Supplementary Information: The Radiogenomic and Spatiogenomic Landscapes of Glioblastoma and Their Relationship to Oncogenic Drivers

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

* These authors jointly supervised this work

[1] AI²D Center for AI and Data Science for Integrated Diagnostics, University of Pennsylvania, Philadelphia, PA, USA

[2] Center for Data-Driven Discovery in Biomedicine (D³b), Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[3] Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4] Department of Bioengineering, School of Engineering, Santa Clara University, Santa Clara, CA, USA

[5] Rutgers Cancer Institute of New Jersey, Rutgers the State University of New Jersey, New Brunswick, NJ, USA

[6] Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

[7] Penn Genomic Analysis Core, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[8] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[9] Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[10] Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[11] Department of Neurological Surgery, Weill Cornell Medicine, New York, NY

[12] *Glioblastoma Translational Center of Excellence, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA*

[13] *Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*

[14] *Hellenic Pasteur Institute, Athens, Greece*

[15] *Rutgers Cancer Institute of New Jersey, Rutgers the State University of New Jersey, New Brunswick, NJ, USA*

[16] *Department of Pathology & Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA*

[17] *Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA*

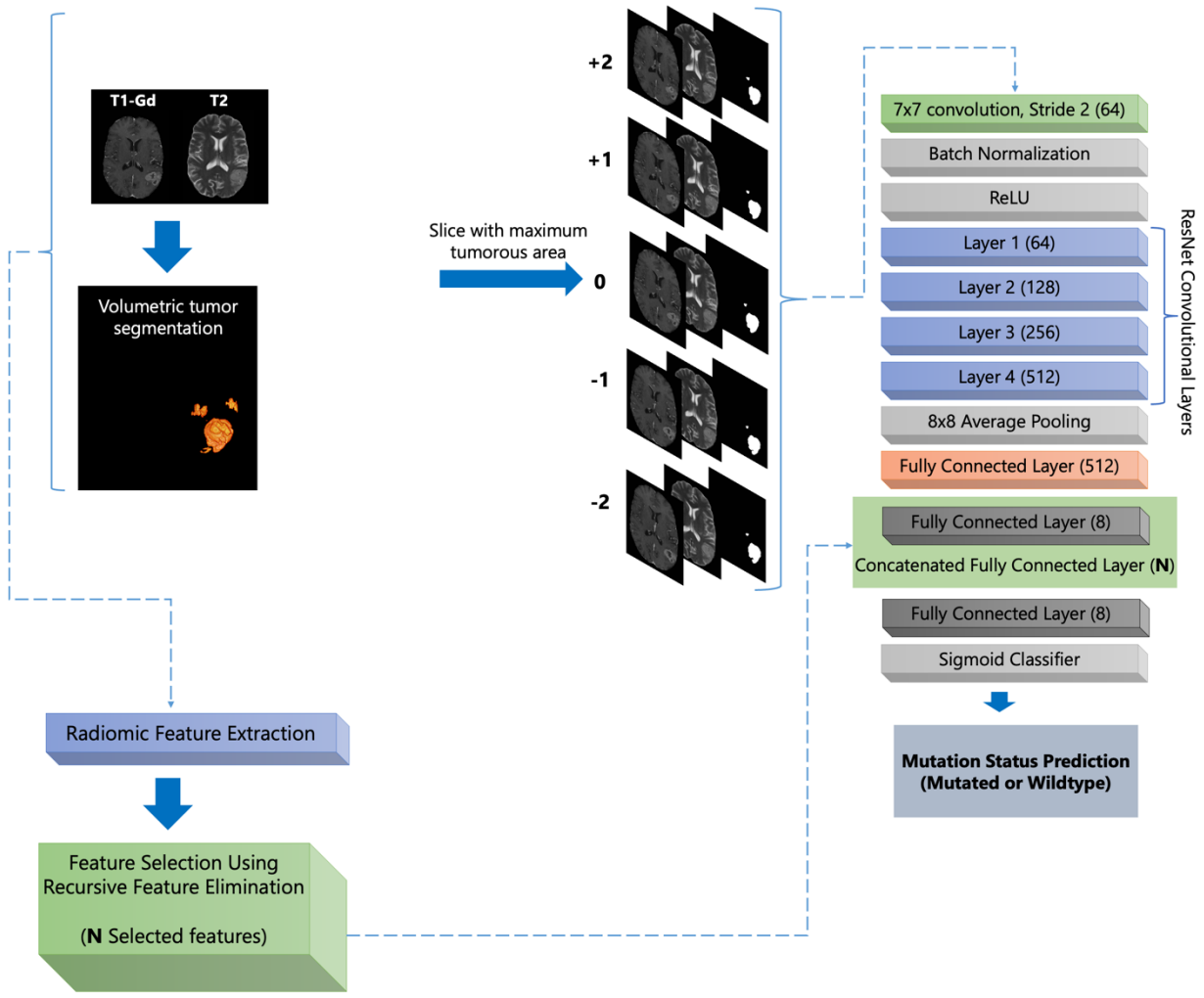
Supplementary Table 1. A summary of demographic and clinical characteristics and tumor genomics of the included IDH-wildtype patients in this study.

	Cohort (conventional MRI)		**Cohort (conventional and advanced MRI)**	
Demographics				
No. of Patients	357		228	
Mean Age (years)	63.2		62.4	
Age Range (years)	20.7 – 89.6		22.0 – 89.6	
No. of Females	132		94	
Survival (months)				
Mean ± Std	15.8 ± 14.1		16.9 ± 14.1	
No. of Censored Patients	92		58	
MGMT Methylation Status				
Methylated	95		48	
Unmethylated	140		73	
Indeterminate or Not Available	122		107	
Driver Genes				
	Co-occurring	**Exclusive**	**Co-occurring**	**Exclusive**
$EGFR_{mut}$ | $EGFR_{wt}$	83 | 274	36 | 43	62 | 166	25 | 24
$NF1_{mut}$ | $NF1_{wt}$	62 | 295	13 | 43	40 | 188	10 | 24
$PTEN_{mut}$ | $PTEN_{wt}$	144 | 213	42 | 43	86 | 142	21 | 24
$TP53_{mut}$ | $TP53_{wt}$	103 | 254	27 | 43	70 | 158	19 | 24
Core signaling pathways				
	Co-occurring	**Exclusive**	**Co-occurring**	**Exclusive**
RTK_{mut} | RTK_{wt}	107 | 250	45 | 43	81 | 147	32 | 24
$PI3K_{mut}$ | $PI3K_{wt}$	185 | 172	68 | 43	114 | 114	37 | 24
$MAPK_{mut}$ | $MAPK_{wt}$	67 | 290	-	43 | 185	-
$P53_{mut}$ | $P53_{wt}$	105 | 252	-	72 | 156	-

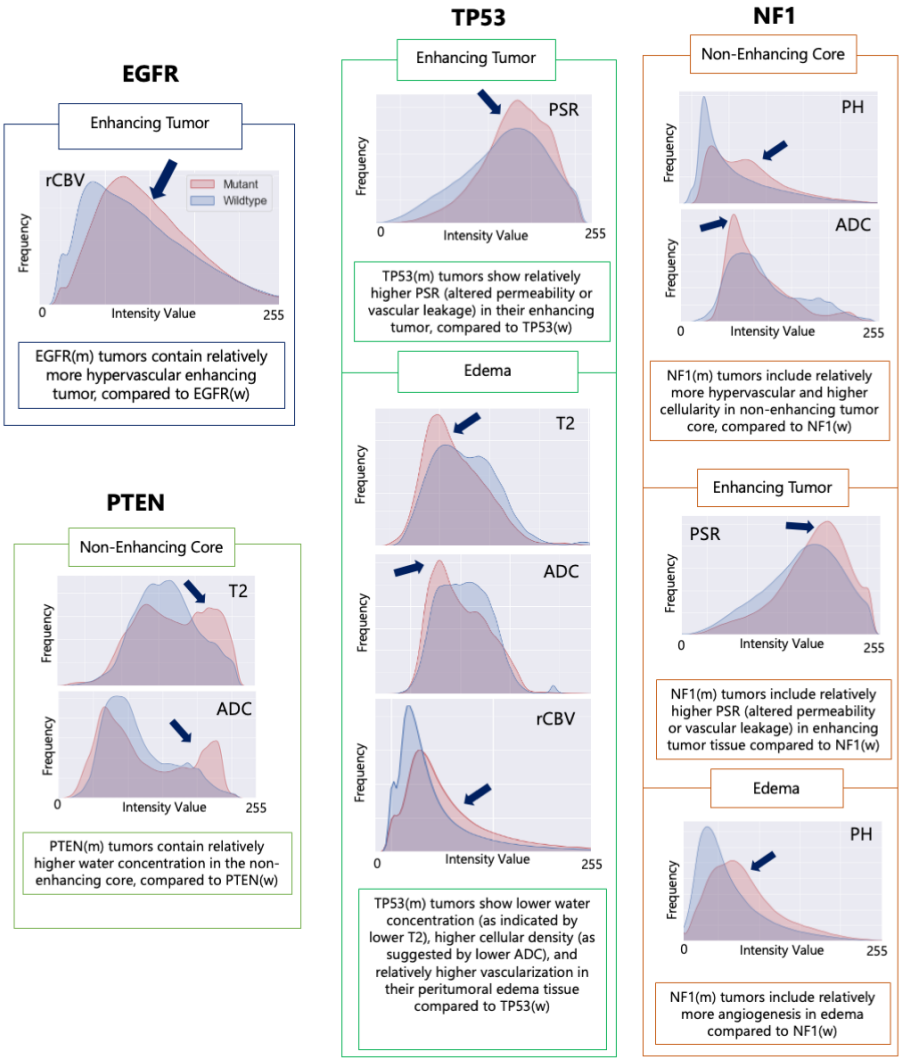
Numbers in parentheses indicate percentages. Std = standard deviation. AUC = Area Under the receiver operating characteristic (ROC) Curve. CI = Confidence Interval

Supplementary Table 2. A summary of the copy number variations (CNVs) in the key driver pathways in the cohort.

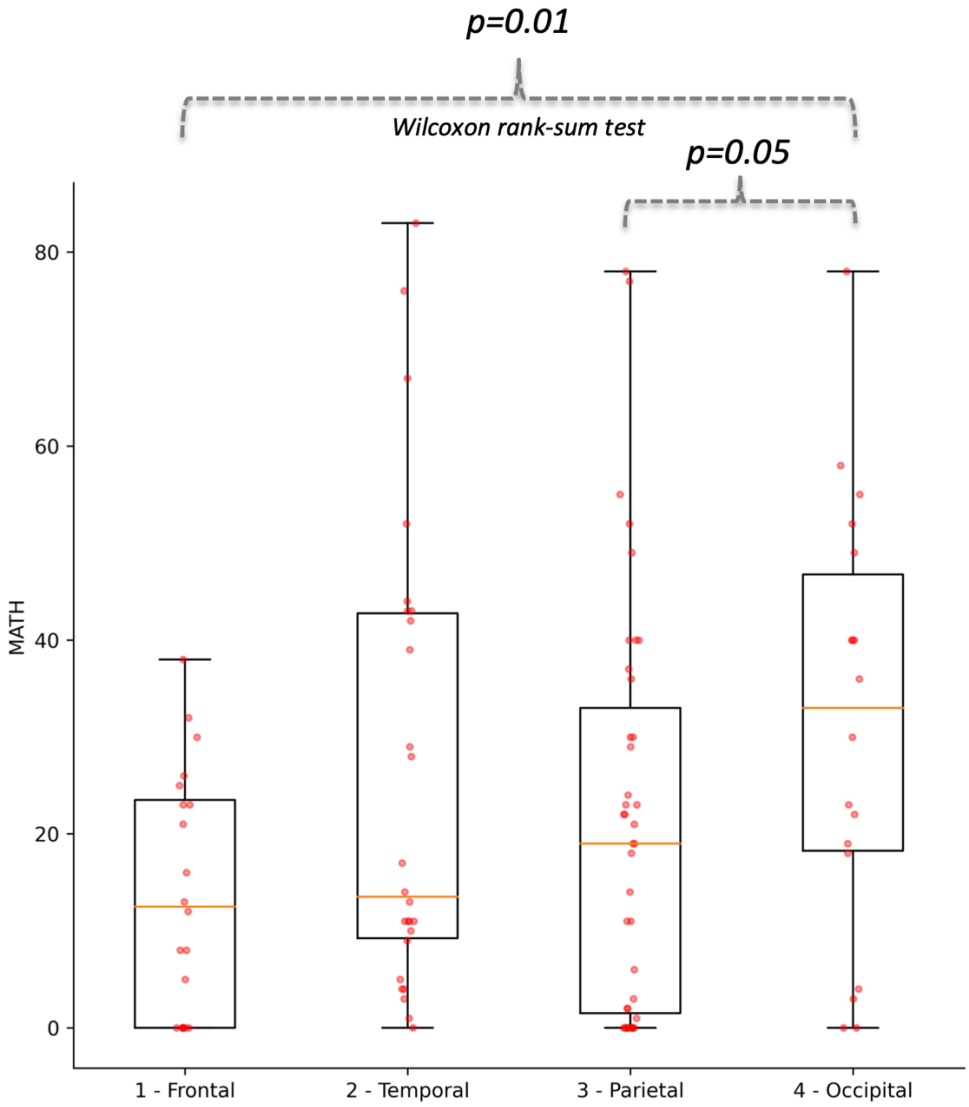
Pathways	Alterations	No. of Patients with Alteration (Frequency)
RTK	***EGFR* gain**	127/177 (71.7%)
	***MET* gain**	145/177 (81.9%)
	***PDGFRA* gain**	20/177 (11.3%)
	***FGFR2* loss**	110/177 (62.1%)
PI3K	***PTEN* loss**	37/177 (20.9%)
	***PIK3CA* gain**	15/177 (8.5%)
	***PIK3R1* gain**	22/177 (12.4%)
P53	***TP53* loss**	41/177 (23.2%)
	***MDM4* gain**	18/177 (10.2%)
RB1	***RB1* loss**	3/177 (1.7%)
	***CDKN2A* loss**	47/177 (26.6%)



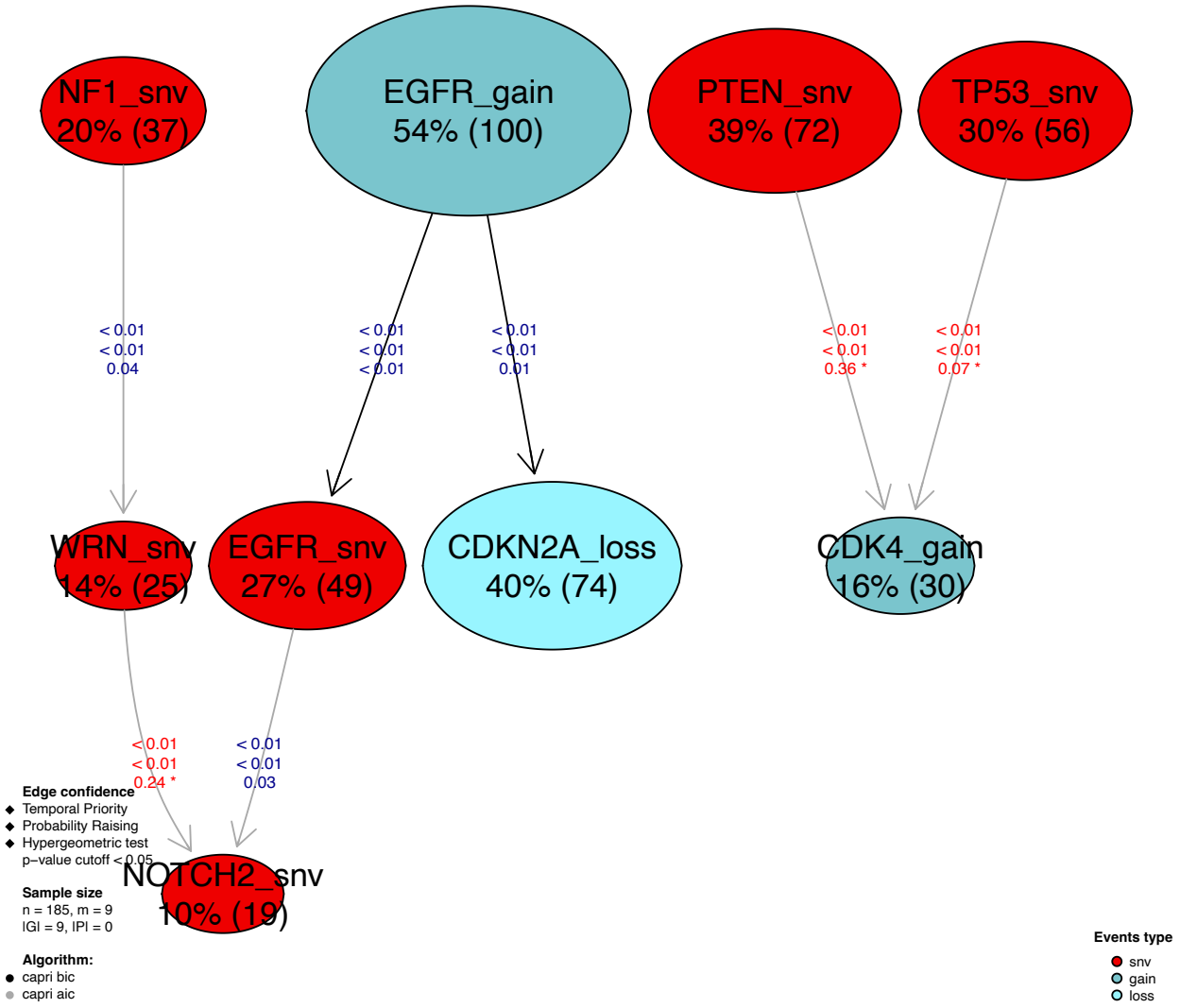
Supplementary Figure 1. Modified 34-layer ResNet architecture with an initial 7×7 convolution followed by four layers (comprising 3, 4, 6, and 3 residual blocks, each with two 3×3 convolutions). Input to the model included slices from T1-Gd and T2 structural images, along with tumor segmentation masks (three image volumes). After pretraining, weights from the initial layers up to layer 3 were transferred to the CNN classifier but remained fixed. Radiomic features, including spatial location, were fed into fully connected layers after layer 4, which, along with layer 4 weights, were trained for classification.



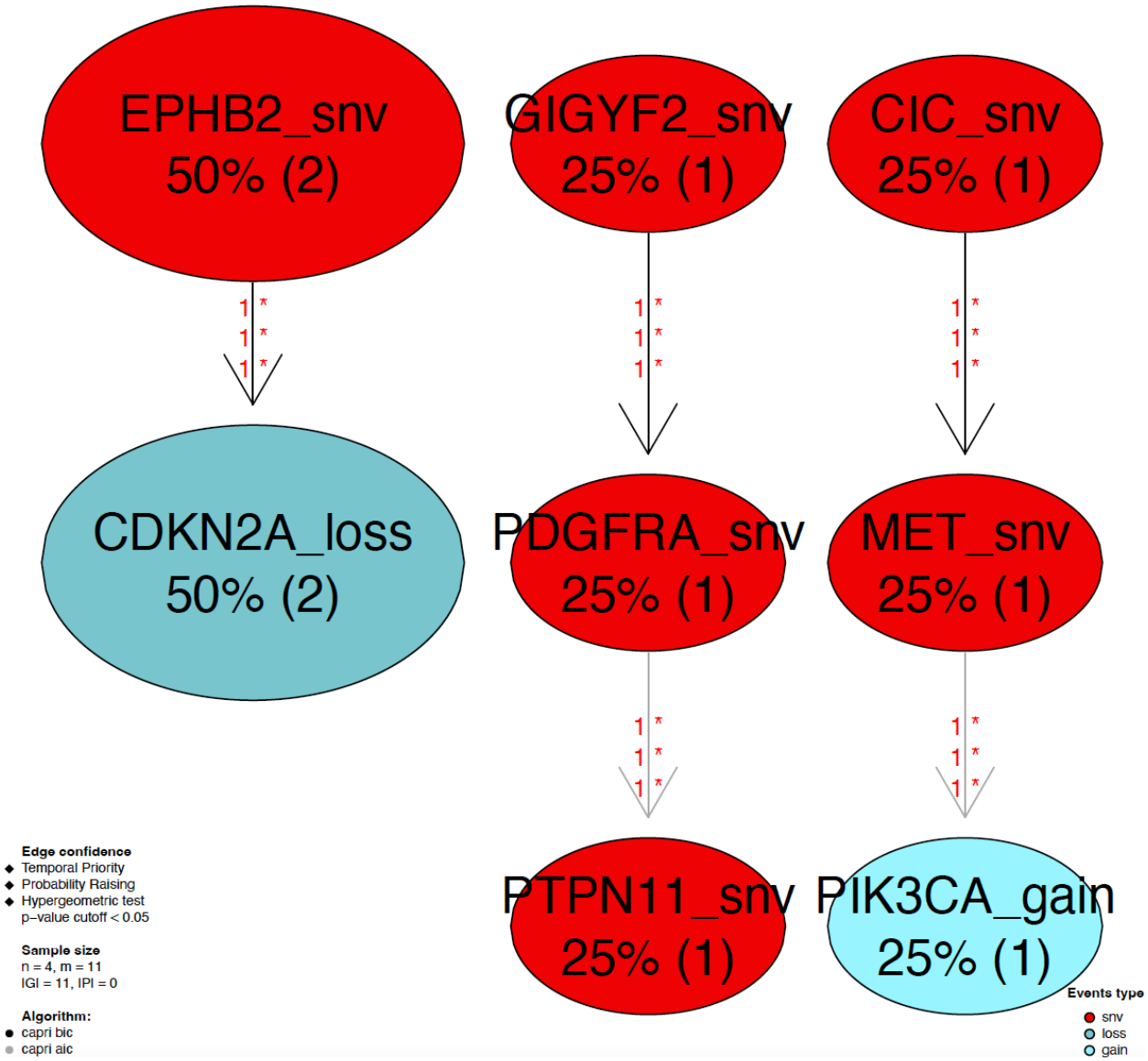
Supplementary Figure 2. Representative imaging characteristics in different tumorous subregions for the tumors with mutations in key driver genes in comparison to their wildtype counterparts [rCBV = relative cerebral blood volume; PSR = Percentage Signal Recovery; ADC = Apparent Diffusion Coefficient; PH = Peak Height]. The annotation (m) or (w) after the names of genes, e.g., EGFR(m) or EGFR(w), in the boxes denote mutant or wildtype tumors, respectively. Edema refers to peritumoral edema, which is an infiltrated zone surrounding the tumor comprised of vasogenic edema and infiltrating tumor cells diffused in the adjacent normal brain tissue.



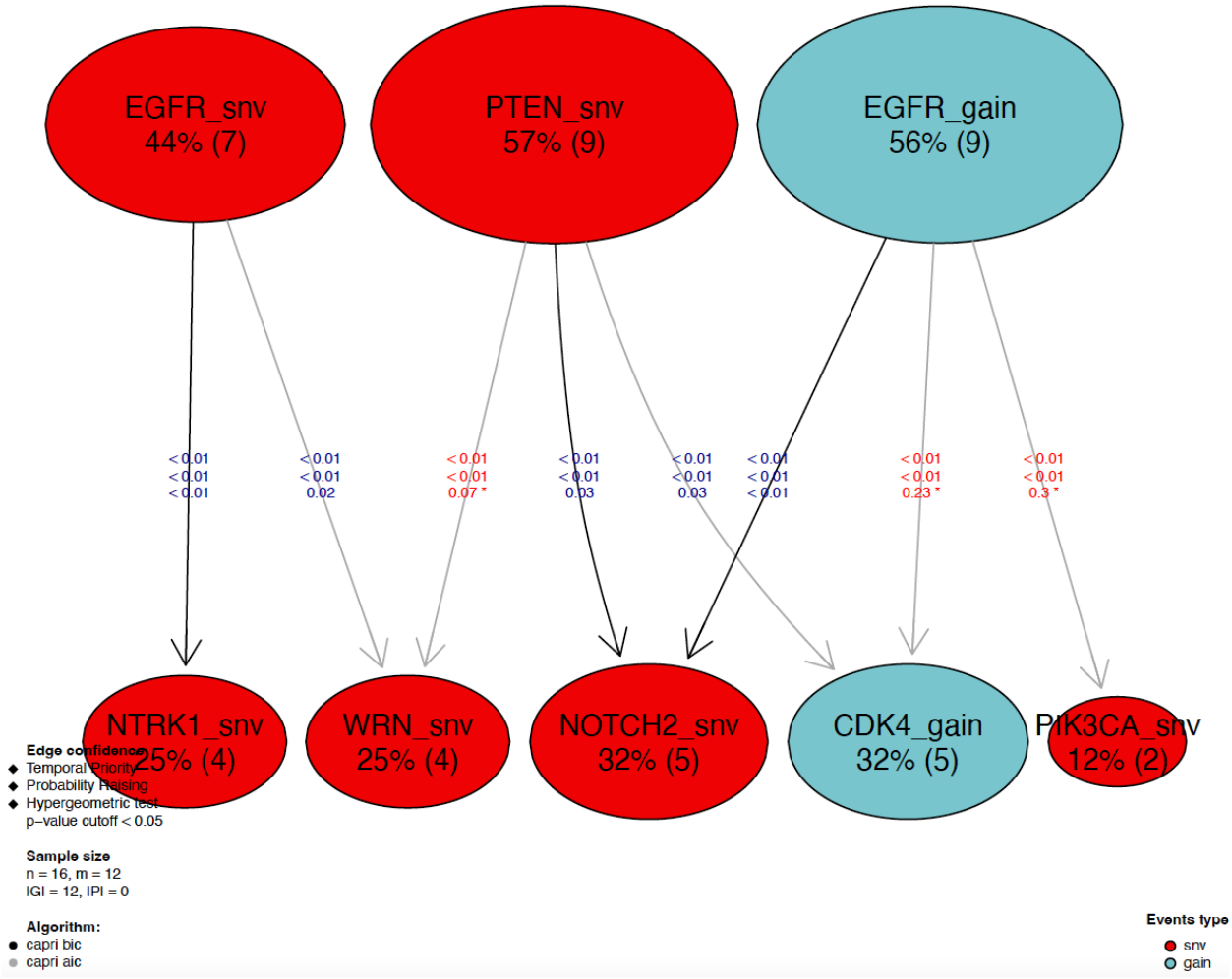
Supplementary Figure 3. Mutant Allele Tumor Heterogeneity (MATH) scores across frontal, temporal, parietal, and occipital brain regions. The differences in MATH scores were determined using Wilcoxon rank-sum test, indicating a statistically significant difference in intra-tumoral heterogeneity between the frontal and occipital regions (p = 0.01).



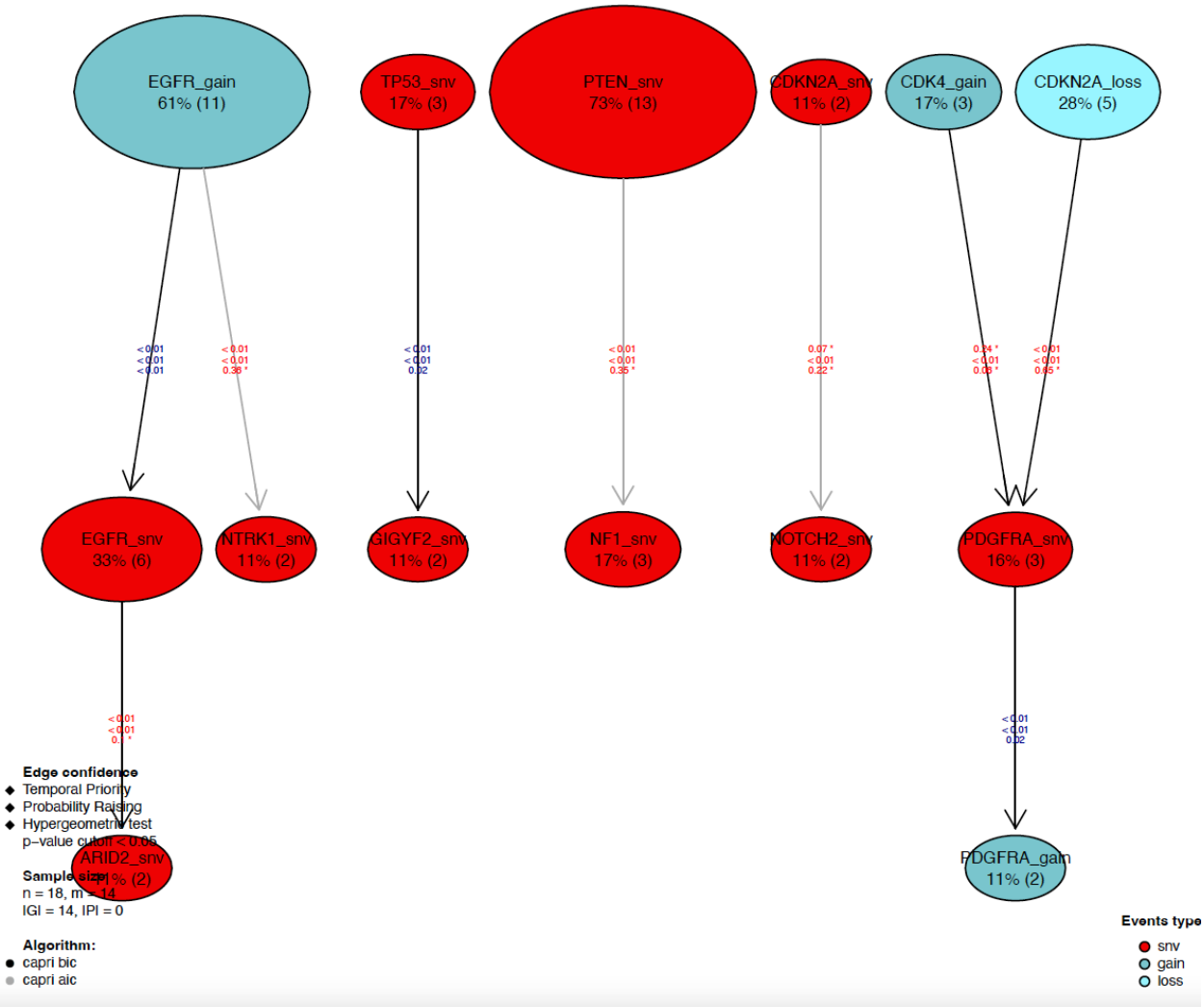
Supplementary Figure 4. Oncogenic drivers in de novo GBMs based on mutated genes in six signaling pathways present in over 10% of samples. The model identified early drivers, including EGFR amplification and somatic mutations in NF1, PTEN, and TP53. The suffix for each gene name indicates somatic point mutation (snv), copy number gains (gain) or copy number losses (loss) observed in the cohort. The corresponding relative and absolute frequencies in the cohort are listed as well.



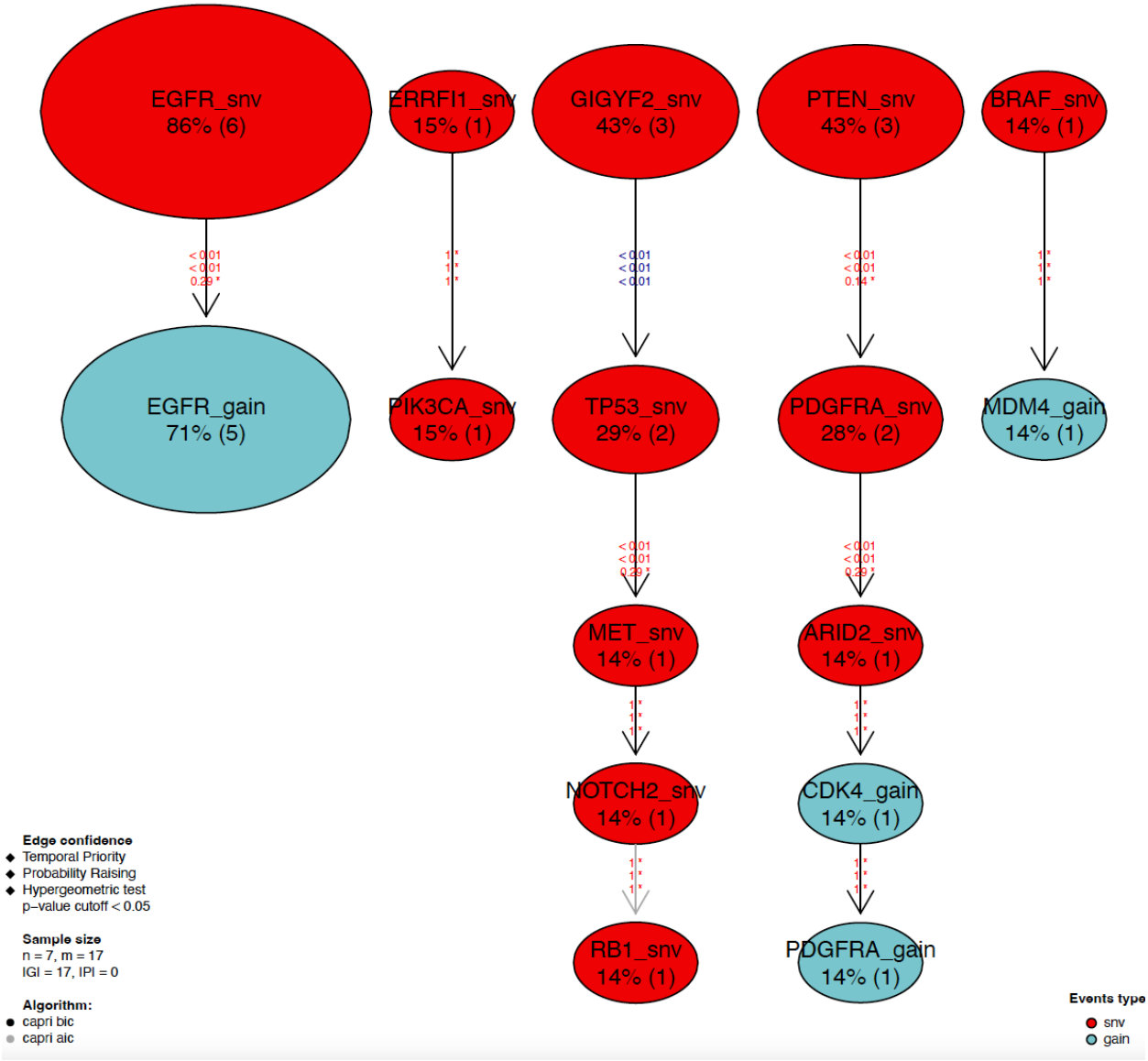
Supplementary Figure 5. Oncogenic drivers in Group 1 tumors. The model identified early drivers and subsequent genomic alteration events along three main trajectories. The suffix for each gene name indicates somatic point mutation (snv), copy number gains (gain) or copy number losses (loss) observed in the cohort. The corresponding relative and absolute frequencies in the cohort are listed as well.



Supplementary Figure 6. Oncogenic drivers in Group 2 tumors. The model identified early drivers and subsequent genomic alteration events. EGFR and PTEN alterations were early events. The suffix for each gene name indicates somatic point mutation (snv), copy number gains (gain) or copy number losses (loss) observed in the cohort. The corresponding relative and absolute frequencies in the cohort are listed as well.



Supplementary Figure 7. Oncogenic drivers in Group 3 tumors. The model identified early drivers and subsequent genomic alteration events. The suffix for each gene name indicates somatic point mutation (snv), copy number gains (gain) or copy number losses (loss) observed in the cohort. The corresponding relative and absolute frequencies in the cohort are listed as well.



Supplementary Figure 8. Oncogenic drivers in Group 4 tumors. The model identified early drivers and subsequent genomic alteration events. The suffix for each gene name indicates somatic point mutation (snv), copy number gains (gain) or copy number losses (loss) observed in the cohort. The corresponding relative and absolute frequencies in the cohort are listed as well.