

<https://doi.org/10.1038/s43856-025-00767-0>

# The radiogenomic and spatiogenomic landscapes of glioblastoma and their relationship to oncogenic drivers

Check for updates

Anahita Fathi Kazerooni <sup>1,2,3</sup>, Hamed Akbari <sup>4</sup>, Xiaoju Hu <sup>5</sup>, Vikas Bommineni <sup>1</sup>, Dimitris Grigoriadis <sup>6</sup>, Erik Toorens <sup>7</sup>, Chiharu Sako <sup>8</sup>, Elizabeth Mamourian <sup>8</sup>, Dominique Ballinger <sup>9</sup>, Robyn Sussman <sup>9</sup>, Ashish Singh <sup>1,8</sup>, Ioannis I. Verginadis <sup>10</sup>, Nadia Dahmane <sup>11</sup>, Constantinos Koumenis <sup>10</sup>, Zev A. Binder <sup>3,12</sup>, Stephen J. Bagley <sup>10,13</sup>, Suyash Mohan <sup>8</sup>, Artemis Hatzigeorgiou <sup>6,14</sup>, Donald M. O'Rourke <sup>3,12</sup>, Tapan Ganguly <sup>7</sup>, Subhajyoti De <sup>5</sup>, Spyridon Bakas <sup>15,16</sup>, MacLean P. Nasrallah <sup>1,9,17</sup> & Christos Davatzikos <sup>1,8,17</sup> ✉

## Abstract

**Background** Glioblastoma is a highly heterogeneous brain tumor, posing challenges for precision therapies and patient stratification in clinical trials. Understanding how genetic mutations influence tumor imaging may improve patient management and treatment outcomes. This study investigates the relationship between imaging features, spatial patterns of tumor location, and genetic alterations in IDH-wildtype glioblastoma, as well as the likely sequence of mutational events.

**Methods** We conducted a retrospective analysis of 357 IDH-wildtype glioblastomas with pre-operative multiparametric MRI and targeted genetic sequencing data. Radiogenomic signatures and spatial distribution maps were generated for key mutations in genes such as *EGFR*, *PTEN*, *TP53*, and *NF1* and their corresponding pathways. Machine and deep learning models were used to identify imaging biomarkers and stratify tumors based on their genetic profiles and molecular heterogeneity.

**Results** Here, we show that glioblastoma mutations produce distinctive imaging signatures, which are more pronounced in tumors with less molecular heterogeneity. These signatures provide insights into how mutations affect tumor characteristics such as neovascularization, cell density, invasion, and vascular leakage. We also found that tumor location and spatial distribution correlate with genetic profiles, revealing associations between tumor regions and specific oncogenic drivers. Additionally, imaging features reflect the cross-sectionally inferred evolutionary trajectories of glioblastomas.

**Conclusions** This study establishes clinically accessible imaging biomarkers that capture the molecular composition and oncogenic drivers of glioblastoma. These findings have potential implications for noninvasive tumor profiling, personalized therapies, and improved patient stratification in clinical trials.

## Plain language summary

Glioblastoma is a type of brain cancer that grows and spreads quickly, making treatment challenging. The changes that lead to cancer and parts of the brain in which it grows are also very variable. We explored how specific genetic changes in glioblastoma influence its appearance on brain scans and how tumor location within the brain relates the genetic changes seen. We applied computational models to the imaging data to identify patterns in where cancers with particular genetic changes are found in the brain. These findings could help doctors predict genetic changes in tumors without the need for invasive procedures, improving patient selection for targeted therapies and clinical trials.

Most cancers harbor diverse tumor subpopulations with variable genetic and epigenetic properties, resulting in spatial and molecular intratumoral heterogeneity. Malignant brain tumor progression and heterogeneity may be driven by the selection of specific cancer stem cells (CSCs) over time, influenced by their genetic, epigenetic and tumor microenvironment (TME)

interactions<sup>1–4</sup>. CSCs may leave the tumor bed, acquire additional mutations, and contribute to tumor progression in distant areas. Within the TME, some CSCs grow and populate in highly vascular, nutrition-sufficient regions, while others adapt to hypoxic, nutrition-depleted environments<sup>5</sup>. This co-existence of diverse subpopulations and their adaptation to

A full list of affiliations appears at the end of the paper. ✉ e-mail: [Christos.Davatzikos@penncmedicine.upenn.edu](mailto:Christos.Davatzikos@penncmedicine.upenn.edu)

environmental forces creates a robust tumor landscape resistant to standard treatments. Consequently, therapies that fail to eliminate tumor stem cells are unlikely to be effective in eradicating the tumor. The diversity across the tumor landscape, as reflected in radiographic phenotypes, allows tumor cells to adapt and influences their clinical behavior and response to therapy.

Glioblastoma (GBM), as defined by the 2021 WHO central nervous system (CNS) tumor classification, is the most common primary malignant brain tumor in adults, with a grim prognosis<sup>6,7</sup>. GBM cells extensively infiltrate brain parenchyma, making surgical cure unachievable even with maximal safe resection and adjuvant chemoradiotherapy<sup>8,9</sup>. Furthermore, GBM comprises a robust evolutionary ecosystem with multiple genetically distinct clonal populations, resulting in intratumor heterogeneity<sup>10,11</sup>. Standard treatment (Stupp protocol<sup>12</sup>), i.e., surgical resection followed by chemoradiotherapy with temozolomide (TMZ), may select resistant subclones from the residual tumor, leading to rapid tumor regrowth and treatment failure<sup>10,11,13</sup>.

Advances in high-throughput sequencing enable the mapping of genomic alterations and driver mutations, providing opportunities for clinical trial enrollment or development of personalized therapies beyond the standard-of-care<sup>14</sup>. However, sequencing is generally performed on a small portion of the tumor, which may not reflect the full tumor's heterogeneous characteristics<sup>15,16</sup>. Multiple mutations can be detected in a single specimen of a sequenced tumor. However, it remains challenging to identify primary driver genes responsible for the global radiographic phenotype, as mutations are unevenly distributed across the tumor. Furthermore, sequencing outcomes depend on the assay type, coverage depth, library comprehensiveness, and analysis techniques, complicating the development of standardized pipelines for multi-institutional clinical trials<sup>17</sup>.

Machine learning (ML)-based radiogenomic analyses of multiparametric MRIs (mpMRI) offer a promising noninvasive approach to explore tumor genomics, spatial heterogeneity, and sampling bias<sup>18,19</sup>. This study investigates how GBM molecular heterogeneity impacts imaging features through radiogenomics, spatiogenomics (relationship between spatial location patterns of tumor occurrence to mutational composition) and the cross-sectionally estimated molecular evolutionary landscapes. We generate radiogenomic signatures from computational analyses of mpMRI scans to predict alterations in driver genes or core signaling pathways using a targeted sequencing panel. These signatures provide insights into TME properties, such as neoangiogenesis, cellularity, hypoxia, and infiltration, serving as surrogate in vivo biomarkers for de novo (primary, i.e., Grade 4 tumor at diagnosis) *IDH*-wildtype GBM. We hypothesize that tumors with simpler genomic landscapes exhibit more distinctive radiogenomic signatures than heterogeneous tumors and that spatial patterns of tumor locations correlate with molecular signatures or sequence of events, driven by cells of origin and local TME. Additionally, we explore imaging phenotypes associated with tumor clonal evolution.

ML analysis of mpMRI scans identifies distinctive in vivo imaging signatures associated with key driver mutations, including *EGFR*, *PTEN*, *NF1*, and *TP53*, and their associated pathways in GBM, with more pronounced differentiation in homogeneous tumors harboring a mutation in one of these key driver genes or pathways. These imaging signatures offer insights into how various mutations influence tumor phenotype as quantified by imaging features. Additionally, the spatial distribution of tumors varies based on their molecular characteristics. Finally, distinct imaging and spatial characteristics are associated with cross-sectionally estimated evolutionary trajectories of the tumors.

## Methods

### Study cohort

This retrospective HIPAA-compliant study was approved by the IRB of the University of Pennsylvania (IRB protocol 706564). All patients provided their informed consent at the time of imaging. A retrospective cohort of  $n = 585$  patients with GBM tumors who had undergone surgery at the Hospital of the University of Pennsylvania (HUP) between 2006 to 2019 was reviewed. The patients who had undergone maximal safe or partial

resection, were histopathologically confirmed with de novo glioblastoma tumors, received radiotherapy, and concomitant and adjuvant chemotherapy with temozolomide, and at the very least had four conventional MRI scans, including pre- and post-gadolinium T1-weighted (T1, and T1-Gd), T2-weighted (T2), and T2 fluid attenuation inversion recovery (T2-FLAIR), acquired pre-operatively on a 3 T MRI scanner (Siemens, Tim Trio, Erlangen, Germany) for a first-occurrence brain tumor were included. A cohort of 374 patients meeting the inclusion criteria was collected. Of these, 244 had additional advanced MRI sequences available, including diffusion tensor imaging (DTI) and dynamic susceptibility contrast (DSC) MRI with preload administration. All measurements described below were conducted on separate samples and patients.

### Next generation sequencing (NGS)

Given that our imaging data and tissue samples were from patients diagnosed between 2006–2019, we reclassified the cohort based on our molecular data. We retrospectively sequenced all tissue samples included in this study using the NGS panels described below. This sequencing ensured that the majority of the included subjects met the definition of GBM according to the 2021 WHO classification of tumors of the central nervous system<sup>20</sup>.

### NGS panel 1

Sequencing was performed for  $n = 183$  subjects using an in-house research NGS panel. A custom AmpliSeq sequencing panel was designed to target the coding sequence of 45 genes and 4 additional mutation loci of interest using the Ion AmpliSeq Designer portal (ThermoFisher, <https://ampliseq.com/>). The design consists of 2130 amplicons covering 206 kbp, with amplicon size range optimized for FFPE DNA (125–175 bp). Barcoded libraries were prepared from nanogram quantities of FFPE gDNA using the Ion AmpliSeq Library Kit 2.0 (ThermoFisher cat. no. 4480442). Libraries were quantified using the Ion Library TaqMan Quantitation Kit (cat. no. 4468802) eliminating the need for amplification of the final library. Library concentrations were normalized and pooled in equimolar ratios for multiplex sequencing on an Ion Torrent S5 sequencing system using the Ion 540 Kit-OT2 and Ion 540 Chip (cat. nos. A27753 and A27766).

Signal processing and base calling, followed by alignment of sequence reads to the hg19 reference genome, were performed on instrument using Torrent Suite Software version 5.8. Library performance was assessed using the Torrent Suite coverage analysis plugin. Aligned reads in BAM format were uploaded to the cloud-based Ion Reporter Software (ThermoFisher, <https://ionreporter.thermofisher.com/>) for variant calling using the version 5.10 default single sample somatic variant calling parameters. Ion Reporter analyses were downloaded to facilitate further processing of the resulting VCF files.

Variant calls were normalized, multiallelic sites split, and NOCALL or reference calls removed using BCFtools. Normalized variants were annotated with Annovar then filtered on variant quality metrics and annotations, again with BCFtools. Variants with variant quality score <30, read depth <100, allele frequency <0.05, number of alternate reads <5, strand bias >0.7, as well as indels called in a homopolymer greater than 8 bp in length, were removed. Also removed were any variants identified as likely artifacts by (1) low population frequency (gnomAD exome or genome frequency <1%) AND (2) presence in more than 2 of 10 HapMap samples obtained from the Coriell Institute and sequenced by the same means. The 10 control samples include the following DNA sample obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: NA12342. Annotation filters retained only exonic and splicing variants with gnomAD exome or genome population frequency less than 1%.

Target genes in this sequencing panel included *TP53*, *PTEN*, *ATRX*, *EGFR*, *VEGF*, *PDGFRA*, *PIK3CA*, *PIK3R1*, *NF1*, *PDL1*, *CTLA4*, *HIF1A*, *MDM4*, *RBI*, *STAT1*, *CD70*, *CIC*, *FUBP1*, *CDK4*, *ACADL*, *TRIM26*, *SMAD1*, *ARID2*, *CDKN1B*, *CREBZF*, *DNMT3A*, *EPHB2*, *ERF11*, *FGFR2*, *GIGYF2*, *KDR*, *KRAS*, *MAP3K1*, *MET*, *NIPBL*, *NOTCH2*, *NRAS*, *NTRK1*, *PTBP1*, *PTPN11*, *SETD2*, *SMARCB1*, *TP63*, *WRN*. The hotspots included *IDH1* p.R132, *IDH2* p.R172, *BRAF* p.V600, *H3F3A* p.K28, *TERT* (chr5:1295151-1295315). *H3F3A* and *TERT* were excluded, the former was

not effectively covered due to presence of pseudogene that was ignored by the variant caller, and the latter had poor amplicon performance.

## NGS panel 2

Genetic data was available for  $n = 191$  patients through a clinical sequencing panel of 153 actionable and prognostic genes for solid tumors, implemented and validated at our institution for clinical assessment of the resected tumors. The panel has a full coverage of all included genes with its Agilent Haloplex design. Full description of this panel and the in-house data processing bioinformatics pipeline can be found in the study carried out by Nasrallah et al.<sup>21</sup>.

## Final mutational data

A total of 27 genes in both panels were included for  $n = 374$  subjects and used for our radiogenomic analysis: *ARID2*, *ATRX*, *BRAF*, *CDKN2A*, *CIC*, *DNMT3A*, *EGFR*, *FGFR2*, *FUBP1*, *IDH1*, *IDH2*, *KDR*, *KRAS*, *MDM4*, *MET*, *NF1*, *NOTCH2*, *NTRK1*, *PDGFRA*, *PIK3CA*, *PIK3R1*, *PTEN*, *PTPN11*, *RBI*, *SETD2*, *SMARCB1*, *TP53*. We excluded patients with mutations in *IDH1* or *IDH2* ( $n = 16$ ). The histology of the tumors from the remaining patients ( $n = 357$ ) exhibited grade 4 features, aligning with the current WHO definition of glioblastoma. Among these 357 patients, 290 were aged 54 or older. Of the remaining 68 cases in our cohort, all but 13 (all *IDH1/2* wildtype) had mutations in one or more of the following genes: *EGFR*, *NF1*, *PTEN*, *RBI*, *PTPN11*, *PIK3CA*, or *PDGFRA*, suggesting a high likelihood of consistency with the current biological concept of glioblastoma.

A cohort of  $n = 357$  patients with *IDH*-wildtype GBM (a summary of demographics, clinical, and genomic characteristics can be found in Supplementary Table 1) was considered for generating the population atlases of spatial locations and generating radiogenomics signatures based on conventional MRI (T1, T1-Gd, T2, and T2-FLAIR) scans, of which a subset of  $n = 228$  patients had advanced imaging scans (DTI and DSC-MRI) along with conventional MRI.

We considered the GBM pathways introduced in the seminal paper by Brennan et al.<sup>22</sup>, including RTK pathway (with *EGFR*, *PDGFRA*, *MET*, *FGFR2* genes), PI3K pathway (including *PIK3CA*, *PIK3R1*, *PTEN* genes), MAPK pathway (*NF1*, *BRAF*), P53 pathway (*TP53*, *MDM4*), RB1 pathway (*CDKN2A*, *RBI*), and ChrMod pathway (*ATRX*).

## Mutational signature analysis

We extracted mutational signatures to explore characteristic molecular underpinnings for the groups of patients with specific spatial predisposition. The analysis was performed in accordance with the method proposed by Alexandrov et al.<sup>23</sup>. De novo global-signature discovery was performed using sigProfilerExtractor<sup>23</sup> with 100 iterations to extract each signature and 96 trinucleotide sequence contexts were included.

## Mutant-allele tumor heterogeneity (MATH) score; a measure of molecular heterogeneity

Mutant-Allele Tumor Heterogeneity (MATH) score utilizes bulk NGS data to provide a quantitative measurement of intratumor heterogeneity<sup>24</sup>. The MATH score is determined by computing the width of the distribution of mutant-allele fractions (MAF) among tumor loci<sup>24</sup>. In the clonal evolution of the tumor, a locus with earlier mutation has higher MAF and therefore MATH score, while later events result in lower MAF and MATH score<sup>24</sup>. We imported the clinical data and NGS sequencing results for 183 GBMs. MATH score for each tumor was calculated from the median absolute deviation (MAD) and the median of its mutant-allele fractions at tumor-specific mutated loci according to the following formula<sup>24</sup>:

$$MATH = 100 \times MAD / median \quad (1)$$

## Tumor evolutionary trajectories

**Analysis of copy number variation.** CNVs were called from the Ion Torrent sequence read alignments (BAM files) using CNVkit (version

0.9.9)<sup>25</sup>. An FFPE CNV baseline was bootstrapped using ten HapMap control samples obtained from Coriell and FFPE tissue derived DNA from 14 lung metastases to the brain, all sequenced with the AmpliSeq panel. The HapMap samples were screened for the absence of CNVs in panel genes in DGV, but an FFPE baseline is desirable to more accurately represent the variability in sequencing coverage expected from the cases. An initial baseline was generated using the HapMap samples, against which the lung metastases were analyzed. The lung metastases were then filtered based on CNV calls to retain 6 samples for which all calls were either single copy gain or loss, and no two samples contained a call in the same gene. These six samples were used to generate an FFPE CNV baseline for analysis of the AmpliSeq cohort.

The case and control sample BAM files were first processed using CNVkit's 'batch' command in 'targeted amplicon sequencing' mode with the '-drop-low-coverage' flag, using the UCSC hg19 refFlat.txt file for gene annotation. CNVkit splits the targeted regions into bins and calculates coverage tables for each sample. A pooled copy number reference is generated from the control samples' coverage tables then used to correct case sample coverage for regional biases and calculate normalized copy number ratios.

CNVkit's 'segment' command uses the 'circular binary' method to generate copy number segments from bins. Because the 45 panel genes are sparsely distributed across 18 chromosomes, this approach resulted in segments combining genes separated by large genomic distances. To avoid this behavior bins were instead segmented per gene using the 'genometrics' command. Segments consisting of less than 5 bins were dropped, and a copy number ratio threshold of 0 was used to prevent copy number calls at this stage. Segments were then annotated with additional summary statistics using the 'segmetrics' command to support downstream QC and filtering.

Using the 'call' command, copy number ratios were adjusted for tumor cellularity and converted to integer copy number values using the 'threshold' method. Copy number ratio thresholds of -1 and 0.58 were selected to correspond to strict copy loss and gain respectively<sup>26,27</sup>. Germline variant allele frequencies were input during the 'call' step resulting in output annotated with the average b-allele frequencies of heterozygous SNVs in each segment. Finally, CNVs were filtered using a custom script to remove calls for which the log<sub>2</sub> ratio 95% confidence interval contains 0, the balanced case. The specific details about the statistics of copy number variations in the patient population are provided in Supplementary Table 2.

The default for CNVkit is to calculate log<sub>2</sub> ratios and copy number with respect to a diploid chrX and haploid chrY, infer sample sex based on chrX and chrY coverage, and double chrX coverage for male samples. Because this panel targets only one gene on chrX and none on chrY, sex inference was not always accurate. Additionally, the 'genometrics' and 'call' commands are not typically used in conjunction, and both attempt to determine sex, which can result in unexpected behavior. To address this all samples were treated as female at the 'genometrics' step while sex was explicitly asserted in the 'call' command.

## Prediction of evolutionary trajectories

The oncoprint function implemented in maftools package<sup>28</sup> was used for visualization of somatic mutations and pathway level alterations in the GBM samples. We performed the evolutionary analysis using the CAPRI algorithm implemented in TRONCO<sup>23</sup> with its default parameters. Both Akaike and Bayesian information criteria were used for regularization to prevent overfitting. The evolutionary trajectories were drawn using sem-supervised approach to improve clarity and interpretability. We used chordDiagram function implemented in circlize package<sup>29</sup> to visualize the co-mutated genes in GBM tumor samples. ggplot2 package<sup>30</sup> was used to visualize the allele frequency for all mutate genes. All analysis was conducted in R 3.6.0 (2019-04-26).

## Radiogenomic signatures

We trained and evaluated the performances of ML and DL classifiers in predicting mutation status, where the presence of a mutation was assigned a

label of 1, and 0 otherwise. Radiogenomic signatures were calculated from the output probabilities of the trained classifiers in predicting the mutation status in each patient.

### Image pre-processing

Pre-processing of multi-parametric MRI (mpMRI) scans was performed using the CaPTk software<sup>31,32</sup> (<https://www.med.upenn.edu/cbica/captk/>) with the following steps: (1) conversion of MRI volumes from DICOM to Neuroimaging Informatics Technology Initiative (NIfTI) format; (2) re-orientation to a reference coordinate system (here, left-posterior-superior (LPS)); (3) co-registration and resampling to an isotropic resolution of 1mm<sup>3</sup> based on a common anatomical SRI24 atlas<sup>33</sup>, performed using the Greedy software<sup>34</sup>; (4) skull-stripping of the MRI volumes using a pretrained DeepMedic algorithm<sup>35</sup> available through CaPTk, followed by manual revisions where necessary.

Automatic segmentation of tumorous subregions, i.e., enhancing tumor (ET), nonenhancing tumor core (NC), and peritumoral edema (ED), was applied using DeepMedic followed by manual revisions where necessary. An additional segmentation label, namely the whole tumor, was produced by taking the union of all three subregions (WT = ET + NC + ED). Finally, the image intensities were rescaled in the intensity range of 0–255 after removing outliers.

For radiogenomics signatures, in addition to the structural images described above, quantification of DTI scans was performed to generate maps of axial diffusivity (DTI-AD), trace (DTI-TR), radial diffusivity (DTI-RD), and fractional anisotropy (DTI-FA), and DSC-MRI perfusion scans to create maps of peak height (DSC-PH), percentage of signal recovery (DSC-PSR) after considering leakage correction<sup>36,37</sup>, and an automatically extracted proxy to relative cerebral blood volume (DSC-rCBV). For DSC-MRI, principal component analysis (PCA) was performed to reduce the dimensionality of perfusion time series to 5 components that represent the temporal dynamics of tissue perfusion. Quantification of perfusion PCA was performed using CaPTk's Perfusion PCA Parameter Extractor ([https://cbica.github.io/CaPTk/PCA\\_Extraction.html](https://cbica.github.io/CaPTk/PCA_Extraction.html))<sup>38</sup>. As a result, voxel-wise parametric maps of five PCAs (DSC-PC1, DSC-PC2, DSC-PC3, DSC-PC4, and DSC-PC5) were generated.

### Atlases of spatial location

We applied the Glioma Image Segmentation and Registration (GLISTR) software<sup>39</sup> to generate population atlases of spatial distributions. GLISTR simultaneously co-registers the structural MRI scans (T1, T1-Gd, T2, T2-FLAIR images) of each patient to the reference atlas, and segments the brain tissue, including tumorous regions of peritumoral edema, enhancing tumor and non-enhancing necrosis core. As a result of applying this method to patient scans, the images and segmentations are warped into a common atlas space, accounting for the mass effect of the tumor.

### Radiomic feature extraction

Radiomic feature extraction was performed using the CaPTk software<sup>31,32</sup>, which offers functionalities to extract a wide array of radiomic features compliant with the Image Biomarker Standardization Initiative (IBSI)<sup>40</sup> (see: [https://cbica.github.io/CaPTk/tr\\_FeatureExtraction.html](https://cbica.github.io/CaPTk/tr_FeatureExtraction.html)). The masks of tumor segmentations, i.e., ET, NC, ED, and WT were overlaid on each of the MRI scans or the DTI or DSC-derived parametric maps (T1, T1-Gd, T2, T2-FLAIR, DTI-AD, DTI-FA, DTI-RD, DTI-ADC, DSC-rCBV, DSC-PH, DSC-PSR, DSC-PC1 to PC5) to extract radiomic features of shape, volume, intensity, first-order histogram, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighborhood gray tone difference matrix (NGTDM), local binary pattern (LBP), and Collage features. The extracted radiomic features were normalized using z-scoring. Furthermore, for each patient,  $n = 9$  loci features were generated by overlaying the tumor core (ET + NC) on nine brain regions, i.e., frontal, temporal, parietal, and occipital lobes, basal ganglia, insula, corpus callosum, cerebellum, and brain stem, and calculating the average probability of the tumor core belonging to each of these regions.

For  $n = 228$  subjects that had all the conventional and advanced MRI scans, a total of  $n = 5,569$  features, and for  $n = 357$  subjects that only had conventional MRI scans,  $n = 2,477$  features were extracted from multi-parametric MRI scans.

Pairwise correlation was performed on the extracted radiomic features using Pearson's method and in highly correlated pair of features ( $r > 0.85$ ), one of the features was eliminated. The following model training methods were carried out in the discovery cohorts for each of the classification scenarios, and the trained model was independently tested on the replication cohorts.

### Deep learning-based radiogenomic signatures

In the discovery cohort, the remaining radiomic features were ranked using recursive feature elimination with random forests and the optimal number of features was determined through nested cross-validation (10 folds for the outer and 5 folds for the inner folds). The selected feature subset was imported into the CNN classifier, as described below, for prediction of mutation status in the key driver genes, i.e., *EGFR*, *NF1*, *PTEN*, and *TP53*, and pathways, i.e., *RTK* and *PI3K*.

In the discovery cohort, we trained a CNN classifier based on a modified 34-layer ResNet architecture<sup>41</sup>, with an initial  $7 \times 7$  convolution and four layers (layers 1 to 4, composed of 3, 4, 6, and 3 residual blocks respectively; each residual block comprises two  $3 \times 3$  convolution)<sup>42</sup>. Slices of two structural images, i.e., T1-Gd and T2, along with slices of tumor segmentations (overall three image volumes) were passed as inputs for training the ResNet model. Specifically, the structural images and segmentation masks were cropped with margins of 8 pixels from the boundaries of brain tissue and resampled to  $128 \times 128 \times 128$  voxels. Intensity normalization of the brain tissue was performed using z-scoring for all structural MRI scans. Five axial slices were identified from the tumor segmentations, where the axial slice in the center included the maximum tumor mask area; 2 upper and 2 lower slices were also identified along with the center slice. After concatenating all three volumes together, the five slices were finally extracted from this concatenated volume. The discovery cohort was split into 70–30% training-validation sets. The ResNet model was pretrained with input size of  $5 \times 3 \times 128 \times 128 \times 128$  for each subject in the training set (Supplementary Fig 1). After pretraining the ResNet, the weights from the initial  $7 \times 7$  convolution to layer 3 were fed to the CNN classifier and were not used in any further training process. The radiomic features (including spatial location features) were imported as inputs to the fully connected layers that followed layer 4, and along with the weights from layer 4 were used to train the CNN classifier. Input images were further augmented using rotation. Optimization was performed using Adam optimizer (initial learning rate ( $x_0$ ) = 0.0001), with L2 penalty of 0.01, the cross-entropy cost function, and an exponential learning rate decay ( $x_{epoch} = x_0 * 0.99^{epoch}$ ). For both pre-training and training phases, the maximum training epoch was set to 75, and the model was saved when minimum validation loss was achieved.

### Machine learning-based radiogenomic signatures

We also trained classical machine learning models, using support vector machines (SVM), with radiomic features for prediction of mutations in each of the key driver genes and pathways. For the patients in the discovery cohort, feature selection and classifier training were performed using Least Absolute Shrinkage and Selection Operator (LASSO) feature selection approach wrapped with SVM with a linear kernel. This method was applied in a nested cross-validation (nested-CV) schema with 5-fold CV in the inner loop for feature subset selection and with 10-fold CV in the outer loop for model optimization and hyperparameter selection to avoid data overfitting and ensure generalizability to the unseen replication cohort.

### Classification scenarios

Radiogenomic signatures of co-occurring mutations in key driver genes, i.e., *EGFR*, *PTEN*, *TP53*, *NF1*, and key pathways, i.e., *RTK*, *PI3K*, *P53*, and *MAPK* were generated based on (1) radiomic features extracted from multi-parametric conventional MRI scans using deep learning and SVM

**Table 1 | Performance of radiogenomic classifiers in the test set of the discovery cohort**

	Deep (conventional MRI)		SVM (conventional MRI)		SVM (conventional and advanced MRI)	
	AUC (95% CI)	Balanced accuracy (%)	AUC (95% CI)	Balanced accuracy (%)	AUC (95% CI)	Balanced accuracy (%)
<i>EGFR</i>	0.75 (0.65 – 0.79)	74.7	0.68 (0.43 – 0.90)	68.1	0.78 (0.45 – 1)	83.3
<i>PTEN</i>	0.72 (0.67 – 0.76)	71.1	0.73 (0.46 – 1)	75.0	0.79 (0.57 – 1)	82.6
<i>TP53</i>	0.75 (0.65 – 0.80)	75.1	0.62 (0.38 – 0.87)	68.6	0.76 (0.46 – 0.97)	76.9
<i>NF1</i>	0.75 (0.65 – 0.81)	76.9	0.69 (0.33 – 1)	77.1	0.80 (0.41 – 1)	83.3
<i>RTK</i>	0.73 (0.68 – 0.77)	70.3	0.71 (0.51 – 0.92)	73.8	0.75 (0.40 – 1)	76.2
<i>PI3K</i>	0.71 (0.67 – 0.76)	70.6	0.69 (0.49 – 0.89)	72.4	0.82 (0.61 – 1)	82.6
<i>P53</i>	0.73 (0.68 – 0.77)	71.4	0.66 (0.44 – 0.81)	70.2	0.78 (0.51 – 0.95)	79.2
<i>MAPK</i>	0.76 (0.70 – 0.82)	69.5	0.72 (0.48 – 0.98)	70.0	0.81 (0.49 – 1)	84.4

The names of the genes and pathways are indicated in italic font.

The performance metrics are reported as the area under the receiver operating characteristics (ROC) curve (AUC) and balanced accuracy. These metrics were calculated for key driver genes and pathways. The deep learning model utilized conventional MRI data, while the SVM classifier was trained on either conventional MRI alone or a combination of conventional and advanced MRI.

classification methods, and (2) the features calculated from multi-parametric conventional and advanced MRI scans using SVM classification.

### Histograms of imaging features

Histograms of the most predictive features were generated for the mutant and wildtype tumors for each of classification scenarios, to gain a deeper understanding about the underlying biological processes that follow certain genetic mutations. Histograms indicate the frequency (y-axis) of a feature value (x-axis) based on the data from all patients.

### Statistics and reproducibility

In all DL and ML classification scenarios, the data was randomly split using a stratified approach to ensure the proportion of mutated to wildtype tumors was preserved in both the discovery and replication cohorts. The split assigned 75% of the data to the discovery cohort and 25% to an independent (unseen) replication cohort. For the DL and ML models utilizing conventional MRI, the discovery and replication cohorts included 268 and 89 patients, respectively. For the ML model incorporating both conventional and advanced MRI, the discovery and replication cohorts consisted of 171 and 57 patients, respectively. The performance of our predictive ML and DL models in stratifying mutation status was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC) and balanced accuracy metrics. Pairwise correlation of the extracted radiomic features was assessed using Pearson's correlation coefficient. To analyze the separability of DL and ML signatures, we compared tumors with single alterations (i.e., exclusive tumors) to those with co-occurring multiple alterations using Cohen's D effect size. Additionally, cosine similarity metrics were calculated to compare molecular signatures across tumor groups stratified by spatiogenomic tumor location, thereby quantifying and validating distinct mutational profiles among these groups. Differences in MATH scores across brain regions were analyzed using the Wilcoxon rank-sum test.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Results

### Radiogenomic markers are more distinctive in tumors with single gene/pathway alterations compared to tumors with multiple genes/pathways altered

Radiogenomic signatures were generated using deep learning (DL) and conventional machine learning (ML) models to predict the presence or absence of alterations, which in this analysis refers to point mutations or insertions/deletions, in the key driver genes and their associated pathways (Supplementary Table 1). Throughout this section, for brevity, we will refer to a tumor as mutant/wildtype if it does/does not have alterations in a certain gene or pathway of interest, respectively. For co-occurring radiogenomic

signatures, the tumors are considered as mutant/wildtype for a gene or pathway if they do/do not have alterations in a gene/pathway of interest, respectively, regardless of the presence or absence of alterations in other genes or pathways.

The cohorts were divided into discovery and independent replication sets to validate the reproducibility of the generated signatures in unseen data. A summary of diagnostic performance of our predictive models in terms of area under the receiver operating characteristic (ROC) curve (AUC) and balanced accuracy, is provided in Tables 1 and 2. We found high reproducibility of our radiogenomic signatures in independent replication cohorts. DL and support vector machine (SVM) models based only on features computed from conventional MRI scans resulted in comparable distinctive power for most of co-occurring signatures, while the SVM models with features from conventional and advanced MRI achieved better performances (Table 2). This finding highlights the importance of physiological MRI scans, namely diffusion tensor imaging (DTI) and dynamic susceptibility contrast enhanced (DSC-) perfusion MRI, in characterizing the pathophysiological processes that are derived from molecular alterations in the TME<sup>43,44</sup>.

We postulated that radiogenomic signatures become more distinctive in tumors that have mutations in one of the key driver genes/pathways in their mutational landscape, compared to those with mutations in multiple genes/pathways. To test this hypothesis, the distinction of our generated radiogenomic signatures for single-alteration or exclusive tumors was investigated. For such radiogenomic signatures, in contrast to the multiple-alterations or co-occurring situation, the mutant tumors only had mutations in one gene/pathway of interest, and wildtype tumors had no mutations in any of the genes in the main six pathways (described in the Methods section). The performances of most radiogenomic signatures showed improvement in discrimination of exclusive mutant from wildtype tumors (as indicated in Table 2).

The separability of the signatures for tumors with single alterations, i.e., exclusive tumors, is presented in Fig. 1a–c, as a comparison with the values of the signatures for co-occurring tumors with multiple alterations, along with their Cohen's D effect sizes. These results show larger effect sizes for discrimination of tumors with and without mutations for the exclusive cases, in contrast with the co-occurring tumors for most of the predictive models. This finding suggests that the imaging signatures become clearer and can differentiate the mutant and wildtype tumors better than the tumors with mixed mutation compositions. However, even in the presence of co-occurring mutations, the imaging signatures carry fairly strong distinctive power, as reflected by the effect sizes.

In the histograms presented in Supplementary Fig 2, the imaging characteristics that are most discriminative between tumors that do and do not have mutations in the various genes are summarized. As the results show, *EGFR* mutant (*EGFR(m)*) tumors exhibit hypervascularity (increased relative cerebral blood volume (rCBV)) in their enhancing tumor region compared to the tumors without *EGFR* changes. Mutations in *TP53* are

**Table 2 | Performance of radiogenomic classifiers across all subjects in the discovery and replication sets**

	Deep (conventional MRI)		SVM (conventional MRI)		SVM (conventional and advanced MRI)	
	AUC (95% CI)	Balanced accuracy (%)	AUC (95% CI)	Balanced accuracy (%)	AUC (95% CI)	Balanced accuracy (%)
<b>Discovery cohort</b>						
<i>EGFR (co-occurring)</i>	0.79 (0.73–0.87)	76.1	0.72 (0.64–0.79)	70.3	0.85 (0.78–0.93)	84.9
<i>PTEN (co-occurring)</i>	0.70 (0.63–0.76)	77.8	0.68 (0.61–0.74)	67.2	0.87 (0.81–0.93)	82.8
<i>TP53 (co-occurring)</i>	0.73 (0.66–0.80)	76.5	0.74 (0.67–0.81)	73.5	0.83 (0.76–0.91)	81.3
<i>NF1 (co-occurring)</i>	0.77 (0.69–0.85)	75.3	0.77 (0.69–0.86)	78.3	0.86 (0.77–0.95)	83.0
<i>EGFR (exclusive)</i>	0.97 (0.93–1)	98.4	0.83 (0.72–0.94)	82.6	0.90 (0.80–1)	92.5
<i>PTEN (exclusive)</i>	0.92 (0.85–0.99)	88.9	0.79 (0.68–0.91)	77.6	0.91 (0.79–1)	87.1
<i>TP53 (exclusive)</i>	0.96 (0.89–0.1)	93.8	0.80 (0.66–0.94)	85.2	0.77 (0.59–0.95)	74.7
<i>RTK (co-occurring)</i>	0.71 (0.64–0.78)	71.4	0.78 (0.70–0.84)	76.5	0.84 (0.78–0.91)	83.5
<i>PI3K (co-occurring)</i>	0.77 (0.71–0.83)	75.4	0.69 (0.62–0.75)	67.4	0.85 (0.80–0.91)	81.9
<i>P53 (co-occurring)</i>	0.71 (0.64–0.78)	76.1	0.70 (0.63–0.78)	69.1	0.82 (0.75–0.90)	84.1
<i>MAPK (co-occurring)</i>	0.75 (0.67–0.83)	69.0	0.72 (0.63–0.81)	71.2	0.81 (0.71–0.90)	80.8
<i>RTK (exclusive)</i>	0.70 (0.56–0.85)	75.0	0.74 (0.63–0.86)	75.2	0.82 (0.70–0.95)	82.5
<i>PI3K (exclusive)</i>	0.78 (0.66–0.89)	79.5	0.78 (0.68–0.88)	80.2	0.84 (0.72–0.95)	78.7
<b>Replication Cohort</b>						
<i>EGFR (co-occurring)</i>	0.75 (0.62–0.88)	75.0	0.77 (0.65–0.90)	77.6	0.84 (0.71–0.98)	87.6
<i>PTEN (co-occurring)</i>	0.71 (0.60–0.82)	76.2	0.62 (0.50–0.74)	63.2	0.86 (0.75–0.97)	81.7
<i>TP53 (co-occurring)</i>	0.70 (0.57–0.82)	71.6	0.71 (0.60–0.84)	72.8	0.79 (0.65–0.93)	82.4
<i>NF1 (co-occurring)</i>	0.75 (0.60–0.89)	75.2	0.81 (0.68–0.94)	75.6	0.85 (0.69–1)	85.7
<i>EGFR (exclusive)</i>	0.93 (0.81–1)	94.3	0.89 (0.74–1)	84.3	0.92 (0.74–1)	92.9
<i>PTEN (exclusive)</i>	0.98 (0.94–1)	99.1	0.73 (0.51–0.94)	77.3	0.91 (0.73–1)	87.5
<i>TP53 (exclusive)</i>	0.94 (0.83–1)	94.4	0.91 (0.77–1)	90.9	0.70 (0.40–0.99)	75.5
<i>RTK (co-occurring)</i>	0.72 (0.60–0.84)	71.3	0.74 (0.62–0.86)	73.8	0.87 (0.77–0.98)	88.4
<i>PI3K (co-occurring)</i>	0.74 (0.63–0.84)	75.2	0.61 (0.49–0.72)	64.0	0.78 (0.66–0.90)	77.3
<i>P53 (co-occurring)</i>	0.76 (0.64–0.81)	78.3	0.75 (0.63–0.87)	70.5	0.78 (0.64–0.92)	82.4
<i>MAPK (co-occurring)</i>	0.82 (0.78–0.84)	74.4	0.77 (0.63–0.91)	76.8	0.84 (0.67–0.99)	87.9
<i>RTK (exclusive)</i>	0.69 (0.15–1)	83.3	0.75 (0.54–0.96)	81.8	0.88 (0.69–1)	85.4
<i>PI3K (exclusive)</i>	0.75 (0.28–1)	87.5	0.83 (0.68–0.98)	85.0	0.90 (0.73–1)	95.0

The names of the genes and pathways are indicated in italic font.

Performance metrics, including the area under the receiver operating characteristic (ROC) curve (AUC) and balanced accuracy, are presented for radiogenomic signatures of key driver genes and pathways. These metrics were evaluated for all patients, including those with co-occurring tumors, as well as for mutually exclusive tumors. The signatures were derived using a deep learning model trained on conventional MRI and an SVM classifier trained on either conventional MRI alone or a combination of conventional and advanced MRI.

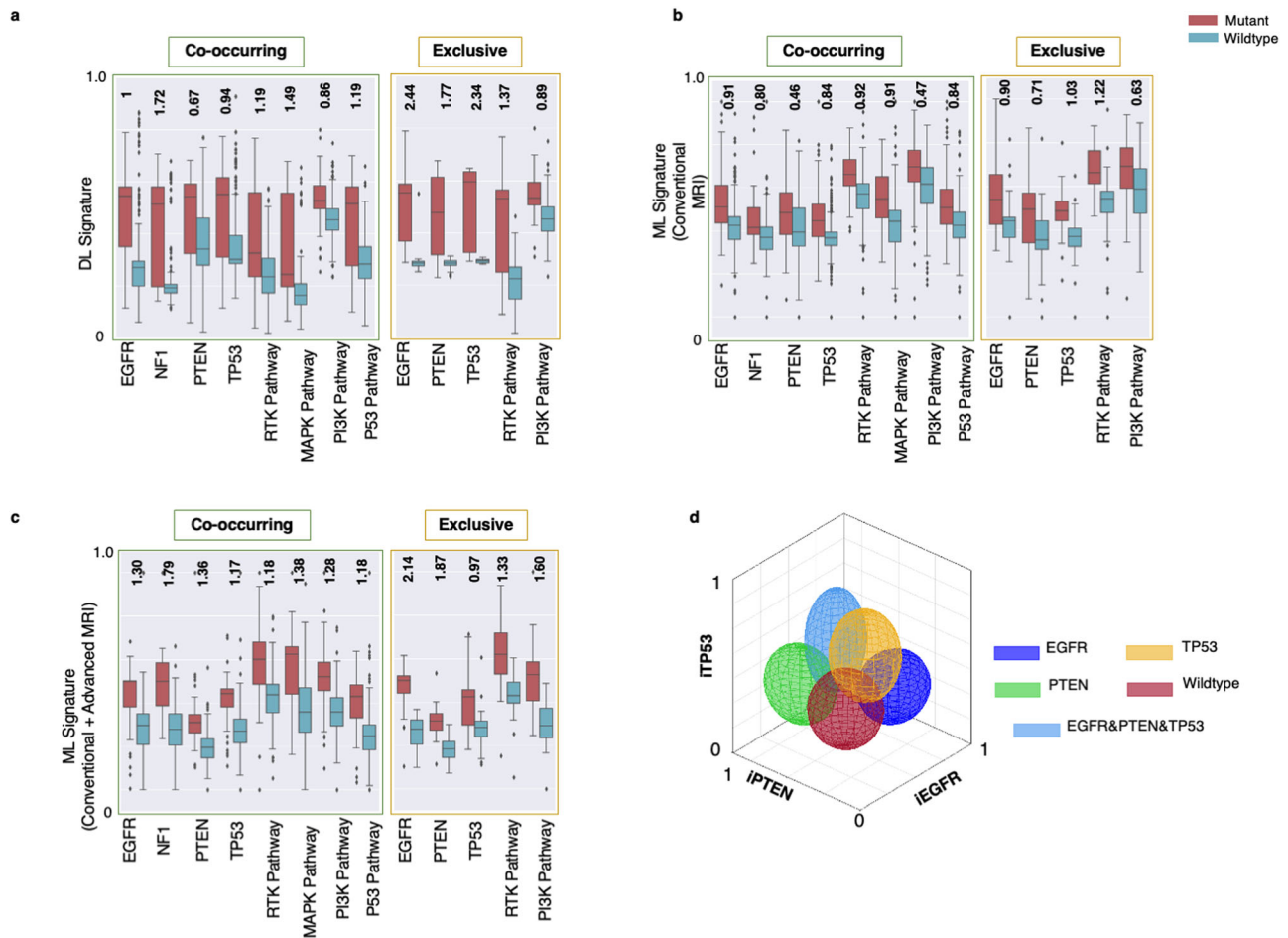
associated with altered permeability and vascular leakage (increased percent signal recovery (PSR)) in the enhancing tumor, decreased water concentration (decreased T2), higher cell density (lower apparent diffusion coefficient (ADC)), and higher neovascularization (elevated rCBV) in their peritumoral edema region. The so-called peritumoral edema manifests as a region of high signal intensity on T2-weighted and T2-FLAIR MRI sequences surrounding the enhancing tumor mass and presents a combination of vasogenic edema and infiltrating glioma cells. Tumors with mutations in *PTEN* show relatively more water concentration (increased T2 signal) and facilitated diffusion (higher ADC) in the non-enhancing tumor region. *NF1*-mutated tumors have relatively more pronounced vascularization (increased peak height (PH)) and reduced diffusion resulting from higher cellularity (lower ADC) in the non-enhancing tumor region. Furthermore, mutations in *NF1* lead to altered permeability and vascular leakage (increased PSR) in the enhancing tumor and higher angiogenesis (increased PH) in the peritumoral edema.

The scatterplots in Fig. 1d illustrate the imaging signatures (denoted by *iEGFR*, *iPTEN*, and *iTP53*) for our ML models created based on conventional and advanced MRI sequences for prediction of mutations in *EGFR*, *PTEN*, and *TP53* genes. This figure demonstrates the discrimination of tumors for which our sequencing panel identified mutations in a single gene

(related to purer molecular landscape) in comparison to tumors with no mutations detected or the tumors with mutations in multiple genes. Radiogenomic signature values for mixed tumors are typically scattered across the plot, whereas values for tumors harboring a single mutation tend to cluster within a specific region. These findings further confirm that tumors with multiple mutations or a more heterogeneous molecular landscape exhibit imaging characteristics that reflect the combined signatures of their co-existing mutations.

### Spatio-genomic landscape of IDH-wildtype GBM: spatial distributions are associated with specific molecular mutations and pathways

To test whether spatial location of the tumor is associated with its mutational profile, statistical atlases capturing the probability that a tumor with certain mutation(s) occurs at a particular brain region (see Methods) were generated. Figure 2 depicts spatial maps of the frequency of occurrence of 357 IDH-wildtype GBMs with mutations in each of the key driver genes, i.e., *EGFR*, *PTEN*, *NF1*, and *TP53* (Fig. 2a, Supplementary Video 1), as well as their associated signaling pathways, i.e., *RTK*, *PI3K*, *MAPK*, and *P53* (Fig. 2c, Supplementary Video 2). Furthermore, we generated atlases of spatial patterns for wildtype tumors, i.e., tumors without any mutations in any of



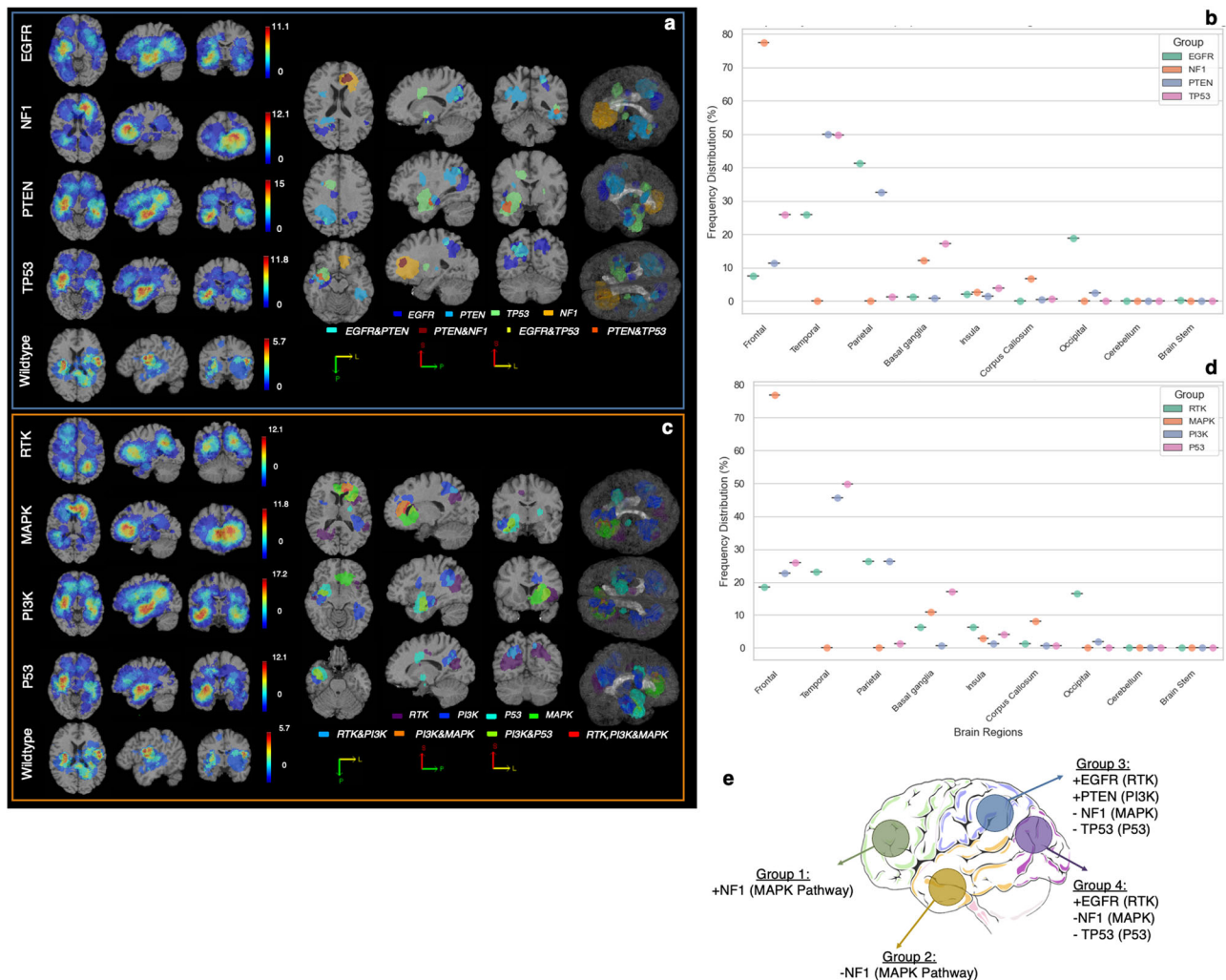
**Fig. 1 | Imaging signatures of IDH-wildtype GBM tumors.** Signatures generated using deep learning (DL; subfigure a;  $n = 357$  independent samples), and conventional machine learning (ML) using only conventional MRI (subfigure b;  $n = 357$  independent samples) as well as using both conventional and advanced MRI (subfigure c;  $n = 228$  independent samples) in mutant (red boxes) and wildtype (blue boxes) tumors. These signatures in Panels (a & c) are shown for co-occurring (left) and exclusive (right) mutations and pathways. The numbers indicated at the top of each pair of mutant-wildtype box-whisker plots indicate the Cohen's D effect sizes for the signatures generated by DL, ML (conventional MRI), ML (conventional and advanced MRI) models for co-occurring and mutually exclusive tumors. Error bars in the box-and-whisker plots represent the range of the data excluding outliers, with whiskers extending to the smallest and largest values within 1.5 times the

interquartile range (IQR) from the lower and upper quartiles. The outliers are represented as individual points outside this range. **d** A representative example of the degree of expression of imaging signatures, quantitatively measured along respective axes, for mutations in the EGFR (dark blue), PTEN (green), and TP53 (yellow) genes, wildtype tumors (red), and tumors with a combination of EGFR, PTEN, and TP53 mutations (light blue) [iEGFR, iPTEN, and iTP53 = imaging signatures of EGFR, PTEN, and TP53 mutations, respectively]. The colors of the ellipsoids represent the molecularly defined mutation status, and the shape of the ellipsoids reflects the spread of values of respective imaging signatures. For example, molecularly defined EGFR mutant tumors (dark blue) tend to express the iEGFR signature, i.e. they have higher values along the iEGFR axis.

the six main pathways, *RTK*, *PI3K*, *MAPK*, *P53*, and *RBI* pathways, and *ChrMod* regulation (including chromatin modifier genes, such as *ATRX*)<sup>22</sup>. Tumors with mutations in certain genes/pathways or their combinations show predilection for specific brain regions, as demonstrated in the atlases on the right side of panels (A) and (C) in Fig. 2. We found that *EGFR<sub>m</sub>* tumors to be more frequent in the left parietal, right temporal, and right occipital lobes, while *PTEN<sub>m</sub>* cases tend to occur more frequently in the left and right temporal, and right parietal lobes (Fig. 2a, b). *TP53<sub>m</sub>* is more likely to occur in the right temporal and right frontal lobes, and *NF1<sub>m</sub>* has a predilection for the left frontal lobe (Fig. 2a, b). As the results in Fig. 2(c, d) suggest, subjects with activated *RTK* pathway are more frequent in the right temporal, left parietal, and left frontal lobes, tumors with mutations in the *PI3K* pathway tend to occur in the left and right temporal and left frontal lobes, *P53* activated pathway subjects are more prominent in the right temporal, right frontal lobes, and right basal ganglionic regions, and *MAPK* activated tumors are more likely to grow in the left frontal lobe.

Based on a population-based approach, four distinct brain regions were derived, representing relatively high prevalence of certain mutations in

those regions. A summary of the spatiogenomic markers is provided in Fig. 2e where four main groups of tumor occurrence propensity are specified. To explore characteristic molecular underpinnings for these groups of tumors we extracted mutational signatures (Fig. 3a). We used cosine similarity metric (CM) as a measure of closeness among the four mutational profiles (Fig. 3b). This value ranges from 0 (completely different mutational signatures) to 1 (identical signatures). Mutational signatures were significantly different (CM = 0.35) between Group 1 (frontal lobe) and Group 4 (occipital lobe), which respectively include or lack *NF1* (*MAPK*) mutations. The differences between the spatiogenomic groups were further examined through calculation of MATH (Mutant-Allele Tumor Heterogeneity; See Methods Section) scores, which quantitatively measure intra-tumor heterogeneity<sup>24</sup>. During the clonal evolution of a tumor, loci with earlier mutations exhibit higher mutant allele fractions (MAFs) and thus higher MATH scores, whereas loci with later mutations have lower MAFs and MATH scores. Notably, our results show that Group 4 tumors in the occipital lobe, which reflect these early mutation patterns, demonstrated significantly higher MATH scores compared to Group 1 tumors in the



**Fig. 2 | Spatiogenomic landscape of IDH-wildtype GBM.** **a** Spatial maps of frequency of occurrence for the tumors with mutations in key driver genes, including EGFR, NF1, PTEN, and TP53, and **(c)** their associated signaling pathways, i.e., RTK, MAPK, PI3K, and P53. The spatial map of wildtype tumors, i.e., tumors with no mutations in any of the genes involved in the main six signaling pathways (RTK, MAPK, PI3K, P53, RB1, ChrMod) is also shown. In each of subfigures **(a)** and **(c)**, the maps on the right side indicate the regions with higher likelihood of occurrence of mutations in a gene/pathway or a combination of genes/pathways. The box plots

shown in subfigures **(b)** and **(d)** indicate frequency distribution for occurrence of tumors carrying each of the key driver genes/pathways in different brain regions. The basal ganglia label consists of multiple brain regions, i.e., putamen, caudate nucleus, globus pallidus, subthalamic nucleus, nucleus accumbens, internal capsule, and thalamus. All images are displayed in the radiological convention orientation. Subfigure **(e)** summarizes the findings, illustrating four main groups of regions, towards which the tumors with specific gene mutations show higher predilection.

frontal lobe ( $p = 0.013$ ; Supplementary Fig 3). The MATH scores for Group 4 tumors were lower compared to Group 2 and Group 3 tumors located in the temporal and parietal lobes, though this difference did not reach statistical significance ( $p = 0.05$ ; Supplementary Fig 3).

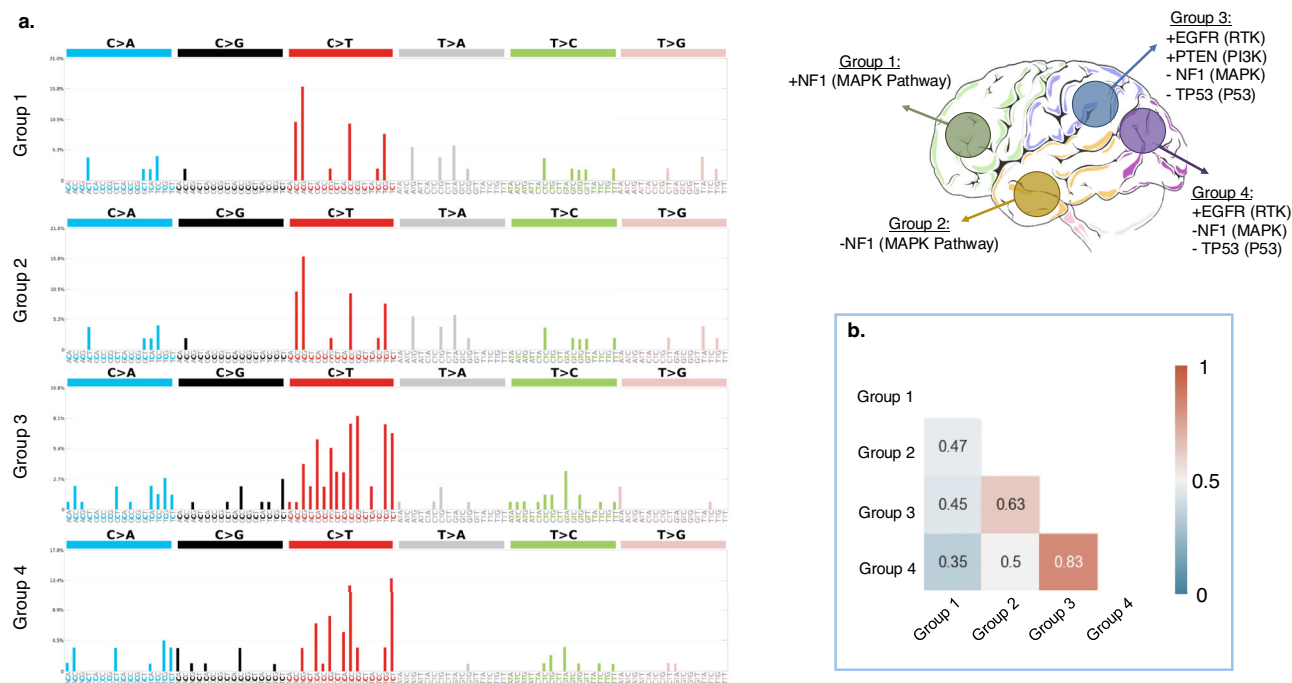
**Molecular underpinnings of primary IDH-wildtype GBM with spatial anatomical predilection**

To elucidate the links between tumor progression or evolution with radiographic appearances, 177 samples with available single nucleotide variants (SNVs) and copy number variation (CN) information were analyzed. Only somatic mutations with MAFs of 5% or greater were included. Figure 4 and Supplementary Table 2 depict a summary of common oncogenic mutations, pathway-level alterations, and trajectory evolution model of GBM. The high copy number gains and losses were defined as  $\geq 5$  copies and  $\leq 1$  copy, respectively.

The oncoplot (Fig. 4a) shows that the most prevalent somatic genomic alterations present in these 177 samples were EGFR CN gain (54%), CDKN2A CN loss (40%) and PTEN somatic mutation (39%). These findings are in line with reports that EGFR amplification and rearrangement are early

events in tumorigenesis<sup>45</sup>. The pathways with the most frequent genomic alterations were receptor tyrosine kinase (RTK, 74%) pathway, which included EGFR, followed by PI3K pathway (54%) which included PTEN, followed by RB1 pathway (48%) including CDKN2A (Fig. 4b). Overall, more than 90% of the tumors have RTK and/or PI3K alterations (Fig. 4b). The circo plot (Fig. 4d) shows that the most co-mutated alterations were EGFR snv - EGFR amplification, EGFR amplification - CDKN2A deletion, and EGFR amplification - PTEN snv.

Next, we examined whether the tumors differ in their order of events of somatic genomic alterations during GBM evolution using a Bayesian inference approach implemented in TRONCO<sup>23</sup>. In brief, this approach combines techniques for sample stratification, driver selection, and identification of fitness-equivalent exclusive alterations to predict likely temporal order of mutational events during disease progression based on Suppes’s probabilistic causation. The trajectory of GBM progression was inferred based on mutated genes included in the six signaling pathways that were present in more than 10% of all samples. The model captured the early drivers including EGFR amplification, and somatic mutations in NF1, PTEN and TP53 (Supplementary Fig 4). EGFR amplification was accompanied by



**Fig. 3 | Mutational signatures for the tumors in the four groups of spatial locations.** The four groups from Fig. 2e are illustrated on the top right side of this figure. **a** Mutational signatures; **b** cosine similarity metric (CSM) for the signatures of the

four groups, suggestive of least similarity between tumors in groups 1 and 4. MATH scores for the tumors in these four groups showed significant differences in the mean MATH scores between groups 1 and 4 ( $p = 0.013$ ).

elevated allele frequency and mutations in *PTEN* and *TP53* also showed higher allele frequencies (Fig. 4e, f). *EGFR* point mutations were detectable at a moderately low allele frequency (Fig. 4e).

Next, we postulated that the differences in oncogenic alterations and evolutionary trajectory are manifested in tumor phenotypes, ultimately reflected in radiogenomic features and therefore investigated the evolution patterns for the tumors originating in the distinctive brain regions of Groups 1-4 of Fig. 2. Figure 5 provides the evolutionary trajectories for tumors in the four brain regions shown in the middle (regions of relatively distinct molecular profiles). As suggested by the results, *NF1* alterations (100%) were identified as the most frequent and potentially early event in Group 1. In contrast, *PTEN* mutations (56%), *EGFR* amplification (56%), and *EGFR* mutations (44%) were the most common events in Group 2, followed by *TP53* (38%) and *NOTCH2* (31%) mutations. *PTEN* mutation and *EGFR* amplification were the likely early and driver mutations in Group 2 tumors, but it was not possible to ascertain the preferential precedence between those two with statistical significance. In Group 3, *PTEN* mutations (73%), *EGFR* amplification (61%), and *EGFR* mutations (33%) were the most frequent events, and the evolutionary analysis identified *PTEN* mutation and *EGFR* amplification as the likely early driver events. In addition, *TP53* and *CDKN2A* mutations were also captured as additional driver mutations. The most frequently occurring alterations in Group 4 were *EGFR* mutations (86%), *EGFR* amplification (71%) and *GIGY2* mutations (43%). The evolutionary trajectories also indicated the possibility of *EGFR* and *PTEN* mutations as early events in the progression of tumor growth, but sample sizes were small and inferences did not reach significance level ( $p > 0.05$ ). The derived evolutionary trajectories for all the four groups are provided in more detail in (Supplementary Fig 5-8).

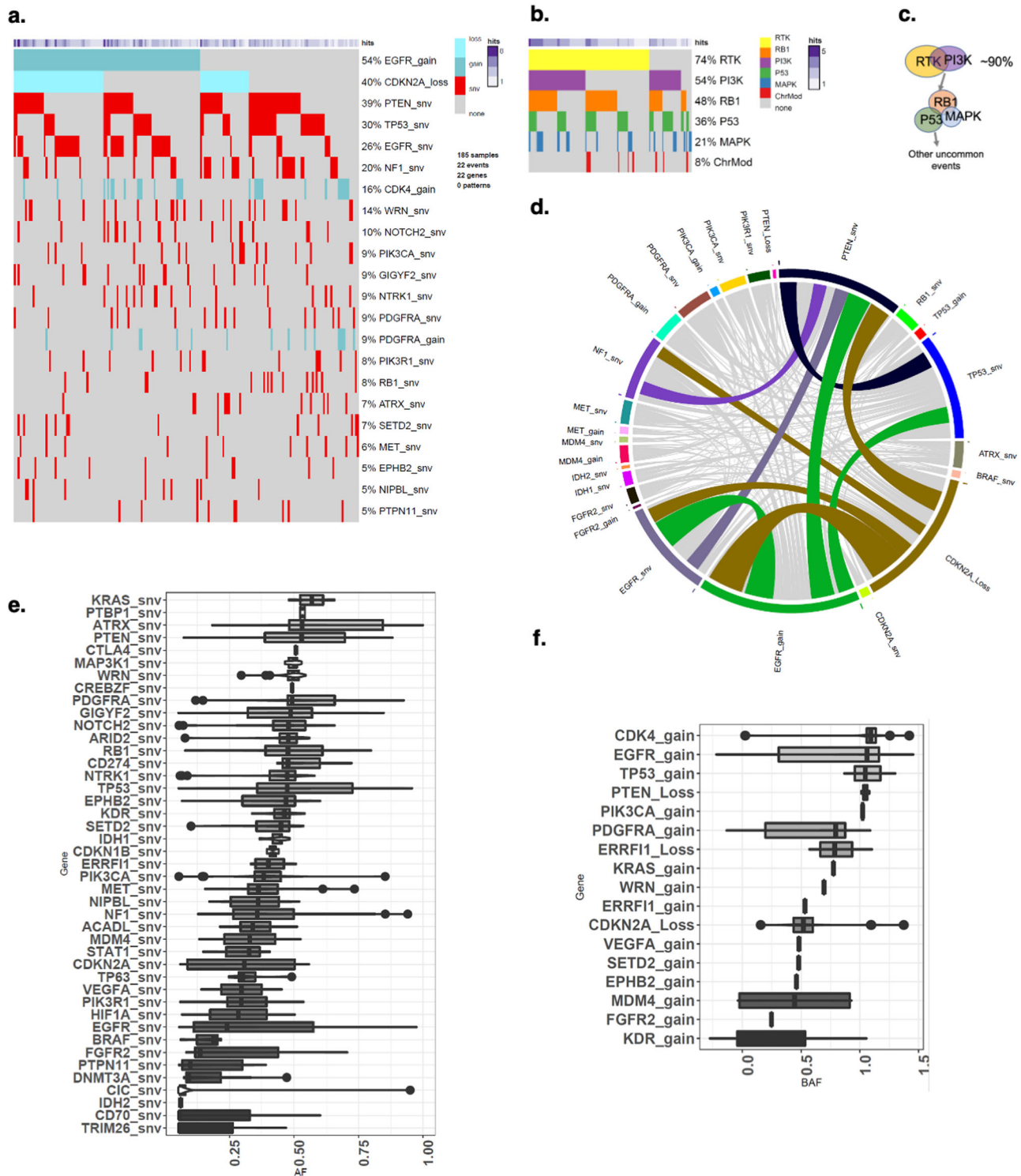
## Discussion

Cancer arises from genetic and epigenetic alterations in primary founding clones<sup>46</sup>, with subclones developing during tumor progression, contributing to intratumoral heterogeneity<sup>46</sup>. Like other cancers, GBM formation and progression involves oncogene activation, tumor suppressor inactivation, DNA repair deregulation, genomic copy number changes, and epigenetic alterations<sup>47</sup>. These alterations occur across subpopulations within the same

tumor, giving rise to intratumor heterogeneity, influencing imaging characteristics. Our study explores the relationship between neuroimaging patterns, or radiophenotypes, and the molecular underpinnings of *IDH*-wildtype GBM revealing how mutations influence radiographic characteristics of the tumor. We attempted to bridge genomic and imaging scales by mapping the molecular landscape of *IDH*-wildtype GBM through spatio-genomic analysis and multivariate quantification of clinically-obtained mpMRI scans.

We developed imaging-AI models using radiomic features extracted from the entire tumor and its subregions, predicting mutations sequenced from single tumor specimens. These models revealed associations between in vivo imaging characteristics and underlying genetics, identifying non-invasive radiogenomic signatures of driver genes or pathways and assessing the impact of molecular heterogeneity on these imaging signatures. Unlike previous studies on GBM radiogenomics<sup>43,44,48-51</sup> that focused on single targets, we explored co-occurring mutations simultaneously, demonstrating that AI-driven radiogenomics analyses can capture phenotypic expressions of driver genes alterations.

Our results demonstrate that molecular heterogeneity is reflected in radiophenotypes, with less heterogeneous tumors exhibiting more distinctive radiogenomic signatures. These in vivo phenotypic biomarkers provide insights into tumor ecosystems, including its molecular heterogeneity<sup>52</sup>, aiding clinical management and trial stratification, particularly for molecularly targeted therapies. Tumors with *EGFR* mutations showed imaging features consistent with increased angiogenesis in enhancing tumor regions. This finding is in agreement with the involvement of amplified and/or overexpressed growth factors, or oncogenes, including *PDGF*, *EGFR*, *MET*, and *FGFR*, in overgrowth of high-grade gliomas<sup>53</sup>. The *EGFR* gene, as one of the most frequently altered genes in the *RTK* pathway, encodes the EGFR protein, which is a transmembrane tyrosine kinase receptor, and is critical in regulating cell division and death. Ligand-receptor binding transmits signals that result in activation of a series of intracellular signaling cascades affecting apoptosis, angiogenesis, and invasion<sup>53</sup>. Alterations in *EGFR* can lead to constitutive activation of the signaling pathway<sup>53</sup>, and the resulting phenotype is detectable by our imaging analysis.



**Fig. 4 | A summary of common oncogenic mutations, pathway-level alterations, and trajectory evolution model of GBM. a** Heatmap showing the somatic mutations including single nucleotide variant (snv) and copy number changes identified from target exon sequencing of  $n = 177$  independent tumors. Snv, CN deletion and amplification are separately denoted by red, blue and dark blue color. **b** Heatmap showing the six mutation pathways including 35 mutated genes in 177 tumor samples. The bar graph on the top depicts the number of tumors. **c** Schematic representation of evolutionary trajectory model of GBM progression that shows the estimated order of events at the pathway level. **d** Circos plot showing the co-mutated

genes shared by tumors; colored connections present co-mutated genes shared by more than 20% of tumors and the grey color connections denote co-mutated genes shared by less than 20% of samples. The width of the connection represents the sample size. **e** The allele frequency for all sequenced genes, **f** BAF allele frequency for genes with CN alterations (gains and losses). Error bars in the box-and-whisker plots represent the range of the data excluding outliers, with whiskers extending to the smallest and largest values within 1.5 times the interquartile range (IQR) from the lower and upper quartiles. The outliers are represented as individual points outside this range.



neo-angiogenesis in peritumoral edema). This tumor subregion is formed as a biological response to diffuse infiltration of glioma cells into normal tissue surrounding the tumor, which causes the release of angiogenic and vascular permeability factors<sup>49</sup>. As resection of peritumoral edema, the propagating front of the tumor, is not usually a part of standard of care treatment for GBMs, future tumor recurrence becomes inevitable. Better understanding of radiogenomic characteristics of peritumoral edema can potentially aid in adopting alternative local therapies such as extended resection, intensified radiation, focused ultrasound, intra-operative viral injections, or convection enhanced intratumoral delivery of therapeutic agents.

We also observed spatial patterns in tumor locations associated with specific mutations, suggesting contributions from region-specific precursor cells. We found clusters of tumors in perisylvian regions and subcortical white matter with extension to superficial grey matter, which aligns with the previous findings about presence of neural CSCs in regions of the brain, such as the dentate gyrus<sup>64</sup>. GBMs likely arise from neural progenitor cells or stem cells, specifically astrocyte-like neural CSCs in the subventricular zone with low-level somatic mutations, which are the likely origin of *IDH*-wildtype GBM<sup>65</sup>.

We show that molecular heterogeneity, to the extent that can be measured by our targeted sequencing panel and MATH, and the similarity of the mutational signature for the group of tumors located in the frontal lobe are statistically lower ( $p = 0.01$ ) than heterogeneity and signatures of the tumors clustered in the occipital lobe, and lower than the clusters of tumors in the temporal and parietal lobes (although not found statistically significant,  $p = 0.05$ ). This distinction in the molecular composition of the tumors in the discovered tumor region groups was also observed by analyzing the evolutionary trajectories, which were estimated via statistical modeling of the pre-surgical genomic data. As the findings suggest, the tumors in the cluster concentrated around the frontal lobe may undertake a different progression path compared to tumors that occur in other regions, especially those arising in proximity to the occipital lobe.

It is proposed in the literature that tumor progression and heterogeneity are governed by selection of specific CSCs over time, due to factors including their genetic, epigenetic and TME interactions<sup>1-4</sup>. The progenitor CSCs may migrate from their origin and acquire additional somatic mutations, ultimately resulting in formation and progression of the tumor in distant brain parenchyma. Interactions with the TME can give rise to growth of some CSCs in highly vascular regions, or adaptation of some subpopulations to more hypoxic TME<sup>5</sup>. Numerous studies have reported on the role of the TME in forming the tumor phenotype and giving rise to its spatial and temporal heterogeneity<sup>66</sup>. The distinction in spatiogenomic signatures of tumors located in different brain regions, as supported by their modeled temporal evolution path, further highlights the contribution of TME in building the underlying molecular landscape of GBM.

In a previous study on the spatial tumor location of GBMs, tumors with amplification in *EGFR* and mutation in *EGFR VIII* were found to occur in the left temporal lobe<sup>67</sup>. In another study, the right frontotemporal region was reported to be absent in tumors with mutation in *EGFR VIII*<sup>64</sup>. We found that *EGFR* mutant tumors, including those with *EGFR VIII* mutations, more frequently occur in the left parietal, right temporal, and right occipital lobes. The differences in the reported tumor locations could be attributed to the inclusion of different *EGFR* alterations in our study. *EGFR* amplification, reported in the study by Ellingson et al.<sup>67</sup>, occurs in 57.4% of primary GBMs<sup>23</sup>, leading to a higher frequency than somatic mutations and resulting in different findings. Additionally, varying thresholds and sequencing approaches could contribute to these differences. The smaller sample size in the study by Billelo et al.<sup>64</sup> could also be a factor in the absence of mutations of *EGFR VIII* in the right temporal region.

As our study did not investigate the spatial location of tumors with or without *PTEN* loss, it is difficult to compare our finding that *PTEN* mutant tumors more frequently occur in the left and right temporal, and right parietal lobes with the findings reported by Ellingson et al.<sup>67</sup>, who reported that tumors without *PTEN* loss most frequently occurred in the frontal lobe. Further research is needed to reconcile these differences and understand the

underlying factors influencing the spatial distribution of *PTEN* and *EGFR* mutant tumors.

Our study has several limitations. While the sample size for this study was larger than several other radiogenomic studies of GBM, it would benefit from a larger cohort, in which additional pathway level alterations can be detected and be correlated to the radiogenomic and spatiogenomic characteristics. Furthermore, this study did not include data from external institutions. The MR images from public databases are often of low quality and lack the necessary imaging data required for our multiparametric MRI predictive models. This limitation highlights the need for further validation of our predictive models to ensure their generalizability to multi-institutional data. Future studies should focus on acquiring higher-quality and more comprehensive datasets from diverse sources to enhance the robustness and applicability of the models.

Furthermore, limited sampling of biospecimen for genomic profiling and targeted gene panel sequencing may not capture certain classes of genomic alterations (e.g. fusions) and underappreciate intratumor genetic heterogeneity in driver mutations. Formalin fixation of the tumor samples that is carried out for diagnostic tests in molecular pathology can cause DNA damage, such as fragmentation and non-reproducible sequencing artifacts<sup>47</sup>. In our analysis of mutational signatures, disproportionate levels of C > T changes that appeared in all our four signatures may have been a result of these artifacts. However, as we used cosine similarity metric to determine the distinct patterns of mutational signatures, we do not believe that these artifacts that were similar among all four signatures influenced the derived conclusions.

While our sequencing panel ensured that the majority of tumors met the GBM criteria recommended by the 2021 WHO CNS tumor classification, the absence of *TERT* among the sequenced genes in our NGS panels and copy number data means that a few non-GBM patients may have been included in our cohort. However, given our thorough model training using nested cross-validation and independent testing approaches, the impact of these few non-GBM patients on the robustness of our models is expected to be minimal.

Additionally, our study demonstrated differences in imaging characteristics of tumors with different molecular drivers. However, these conclusions are based on the analysis of retrospective clinical data. Our study was limited to post-operative biopsies collected from resected tumors, making it impossible to obtain localized biopsies from different tumor subregions to directly associate their imaging phenotypes with genomic markers. We developed imaging-AI predictive models of gene or pathway mutations based on the imaging characteristics derived from the whole tumor and its subregions, including the enhancing tumor, non-enhancing tumor core, and peritumoral edema. However, the mutations were analyzed based on a single tissue sample typically cut from the enhancing core of the tumor. Validation of the characteristics of the peritumoral edema and non-enhancing core requires the collection of biopsy samples from these regions and the analysis of the gene mutations present within those areas, as carried out in a few other studies<sup>68-71</sup>. Alternatively, preclinical experiments can be conducted to confirm the observed imaging-molecular associations.

Lastly, it was not possible to directly associate changes in PSR levels in tumors with or without *TP53* or *NF1* mutations to specific biological characteristics like tumor permeability and vascular leakage. This is primarily because PSR variations are influenced by several factors, including imaging acquisition parameters (e.g., flip angle, echo time (TE), preload injection) and vascular bed architecture, such as compact extracellular-extravascular space (EES), rapid cerebral blood flow, and variations in capillary leakiness<sup>36,37,72,73</sup>.

In conclusion, we have elucidated the links between regional variations expressed as radiophenotypes and the molecular landscape of *IDH*-wildtype GBM. Prompted by the growing advances in computational analytics in the radiology – neuro-oncology domain and evidence for the potential of imaging to reveal molecular characteristics of the tumors, we comprehensively analyzed the radiogenomic, spatiogenomic, and evolutionary landscapes of GBMs to provide better understanding of the links between radio-

phenotypes and the underlying molecular heterogeneity of GBMs. We found distinct radiogenomic signatures of mutations in several genes and signaling pathways that became more pronounced when the tumors were less molecularly heterogeneous. Importantly, these signatures offered mechanistic insights into how various mutations shape the phenotype of the tumor and its surrounding brain tissue via changes in cell density and neo-vascularization. We further discovered spatiogenomic signatures from tumors that carry certain mutations, suggesting the association of molecular composition with the brain regions they arise from. We showed that these spatial patterns may emerge as a result of differences in TMEs in different brain regions that give rise to distinct spatial and temporal heterogeneity and define molecular composition of the tumors. The findings of this study can assist in noninvasive identification of patients who may benefit the most from molecular targeted therapies, as well as allow for longitudinal monitoring of mutational changes during treatment.

### Data availability

The MRI sequences, image acquisition parameters, and clinical variables used in this study can be found in the UPenn-GBM collection in the Cancer Imaging Archive (<https://doi.org/10.7937/TCIA.709X-DN49>)<sup>74</sup>. The derived data (radiomic features and mutation status) that support the findings of this study are available on reasonable request from the corresponding author (C. D.). However, the raw data (images and genomic variables) are not distributable, as the patients were not consented for public release of their data. The numerical data used to create the figures in this manuscript are available in the Supplementary Information/Source Data file on Figshare, <https://doi.org/10.6084/m9.figshare.28243670>. All oligonucleotide sequences used in this study, including primers and probes are provided on Figshare, <https://doi.org/10.6084/m9.figshare.28243673>. Additional details are available upon reasonable request from the corresponding author.

### Code availability

All image processing and feature extraction tools used in this study are freely available for public use (CaPTk, <https://www.cbica.upenn.edu/captk>). The predictive models using SVM ML classification based on conventional and advanced MRI sequences along with the steps for image processing and feature extraction can be accessed at [https://github.com/anahitafk/radiogenomic\\_adult\\_GBM](https://github.com/anahitafk/radiogenomic_adult_GBM), <https://doi.org/10.5281/zenodo.14582800><sup>75</sup>.

Received: 14 September 2023; Accepted: 12 February 2025;  
Published online: 01 March 2025

### References

1. Galli, R. et al. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* **64**, 7011–7021 (2004).
2. Singh, S. K. et al. Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).
3. Tirosh, I. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* **539**, 309–313 (2016).
4. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
5. Lauko, A., Lo, A., Ahluwalia, M. S. & Lathia, J. D. In Seminars in cancer biology. 162–175 (Elsevier).
6. Stupp, R. et al. Maintenance therapy with tumor-treating fields plus temozolomide vs temozolomide alone for glioblastoma: a randomized clinical trial. *Jama* **314**, 2535–2543 (2015).
7. Ostrom, Q. T., Cioffi, G., Waite, K., Kruchko, C. & Barnholtz-Sloan, J. S. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. *Neuro-oncology* **23**, iii1–iii105 (2021).
8. Lombardi, M. Y. & Assem, M. Glioblastoma Genomics: A Very Complicated Story. In *Glioblastoma* (ed. De Vleeschouwer, S.), Chapter 1 (Codon Publications, Brisbane (AU), 2017).

9. Claes, A., Idema, A. J. & Wesseling, P. Diffuse glioma growth: a guerilla war. *Acta Neuropathol.* **114**, 443–458 (2007).
10. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
11. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl Acad. Sci.* **110**, 4009–4014 (2013).
12. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
13. Meyer, M. et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc. Natl Acad. Sci.* **112**, 851–856 (2015).
14. Bagley, S. J. et al. Glioblastoma clinical trials: current landscape and opportunities for improvement. *Clin. Cancer Res.* **28**, 594–602 (2022).
15. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
16. Prados, M. D. et al. Toward precision medicine in glioblastoma: the promise and the challenges. *Neuro-Oncol.* **17**, 1051–1063 (2015).
17. Brandner, S. et al. MGMT promoter methylation testing to predict overall survival in people with glioblastoma treated with temozolomide: a comprehensive meta-analysis based on a Cochrane Systematic Review. *Neuro-Oncol.* **23**, 1457–1469 (2021).
18. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
19. Fathi Kazerooni, A., Bakas, S., Saligheh Rad, H. & Davatzikos, C. Imaging signatures of glioblastoma molecular characteristics: a radiogenomics review. *J. Magn. Reson. Imaging* **52**, 54–69 (2020).
20. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncol.* **23**, 1231–1251 (2021).
21. Nasrallah, M. P. et al. Molecular neuropathology in practice: clinical profiling and integrative analysis of molecular alterations in glioblastoma. *Acad. Pathol.* **6**, 2374289519848353 (2019).
22. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
23. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
24. Mroz, E. A. & Rocco, J. W. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral. Oncol.* **49**, 211–215 (2013).
25. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Computational Biol.* **12**, e1004873 (2016).
26. Grasso, C. et al. Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. *J. Mol. Diagnostics* **17**, 53–63 (2015).
27. Nikiforova, M. N. et al. Targeted next-generation sequencing panel (GlioSeq) provides comprehensive genetic profiling of central nervous system tumors. *Neuro-Oncol.* **18**, 379–387 (2015).
28. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
29. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. “Circlize” implements and enhances circular visualization in R. (2014).
30. Wickham, H., Chang, W. & Wickham, M. H. Package ‘ggplot2’. *Create elegant data visualisations using the grammar of graphics. Version 2*, 1–189 (2016).
31. Davatzikos, C. et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *J. Med. Imaging* **5**, 011018–011018 (2018).
32. Pati, S. et al. in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5*. 380–394 (Springer).

33. Rohlfing, T., Zahr, N. M., Sullivan, E. V. & Pfefferbaum, A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum. brain Mapp.* **31**, 798–819 (2010).
34. Yushkevich, P. A. et al. IC-P-174: fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions In 3 Tesla and 7 Tesla T2-Weighted MRI. *Alzheimer's. Dement.* **12**, P126–P127 (2016).
35. Kamnitsas, K. et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017).
36. Boxerman, J., Schmainda, K. & Weisskoff, R. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *Am. J. Neuroradiol.* **27**, 859–867 (2006).
37. Cha, S. et al. Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *Am. J. Neuroradiol.* **28**, 1078–1084 (2007).
38. Akbari, H. et al. Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. *Radiology* **273**, 502–510 (2014).
39. Gooya, A. et al. GLISTR: glioma image segmentation and registration. *IEEE Trans. Med. Imaging* **31**, 1941–1954 (2012).
40. Zwanenburg, A. et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).
41. He, K., Zhang, X., Ren, S. & Sun, J. In Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778.
42. Choi, Y. S. et al. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. *Neuro-Oncol.* **23**, 304–313 (2021).
43. Park, J. E. et al. Prediction of core signaling pathway by using diffusion-and perfusion-based MRI radiomics and next-generation sequencing in isocitrate dehydrogenase wild-type glioblastoma. *Radiology* **294**, 388–397 (2020).
44. Akbari, H. et al. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro-Oncol.* **20**, 1068–1079 (2018).
45. Del Vecchio, C. et al. EGFRvIII gene rearrangement is an early event in glioblastoma tumorigenesis and expression defines a hierarchy modulated by epigenetic mechanisms. *Oncogene* **32**, 2670–2681 (2013).
46. Gerlinger, M. & Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* **103**, 1139–1143 (2010).
47. Langie, S. A. et al. Causes of genome instability: the effect of low dose chemical exposures in modern society. *Carcinogenesis* **36**, S61–S88 (2015).
48. Kickingereder, P. et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology* **281**, 907–918 (2016).
49. Bakas, S. et al. In vivo detection of EGFRvIII in glioblastoma via perfusion magnetic resonance imaging signature consistent with deep peritumoral infiltration: The  $\phi$ -index. *Clin. Cancer Res.* **23**, 4724–4734 (2017).
50. Binder, Z. A. et al. Epidermal growth factor receptor extracellular domain mutations in glioblastoma present opportunities for clinical imaging and therapeutic development. *Cancer cell* **34**, 163–177.e167 (2018).
51. Rathore, S. et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci. Rep.* **8**, 5087 (2018).
52. Gatenby, R. A., Grove, O. & Gillies, R. J. Quantitative imaging in cancer evolution and ecology. *Radiology* **269**, 8–14 (2013).
53. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol. cancer* **17**, 1–13 (2018).
54. Molinaro, A. M., Taylor, J. W., Wiencke, J. K. & Wrensch, M. R. Genetic and molecular epidemiology of adult diffuse glioma. *Nat. Rev. Neurol.* **15**, 405–417 (2019).
55. Streitberger, K.-J. et al. How tissue fluidity influences brain tumor progression. *Proc. Natl. Acad. Sci.* **117**, 128–134 (2020).
56. Benitez, J. A. et al. PTEN regulates glioblastoma oncogenesis through chromatin-associated complexes of DAXX and histone H3. *Nat. Commun.* **8**, 15223 (2017).
57. Louis, D. N. Molecular pathology of malignant gliomas. *Annu. Rev. Pathol. Mech. Dis.* **1**, 97–117 (2006).
58. Furnari, F. B. et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.* **21**, 2683–2710 (2007).
59. Cancer Genome Atlas Research Network et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
60. Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N. & Upadhyaya, M. The NF1 somatic mutational landscape in sporadic human cancers. *Hum. Genom.* **11**, 1–19 (2017).
61. Verhaak, R. G. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–NF110 (2010).
62. Ahn, S.-H. et al. Necrotic cells influence migration and invasion of glioblastoma via NF- $\kappa$ B/AP-1-mediated IL-8 regulation. *Sci. Rep.* **6**, 24552 (2016).
63. Kim, Y. et al. Perspective of mesenchymal transformation in glioblastoma. *Acta Neuropathologica Commun.* **9**, 1–20 (2021).
64. Bilello, M. et al. Population-based MRI atlases of spatial distribution are specific to patient and tumor characteristics in glioblastoma. *NeuroImage: Clin.* **12**, 34–40 (2016).
65. Lee, J. H. et al. Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* **560**, 243–247 (2018).
66. Dirkse, A. et al. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat. Commun.* **10**, 1787 (2019).
67. Ellingson, B. et al. Probabilistic radiographic atlas of glioblastoma phenotypes. *Am. J. Neuroradiol.* **34**, 533–540 (2013).
68. Barajas, R. F. Jr. et al. Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging. *Radiology* **254**, 564–576 (2010).
69. Hu, L. S. et al. Integrated molecular and multiparametric MRI mapping of high-grade glioma identifies regional biologic signatures. *Nat. Commun.* **14**, 6066 (2023).
70. Hu, L. S. et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro-Oncol.* **19**, 128–137 (2017).
71. Fathi Kazerooni, A. et al. Characterization of active and infiltrative tumorous subregions from normal tissue in brain gliomas using multiparametric MRI. *J. Magn. Reson. Imaging* **48**, 938–950 (2018).
72. Mangla, R. et al. Percentage signal recovery derived from MR dynamic susceptibility contrast imaging is useful to differentiate common enhancing malignant lesions of the brain. *Am. J. Neuroradiol.* **32**, 1004–1010 (2011).
73. Hartmann, M. et al. Distinguishing of primary cerebral lymphoma from high-grade glioma with perfusion-weighted magnetic resonance imaging. *Neurosci. Lett.* **338**, 119–122 (2003).
74. Bakas, S. et al. The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Sci. Data* **9**, 453 (2022).
75. Fathi Kazerooni, A. et al. The Radiogenomic and Spatiogenomic Landscapes of Glioblastoma and Their Relationship to Oncogenic Drivers. *Zenodo* <https://doi.org/10.5281/zenodo.14582800> (2024).

## Acknowledgements

Research reported in this publication was partly supported by the National Institutes of Health (NIH) under award numbers NIH/NINDS: R01NS042645, NIH/NCI: U24CA189523, and NIH/NCI: U01CA242871, NIH/NCI: R21248122, NIH/NIGMS: R01GM129066 and NIH/NIGMS: R35GM149224. The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH.

## Author contributions

Conception and design: A Fathi Kazerooni, H Akbari, X Hu, S De, D Grigoriadis, A Hatzigeorgiou, S Bakas, M Nasrallah, C Davatzikos. Development of methodology: A Fathi Kazerooni, H Akbari, X Hu, D Grigoriadis, V Bommineni, E Toorens, T Ganguly, S De, S Bakas, M Nasrallah, C Davatzikos. Acquisition of data: C Sako, E Mamourian, E Toorens, T Ganguly, R Sussman, S Bakas, D Ballinger, S J Bagley, Z A Binder, S Mohan. Processing of data: C Sako, E Mamourian, A Fathi Kazerooni, E Toorens, A Singh. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A Fathi Kazerooni, H Akbari, X Hu, S De, I Verginadis, D Grigoriadis, N Dahmane, D M O'Rourke, C Koumenis, A Hatzigeorgiou, M Nasrallah, C Davatzikos. Writing, review, and/or revision of the manuscript: A Fathi Kazerooni, H Akbari, X Hu, V Bommineni, D Grigoriadis, E Toorens, C Sako, E Mamourian, D Ballinger, R Sussman, A Singh, I Verginadis, N Dahmane, C Koumenis, Z A Binder, S J Bagley, S Mohan, A Hatzigeorgiou, D M O'Rourke, T Ganguly, S De, S Bakas, M Nasrallah, C Davatzikos

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-00767-0>.

**Correspondence** and requests for materials should be addressed to Christos Davatzikos.

**Peer review information** *Communications Medicine* thanks Zhenyu Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

<sup>1</sup>AI2D Center for AI and Data Science for Integrated Diagnostics, University of Pennsylvania, Philadelphia, PA, USA. <sup>2</sup>Center for Data-Driven Discovery in Biomedicine (D3b), Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>3</sup>Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>4</sup>Department of Bioengineering, School of Engineering, Santa Clara University, Santa Clara, CA, USA. <sup>5</sup>Rutgers Cancer Institute of New Jersey, Rutgers the State University of New Jersey, New Brunswick, NJ, USA. <sup>6</sup>Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece. <sup>7</sup>Penn Genomic Analysis Core, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>8</sup>Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>9</sup>Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>10</sup>Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>11</sup>Department of Neurological Surgery, Weill Cornell Medicine, New York, NY, USA. <sup>12</sup>Glioblastoma Translational Center of Excellence, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA. <sup>13</sup>Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>14</sup>Hellenic Pasteur Institute, Athens, Greece. <sup>15</sup>Department of Pathology & Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>16</sup>Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>17</sup>These authors jointly supervised this work: MacLean P. Nasrallah, Christos Davatzikos. ✉ e-mail: [Christos.Davatzikos@penmedicine.upenn.edu](mailto:Christos.Davatzikos@penmedicine.upenn.edu)