecular nomogram, the radiomics-clinicomolecular nomogram incorporated molecular subgroups, clinical risk factors and the radiomics signature based on Cox analysis.

According to the guideline reported by Steyerberg EW et al. [23], the incremental value of the radiomics signature to the clinicomolecular risk factors was assessed in terms of discrimination, calibration, reclassification and clinical usefulness. Specifically, the C-index was calculated with R package survival and used as a discrimination measure [24]. Calibration curves were plotted with R package rms to assess the agreement between the predicted OS/PFS based on proposed nomograms and the observed survival outcomes [25]. The net reclassification improvement (NRI) was calculated using R package survIDINRI to evaluate the usefulness improvement added by the radiomics signature [26]. The Akaike information criterion (AIC) was computed using R package stats to assess the risk of model overfitting. A decision curve analysis was performed with R package rmda to confirm the clinical usefulness of the radiomics-clinicomolecular nomogram [27].

Identification of biological pathways associated with radiomics signature: Based on the radiogenomics analysis cohort (GSE151519) with both MRI and RNA-seq data, we further investigated the possible biological pathways underlying the radiomics features. First, differentially expressing genes (DEGs) between the high- and low-risk groups stratified by the radiomics signature were identified using an R package DESeq2. Then, significant DEGs with false discovery rate (FDR)-adjusted $P < 0.25$ and $|\log_2(\text{Fold Change})| > 0.10$ were enriched to find overrepresented pathways using an R package clusterProfiler by querying the following annotated databases: Gene Ontology (GO) Biological Process, Kyoto Encyclopedia of Genes and Genomes (KEGG), Pathway Interaction Database (PID), and Reactome. FDR-adjusted hypergeometric $P < 0.05$ indicated significant enrichment. Then, a sample-based gene set variation analysis (GSVA) was performed on each enriched pathway to calculate a patient-specific GSVA score that quantified the pathway activity [28]. A Pearson correlation was used to assess if the pathway GSVA score was significantly associated (FDR-adjusted *P* value < 0.10) with the radiomics signature. The significantly correlated pathways were used to annotate the radiomics signature. Finally, the collective prognostic value of the radiomics-correlated pathways was assessed on a public cohort (GSE85218) by Cox regression using the average expression value of the genes contained in all significantly correlated pathways. Based on a cutoff value of the average expression, the patients in the public cohort were stratified into two risk groups, where log-rank $P < 0.05$ indicated the prognostic significance.

2.10. Role of funding source

Funders did not play any part in study design, data collection, data analyses, interpretation, or writing of the manuscript.

3. Results

3.1. Patient characteristics

According to the selection criteria, a total of 166 patients were included in the current study. The distribution of molecular subgroups of the current cohort was depicted in **Supplementary Figure S2.** Between the training and testing data sets, there were no significant differences in OS and PFS (Mean OS: training data set, 26.10 months; testing data set, 25.60 months; log-rank *P*-value, 0.5807. Mean PFS: training data set, 24.36 months, testing data set, 23.23 months; log-rank *P*-value, 0.5448). Furthermore, the distribution of clinical characteristics (subgroup [Chi-square $P = 0.4013$], extent of resection [Chi-square $P = 0.8669$], radiation [Chi-square $P = 0.7191$],

chemotherapy [Chi-square P = 0.3523], sex [Chi-square P = 0.0908], age [Wilcoxon's P = 0.4795], KPS [Wilcoxon's P = 0.5966]) was balanced between the training and testing data sets (**Supplementary Table S3**).

3.2. Feature selection and radiomics signature construction

Three-stage feature selection method was performed. After the intra-rater and inter-rater robustness tests, 1197 out of 5929 features remained (**Supplementary Figure S3**). After the univariate selection, 113 of 1197 radiomic features were selected for further analysis. Then, based on the selected features, the LASSO Cox model was built to fit the training data set. The detailed description of the radiomics features used in this study can be found in **Supplementary material**. Eleven radiomics features f_1–f_{11} with non-zero coefficients were selected by LASSO, as shown in **Supplementary Table S4**. The LASSO feature selection and signature construction were described in **Supplementary Figure S4A-B**. Finally, a radiomics signature was constructed by linearly combining the eleven features weighted by their coefficients as

Radiomics signature = –0.17755270 • f_1 –0.12569200 • f_2 + 0.03630490 • f_3 + 0.34856512 • f_4 –0.01929311 • f_5 + 0.42052760 • f_6 + 0.12144314 • f_7 –0.05861710 • f_8 –0.01884836 • f_9 –0.19943583 • f_{10} –0.19353807 • f_{11}. The features of f_1 to f_{11} were described in **Supplementary Table S4**.

The optimum cutoff value generated by the X-tile software was 0.10. Using this cutoff value, patients were divided into a high-risk subgroup (radiomics signature ≥ 0.10) and a low-risk subgroup (radiomics signature < 0.10). The signature value distribution was shown in **Supplementary Figure S4C**.

3.3. Validation of the radiomics signature

The association of radiomics signature with both OS (log-rank P = 0.0031; hazard ratio (HR) = 5.265, 95% confidence interval [CI]: 2.414, 11.48) and PFS (log-rank P = 0.0033; HR = 3.816, CI: 1.813, 8.033) was significant in training data set, as shown in Fig. 2A and 2B, respectively. The significant association was then demonstrated in the testing data set for both OS (log-rank P = 0.0086; HR = 4.3558, 95% CI: 1.66, 11.43) and PFS (log-rank P = 0.017; HR = 2.4636, 95% CI: 1.061, 5.722), as shown in Fig. 2C and 2D, respectively. The log-rank P value revealed significant difference in both OS and PFS distributions of the low-risk and high-risk subgroups. The radiomics signature was identified as an independent risk factor by the Cox regression analysis (HR: 4.7372, 95% CI: 2.574, 8.72, P = 2e-04). The Kaplan-Meier analyses for both OS and PFS in patient subgroups stratified by clinicomolecular risk factors were shown in the **Supplementary Figure S5** and **Supplementary Figure S6,** respectively, where significant associations (log-rank P < 0.05) were found in several subgroups.

3.4. Assessment of the incremental value of the radiomics signature

The clinicomolecular nomogram and radiomics-clinicomolecular nomogram for individual OS prediction were shown in Fig. 3A and

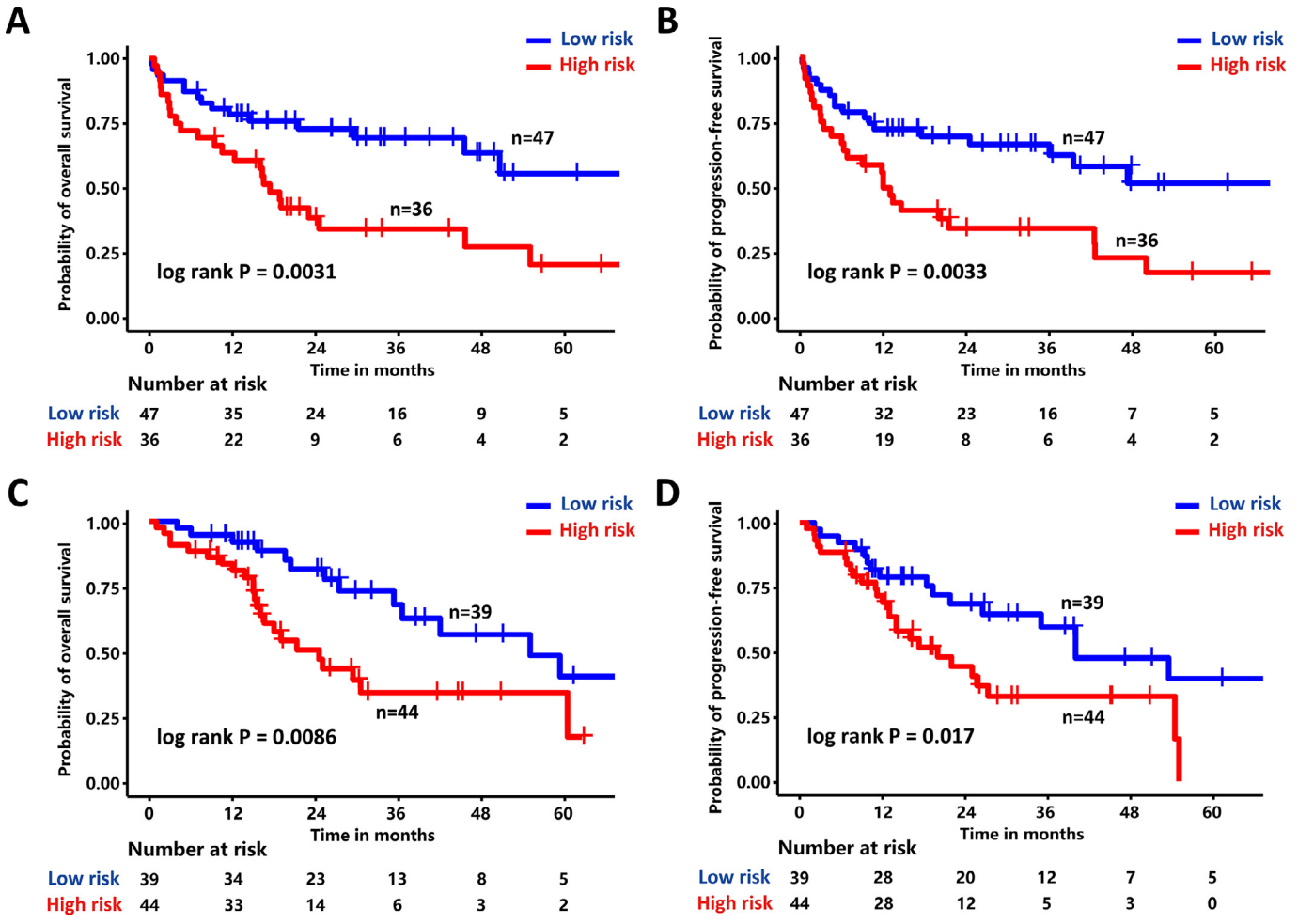

Fig. 2. Kaplan-Meier analysis according to the radiomics signature for OS (A) and PFS (B) in the training data set, and for OS (C) and PFS (D) in the testing data set. The significant associations of the radiomics signature with both OS and PFS were demonstrated. The numbers of patients at risk for each time step are shown in the bottom.

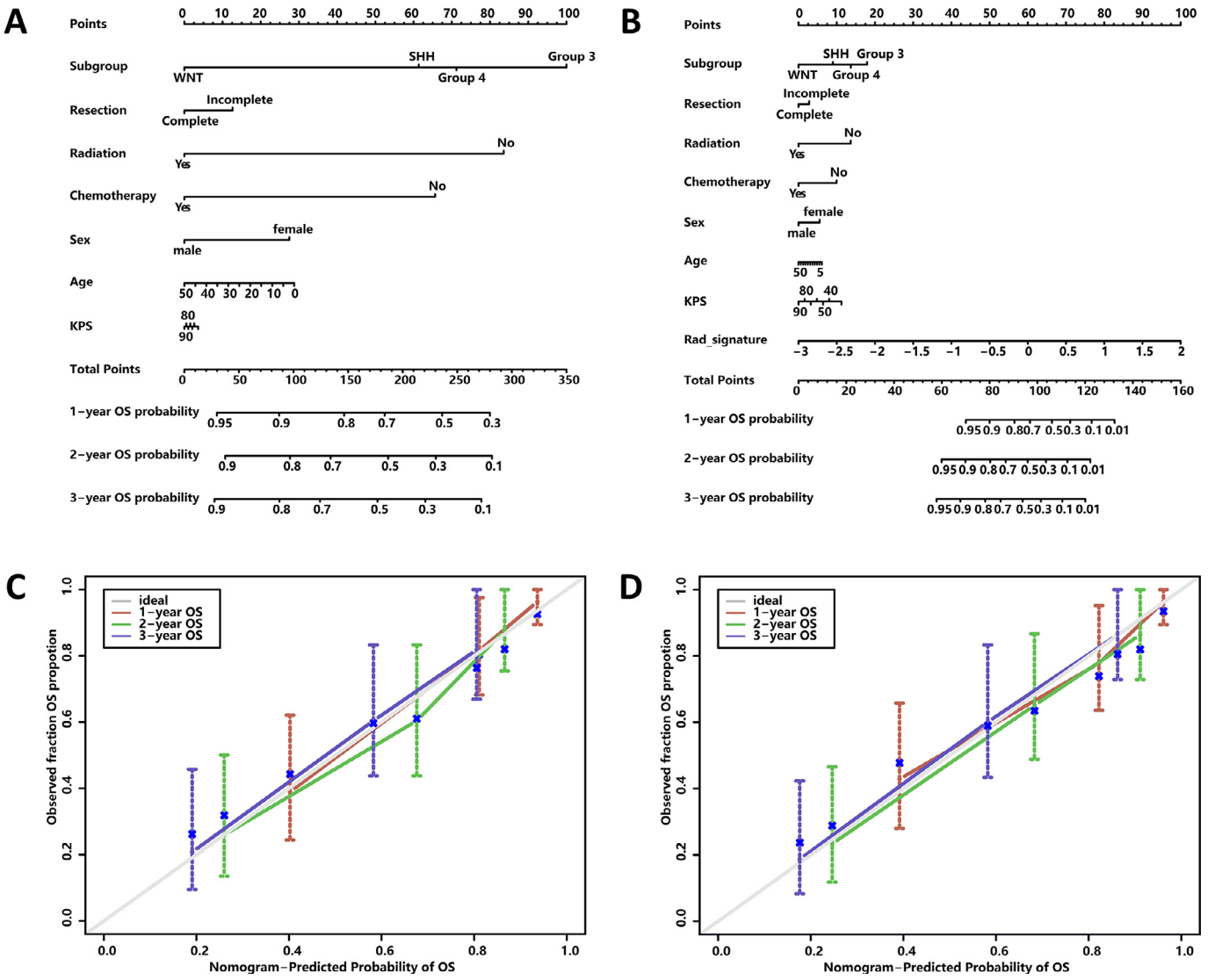

Fig. 3. The clinicomolecular nomogram (A) and the radiomics-clinicomolecular nomogram (B) for predicting the 1-, 2-, and 3-year OS outcomes, along with the calibration curves for assessment of the clinicomolecular nomogram (C) and the radiomics-clinicomolecular nomogram (D).

3B, respectively. The calibration curves of the clinicomolecular and radiomics-clinicomolecular nomograms for the probability of 1-, 2-, or 3-year death were depicted in Fig. 3C and 3D, respectively. Compared with the calibration curve of the clinicomolecular nomogram, the calibration curve of the radiomics-clinicomolecular nomogram demonstrated better agreement between the prediction and observation.

Similarly, we established a clinicomolecular nomogram and a radiomics-clinicomolecular nomogram for individualized PFS prediction, as shown in **Supplementary Figure S7A** and **S7B**, respectively. The calibration curve of the radiomics-clinicomolecular nomogram demonstrated better agreement between prediction and observation than that of the clinicomolecular nomogram, as shown in **Supplementary Figure S7C-D**.

The C-indices and AIC values were summarized in Table 1 for the radiomics signature, the clinicomolecular signature and the radiomics-clinicomolecular signature in both training and testing data sets. For OS prediction, the radiomics signature achieved a C-index of 0.677 (95% CI: 0.600, 0.754) in the training data set and 0.649 (95% CI: 0.567, 0.731) in the testing data set. The clinicomolecular signature achieved a C-index of 0.769 (95% CI: 0.697, 0.842) in the training data set and 0.725 (95% CI: 0.642, 0.809) in the testing data set. By combining the radiomics signature and the clinicomolecular risk factors, the discriminative performances in terms of C-index improved to 0.817 (95% CI: 0.759, 0.874) in the training data set and 0.762 (95% CI: 0.688, 0.836) in the testing data set. For PFS prediction, the C-index of the radiomics signature achieved 0.658 (95% CI: 0.581, 0.735) in the training data set and 0.593 (95% CI: 0.507, 0.679) in the testing data set. The clinicomolecular signature achieved a C-index of 0.749 (95% CI: 0.678, 0.821) in the training data set and 0.691 (95% CI: 0.604, 0.779) in the testing data set. After integrating the radiomics signature with clinicomolecular risk factors, the C-index increased to 0.787 (95% CI: 0.720, 0.855) in the training data set and 0.697 (95% CI: 0.612, 0.783) in the testing data set. Among all prediction models, the radiomics-clinicomolecular models achieved the lowest AIC values, indicating its better reliability against overfitting. The Kaplan-Meier analysis results of OS and PFS for both the clinicomolecular model and the radiomics-clinicomolecular model were shown in **Supplementary Figure S8A-H**, respectively, where significant prognostic associations (log-rank $P < 0.05$) were found for both models.

The integration of the radiomics signature with the clinicomolecular risk factors generated a total NRI of 0.298 (95% CI: 0.021, 0.496, $P = 0.022$) for OS prediction and a total NRI of 0.252 (95% CI: 0.012, 0.509, $P = 0.026$) for PFS prediction, implying the improved classification performance for both OS and PFS prediction. The decision curves

Table 1
C-indices and AIC values for OS and PFS prediction in both training and testing data sets. CM and R-CM are short for clinicomolecular and radiomics-clinicomolecular, respectively.

Model	C-index		AIC	
	Training data set			
	OS	PFS	OS	PFS
Radiomics signature	0.677 (0.600 0.754)	0.658 (0.581 0.735)	291.1901	334.4212
CM nomogram	0.769 (0.697 0.842)	0.749 (0.678 0.821)	277.4064	318.3184
R—CM nomogram	0.817 (0.759 0.874)	0.787 (0.720 0.855)	261.6403	305.9829
	Testing data set			
	OS	PFS	OS	PFS
Radiomics signature	0.649 (0.567 0.731)	0.593 (0.507 0.679)	257.0576	310.8909
CM nomogram	0.725 (0.642 0.809)	0.691 (0.604 0.779)	242.5913	296.3209
R-CM nomogram	0.762 (0.688 0.836)	0.697 (0.612 0.783)	239.7708	294.9218

for both OS and PFS prediction shown in Fig. 4 demonstrated the clinical usefulness of the prediction models, indicating that the radiomics-clinicomolecular nomogram achieved a higher overall net benefit than the clinicomolecular nomogram within most range of threshold probabilities.

3.5. *Identification of biological pathways associated with radiomics signature*

In our MB cohort with both data of RNA-seq and MRI (GSE151519, 7 male and 10 female, age range: 2.0–26.0 years, median age: 9.0 years), 559 DEGs were identified to be differentially expressed between the high- and low-risk groups stratified by the radiomics signature, as shown by a volcano plot in Fig. 5A. The enrichment analysis based on the DEGs identified the key biological pathways, as shown in Fig. 5B. A complete list of all enriched pathways was provided in **Supplementary Table S5**. The radiomics signature was found to be significantly associated with up regulation of genes involved in WNT signaling, P53 pathway, PI3K/AKT pathway, interleukin-2 pathway, and membrane protein activities, as shown in Fig. 5C. The collective prognostic performance of the average expression value of genes contained in the nine pathways was evaluated in the GSE85218 cohort (From all 763 patients, 612 patients with available survival data were included while the other 151 patients without available survival data were excluded; for the included 612 patients: 392 male and 207 female and 13 unavailable, age range: 0.24–56.8 years, median age: 8.0 years, survival range: 0.16–300 months, mean survival: 59.61 months). The results of Kaplan-Meier analysis (Fig. 5D) showed that patients in the public cohort can be stratified into two risk groups in terms of OS (log-rank P = 0.016; HR = 0.6322, 95% CI: 0.3789, 1.055) with a cutoff value of 5.83. To further reveal the radiomics-pathway-prognosis relevance, the feature maps of 11 radiomics features of two representative patients in the low- and high-risk groups were presented respectively in Fig. 5E. The delineated tumor was overlapped in red on the MR images. Radiomics feature maps of the selected 11 features were presented. The figures from left to right were: rT1, rT1c, rT2, rFLAIR, rADC image overlapped with the tumor contour, f_1–f_{11} features (as defined in **Supplementary Table S4**). The feature maps as a visualization tool can reveal the intratumoral variations of the imaging patterns between the high- and low-risk patients, showing the association of the radiomics signature with patient prognosis. Detailed description of the meanings of the 11 features can be found in **Supplementary material**.

4. Discussion

Current risk stratification for patients with MB is based on clinical factors (age, extent of resection, etc.) or molecular subgroups [3]. In this study, we utilized maching-learninng approaches to develop and validate a radiomics signature for prediction of OS and PFS from preoperative MRI (T1, T1c, T2, FLAIR and ADC) and a radiomics-clinicomolecular prognostic nomogram for assessing the added value of the radiomics signature over existing risk factors in MB patients.

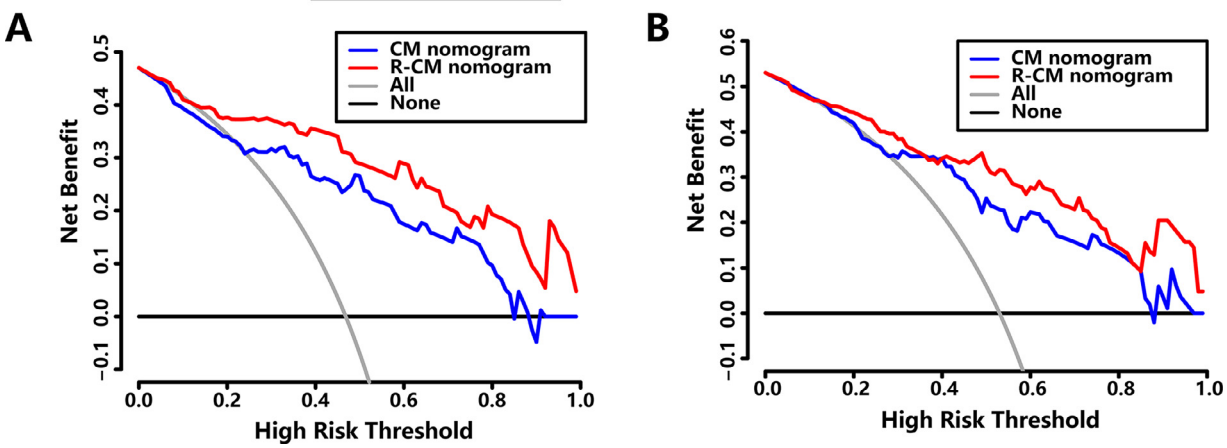

Fig. 4. Decision curve analysis (DCA) for radiomics-clinicomolecular (R-CM) nomogram and clinicomolecular (CM) nomogram to estimate the OS (A) and PFS (B). The x-axis represents the threshold probability and the y-axis measures the net benefit.

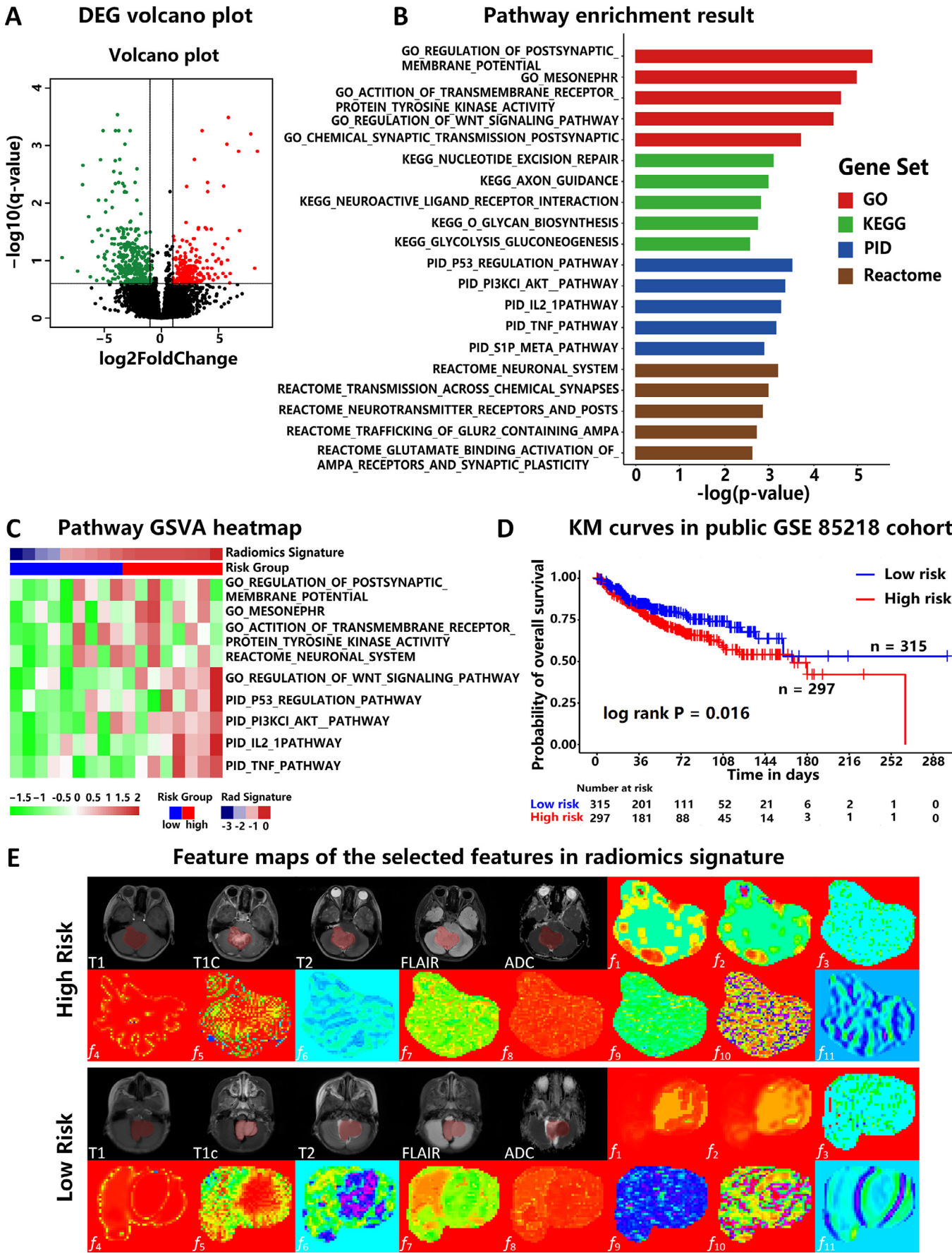

Fig. 5. A summary of the radiogenomics analysis results showing the radiomics-transcriptomics-prognosis association in MB. (A) Volcano plot of the differentially expressed genes (DEGs) between risk groups stratified by the radiomics signature. The vertical line is at $|\log_2 \text{Fold Change}| = 0.10$ while the horizontal line at false discovery rate (FDR) = 0.25. The red

Moreover, paired data of radiomics and transcriptomes of MB were investigated for revealing possible signaling pathways associated with radiomic patterns.

In recent years, an increasing number of reports have demonstrated that radiomics analysis on MRI may have the potential to differentiate diagnostic ambiguity, to monitor response to adjuvant therapies, to characterize the genotypes of CNS tumors, and to improve prognostic models [8,9,29]. Typically, gliomas are one of the most studied tumors. On the one hand, studies have revealed that machine learning based radiomics are able to predict molecular markers such as IDH mutations [30] and molecular subtypes in gliomas [31]. On the other hand, advancement has also been achieved in survival prediction of gliomas by radiomics [9]. As for MB, Dasgupta A et al. [10] have used 19 conventional MRI features in 111 patients with MB for prediction of molecular subgroups, and Iv M et al. [11] developed and validated a machine-learning radiomic method that identifies MRI features predictive of molecular subgroups of MB. To our knowledge, the relationships between radiomic features and the survivals of MB patients has not been investigated.

The present study revealed that radiomic signatures from MRI has independent and incremental prognostic value with respect to the putative molecular and clinical parameters for individualized survival prediction in patients with MB. Previous studies have reported the heterogeneity within the four molecular subgroups of MB. For example, Cavalli FMG et al. [15] identified 12 clinical and biological relevant subtypes within the frame of the four molecular subgroups of MB by integrative clustering. It is reasonable to infer that the existing molecular grouping still has room to be refined, and radiomics may identify radiographic phenotypes that could offer distinct and complementary prognostic information beyond molecular subgroups. Two intensity features and nine texture features were selected for constituting the signature, while no shape features were selected. This result emphasized the prognostic value of intensity/texture heterogeneity within tumor for MB. The feature maps in Fig. 5E indicate that the selected features can capture subtle risk-relevant intratumoral heterogeneity at the radiological level and thus are associated with patient prognosis. Detailed description of the meanings of the eleven features can be found in **Supplementary material**. Note that by incorporating the radiomics signature into the clinicomolecular model, the C-index in the testing data set for PFS prediction increased slightly by 0.06, less than that for OS prediction of 0.37 (Table 1).

One of the major obstacles lies in radiomic research is the obscurity associated with the underlying biological explanations of radiomic features. In the current study, a radiogenomic analysis of MB based on integrative analysis of radiomic and transcriptomic profiles indicated that a high-risk radiomic phenotype is significantly associated with several dysregulated signaling pathways, most of which were demonstrated to play important parts in the carcinogenesis or progression of MB. For example, WNT signaling cascade is well known in MB as it is the dominating pathway of WNT subgroup [3-5], while PI3K/AKT pathway was found to have cross-talk with WNT pathway in MB pathophysiology [32]. Moreover, PI3K/AKT pathway was also revealed to be activated in MB cell proliferation [33] and related to poor prognosis in Group 3 and 4 subgroups [34]. In addition, as upstream molecules of PI3K/AKT pathway, transmembrane receptor tyrosine kinases such as ERBB4 were identified as key proteins in the aberrant signaling specific to Group 4 MB [35], and EphB2 was revealed to play a pivotal role in MB cell invasion [36]. As for p53 regulation pathway, IL2 pathway and TNF pathway, these signaling pathways are canonical pathways in carcinogenesis and significantly related to SHH, Group 3 and Group 4 subgroups of MB [34]. Furthermore, the mean expression of the pathway genes was found to be significantly associated with overall survival, as assessed in a public GSE85218 cohort [15], demonstrating that the radiomics-associated pathway genes may be involved in key biological processes that contribute to MB prognosis.

Admittedly, this study has limitations. First, the relative higher proportion of Group 3 (44%) in our cohort (n = 166) compared to that (27%) reported by the international meta-analysis (n = 550) [3] may be explained by the relatively limited sample size of this study. Further investigations with larger sample size from multiple institutions are necessary to substantiate our findings. Second, this study lacks volumetric MRI data, which could introduce significant interpolation when 5 mm thick slices were resampled to 1 mm thickness. Third, despite our study included the most common five MR sequences, incorporating advanced MRI sequences such as magnetic resonance spectroscopy (MRS), and dynamic susceptibility contrast (DSC) perfusion may provide more information and boost prognostication performance. Fourth, the sample size of fresh tumor sepcimens used for RNA-sequencing (n = 17) was limited, and larger sample size for radiogenomic analysis was needed in future studies.

In conclusion, the current study demonstrated proof-of-concept results for integrating radiomics into nomograms comprising of molecular subgroups and clinical factors for improving prognostication performance of MB patients. Notably, we revealed that the radiomic features, which are associated with distinct biological pathways, could bring considerably incremental values to the existing clinicomolecular factors for survival prediction in MB patients.

Contributors

All authors read and approved the final version of the manuscript.

Research conception, design and manuscript revision: Zhenyu Zhang, Jing Yan, Zhicheng Li, Xianzhi Liu, Kay Ka-Wai Li, Ho-Keung Ng, Binke Yuan, Jingliang Cheng and Wencai Li

Data processing, statistical analysis, drafting of the manuscript: Zhenyu Zhang, Jing Yan, Zhicheng Li, Shenghai Zhang and Lei Liu

Detection of molecular subgroups of medulloblastoma: Kay Ka-Wai Li, Ho-Keung Ng, Weiwei Wang and Li Wang.

Delineation volume of interest (VOI) of the tumor contours: Jing Yan and Zhenyu Zhang

Download the MRI data from the picture archiving and communication system (PACS) and sorting the MRI data: Jing Yan, Ke Li and Xiangxiang Wang

Acquisition of tissue specimens, clinical data and survival data: Ke Li, Yunbo Zhan, Wenchao Duan, Dongling Pei, Haibiao Zhao, Tao Sun, Chen Sun, Wenqing Wang, Zhen Liu, Xuanke Hong and Yu Guo

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 81702465, 61571432 and U1804172 and U1904148), the Science and Technology Program of Henan Province (No. 182102310113, 192102310123 and 192102310050), the

and green dots represent DEGs found to be upregulated and downregulated, respectively. (B) Top five enriched pathways in Gene Ontology (GO) Biological Process (red), Kyoto Encyclopedia of Genes and Genomes (KEGG, green), Pathway Interaction Database (PID, blue), and Reactome (brown). (C) A heatmap of the gene set variation analysis (GSVA) score of enriched pathways significantly correlated with the radiomics signature. (D) Kaplan-Meier curves based on the average expression value of the genes contained in the radiomics signature-correlated pathways for OS prediction in the public GSE85218 cohort. (E) MR images (row 1 and row 3) and corresponding feature maps in two patients who were classified into the high-risk group (top two rows, OS = 2 months, Radiomics signature score = 0.3752) and low-risk group (bottom two rows, OS = 44.3 months, Radiomics signature score = −1.1124).

Research Projects of Henan Higher Education (No.18A320077), and Shenzhen Basic Research Program (JCYJ20170413162354654).

Data Sharing

The sequencing data is deposited in GEO databases, and that accession number is GSE151519 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151519). The publicly created dataset is available.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.103093.

References

[1] Leece R, Xu J, Ostrom QT, et al. Global Incidence of malignant brain and other central nervous system tumors by histology, 2003–2007. Neuro Oncol 2017;19(11):1553–64.
[2] Thompson EM, Hielscher T, Bouffet E, et al. Prognostic value of medulloblastoma extent of resection after accounting for molecular subgroup: a retrospective integrated clinical and molecular analysis. Lancet Oncol 2016;17(4):484–95.
[3] Taylor MD, Northcott PA, Korshunov A, et al. Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol 2012;123(4):465–72.
[4] Kool M, Korshunov A, Remke M, et al. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, group 3, and group 4 medulloblastomas. Acta Neuropathol 2012;123(4):473–84.
[5] Gibson P, Tong Y, Robinson G, et al. Subtypes of medulloblastoma have distinct developmental origins. Nature 2010;468(7327):1095–9.
[6] Northcott PA, Buchhalter I, Morrissy AS, et al. The whole-genome landscape of medulloblastoma subtypes. Nature 2017;547(7663):311–7.
[7] Louis DN, Perry A, Reifenberger G, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 2016;131(6):803–20.
[8] Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA-Cancer J Clin 2019;69(2):127–57.
[9] Kickingereder P, Neuberger U, Bonekamp D, et al. Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma. Neuro Oncol 2018;20(6):848–57.
[10] Dasgupta A, Gupta T, Pungavkar S, et al. Nomograms based on preoperative multiparametric magnetic resonance imaging for prediction of molecular subgrouping in medulloblastoma: results from a radiogenomics study of 111 patients. Neuro Oncol 2019;21(1):115–24.
[11] Iv M, Zhou M, Shpanskaya K, et al. MR imaging-based radiomic signatures of distinct molecular subgroups of medulloblastoma. Am J Neuroradiol 2019;40(1):154–61.
[12] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[13] Itakura H, Achrol AS, Mitchell LA, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med 2015;7(303) 303-138.
[14] Tixier F, Cheze-le-Rest C, Schick U, et al. Transcriptomics in cancer revealed by positron emission tomography radiomics. Sci Rep 2020;10:5660.
[15] Cavalli FMG, Remke M, Rampasek L, et al. Intertumoral heterogeneity within medulloblastoma subgroups. Cancer Cell 2017;31(6):737–54.
[16] Northcott PA, Shih DJ, Remke M, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol 2012;123(4):615–26.
[17] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 2008;5:621–8.
[18] Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77(21):e104–7.
[19] Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardisation initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295(2):328–38.
[20] Tibshirani RJ, Sim. The lasso method for variable selection in the Cox model. Stat Med 1997;16(4):385–95.
[21] Camp RL, Dolled-Filhart M, Rimm DLJCcr. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10(21):7252–9.
[22] Buyske S, Fagerstrom R, Ying ZJJotASA. A class of weighted log-rank tests for survival data when the event is rare. J Am Stat Assoc 2000;95(449):249–58.
[23] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models a framework for traditional and novel measure. Epidemiology 2010;21(1):128–38.
[24] Harrell FE, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. JAMA 1982;247(18):2543–6.
[25] Kattan MW. Judging new markers by their ability to improve predictive accuracy. J Natl Cancer I 2003;95(9):634–5.
[26] Pencina MJ, D'Agostino Sr RB, Steyerberg EWJSim. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30(1):11–21.
[27] Vickers AJ, Elkin EBJMDM. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26(6):565–74.
[28] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
[29] Macyszyn L, Akbari H, Pisapia JM, et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro Oncol 2016;18(3):417–25.
[30] Zhang B, Chang K, Ramkissoon S, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. Neuro Oncol 2017;19(1):109–17.
[31] Lu CF, Hsu FT, Hsieh KL, et al. Machine learning-based radiomics for molecular subtyping of gliomas. Clin Cancer Res 2018;24(18):4429–36.
[32] Baryawno N, Sveinbjörnsson B, Eksborg S, et al. Small-molecule inhibitors of phosphatidylinositol 3-kinase/Akt signaling inhibit Wnt/beta-catenin pathway cross-talk and suppress medulloblastoma growth. Cancer Res 2010;70(1):266–76.
[33] Hartmann W, Digon-Söntgerath B, Koch A, et al. Phosphatidylinositol 3′-kinase/AKT signaling is activated in medulloblastoma cell proliferation and is associated with reduced expression of PTEN. Clin Cancer Res 2006;12(10):3019–27.
[34] Park AK, Lee JY, Cheong H, et al. Subgroup-specific prognostic signaling and metabolic pathways in pediatric medulloblastoma. BMC Cancer 2019 Jun 11;19(1):571.
[35] Forget A, Martignetti L, Puget S, et al. Aberrant ERBB4-SRC signaling as a hallmark of group 4 medulloblastoma revealed by integrative phosphoproteomic profiling. Cancer Cell 2018;34(3):379–395.e7.
[36] Sikkema AH, den Dunnen WF, Hulleman E, et al. EphB2 activity plays a pivotal role in pediatric medulloblastoma cell adhesion and invasion. Neuro Oncol 2012;14(9):1125–35.