

**Title:** Prediction of H3K27M-mutant brainstem glioma by amide proton transfer weighted imaging and its derived radiomics

**Running title:** Brainstem glioma H3K27M-mutant prediction by APTw imaging

**Author list:** Zhizheng Zhuo, PhD<sup>1\*</sup>; Liying Qu, MS<sup>1\*</sup>; Peng Zhang, MD, PhD<sup>2\*</sup>; Yunyun Duan, MD, PhD<sup>1</sup>; Dan Cheng, MD<sup>1</sup>; Xiaolu Xu, MD, PhD<sup>1</sup>; Ting Sun, PhD<sup>1</sup>; Jinli Ding, PhD<sup>1</sup>; Cong Xie, MS<sup>1</sup>; Xing Liu, MD, PhD<sup>3</sup>; Sven Haller, MD, PhD<sup>4</sup>; Frederik Barkhof, MD, PhD<sup>5,6</sup>; Liwei Zhang, MD, PhD<sup>2#</sup>; Yaou Liu, MD, PhD<sup>1#</sup>

**Affiliation:**

1. Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, 10070

2. Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, 10070

3. Department of Pathology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, 10070

4. Department of Imaging and Medical Informatics, University Hospitals of Geneva and Faculty of Medicine of the University of Geneva, Geneva, Switzerland

5. UCL Institutes of Neurology and Healthcare Engineering, London, UK

6. Department of Radiology & Nuclear Medicine, Amsterdam University Medical Centers, The Netherlands

\*Zhizheng Zhuo, Liying Qu and Peng Zhang contributed equally to this study.

# Liwei Zhang and Yaou Liu were both corresponding authors, who contributed equally to this study.

Primary corresponding to: Yaou Liu, MD, PhD, Department of Radiology, Beijing Tiantan Hospital, Capital Medical University. No.119, the West Southern 4th Ring Road, Fengtai District, Beijing, 100070, China. Tel: 86 13810345782; Fax: 86 10 59975358. E-mail: [liuyaou@bjtth.org](mailto:liuyaou@bjtth.org).

**Word count:** 4622 including words in abstract, text, references, and figure legends.

## **Abstract**

### **Purpose**

H3K27M-mutant associated brainstem glioma (BSG) carries a very poor prognosis. We aimed to predict H3K27M mutation status by amide proton transfer weighted (APT<sub>w</sub>) imaging and radiomic features.

### **Methods**

Eighty-one BSG patients with APT<sub>w</sub> imaging at 3T MRI and known H3K27M status were retrospectively studied. APT<sub>w</sub> values (mean, median and max) and radiomic features within manually delineated 3D tumor masks were extracted. Comparison of APT<sub>w</sub> measures between H3K27M-mutant and wildtype groups was conducted by two-sample Student's T/Mann-Whitney U test and receiver operating characteristic curve (ROC) analysis. H3K27M-mutant prediction using APT<sub>w</sub>-derived radiomics was conducted using a machine-learning algorithm (Support Vector Machine) in randomly selected train (n=64) and test (n=17) sets. Sensitivity analysis with additional random splits of train and test sets, 2D tumor masks and other classifiers were conducted. Finally, a prospective cohort including 29 BSG patients was acquired for validation of the radiomics algorithm.

### **Results**

BSG patients with H3K27M-mutant were younger and had higher max APT<sub>w</sub> values than those with wildtype. APT<sub>w</sub>-derived radiomic measures reflecting tumor heterogeneity could predict H3K27M mutation status with an accuracy of 0.88, sensitivity of 0.92 and specificity of 0.80 in the test set. Sensitivity analysis confirmed the predictive ability (accuracy range: 0.71-0.94). In the independent prospective validation cohort, the algorithm reached an accuracy of 0.86, sensitivity of 0.88 and specificity of 0.85 for predicting H3K27M-mutation status.

### **Conclusion**

BSG patients with H3K27M-mutant had higher max APT<sub>w</sub> values than those with wildtype. APT<sub>w</sub>-derived radiomics could accurately predict a H3K27M-mutant status in BSG patients.

### **Keywords:**

Brainstem glioma; H3K27M-mutant; Amide proton transfer weighted imaging; Radiomics; Machine learning

## Introduction

Brainstem gliomas (BSGs) are a heterogeneous group of tumors involving midbrain, pons and medulla. Genetic characterization is able to identify BSG patients with a poorer prognosis in those harboring a methionine in histone H3 at lysine 27 (H3K27) mutation [1], which causes an oncogene expression through global depletion of the repressive modification H3 lysine 27 trimethylation (H3K27me3) [2]. Accurate identification of H3K27M status contributes to diagnostic accuracy, improves determination of prognosis and treatment response, and allows identification of potential therapeutic targets [1-3].

Invasive stereotactic biopsy can identify the H3K27M-mutant status with a high sensitivity and specificity, but carries a significant risk of complication [4]. Additionally, the inherent heterogeneity of BSG poses the risk of a biopsy bias, while the alternative cerebrospinal fluid (CSF) derived tumor DNA analysis approach might be inconclusive if the tumor is not directly adjacent to the CSF [3-5]. Non-invasive conventional magnetic resonance imaging (MRI) is able to depict BSG location, morphology, diffusion, perfusion and metabolic characteristics, but is unable to predict genetic mutation status with insufficient specificity (<75%) [6-9]. Therefore, the accurate non-invasive determination of genetic H3K27 mutation in BSG remains a challenge in clinical practice.

Amide proton transfer weighted (APT<sub>w</sub>) imaging is a novel MRI technique, which generates image contrast based on the endogenous amide protons in mobile cellular proteins and peptides [10], providing a promising avenue to explore the molecular metabolism associated with the cellular proliferating and gene expression. To fully appreciate tumor heterogeneity, radiomic features from conventional MRI have been successfully used to predict genetic alterations (e.g., isocitrate dehydrogenase [IDH] mutation, H3K27M) in brain tumors [11-14]. So far, there have been no studies using APT<sub>w</sub> and its derived radiomics to predict H3K27M mutation status in BSG.

In this study, we examined whether APT<sub>w</sub> and APT<sub>w</sub>-derived radiomic features could characterize the BSG metabolic heterogeneity and predict H3K27M mutation status. To this end, a retrospective study was conducted to build and test a prediction model to predict H3K27M-mutation status, which was validated in an independent prospective cohort.

## **Materials and Methods**

### **Ethics approval and study design**

This study was approved by the Animal and Human Ethics Committee of Beijing Tiantan Hospital, Capital Medical University. All the patients or legal guardians provided written informed consents.

In this study, data from a retrospective cohort was examined to build and test the H3K27M-mutant prediction model and an independent prospective cohort was then studied to validate the prediction model (**Figure 1**).

### **Retrospective cohort**

From December 2018 to June 2020, a total of 87 patients diagnosed with BSG (aged 2-65 years old) who had both APTw imaging and available H3K27M status were retrospectively studied (representative patients were showed in **Figure 2A**). Exclusion criteria included: (1) patients with post-treatment (surgery, radiotherapy or chemotherapy) APTw data; (2) patients with poor imaging quality (**Figure 1**).

### **Prospective cohort**

From June 2020 to December 2020, 30 patients with clinically and radiologically suspected BSG were prospectively recruited for APTw imaging. Inclusion criteria included: (1) patients without prior treatment history; (2) patients would receive biopsy or surgical resection; (3) patients would undergo H3K27M status testing. Exclusion criteria included: (1) patients were not pathologically diagnosed as glioma; (2) patients with poor image quality (**Figure 1**).

### **MRI Acquisition**

MRI acquisition was performed on a 3T MR scanner (Ingenia CX, Philips Healthcare, Best, the Netherlands) with a 32-channel head receiver coil. The MR protocol included T1w, T2w, T2-FLAIR, diffusion weighted imaging and contrast enhanced T1w. For APTw, the parameters of the 3D turbo spin echo were: repetition time 5864 ms, echo time 8.8 ms, flip angle = 90°, acquired voxel size= 2 mm×2 mm×6 mm, 7 slices acquired. A saturation pulse was applied at +3.5 ppm (relative to the water resonance frequency) to saturate amide protons, saturation B1 amplitude=2 uT, saturation duration=2s. Six

additional image volumes at critical frequencies ( $\pm 3.1$  ppm,  $-3.5$  ppm,  $\pm 3.9$  ppm and  $-1560$  ppm) were acquired for the Z-spectrum normalization and interpolation, and three acquisition were performed at  $+3.5$  ppm with different echo time shifts to obtain a Dixon-type B0 field map to correct for B0 inhomogeneities in the Z-spectrum frequency domain [15]., parallel acceleration factor (SENSE) = 1.6, acquisition time = 1min54s. Details of conventional pulse-sequences can be found in **Supplemental Table S1**.

### **APT<sub>w</sub> image calculation**

APT<sub>w</sub> images were automatically calculated using the embedded post-processing pipeline on the MR console.

An MTR<sub>asym</sub> analysis (asymmetry with respect to the water frequency) was conducted by a voxel-by-voxel analysis to distinguish the APT signal from background effects (e.g., direct water saturation and magnetization transfer contributions from semi-solid tissue components) [16].

First, the Z-spectrum was aligned per voxel using the B0 field map to correct for B0 inhomogeneity artifacts . Next, asymmetry was calculated by subtracting the positive frequency side signal  $S[+\Delta\omega]$  from the negative side signal  $S[-\Delta\omega]$  and normalized to the unsaturated image signal  $S_0$  (See Equation 1). The resulting MTR<sub>asym</sub> value at  $+3.5$  ppm is presented as percent level (relative to  $S_0$ ) in the final APT<sub>w</sub> images (Equation 2).

$$\text{MTR}_{\text{asym}}(\%) = (S_{-\Delta\omega} - S_{+\Delta\omega})/S_0 \text{ (Equation 1)}$$

$\text{MTR}_{\text{asym}}(\%)$  = Magnetization Transfer Ratio asymmetry in percent,  $S_{-\Delta\omega}$  = signal at negative frequency offset  $-\Delta\omega$ ,  $S_{+\Delta\omega}$  = signal at positive frequency offset  $+\Delta\omega$ ,  $S_0$  = signal without radio frequency saturation.

$$\text{APT}_w\% = \text{MTR}_{\text{asym}}[\Delta\omega = +3.5\text{ppm}](\%) \text{ (Equation 2)}$$

$\text{APT}_w\%$  = Amide Proton Transfer weighted percentage,  $\text{MTR}_{\text{asym}}[\Delta\omega = +3.5\text{ppm}](\%)$  = Magnetization Transfer Ratio asymmetry at  $+3.5\text{ppm}$  offset frequency.

### **Tumor segmentation and radiological analysis of tumor**

The solid tumor boundaries (excluding cystic, necrotic and hemorrhagic areas) were manually delineated on T2w images independently by two radiologists (L.Q and D.C, both with 3 years' experience in neuroradiology), who were blinded to the H2K27M status, with contrast-enhanced T1w and T2-FLAIR as reference, using 3D Slicer software (v4.10.2, <https://www.slicer.org/>). In case of discrepancy between the readers, the 3D tumor mask was reviewed, modified (if necessary) and confirmed by a senior radiologist (Y.D with 12 years' experience in neuroradiology). Finally, the overlapping regions of tumor masks drawn by the two radiologists (average dice score=0.87) were used to determine the tumor volume. Using this mask, voxel-wise lesion frequency mapping was performed by normalizing the tumor mask to Montreal Neurological Institute (MNI) coordinates using Statistical Parametric Mapping (SPM12, <https://www.fil.ion.ucl.ac.uk/spm/>) and dividing the summed binary tumor masks by the total number of enrolled patients (**Figure 2B**).

Based on conventional MRI (T1w, T2w and contrast-enhanced T1w), the tumor location (midbrain, pons and medulla) and presence of dorsal exophytic components, hydrocephalus, cystic components, contrast enhancement, and engulfment of basilar artery were assessed by two independent (L.Q and D.C) radiologists; to resolve inconsistencies, a senior radiologist (Y.D) reviewed and determined the final assessments.

### **APT<sub>w</sub> feature extraction**

In order to extract the APT<sub>w</sub> measures and radiomics of BSG, 3D tumor masks were warped to the APT<sub>w</sub> image space by the forward transformation parameters derived from the affine registration of the T2w images to the unsaturated images (S0 image). Mean, median and max APT<sub>w</sub> values within the 3D tumor masks were extracted using an in-house script in MATLAB 2019a (MathWorks, Natick, MA, USA).

### **Radiomics extraction**

Feature Analysis Explorer (FAE v0.3.6, <https://github.com/salan668/FAE>) software was used to extract radiomics parameters. These included shape-based [n=14], first order [n=18], gray level co-occurrence matrix (GLCM) [n=24], gray level dependence matrix (GLDM) [n=14]; gray level run length matrix (GLRLM) [n=16], and gray level size zone matrix (GLSZM) [n=16], and neighboring gray tone



difference matrix (NGTDM) [n=5] of original images and a series of transformation derived images based on wavelet transform, square, square root, Laplacian of Gaussian, logarithm, exponential, gradient and local binary pattern. In total, 1316 radiomics features were obtained within the 3D tumor masks.

### **Histopathological analysis**

Tumor grading was determined according to the 2016 World Health Organization Classification of Tumors of the Central Nervous System [17]. H3K27M status (H3K27M-mutant or wildtype) was determined by immunohistochemical staining using a mutation-specific antibody [18].

### **Statistical analysis**

The statistical analysis was performed using SPSS (SPSS for Windows, version 25.0; IBM, Armonk, NY, USA) and statistics toolbox in Matlab 2019a (MathWorks, Natick, MA, USA).

Categorical variables are displayed as frequencies and tested using Chi-square test. Continuous variables are presented as median and interquartile range (IQR) and tested by Student's T test or Mann-Whitney U test. A p value < 0.05 was deemed statistically significant.

The spatial concordance of the 3D tumor segmentation masks between radiologists was evaluated using the dice score.

Univariable and multivariable logistic regression models (receiver operating characteristic curve [ROC] analysis) were used to evaluate the discriminatory ability of clinical and APTw measures for H3K27M-mutant versus wildtype status.

### **Radiomics prediction model development**

A support vector machine (SVM) approach embedded in FAE (<https://github.com/salan668/FAE>) was adopted for the identification of H3K27M-mutant tumors based on radiomic features.

The retrospective cohort was used to build and test an SVM prediction model. Patients were first

randomly separated into train (80%, n=64 [46 H3K27-mutant and 18 wildtype]) and test (20%, n=17, [12 H3K27 mutant and 5 wildtype]) datasets (**Table 1**). An up-sampling strategy was used to balance the patients with H3K27M-mutant and wildtype tumors in the training set. Z-score transformation was used for feature normalization. Pearson's Correlation Coefficient (PCC) was used for dimensionality reduction by randomly removing one of the paired features with high similarity (PCC>0.9) [19] and Relief-based feature selection [20]. Leave-one-out cross-validation (LOOCV) was used for the model parameter optimization based on the training set. Test data were used for evaluating the model performance using area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value and negative predictive value.

### **Sensitivity analysis and independent validation**

Sensitivity analysis were performed to exclude potential effects of (1) unbalanced separation of train and test datasets by performing 9 additional repeats of the SVM classification procedure; (2) dependency on 3D tumor masks by repeating the radiomics extraction and SVM classification using a 2D tumor mask from the axial slice with the maximum tumor area; (3) spurious selection of classifiers by comparing the current SVM findings with other popular classifiers including adaboost, auto encoder, logistic regression, least absolute shrinkage and selection operator (lasso) regression, linear discriminant analysis, gaussian process, naïve bayes, decision tree and random forest.

The prospective cohort was used as an independent validation dataset to validate the built prediction model.

## Results

### Demographic and clinical characteristics

For the retrospective cohort, 87 BSG patients were eligible. Two patients were excluded due to imaging acquisition after surgery and 4 patients were excluded due to poor APTw image quality. Ultimately, 81 patients including 58 H3K27M-mutant (median age 9 years, female/male=31/27) and 23 wildtype (median age 24 years, female/male=15/8) tumors remained (**Table 1**). Patients with H3K27M-mutant tumors were younger than those with wildtype ( $p=0.009$ ).

### MRI and histopathological findings

For MRI features, a majority (48/58, 83% vs. 8/23, 35%;  $p<0.001$ ) of H3K27M-mutant BSGs were located in the pons, while wildtype BSGs occurred in midbrain (35%), pons (35%) and medulla (30%) with similar frequency. A larger proportion (7/23, 30%) of wildtype BSG cases presented with a dorsal exophytic component compared to those with H3K27M-mutant (4/58, 7%;  $p=0.005$ ), while more H3K27M-mutant BSG cases (41/58, 71%) presented with engulfment of the basilar artery than those with wildtype (6/23, 26%;  $p<0.001$ ).

A majority (49/58, 84% vs. 4/23, 17%;  $p<0.001$ ) of H3K27M-mutant BSG patients were of high WHO grade (IV), while 70% (16/23) of wildtype BSG cases were of low WHO grade (I-II).

No difference was observed between H3K27M-mutant and wildtype patients for gender, lesion volume and other histopathological features and conventional MRI characteristics (e.g., hydrocephalus, cystic component, contrast enhancement) (**Table 1**).

For the prospective validation cohort, 30 BSG patients were enrolled. One patient was excluded due to poor image quality, leaving 29 patients including 16 H3K27M-mutant (median age 8 years, female/male=11/5) and 13 wildtype (median aged 23 years, female/male=7/6) cases available for analysis. Demographics, histopathology, WHO grades, and conventional MRI presentations were largely similar to those in the retrospective cohort (**Table 1**).

### **APTw measures in H3K27M-mutant and wildtype patients**

Higher max APTw values (median value=7.60%) were observed in H3K27M-mutant tumors than those with wildtype (median value=6.14%,  $p=0.02$ ), while no significant difference of mean or median APTw values were observed between groups (**Figure 3A**). However, the max APTw values were insufficient to discriminate H3K27M-mutant from wildtype cases even when combined with clinical information (accuracy=0.63) (**Figure 3A**).

### **Prediction of H3K27M status using APTw-derived radiomics**

Based on the APTw-derived radiomic features identified using SVM in the retrospective cohort, the prediction of H3K27M-mutant tumors achieved an accuracy of 0.99 (63/64) in training set and 0.88 (15/17) in test set respectively (**Figure 3B** and **Table 2**). The radiomic features with the most significant contributions in the final classifier included 18 features (GLSZM [n=11], first order [n=4], GLDM [n=1], GLRLM [n=1] and GLCM [n=1]) based on wavelet-transformed images, 2 shape features (elongation and flatness) based on original images, 2 features (NGTDM and GLSZM) based on logarithm-transformed images, 2 features (GLDM and first order) based on square transformed images, 1 feature (GLDM) based on exponentially-transformed images. When combined with clinical information (e.g., gender and age) and conventional MRI presentations (e.g., tumor locations), the prediction ability did not improve further (accuracy of 0.88 in test dataset).

### **Sensitivity analysis**

Additional 9 random separations of the train and test sets demonstrated a comparable prediction ability (accuracy range of 0.71-0.94) with an average accuracy of 0.87. Radiomic features extracted from 2D tumor masks demonstrated a comparable predictive ability with an average accuracy of 0.84. Non-SVM classifiers achieved a comparable ability to predict H3K27M-mutant with accuracies ranging from 0.77 to 0.87.

### **Prospective independent validation**

Validation of the prediction model based in the independent prospective cohort achieved an accuracy of 0.86, sensitivity of 0.88, specificity of 0.85 and AUC of 0.93 (**Figure 3C**).

## Discussion

In this study, we used APTw imaging and derived radiomic features to predict the H3K27M-mutant status among BSG patients. The primary findings were as follows: (1) BSG patients with a H3K27M-mutant tumor presented at a younger age and higher max APTw value than those with wildtype, but the max APTw values showed an insufficient ability (accuracy of 0.63) to predict H3K27M status; (2) APTw-derived radiomic features showed a good ability to predict H3K27M-mutant status with an accuracy of 0.88, sensitivity of 0.92 and specificity of 0.80 in test dataset; validation in an independent prospective cohort confirmed the findings with an accuracy of 0.86, sensitivity of 0.88 and specificity of 0.85.

H3K27M mutations have been widely reported in diffuse midline glioma (DMG) and integrated into 2016 World Health Organization Classification of Tumors of the Central Nervous System [17], especially for pediatric diffuse intrinsic pontine glioma (DIPG) [21, 22]. The younger age in H3K27M-mutant BSG patients in our cohort is consistent with previous reports [6]. Similarly, our findings of H3K27M-mutant status to be associated with higher WHO grade, more frequently locating at pons and presenting with engulfment of basilar artery, and less frequently presenting with dorsal exophytic component, are consistent with previous findings [23-25].

APTw imaging, as a novel MRI technique, provides semi-quantitative amide proton mapping of the brain tumor, which characterizes the heterogeneous metabolism of proteins and peptides, likely reflective of histopathological and genetic alteration in glioma [26-29]. Previous studies on supratentorial gliomas demonstrated that higher APTw values indicated a high level of protein and peptide metabolism and allowed a semi-quantification of cellular proliferation, associated with a higher tumor grade and genetic mutation (e.g., IDH mutation) [9, 26-29]. The higher max APTw values in BSG patients with H3K27M-mutant observed in the current study mostly likely indicate active tumor cellular proliferation associated with an H3K27M mutation, as the global loss of H3K27 methylation results in increased cell proliferation potential [1, 7, 27-30]. In addition, the alkaline intracellular microenvironment may also increase the APT signal [31-33].

The BSG patients with either H3K27M-mutant or wildtype tumors presented with a heterogeneous APTw

intensity distribution, probably reflective of the known inherent pathological heterogeneity of BSG [2, 5]. Therefore, a single APTw value (mean, median or max) was unable to characterize such heterogeneous tumors. Our study shows that APTw-derived radiomic features (e.g., GLSZM based on wavelet-transformed images) were better able to quantitatively capture the heterogeneity of image intensities across multiple image scales and provide more accurate image biomarkers to predict H3K27-mutant status.

The prediction accuracy of a single MRI modality of APTw (acquisition with less than 2 mins) was 0.88 and 0.86 in the retrospective and prospective cohorts respectively, which was comparable or superior to those in previous radiomic studies in diffuse midline glioma and BSG based on conventional multimodal MRI (accuracy<0.85 with FLAIR, T2w, T1w and/or contrast-enhanced T1w) [11, 12]. These findings support the hypothesis that APTw-derived radiomic features provide novel radiological biomarkers to help BSG staging, and might be contribute to the improvement of patient management and prognosis prediction. Its potential clinical value in the accurate diagnosis, treatment decision, prognosis prediction, evaluation of new therapy targets warrants further clinical validation.

Results of sensitivity analysis support the robustness of the SVM-based prediction models we developed. The randomly separation of train and test datasets, different schemes of feature extraction (3D vs. 2D mask), and different classifiers achieved comparable prediction performances compared to the primary results, which indicated our model is robustly reflecting intrinsic metabolic characteristics (presented by APTw) in BSG and predictive of H3K27M mutation status.

The strengthen of this study was the use of single APTw pulse-sequence and its derived radiomics to predict the H3K27M-mutant in BSG patients in both retrospective and prospective cohorts with a large sample size (110 BSG patients), but there are still some limitations in the current study. First, our study was a cross-sectional single center study using APTw images from one MRI scanner. Further studies using data from other scanners in a multicenter setting are required to validate the current findings and confirm their generalizability. Second, overall survival and treatment effect of the BSG patients was incompletely available; further study is needed to evaluate the ability of APTw imaging to predict BSG

prognosis and treatment effect. Lastly, conventional machine learning methods were used, while deep learning (requiring a larger sample) may be able to achieve a better performance.

## **Conclusion**

BSG patients with H3K27M-mutant present with higher max APTw values than those with a wildtype tumor. The APTw-derived radiomic features have a good ability to predict H3K27M-mutant status, providing novel MRI markers for noninvasive evaluation of genetic alterations in BSG.



### **Authors' contributions**

Data management, data processing, statistical analysis and manuscript drafting: Zhizheng Zhuo.

Data management and lesion segmentation: Liying Qu.

Clinical information acquisition and clinical diagnosis: Peng Zhang.

Data acquisition, lesion segmentation and clinical diagnosis: Yunyun Duan.

Lesion segmentation and clinical evaluation: Dan Cheng.

Manuscript editing: Xiaolu Xu.

Manuscript editing: Ting Sun.

MRI data acquisition: Jinli Ding.

MRI data acquisition: Cong Xie.

Histopathological and molecular information: Xing Liu.

Manuscript editing and review: Sven Haller.

Manuscript editing and review: Frederik Barkhof.

Patient recruitment and clinical diagnosis: Liwei Zhang.

Study design, manuscript editing and final approval of this manuscript: Yaou Liu.

### **Funding**

This study was funded by the National Science Foundation of China (Nos. 81870958 and 81571631), the Beijing Municipal Natural Science Foundation for Distinguished Young Scholars (No. JQ20035), the Special Fund of the Pediatric Medical Coordinated Development Center of Beijing Hospitals Authority (No. XTYB201831). FB was supported by the NIHR biomedical research center at UCLH.

### **Declaration of interest**

Zhizheng Zhuo, Liying Qu, Peng Zhang, Yunyun Duan, Dan Cheng, Xiaolu Xu, Ting Sun, Jinli Ding, Cong Xie, Xing Liu, Sven Haller, Frederik Barkhof, Liwei Zhang and Yaou Liu declared no conflicts of interest.

### **Availability of data and material**

The data would be available if the corresponding author received requirement from qualified researchers.

## Reference

1. Larson, J.D., et al., *Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression*. *Cancer Cell*, 2019. **35**(1): p. 140-155 e7.
2. Nagaraja, S., et al., *Histone Variant and Cell Context Determine H3K27M Reprogramming of the Enhancer Landscape and Oncogenic State*. *Mol Cell*, 2019. **76**(6): p. 965-980 e12.
3. Nikbakht, H., et al., *Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma*. *Nat Commun*, 2016. **7**: p. 11185.
4. Pan, C., et al., *Molecular profiling of tumors of the brainstem by sequencing of CSF-derived circulating tumor DNA*. *Acta Neuropathol*, 2019. **137**(2): p. 297-306.
5. Hoffman, L.M., et al., *Spatial genomic heterogeneity in diffuse intrinsic pontine and midline high-grade glioma: implications for diagnostic biopsy and targeted therapeutics*. *Acta Neuropathol Commun*, 2016. **4**: p. 1.
6. Grimm, S.A. and M.C. Chamberlain, *Brainstem glioma: a review*. *Curr Neurol Neurosci Rep*, 2013. **13**(5): p. 346.
7. Yamasaki, F., et al., *Magnetic resonance spectroscopic detection of lactate is predictive of a poor prognosis in patients with diffuse intrinsic pontine glioma*. *Neuro Oncol*, 2011. **13**(7): p. 791-801.
8. Piccardo, A., et al., *Advanced MR imaging and (18)F-DOPA PET characteristics of H3K27M-mutant and wild-type pediatric diffuse midline gliomas*. *Eur J Nucl Med Mol Imaging*, 2019. **46**(8): p. 1685-1694.
9. Smits, M. and M.J. van den Bent, *Imaging Correlates of Adult Glioma Genotypes*. *Radiology*, 2017. **284**(2): p. 316-331.
10. Zhou, J., et al., *Three-dimensional amide proton transfer MR imaging of gliomas: Initial experience and comparison with gadolinium enhancement*. *J Magn Reson Imaging*, 2013. **38**(5): p. 1119-28.
11. Su, X., et al., *Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain*. *Neuro Oncol*, 2020. **22**(3): p. 393-401.
12. Pan, C.C., et al., *A machine learning-based prediction model of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features*. *Radiother Oncol*, 2019. **130**: p. 172-179.
13. Zhang, B., et al., *Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas*. *Neuro Oncol*, 2017. **19**(1): p. 109-117.
14. Sakai, Y., et al., *MRI Radiomic Features to Predict IDH1 Mutation Status in Gliomas: A Machine Learning Approach using Gradient Tree Boosting*. *Int J Mol Sci*, 2020. **21**(21).
15. Togao, O., et al., *Amide proton transfer imaging of brain tumors using a self-corrected 3D fast spin-echo dixon method: Comparison With separate B0 correction*. *Magn Reson Med*, 2017. **77**(6): p. 2272-2279.
16. Zhou, J., et al., *APT-weighted MRI: Techniques, current neuro applications, and challenging issues*. *J Magn Reson Imaging*, 2019. **50**(2): p. 347-364.
17. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary*. *Acta Neuropathol*, 2016. **131**(6): p. 803-20.
18. Solomon, D.A., et al., *Diffuse Midline Gliomas with Histone H3-K27M Mutation: A Series of 47 Cases Assessing the Spectrum of Morphologic Variation and Associated Genetic*

- Alterations*. Brain Pathol, 2016. **26**(5): p. 569-80.
19. Benesty, J., Chen, J., Huang, Y. and Cohen, I., *Pearson Correlation Coefficient*. In: *Noise Reduction in Speech Processing*. Springer, Heidelberg, 2009: p. 37-40.
  20. Kenji Kira, L.A.R., *A practical approach to feature selection Machine Learning*. Proceedings of the Ninth International Workshop on Machine Learning July 1992 p. 249–256.
  21. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas*. Nat Genet, 2012. **44**(3): p. 251-3.
  22. Hoffman, L.M., et al., *Clinical, Radiologic, Pathologic, and Molecular Characteristics of Long-Term Survivors of Diffuse Intrinsic Pontine Glioma (DIPG): A Collaborative Report From the International and European Society for Pediatric Oncology DIPG Registries*. J Clin Oncol, 2018. **36**(19): p. 1963-1972.
  23. Daoud, E.V., et al., *Adult Brainstem Gliomas With H3K27M Mutation: Radiology, Pathology, and Prognosis*. J Neuropathol Exp Neurol, 2018. **77**(4): p. 302-311.
  24. Aboian, M.S., et al., *Imaging Characteristics of Pediatric Diffuse Midline Gliomas with Histone H3 K27M Mutation*. AJNR Am J Neuroradiol, 2017. **38**(4): p. 795-800.
  25. Fisher, P.G., et al., *A clinicopathologic reappraisal of brain stem tumor classification. Identification of pilocystic astrocytoma and fibrillary astrocytoma as distinct entities*. Cancer, 2000. **89**(7): p. 1569-76.
  26. Sotirios, B., et al., *The role of APT imaging in gliomas grading: A systematic review and meta-analysis*. Eur J Radiol, 2020. **133**: p. 109353.
  27. Paech, D., et al., *Assessing the predictability of IDH mutation and MGMT methylation status in glioma patients using relaxation-compensated multipool CEST MRI at 7.0 T*. Neuro Oncol, 2018. **20**(12): p. 1661-1671.
  28. Joo, B., et al., *Amide proton transfer imaging might predict survival and IDH mutation status in high-grade glioma*. Eur Radiol, 2019. **29**(12): p. 6643-6652.
  29. Park, J.E., et al., *Pre- and Posttreatment Glioma: Comparison of Amide Proton Transfer Imaging with MR Spectroscopy for Biomarkers of Tumor Proliferation*. Radiology, 2016. **278**(2): p. 514-23.
  30. Castel, D., et al., *Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes*. Acta Neuropathol, 2015. **130**(6): p. 815-27.
  31. Zhou, J., et al., *Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI*. Nat Med, 2003. **9**(8): p. 1085-90.
  32. Griffiths, J.R., *Are cancer cells acidic?* Br J Cancer, 1991. **64**(3): p. 425-7.
  33. Anderson, G., *Tumour Microenvironment: Roles of the Aryl Hydrocarbon Receptor, O-GlcNAcylation, Acetyl-CoA and Melatonergic Pathway in Regulating Dynamic Metabolic Interactions across Cell Types-Tumour Microenvironment and Metabolism*. Int J Mol Sci, 2020. **22**(1).

## Figure legends

**Figure 1.** Flowchart of the retrospective cohort for train and test a H3K27M-mutant prediction model and the prospective cohort for independent validation in this study.

**Figure 2.** (A) Representative BSG patients with H3K27M-mutant and wildtype and (B) tumor distribution frequency mapping of the enrolled BSG patients in retrospective cohort. Patient 1. a 7 year-old boy diagnosed with H3K27M-mutant and WHO IV; Patient 2. a 53 year-old male adult diagnosed with wildtype and WHO IV; Patient 3. a 37 year-old female adult diagnosed with H3K27M-mutant and WHO IV; Patient 4. a 7 year-old boy diagnosed with wildtype and WHO II; Patient 5. a 11 year-old girl diagnosed with H3K27M-mutant and WHO IV; WHO, World Health Organization; IHS, immunohistochemical staining; A, anterior; R, right.

**Figure 3.** (A) Clinical and APTw measures between BSG patients with H3K27M-mutant and wildtype and their abilities (age or/and max APTw) to identify H3K27M status based on the retrospective cohort (including both train and test datasets), (B) ROC of SVM classification in train and test sets and the corresponding coefficients of the radiomic features in the prediction model based on the retrospective cohort, and (C) ROC of SVM prediction model validation based on the prospective validation cohort. APTw, amide proton transfer weighted; AUC, area under the curve; ROC, Receiver Operating Characteristic; SVM, support vector machine. N, negative coefficient; P, positive coefficient. Full names of the abbreviations in radiomic feature are listed in **supplemental Table S2**.

## Tables

**Table 1.** Demographics, clinical information and MRI features

	Retrospective cohort									Prospective validation cohort		
	Total retrospective dataset			Train dataset			Test dataset					
	H3K27- mutant (n=58)	wildtype (n=23)	p value	H3K27- mutant (n=46)	wildtype (n=18)	p value	H3K27- mutant (n=12)	wildtype (n=5)	p value	H3K27- mutant (n=16)	wildtype (n=13)	p value
<b>Demographics</b>												
Gender (female/male)	31/27	15/8	0.33	24/22	12/6	0.29	7/5	3/2	0.95	11/5	7/6	0.41
Age (median [IQR], years)	9 (7-21)	24 (5-48)	0.009 <sup>a</sup>	9 (7-14)	18 (5-48)	0.014 <sup>a</sup>	14 (7-33)	29 (8-44)	0.38 <sup>b</sup>	8 (6-11)	23 (7-41)	0.026 <sup>b</sup>
<b>Tumor location</b>												
Midbrain	5 (9%)	8 (35%)	0.004	3 (7%)	6 (33%)	0.006	2 (17%)	2 (40%)	0.30	2 (13%)	4 (31%)	0.23
Pons	50 (86%)	8 (35%)	<0.001	41 (89%)	6 (33%)	<0.001	9 (75%)	2 (40%)	0.17	13 (81%)	3 (24%)	0.002
Medulla	3 (5%)	7 (30%)	0.002	2 (4%)	6 (33%)	0.002	1 (8%)	1 (20%)	0.50	1 (6%)	2 (15%)	0.42
<b>Histopathology (WHO Grade)</b>												
I-II	5 (9%)	16 (70%)	<0.001	3 (7%)	13 (72%)	<0.001	2 (17%)	3 (60%)	0.074	3 (19%)	6 (46%)	0.11
III	4 (7%)	3 (13%)	0.37	2 (4%)	3 (17%)	0.087	2 (17%)	0	0.33	2 (13%)	4 (31%)	0.23
IV	49 (84%)	4 (17%)	<0.001	41 (89%)	2 (11%)	<0.001	8 (67%)	2 (40%)	0.31	11 (69%)	3 (23%)	0.014
<b>MRI characteristics</b>												
Dorsal exophytic component	4 (7%)	7 (30%)	0.005	3 (7%)	5 (28%)	0.017	1 (8%)	2 (40%)	0.12	1 (6%)	2 (15%)	0.029
Hydrocephalus	10 (17%)	5 (22%)	0.64	8 (17%)	4 (22%)	0.60	2 (17%)	1 (20%)	0.87	1 (6%)	0 (0%)	0.63
Cystic component	23 (40%)	9 (39%)	0.97	17 (37%)	6 (33%)	0.87	6 (50%)	3 (60%)	0.71	5 (31%)	4 (31%)	0.19

Contrast enhancement	33 (57%)	14 (61%)	0.74	25 (54%)	11 (61%)	0.51	8 (67%)	3 (60%)	0.79	7 (44%)	5 (38%)	0.98
Engulfment of basilar artery	41 (71%)	6 (26%)	<0.001	33 (72%)	5 (28%)	0.003	8 (67%)	1 (20%)	0.079	12 (75%)	1 (8%)	0.18
Lesion volume (median [IQR], ml)	1167 (621-1707)	1005 (660-1792)	0.87 <sup>a</sup>	1167 (578-1745)	1040 (930-1584)	0.85 <sup>a</sup>	1105 (654-1519)	725 (405-1941)	0.88 <sup>b</sup>	1202 (810-1562)	1217 (765-1435)	0.74 <sup>b</sup>
mean APTw (median [IQR], %)	2.71 (2.17-3.37)	2.68 (1.94-3.10)	0.14 <sup>a</sup>	2.71 (2.18-3.19)	2.50 (1.92-3.04)	0.078 <sup>a</sup>	2.80 (2.14-3.44)	3.00 (2.78-3.41)	0.57 <sup>b</sup>	2.45 (1.97-2.93)	2.63 (1.95-2.91)	0.81 <sup>b</sup>
median APTw (median [IQR], %)	2.66 (2.13-3.27)	2.63 (1.79-3.05)	0.17 <sup>a</sup>	2.66 (2.14-3.24)	2.34 (1.78-2.97)	0.090 <sup>a</sup>	2.81 (2.12-3.34)	3.02 (2.74-3.47)	0.65 <sup>b</sup>	2.43 (1.92-2.91)	2.30 (1.81-2.84)	0.95 <sup>b</sup>
max APTw (median [IQR], %)	7.60 (5.18-9.15)	6.14 (4.55-7.44)	0.021 <sup>a</sup>	7.94 (5.52-9.38)	5.81 (4.55-6.66)	0.002 <sup>a</sup>	6.26 (4.61-7.75)	6.63 (5.77-9.37)	0.51 <sup>b</sup>	7.42 (5.84-9.04)	8.03 (6.60-9.95)	0.37 <sup>b</sup>

Note: IQR, interquartile range; WHO, World Health Organization;

<sup>a</sup> indicated a Student's T test;

<sup>b</sup> indicated a Mann-Whitney U test.

**Table 2.** Prediction of H3K27M-mutant status by support vector machine and sensitivity analysis results.

<b>Groups</b>	<b>Accuracy</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>PPV</b>	<b>NPV</b>	<b>AUC</b>
<b>Primary results</b>						
Train dataset (FN=25)	0.92	0.91	0.94	0.98	0.81	0.94
Test dataset (FN=25)	0.88	0.92	0.80	0.92	0.80	0.85
<b>Sensitivity analysis</b>						
Repeat1 (FN=21)	0.82	0.92	0.60	0.85	0.75	0.72
Repeat2 (FN=30)	0.88	0.83	1	1	0.71	0.95
Repeat3 (FN=25)	0.94	1	0.80	0.92	1	0.83
Repeat4 (FN=21)	0.88	0.83	1	1	0.71	0.93
Repeat5 (FN=27)	0.94	1	0.80	0.92	1	0.92
Repeat6 (FN=19)	0.71	0.58	1	1	0.50	0.80
Repeat7 (FN=29)	0.94	1	0.80	0.92	1	0.95
Repeat8 (FN=5)	0.88	0.83	1	1	0.71	0.95
Repeat9 (FN=30)	0.82	0.83	0.80	0.91	0.67	0.80
SVM (averaged)	0.87	0.88	0.86	0.94	0.79	0.87
2D radiomics (averaged)	0.84	0.83	0.86	0.94	0.71	0.83
AB (averaged)	0.82	0.81	0.84	0.92	0.70	0.81
AE (averaged)	0.85	0.84	0.88	0.95	0.73	0.88
LR (averaged)	0.83	0.83	0.84	0.94	0.71	0.83
LR-lasso (averaged)	0.85	0.87	0.80	0.92	0.74	0.84
LDA (averaged)	0.87	0.88	0.84	0.94	0.78	0.86
GP (averaged)	0.79	0.73	0.96	0.98	0.65	0.82
NB (averaged)	0.77	0.75	0.82	0.93	0.69	0.72
DT (averaged)	0.81	0.87	0.66	0.86	0.72	0.76
RF (averaged)	0.81	0.80	0.80	0.93	0.69	0.78

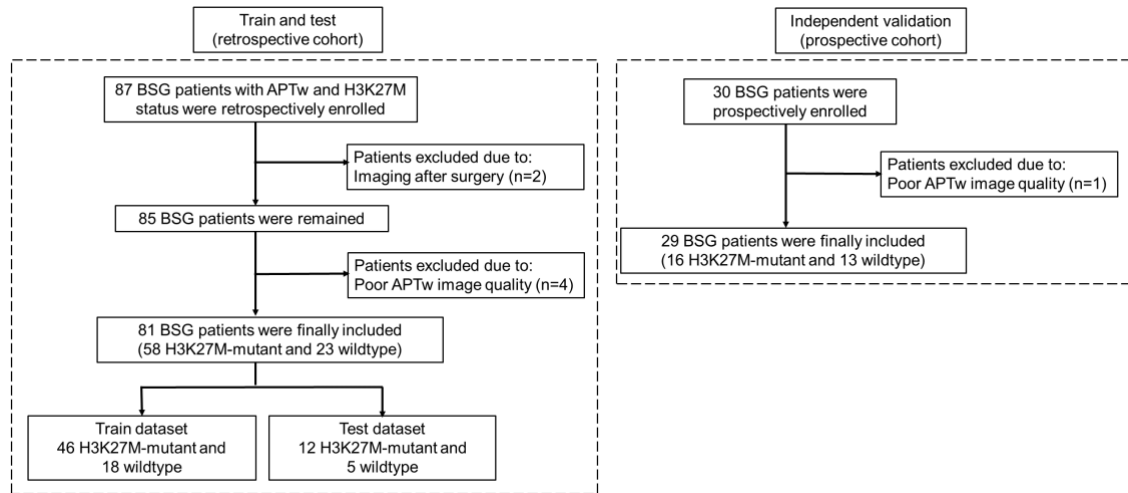
Note: sensitivity analysis results presented here were based on the test data in retrospective cohort.

Averaged means the presentation was the averaged performance of the primary finding and additional 9 repeats on train and test data separation. FN, feature number; PPV, positive predicted value; NPV,

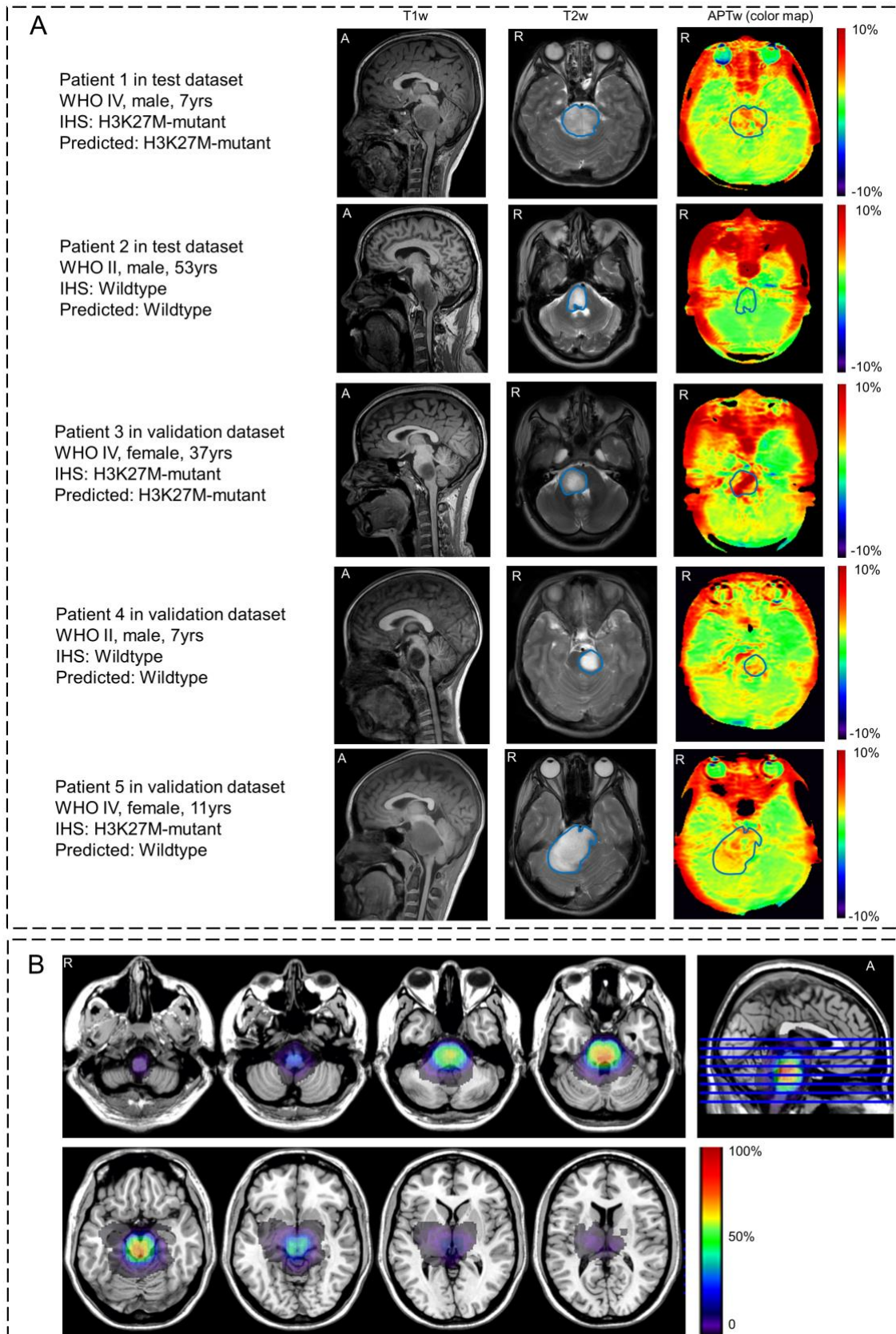
negative predicted value; AUC, area under the curve; AB, adaboost; AE, auto encoder; LR, logistic regression; lasso, Least absolute shrinkage and selection operator regression; LDA, linear discriminant analysis; GP, gaussian process; NB, naïve bayes; DT, decision tree; RF, random forest.



## Figures

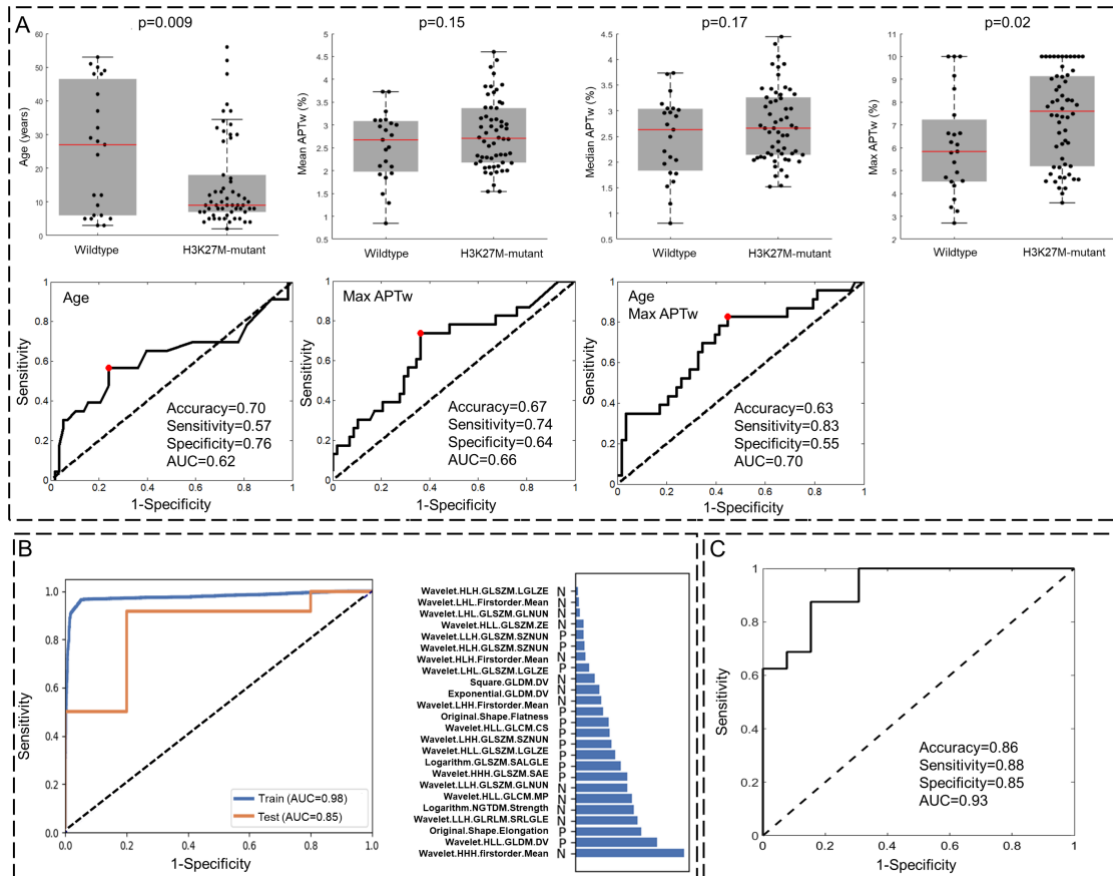


**Figure 1.** Flowchart of the retrospective cohort for train and test a H3K27M-mutant prediction model and the prospective cohort for independent validation in this study.



**Figure 2.** (A) Representative BSG patients with H3K27M-mutant and wildtype and (B) tumor distribution frequency mapping of the enrolled BSG patients in retrospective cohort. Patient 1. a 7 year-

old boy diagnosed with H3K27M-mutant and WHO IV; Patient 2. a 53 year-old male adult diagnosed with wildtype and WHO IV; Patient 3. a 37 year-old female adult diagnosed with H3K27M-mutant and WHO IV; Patient 4. a 7 year-old boy diagnosed with wildtype and WHO II; Patient 5. a 11 year-old girl diagnosed with H3K27M-mutant and WHO IV; WHO, World Health Organization; IHS, immunohistochemical staining; A, anterior; R, right.



**Figure 3.** (A) Clinical and APTw measures between BSG patients with H3K27M-mutant and wildtype and their abilities (age or/and max APTw) to identify H3K27M status based on the retrospective cohort (including both train and test datasets), (B) ROC of SVM classification in train and test sets and the corresponding coefficients of the radiomic features in the prediction model based on the retrospective cohort, and (C) ROC of SVM prediction model validation based on the prospective validation cohort. APTw, amide proton transfer weighted; AUC, area under the curve; ROC, Receiver Operating Characteristic; SVM, support vector machine. N, negative coefficient; P, positive coefficient. Full names of the abbreviations in radiomic feature are listed in **supplemental Table S2**.

## Supplemental materials

**Table S1.** MRI protocols of T1w, T2w, FLAIR and contrast-enhanced T1w.

Protocols	Sequence	FA (degree)	TR (ms)	TE (ms)	IR (ms)	Acquisition voxel size (mm <sup>3</sup> )	In-plane acquisition matrix	Slice number	Acquisition time
T1w (sag)	3D TFE	8	6.6	3		1×1×1	256×256	196	2min20s
T2w (ax)	2D TSE	90	2800	135		1×1×5 (gap 0.5mm)	240×240	24	1min
T2-FLAIR (sag)	3D TSE	90	4800	340	1605	1×1×1	240×240	180	2min28s
DWI (ax)	2D EPI	90	3000	55		1.6×1.6×5 (gap 1mm)	220×220	24	1min02s
Contrast- enhanced T1w (sag)	3D TFE	8	6.6	3		1×1×1	256×256	196	2min20s

Note: DWI, diffusion weighted imaging; FA, flip angle; TR, repetition time; TE, echo time; IR, inversion recovery; FLAIR, fluid attenuated inversion recovery; TFE, turbo field echo; TSE, turbo spin echo; EPI, echo-planar imaging; sag, sagittal; ax, axial.

**Table S2.** Full names of the abbreviations in radiomic features in Figure 3.

Abbreviation	Full name
H	High-pass filter
L	Low-pass filter
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLSZM	Gray Level Size Zone Matrix
NGTDM	Neighboring Gray Tone Difference Matrix
LGLZE	Low Gray Level Zone Emphasis
GLNUN	Gray Level Non-Uniformity Normalized
ZE	Zone Entropy
SZNUN	Size Zone Non-Uniformity Normalized
LGLZE	Low Gray Level Zone Emphasis
DV	Dependence Variance
CS	Cluster Shade
SALGLE	Small Area Low Gray Level Emphasis
SAE	Small Area Emphasis
MP	Maximum Probability
SRLGLE	Short Run Low Gray Level Emphasis