

A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas

Chandan Ganesh Bangalore Yogananda, Bhavya R. Shah,[✉] Maryam Vejdani-Jahromi, Sahil S. Nalawade, Gowtham K. Murugesan, Frank F. Yu, Marco C. Pinho, Benjamin C. Wagner, Bruce Mickey, Toral R. Patel, Baowei Fei, Ananth J. Madhuranthakam, and Joseph A. Maldjian

Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas (C.G.B.Y., B.R.S., M.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W., A.J.M., J.A.M.); Department of Neurological Surgery, University of Texas Southwestern Medical Center, Dallas, Texas (B.M., T.R.P.); Department of Bioengineering, University of Texas at Dallas, Richardson, Texas (B.F.)

Corresponding Author: Joseph A. Maldjian, MD, Department of Radiology, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390–9178 (Joseph.Maldjian@utsouthwestern.edu).

Abstract

Background. Isocitrate dehydrogenase (IDH) mutation status has emerged as an important prognostic marker in gliomas. Currently, reliable IDH mutation determination requires invasive surgical procedures. The purpose of this study was to develop a highly accurate, MRI-based, voxelwise deep-learning IDH classification network using T2-weighted (T2w) MR images and compare its performance to a multicontrast network.

Methods. Multiparametric brain MRI data and corresponding genomic information were obtained for 214 subjects (94 IDH-mutated, 120 IDH wild-type) from The Cancer Imaging Archive and The Cancer Genome Atlas. Two separate networks were developed, including a T2w image-only network (T2-net) and a multicontrast (T2w, fluid attenuated inversion recovery, and T1 postcontrast) network (TS-net) to perform IDH classification and simultaneous single label tumor segmentation. The networks were trained using 3D Dense-UNets. Three-fold cross-validation was performed to generalize the networks' performance. Receiver operating characteristic analysis was also performed. Dice scores were computed to determine tumor segmentation accuracy.

Results. T2-net demonstrated a mean cross-validation accuracy of $97.14\% \pm 0.04$ in predicting IDH mutation status, with a sensitivity of 0.97 ± 0.03 , specificity of 0.98 ± 0.01 , and an area under the curve (AUC) of 0.98 ± 0.01 . TS-net achieved a mean cross-validation accuracy of $97.12\% \pm 0.09$, with a sensitivity of 0.98 ± 0.02 , specificity of 0.97 ± 0.001 , and an AUC of 0.99 ± 0.01 . The mean whole tumor segmentation Dice scores were 0.85 ± 0.009 for T2-net and 0.89 ± 0.006 for TS-net.

Conclusion. We demonstrate high IDH classification accuracy using only T2-weighted MR images. This represents an important milestone toward clinical translation.

Key Points

1. IDH status is an important prognostic marker for gliomas.
2. We developed a non-invasive, MRI based, highly accurate deep-learning method for the determination of IDH status.
3. The deep-learning network utilizes only T2-weighted MR images to predict IDH status, thereby facilitating clinical translation.

Importance of the Study

One of the most important recent discoveries in brain glioma biology has been the identification of IDH mutation status as a marker for therapy and prognosis. The mutated form of the gene confers a better prognosis and treatment response than do gliomas with the non-mutated or wild-type form. Currently, the only reliable way to determine IDH mutation status is to obtain glioma tissue either via an invasive brain biopsy or following open surgical resection. The ability to non-invasively determine IDH mutation status has

significant implications in determining therapy and predicting prognosis. We developed a highly accurate, deep-learning network that utilizes only T2-weighted MR images and outperforms previously published methods. The high IDH classification accuracy of our T2w image-only network (T2-net) marks an important milestone toward clinical translation. Imminent clinical translation is feasible because T2-weighted MR imaging is widely available and routinely performed in the assessment of gliomas.

Isocitrate dehydrogenase (IDH) mutation status has emerged as one of the most important markers for glioma diagnosis and therapy. Gliomas with this mutant enzyme have a better prognosis than tumors of the same grade with wild-type IDH. This observation led the World Health Organization to revise its classification of gliomas in 2016.¹ IDH mutated tumors also have different management and therapeutic approaches than tumors with wild-type mutation status. At the present time, the only way to definitively identify an IDH mutated glioma is to perform immunohistochemistry (IHC) or gene sequencing on a tissue specimen, acquired through biopsy or surgical resection. Because the differences between IDH mutated and IDH wild-type gliomas may have critical treatment implications, there is great interest in attempting to distinguish between these two tumor types prior to surgery. This becomes even more important for brain tumors that are inaccessible for biopsy or resection due to a high risk of severe postoperative complications and impairment.

MR spectroscopy can potentially be used to determine IDH mutation status. The mutant IDH enzyme catalyzes the production of the oncometabolite 2-hydroxyglutarate (2-HG).² MR spectroscopic methods have been developed for identification of 2-HG³⁻⁶ non-invasively in brain tumors. While these methods appear to work well in a research setting, in the busy clinical environment the spectroscopic imaging data are frequently uninterpretable due to artifact, patient motion, poor shimming, small voxel sizes, non-ideal tumor location, or presence of hemorrhage or calcification affecting measurements. Even in the setting of good quality spectra, reliable clinical implementation using 2-HG spectroscopy is further compounded by the recently described high false positive rate of over 20% using this technique in the best hands.⁷

Early determination of IDH mutation status directly impacts treatment decisions. Tumors that appear to be low-grade gliomas but are IDH wild-type are typically treated with early intervention rather than observation. Specific chemotherapeutic interventions are more effective in IDH-mutated gliomas (eg, temozolomide).⁸⁻¹² Additionally, surgical resection of non-enhancing tumor volume (beyond gross total resection of enhancing tumor components) in grades III-IV IDH-mutated tumors has been demonstrated to have a survival benefit.¹³ However, the determination of IDH mutation status continues to be performed using direct tissue sampling. Obtaining tumor-rich tissue samples

for determining IDH status can be a challenge. A report from The Cancer Genome Atlas (TCGA) suggests that only 35% of biopsy samples contain sufficient tumor content for appropriate molecular characterization.¹⁴ The development of a robust non-invasive approach would be beneficial in the care of these patients.

Advances in deep-learning methods are outperforming traditional machine-learning methods in predicting the genetic and molecular biology of tumors based on MRI. For example, Zhang et al used a radiomics approach integrating a support vector machine-based model and multimodal MRI features with an accuracy of 80% for IDH detection.¹⁵ In another study using multimodal MRI, clinical features, and a random forest machine learning algorithm, Zhang et al were able to obtain 86% accuracy in predicting IDH mutation status.¹⁶ In that study, the highest predictive features included age, parametric intensity, texture, and shape features. Recent studies by Chang et al have used deep-learning techniques to non-invasively determine IDH mutation status based on MRI, with accuracies of 94% using the database of The Cancer Imaging Archive (TCIA).¹⁷ Unfortunately, none of these methods are clinically viable, requiring either manual pre-segmentation of the tumor, extensive preprocessing, or multicontrast acquisitions that are frequently affected by patient motion due to the long scan times. Additionally, these existing methodologies use a 2D (slice-wise) classification approach. A known limitation in designing and developing a slice-wise classification model is the *data leakage problem*.^{18,19} Two-dimensional slice-wise models working with cross-sectional imaging data are particularly prone to data leakage because they perform slice randomization across all subjects to generate the training, validation, and testing slices. As a result, adjacent slices from the same subject may be found in the training, validation, or testing data subset. Because adjacent slices often share considerable information, this methodology may artificially boost accuracies by introducing bias into the testing phase. The previously reported studies do not appear to adhere to this caveat, potentially resulting in artificially boosted accuracies.

The purpose of this study is to develop a highly accurate fully automated deep learning IDH classification 3D network using T2-weighted (T2w) images only and compare its performance to a multicontrast 3D network. The use of T2 images only provides strong clinical translation

capability. T2 images are routinely acquired as part of any MRI brain tumor evaluation. These images are robust to motion and can be obtained within 2 minutes. On modern MRI scanners available in most clinical settings, high quality T2w images can be obtained even in the presence of active patient motion using commonly available motion resistant acquisition techniques.²⁰

Materials and Methods

Data and Preprocessing

Multiparametric brain MRI data of glioma patients were obtained from the database of TCIA.²¹ Genomic information was provided from the database of TCGA.²² Only preoperative studies were used. Studies were screened for the availability of IDH status and T2w, T2w–fluid attenuated inversion recovery (FLAIR), and contrast enhanced T1-weighted (T1c) image series. The final dataset included 214 subjects (94 IDH-mutated, 120 IDH wild-type). TCGA subject IDs, IDH mutation status, 1p/19q codeletion status, histology, and clinical variables including age, gender, survival months, and Karnofsky performance scores are listed in [Supplementary Table 1](#). The average age of the cohort was 52 ± 15 years, with 48% female subjects. Histologically, 49% of tumors were glioblastomas, 22% were oligodendrogliomas, 15% were astrocytomas, and 14% were oligoastrocytomas, with 48% of tumors grade IV, 27% grade III, and 24% grade II. In this cohort 56% of tumors were IDH wild-type, 42% were IDH1 mutant, and 1.9% were IDH2 mutant. Since the vast majority of IDH mutations were IDH1, both IDH1 and IDH2 mutants were considered as one group; 86% of the IDH mutated cases did not have 1p/19q codeletions, while 14% did. IDH mutation status provided in TCGA was determined using Sanger sequenced DNA methods and exome sequencing of whole genome amplified DNA. The Sanger method is considered the gold standard in genetic analysis.^{23,24}

Tumor masks for 87 subjects were available through previous expert segmentation.^{25,26} Tumor masks for the remaining 127 subjects were manually drawn and validated by in-house neuroradiologists. The tumor masks were used as the ground truth for the tumor segmentation in the training step. Ground truth whole tumor masks for IDH mutated type were labeled 1 and the ground truth tumor masks for IDH wild-type were labeled 2 ([Fig. 1](#)). Data preprocessing was minimal, including (i) N4BiasCorrection to remove radiofrequency inhomogeneity, (ii) co-registration of multicontrast data to the T1c (for TS-net only), and (iii) intensity normalization to zero-mean and unit variance.²⁷ Preprocessing was developed using the Advanced Normalization Tools (ANTs) software routines²⁸ and took less than 5 minutes per dataset.

Network Details

Two separate networks were developed. These were a T2w image-only network (T2-net) trained only on the T2w images ([Fig. 2A](#)) and a 3-sequence network (TS-net) trained on multicontrast MR data including T2w images,

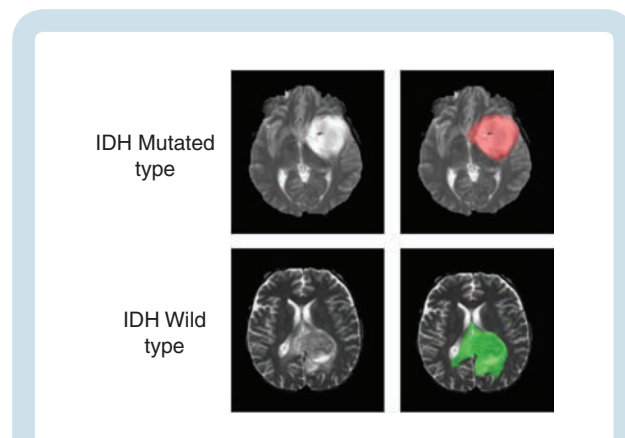


Fig. 1 Ground truth whole tumor masks. Red voxels represent IDH mutated (value of 1) and green voxels represent IDH wild-type (value of 2). The ground truth labels have the same mutation status for all voxels in each tumor.

T2w-FLAIR, and T1c. A 3D $32 \times 32 \times 32$ patch-based training and testing approach was implemented for both networks. Dense-UNets were designed and trained for a voxelwise dual-class segmentation of the whole tumor with classes 1 and 2 representing IDH mutated and IDH wild-type, respectively. The schematics for the network architecture are shown in [Fig. 2B](#). Each network consisted of 7 dense blocks: 3 transition down blocks, 3 transition up blocks, an initial convolution layer, and a final convolution layer followed by an activation layer at the end. Each dense block was made up of 5 layers. Each layer was connected to every other layer in that particular dense block. This dense connection was implemented by concatenating the feature maps from one layer with feature maps from every other layer of that dense block. The input to a dense block was also concatenated with the output of that dense block. Every dense block on the encoder part of the network was followed by a transition down block, while every dense block on the decoder part of the network was preceded by a transition up block. The bottleneck block was used to keep the convolution layers to a smaller number in order to avoid having large convolution layers. With these connecting patterns, all feature maps were reused such that every layer in the architecture received a direct supervision signal.²⁹ A detailed description of the network is given in the [Supplementary Material](#).

Network Implementation and Cross-Validation

To generalize the reliability of the networks, a 3-fold cross-validation was performed on the 214 subjects by randomly shuffling the dataset and distributing it into 3 groups (approximately 70 subjects for each group). During each fold of the cross-validation procedure, the 3 groups are alternated between training, in-training validation, and held-out testing. Group 1 had 72 subjects (32 IDH mutated, 40 IDH wild-type), Group 2 had 71 subjects (31 IDH mutated, 40 IDH wild-type), and Group 3 had 71 subjects (31 IDH mutated, 40 IDH wild-type). The in-training validation set helps improve network performance during

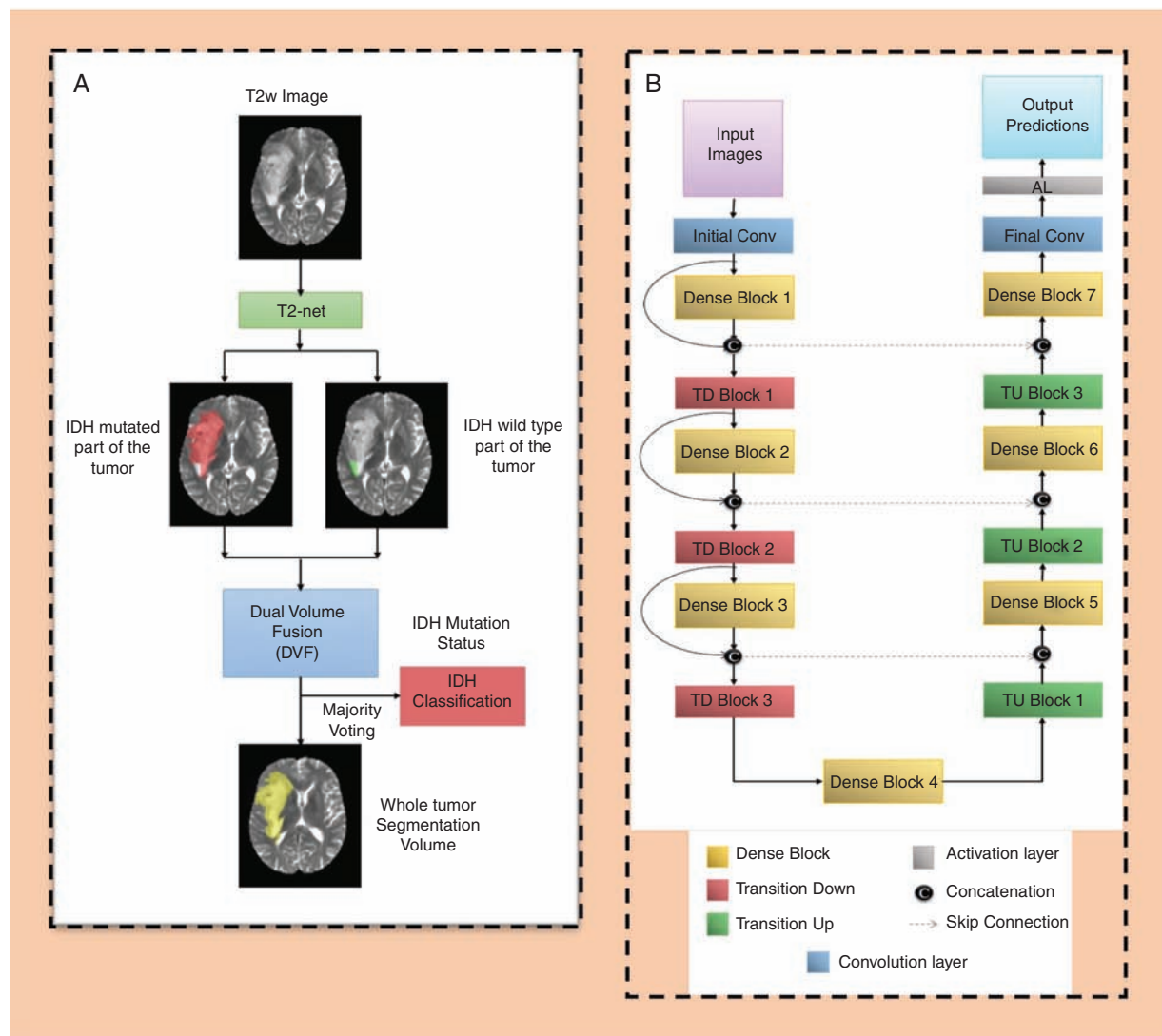


Fig. 2 (A) T2-net overview. Voxelwise classification of IDH mutation status is performed to create 2 volumes (IDH mutated and IDH wild-type). Volumes are combined using dual volume fusion to eliminate false positives and generate a tumor segmentation volume. Majority voting across voxels is used to determine the overall IDH mutation status. (B) Network architecture for T2-net and TS-net. 3D Dense-UNets were employed with 7 dense blocks, 3 transition down blocks, and 3 transition up blocks.

training. Note that each fold of the cross-validation procedure represents a new training phase on a unique combination of the 3 groups. Network performance is only reported, however, on the held-out testing group for each fold (which is never seen by the algorithm during training for that fold). [Supplementary Table 1](#) lists the group membership for each fold of the cross-validation. The in-training validation dataset is used by the algorithm to test performance after each round of training and update model parameters. It is not a true held-out dataset because the algorithm adjusts its performance based on the results in each round from the in-training validation dataset. Once the algorithm has completed all rounds of training, it is evaluated on the true held-out dataset to determine performance.

Seventy-five percent overlapping patches were extracted from the training and in-training validation subjects. To avoid the data leakage problem, no patch from the same subject was mixed with the training, in-training validation, or testing dataset.^{18,19} The data augmentation steps included horizontal flipping, vertical flipping, and random and translational rotation. Data augmentation provided a total of approximately 150 000 patches for training and 150 000 patches for in-training validation. Networks were implemented using Keras³⁰ and Tensorflow³¹ with an adaptive moment estimation optimizer (Adam).³² The initial learning rate was set to 10^{-5} with a batch size of 4 and maximal iterations of 100. Initial parameters were chosen based on previous work with Dense-UNets using brain imaging data and semantic segmentation.^{29,33}

Each network yields 2 segmentation volumes. Volume 1 provides the voxelwise prediction of IDH mutated tumor, and volume 2 identifies the predicted IDH wild-type tumor voxels. A straightforward dual-volume fusion (DVF) approach was developed to combine the 2 segmentation volumes. Both the volumes were combined and the largest connected component was obtained using a 3D connected component algorithm in MatLab. The combined volumes provided a single tumor segmentation map. Majority voting over the voxelwise classes of IDH mutated or IDH wild-type provided a single IDH classification for each subject. Networks were implemented on Tesla P100, P40, and K80 NVIDIA graphic processing units (GPUs). The IDH classification process developed is fully automated, and a tumor segmentation map is a natural output of the voxelwise classification approach.

Statistical Analysis

Statistical analysis was performed in MatLab and R for T2-net and TS-net separately. The accuracy of the 2 networks was evaluated with majority voting (ie, voxelwise cutoff of 50%). This threshold was then used to calculate the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the model for each fold of the cross-validation procedure. A receiver operating characteristic (ROC) curve was also calculated for each fold. A detailed description of the ROC methodology is provided in the [Supplementary Material](#). To evaluate the performance of the networks for tumor segmentation, the Dice score was used. The Dice score determines the amount of spatial overlap between the ground truth segmentation and the network segmentation.

Results

T2-Net

T2-net achieved a mean cross-validation testing accuracy of 97.14% across the 3 folds (97.18%, 97.14%, and 97.10%, SD = 0.04). Mean cross-validation sensitivity, specificity, PPV, NPV, and AUC for T2-net were 0.97 ± 0.03 , 0.98 ± 0.01 , 0.98 ± 0.01 , 0.97 ± 0.01 , and 0.98 ± 0.01 , respectively. The mean cross-validation Dice score for tumor segmentation was 0.85 ± 0.009 . T2-net misclassified 2 cases for each fold (6 total out of 214 subjects). Three subjects were misclassified as IDH mutated, and 3 as IDH wild-type.

Multicontrast TS-Net

The multicontrast TS-net achieved a mean cross-validation testing accuracy of 97.12% across the 3 folds (97.22%, 97.10%, and 97.05%, SD = 0.09). Mean cross-validation sensitivity, specificity, PPV, NPV, and AUC for TS-net were 0.98 ± 0.02 , 0.97 ± 0.001 , 0.97 ± 0.002 , 0.97 ± 0.001 , and 0.99 ± 0.01 , respectively. The mean cross-validation Dice score for tumor segmentation was 0.89 ± 0.006 . TS-net also misclassified 2 cases for each fold (6 total out of 214 subjects). Three subjects were misclassified as IDH

mutated, and 3 as IDH wild-type. The misclassified subjects were not the same as those misclassified by T2-net. Classification accuracies and Dice scores for T2-net and TS-net are presented in [Table 1](#).

ROC Analysis

The ROC curves for each cross-validation fold for T2-net and TS-net are provided in [Fig. 3](#). T2-net and TS-net demonstrated near identical performance curves with extremely high sensitivities and specificities.

Voxelwise Classification

Since these networks are voxelwise classifiers, they perform a simultaneous tumor segmentation. [Figures 4A and B](#) show examples of the voxelwise classification for an IDH wild-type and IDH mutated case using T2-net. The DVF procedure was effective in removing false positives to increase accuracy. The DVF procedure removed approximately 16% and 17% of the classified voxels for T2-net and TS-net, respectively. This procedure improved the Dice scores by approximately 3% for each network. We also computed the voxelwise accuracy for each network. The performance on the IDH wild-type subjects was very similar between the 2 networks, while for IDH mutated the voxelwise accuracies were better for TS-net. For T2-net, the mean voxelwise accuracies were $84.9\% \pm 0.05$ for IDH wild-type and $76.4\% \pm 0.03$ for IDH mutated. For TS-net, the mean voxelwise accuracies were $85.7\% \pm 0.04$ and $84.7\% \pm 0.01$ for IDH wild-type and IDH mutated, respectively.

Training and Segmentation Times

Each network took approximately 2 weeks to train. The trained networks took approximately 3 minutes to segment the whole tumor, implement DVF and predict the IDH mutation status for each subject.

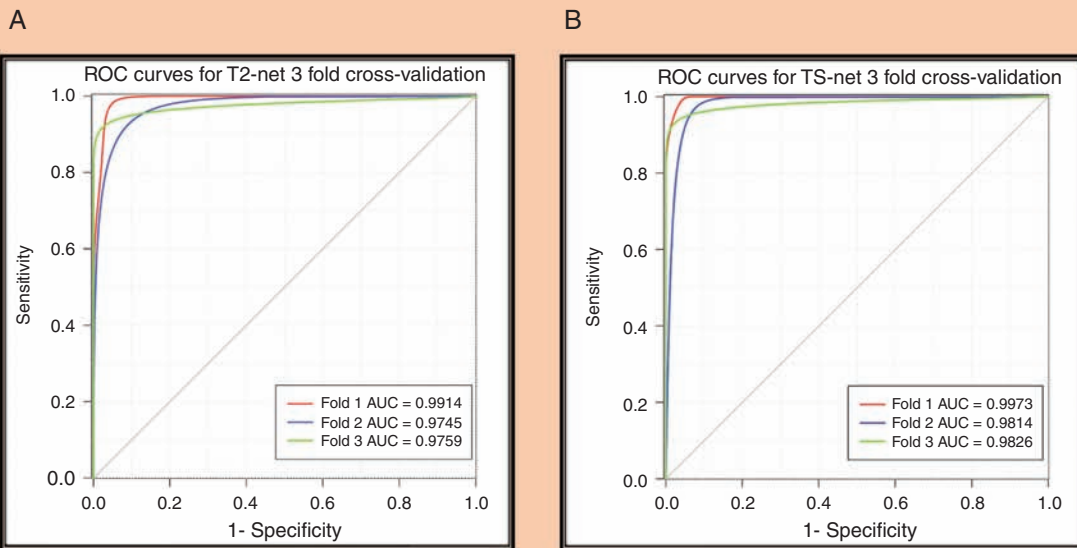
Discussion

We developed 2 deep-learning MRI networks for IDH classification of gliomas based on imaging features alone. Both our T2-net and the multicontrast network outperformed IDH classification algorithms previously reported in the literature.^{15,17,34,35} When comparing the T2-net with the multicontrast network, our results suggest that similar performance can be achieved by using T2w images only. The ability to use only T2w images makes clinical translation much more straightforward and less prone to failures from image acquisition artifacts. The preprocessing used to prepare the data is also minimal. The time required for T2-net to segment the whole tumor, implement DVF, and predict the IDH mutation status for one subject is approximately 3 minutes on a K80 or P40 NVIDIA-GPU.

There are several factors that may explain the higher performance achieved by our networks compared with previously published results. First and foremost is the

Table 1 T2-net and TS-net cross-validation results

Fold Description	T2-net		TS-net	
	% Accuracy	Dice score	% Accuracy	Dice Score
Fold 1	97.18	0.843	97.22	0.88
Fold 2	97.14	0.86	97.10	0.883
Fold 3	97.10	0.857	97.05	0.892
Average	97.14 ± 0.04	0.853 ± 0.009	97.12 ± 0.09	0.885 ± 0.006

**Fig. 3** (A) ROC analysis for T2-net. (B) ROC analysis for TS-net. Separate curves are plotted for each cross-validation fold along with corresponding AUC value.

use of 3D networks, compared with previously reported 2D networks. Additionally, the 3D network architecture is advantageous as the dense connections carry information from all the previous layers to the following layers.²⁹ These types of networks are easier to train and can reduce overfitting.³⁶ The DVF postprocessing step also helps in effectively eliminating false positives while increasing the segmentation accuracy by excluding extraneous voxels that are not connected to the tumor. DVF improved the Dice scores by approximately 3% for each network. The 3D networks interpolate between slices to maintain interslice information more accurately. The networks use minimal preprocessing without any requirement for extraction of pre-engineered features from the images or histopathological data.³⁴

The 3D networks used here are voxelwise classifiers, providing a classification for each voxel in the image. This provides a simultaneous single-label tumor segmentation (eg, the sum of voxels classified as IDH mutated and non-mutated provide the tumor label). The cross-validation

single label whole tumor segmentation performance for these networks provided excellent Dice scores of 0.85 and 0.89 for T2-net and the multicontrast TS-net, respectively. These tumor segmentation Dice scores are similar to the top performers from BraTS2017 tumor segmentation challenge.³⁶

Both T2-net and TS-net achieved similar overall subject classification accuracies. This suggests that the information from the T2w images alone can provide a high classification confidence. For IDH wild-type tumors, both networks incorrectly classified 2 subjects per fold. These 6 subjects were not the same between the networks. In reviewing these cases, there were no discriminating imaging features. The majority of these cases had heterogeneous enhancement, with mixed T2 and FLAIR signal, and surrounding edema. Although T2-net and TS-net demonstrated similar performance on subject-wise IDH classification, the voxelwise performance was different between the networks. TS-net demonstrated similar accuracies in predicting IDH wild-type voxels (85.7% vs 84.9%), and

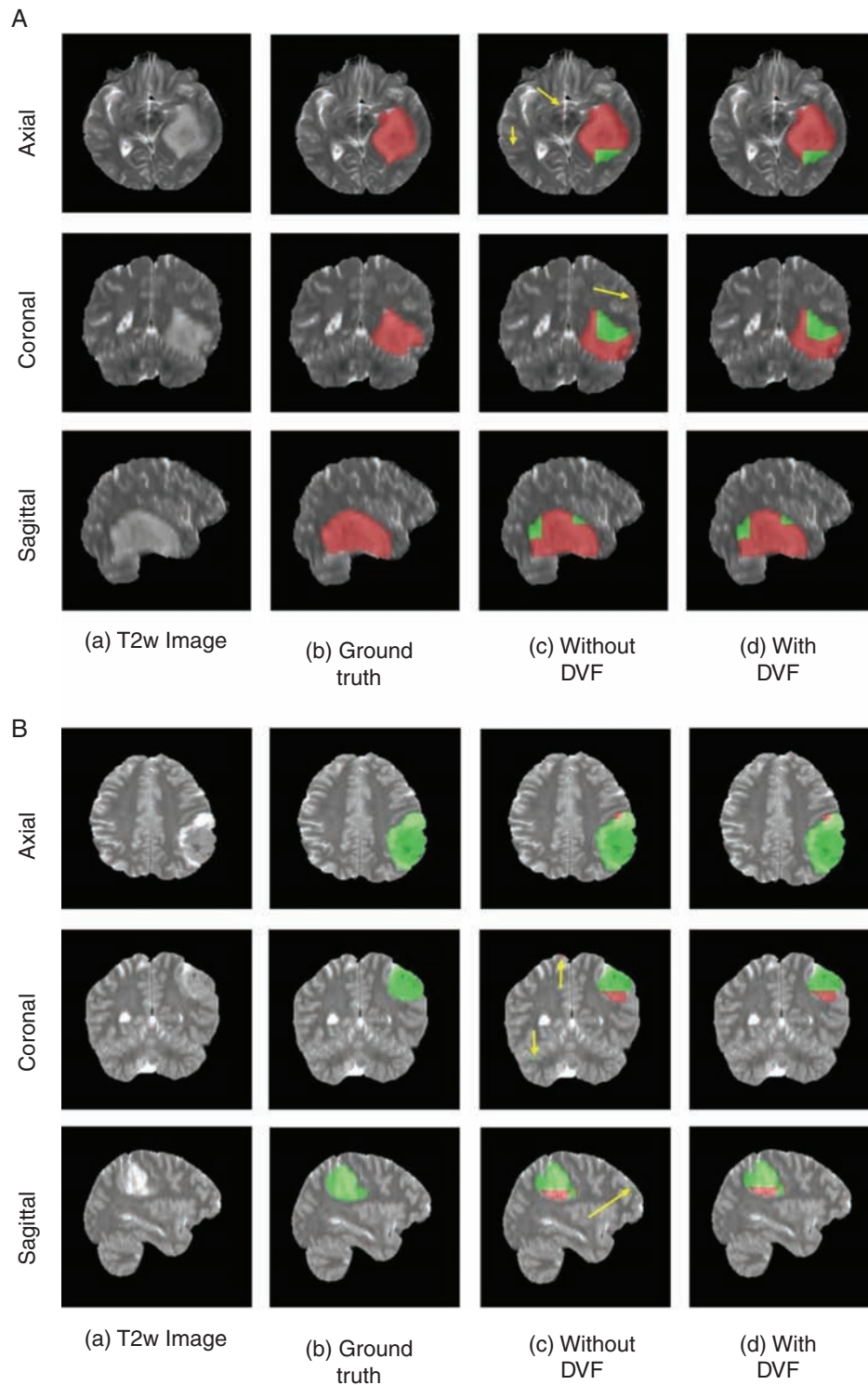


Fig. 4 (A) Example voxelwise segmentation for an IDH mutated tumor. Native T2 image (a). Ground truth segmentation (b). Network output without DVF (c) and after DVF (d). Yellow arrows in (C) indicate false positives. Red voxels correspond to IDH mutated class and green voxels correspond to IDH wild-type. (B) Example voxelwise segmentation for an IDH wild-type tumor. The sharp borders visible between IDH mutated and wild-type result from the patchwise classification approach.

slightly higher accuracies in predicting IDH mutated voxels (84.7% vs 76.4%).

Since these networks are voxelwise classifiers, there are portions within each tumor that are classified as IDH mutated and other areas as IDH wild-type. Heterogeneous genetic expression can occur in gliomas over time and result in varied tumor biology.^{17,37} In the clinical setting, immunohistochemistry (IHC) evaluations are primarily used. IHC uses monoclonal antibodies to detect the most frequent IDH mutations (eg, IDH1-R132H). Different cutoff values have been proposed to determine the IDH status of a tissue sample using IHC methods. While some advocate staining of more than 10% of tumor cells to confer IDH positivity, others suggest that one “strongly” staining tumor cell is sufficient.³⁸ Heterogeneity of staining with IHC has been reported where up to 46% of subjects showed partial uptake.³⁹ In 2011, Perusser et al reported that IDH1-R132H expression may occur in only a fraction of tumor cells.⁴⁰ Heterogeneity of the sample can also affect the sensitivity of genetic testing.⁴¹ IDH heterogeneity and reported false negativity in some gliomas have been explained by monoallelic gene expression, wherein only one allele of a gene is expressed even though both alleles are present. According to Horbinski, sequencing may not always be adequate to identify tumors that are functionally IDH1/2 mutated.^{40,42} Although heterogeneity of IDH status has been reported in histochemical and genomic evaluations of gliomas, we do not make the claim that the deep-learning networks are detecting heterogeneous IDH mutation status in these tumors. Rather, the morphologic expression of IDH mutation status is likely heterogeneous and reflected in the mixed classification outputs of IDH mutated and IDH wild-type within a particular tumor. Regardless, the accuracies using this voxelwise approach well outperform other methods.

Although IHC methods are routinely used in the clinic, several exome sequencing studies have demonstrated that up to 15% of IDH-mutated gliomas remain undetected by traditional IDH1 antibody testing.^{23,24} There are several molecular methods that can be used to determine IDH mutation status from tissue. The current gold standard is the whole genome Sanger DNA sequencing method. This method, however, is limited by the amount of time, cost, and volume of tissue required to perform the genetic analysis. Next-generation sequencing methods such as whole exome sequencing are able to determine mutation status much more rapidly, at decreased cost, and with reduced tissue volumes. However, these methods have false negative rates up to 6% and error rates ~9 times that of whole genome sequencing.⁴³ To further understand the cases that had been misclassified by T2-net, we reviewed the data from these cases in TCGA. There were 3 cases from the cross-fold validation sets that were misclassified by T2-net as IDH mutated. Two of these 3 (TCGA-CS6669, TCGA-020069) demonstrated small tissue volumes obtained during biopsy, limiting molecular characterization. This raises the possibility that the ground truth determination of wild-type for these tumors may have been subject to tissue sampling bias (eg, lack of an appropriate tissue sample, location of sampling).

Another factor that may explain the higher performance achieved by our networks is that previous approaches required multicontrast input, which can be compromised due to patient motion from lengthier examination times

and the need for gadolinium contrast. High-quality T2w images are almost universally acquired during clinical brain tumor diagnostic evaluation. Clinically, T2w images are typically acquired within 2 minutes at the beginning of the exam and are relatively resistant to the effects of patient motion. On modern MRI scanners, high-quality T2w images can even be obtained in the presence of patient motion.²⁰ As such, the ability to use only T2w images is a significant advantage when considering clinical translatability. This method was inspired by a similar approach used for the identification of the status of O⁶-methylguanine-DNA methyltransferase methylation and prediction of 1p/19q chromosomal arm deletion.⁴⁴ Furthermore, our preprocessing steps preserve native image information without the need for any region-of-interest or tumor pre-segmentation procedures. Previous deep-learning algorithms for MRI-based IDH classification used explicit tumor pre-segmentation steps. These were accomplished either by manual delineation of the tumor or by adding a separate deep-learning tumor segmentation network. The use of these pre-segmentation steps adds unnecessary complexity to the classification process, and in the case of manual pre-segmentation, makes them unworkable as a robust automated clinical workflow. Our network uniquely performs a simultaneous tumor segmentation as a natural consequence of the voxelwise segmentation process.

Limitations

Deep-learning studies typically require a very large amount of data to achieve good performance. The number of subjects available from the dataset of TCIA is relatively small compared with the sample sizes typically required for deep learning. Despite this caveat, the data are representative of real-world clinical experience, with multiparametric MR images from multiple institutions, and represent one of the largest publicly available brain tumor databases. Additionally, the acquisition parameters and imaging vendor platforms are diverse across the imaging centers contributing data to TCIA. Although our results show promise for expeditious clinical translation, our algorithm performance will need to be replicated in an independent dataset.

Conclusion

We developed 2 deep-learning MRI networks for IDH classification of gliomas: (i) a T2 network and (ii) a multicontrast network with high accuracy. Both networks outperformed the state-of-the-art algorithms. We also demonstrate similar performance when comparing the T2 network with the multicontrast network. The high accuracy of our network, which utilizes only T2w images, will facilitate imminent clinical translation for this approach.

Supplementary Material

Supplementary data are available at *Neuro-Oncology* online.

Keywords

IDH | deep learning | CNN | MRI | glioma

Funding

Support for this research was provided by NIH/NCI U01CA207091 (A.J.M., J.A.M.).

Acknowledgment

We thank Yin Xi, PhD, statistician for help with the ROC and AUC.

Conflict of interest statement. No conflicts of interest.

Authorship statement. Experimental design: C.G.B.Y., M.V.J., S.S.N., G.K.M., B.C.W., A.J.M., J.A.M. Implementation: C.G.B.Y., M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W., A.J.M., J.A.M. Analysis and interpretation of data: C.G.B.Y., B.R.S., M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.M., T.R.P., B.F., A.J.M., J.A.M. Writing of the manuscript: C.G.B.Y., B.R.S., M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W., B.M., T.R.P., B.F., A.J.M., J.A.M.

References

- 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):803–820.
- Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med.* 2009;360(8):765–773.
- Pope WB, Prins RM, Albert Thomas M, et al. Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. *J Neurooncol.* 2012;107(1):197–205.
- Choi C, Ganji SK, DeBerardinis RJ, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat Med.* 2012;18(4):624–629.
- de la Fuente MI, Young RJ, Rubel J, et al. Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma. *Neuro Oncol.* 2015;18(2):283–290.
- Tietze A, Choi C, Mickey B, et al. Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting. *J Neurosurg.* 2017;128(2):391–398.
- Suh CH, Kim HS, Paik W, et al. False-positive measurement at 2-hydroxyglutarate MR spectroscopy in isocitrate dehydrogenase wild-type glioblastoma: a multifactorial analysis. *Radiology.* 2019;291(3):752–762.
- SongTao Q, Lei Y, Si G, et al. IDH mutations predict longer survival and response to temozolomide in secondary glioblastoma. *Cancer Sci.* 2012;103(2):269–273.
- Okita Y, Narita Y, Miyakita Y, et al. IDH1/2 mutation is a prognostic marker for survival and predicts response to chemotherapy for grade II gliomas concomitantly treated with radiation therapy. *Int J Oncol.* 2012;41(4):1325–1336.
- Mohrenz IV, Antonietti P, Pusch S, et al. Isocitrate dehydrogenase 1 mutant R132H sensitizes glioma cells to BCNU-induced oxidative stress and cell death. *Apoptosis.* 2013;18(11):1416–1425.
- Molenaar RJ, Botman D, Smits MA, et al. Radioprotection of IDH1-mutated cancer cells by the IDH1-mutant inhibitor AGI-5198. *Cancer Res.* 2015;75(22):4790–4802.
- Sulkowski PL, Corso CD, Robinson ND, et al. 2-Hydroxyglutarate produced by neomorphic IDH mutations suppresses homologous recombination and induces PARP inhibitor sensitivity. *Sci Transl Med.* 2017;9(375):eaal2463.
- Beiko J, Suki D, Hess KR, et al. IDH1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro Oncol.* 2014;16(1):81–91.
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008;455(7216):1061–1068.
- Zhang M, Pan Y, Qi X, et al. Identification of new biomarkers associated with idh mutation and prognosis in astrocytic tumors using nanostring nCounter analysis system. *Appl Immunohistochem Mol Morphol.* 2018;26(2):101–107.
- Zhang B, Chang K, Ramkissoon S, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* 2017;19(1):109–117.
- Chang P, Grinband J, Weinberg BD, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am J Neuroradiol.* 2018;39(7):1201–1207.
- Wegmayr V AS, Buhmann J, Petrick N, Mori K, eds. *Classification of brain MRI with big data and deep 3D convolutional neural networks.* SPIE Proceedings, Medical Imaging 2018: Computer-Aided Diagnosis. 2018; 1057501.
- Xinyang Feng JY, Zachary CL, Scott AS, Frank AP. Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification. *bioRxiv.* 2018;2018:456277.
- Nyberg E, Sandhu GS, Jesberger J, et al. Comparison of brain MR images at 1.5T using BLADE and rectilinear techniques for patients who move during data acquisition. *AJNR Am J Neuroradiol.* 2012;33(1):77–82.
- Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* 2013;26(6):1045–1057.
- Ceccarelli M, Barthelemy FP, Malta TM, et al; TCGA Research Network. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell.* 2016;164(3):550–563.
- Cryan JB, Haidar S, Ramkissoon LA, et al. Clinical multiplexed exome sequencing distinguishes adult oligodendroglial neoplasms from astrocytic and mixed lineage gliomas. *Oncotarget.* 2014;5(18):8083–8092.
- Gutman DA, Dunn WD Jr, Grossmann P, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. *Neuroradiology.* 2015;57(12):1227–1237.
- Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging.* 2015;34(10):1993–2024.
- Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data.* 2017;4:170117.

27. Tustison NJ, Cook PA, Klein A, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage*. 2014;99:166–179.
28. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54(3):2033–2044.
29. Jegou S, Drozdal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2017:1175–1183.
30. Chollet F. e. a. Charles PWD. Keras. *GitHub Repository*. 2015.
31. Abadi M, Agarwal A, Barham P. et al. Tensorflow: a system for large-scale machine learning. *OSDI*. 2016:265–284.
32. Kingma DP, Ba JL. Adam: a method for stochastic optimization. *ICLR*. 2015.
33. McKinley R, Meier R, Wiest R. Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation. Paper presented at: International MICCAI Brainlesion Workshop 2018.
34. Delfanti RL, Piccioni DE, Handwerker J, et al. Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: implications for IDH, 1p/19q and ATRX status. *J Neurooncol*. 2017;135(3):601–609.
35. Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clin Cancer Res*. 2018;24(5):1073–1081.
36. Wang GL, Wenqi, Ourselin S, Vercauteren T. Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation. *Front Comput Neurosci*. 2019;13:56–56.
37. Pusch S, Sahn F, Meyer J, Mittelbronn M, Hartmann C, von Deimling A. Glioma IDH1 mutation patterns off the beaten track. *Neuropathol Appl Neurobiol*. 2011;37(4):428–430.
38. Lee D, Suh YL, Kang SY, Park TI, Jeong JY, Kim SH. IDH1 mutations in oligodendroglial tumors: comparative analysis of direct sequencing, pyrosequencing, immunohistochemistry, nested PCR and PNA-mediated clamping PCR. *Brain Pathol*. 2013;23(3):285–293.
39. Agarwal S, Sharma MC, Jha P, et al. Comparative study of IDH1 mutations in gliomas by immunohistochemistry and DNA sequencing. *Neuro Oncol*. 2013;15(6):718–726.
40. Preusser M, Wöhrer A, Sary S, Höftberger R, Streubel B, Hainfellner JA. Value and limitations of immunohistochemistry and gene sequencing for detection of the IDH1-R132H mutation in diffuse glioma biopsy specimens. *J Neuropathol Exp Neurol*. 2011;70(8):715–723.
41. Tanboon J, Williams EA, Louis DN. The diagnostic use of immunohistochemical surrogates for signature molecular genetic alterations in gliomas. *J Neuropathol Exp Neurol*. 2015; 75(1):4–18.
42. Horbinski C. What do we know about IDH1/2 mutations so far, and how do we use it? *Acta Neuropathol*. 2013;125(5):621–636.
43. Wall JD, Tang LF, Zerbe B, et al. Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res*. 2014;24(11):1734–1739.
44. Korfiatis P, Kline TL, Coufalova L, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med Phys*. 2016;43(6):2835–2844.