



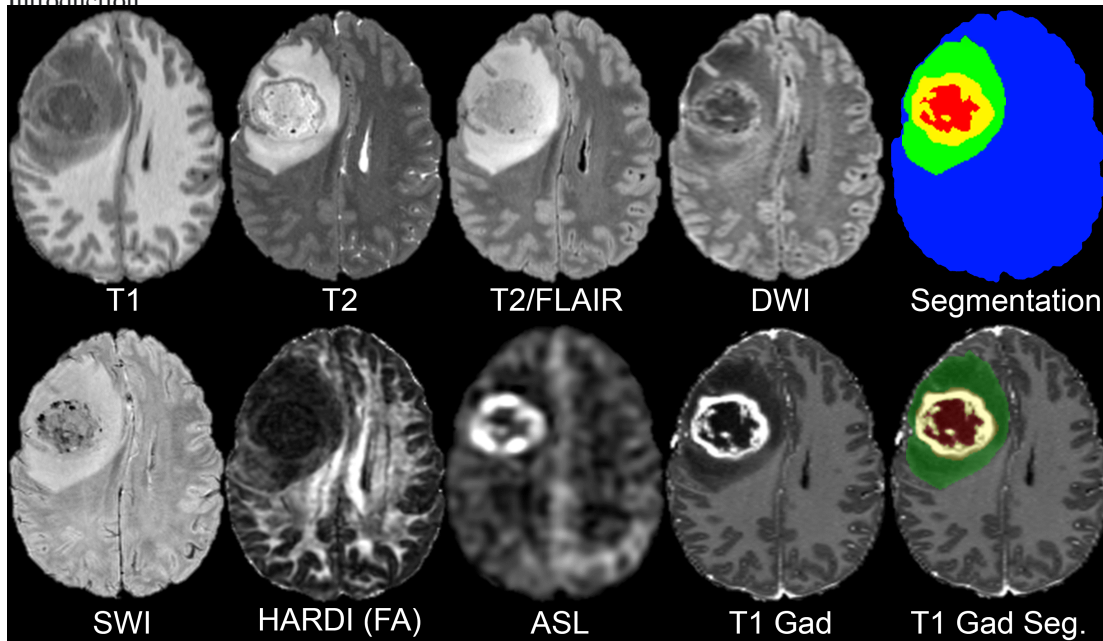
# **The University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM)**

## Summary

### Redirection Notice

This page will redirect to <https://www.cancerimagingarchive.net/collection/ucsf-pdgm/> in about 5 seconds.

#### Introduction



MRI-based artificial intelligence (AI) research on patients with brain gliomas has been rapidly increasing in popularity in recent years in part due to a growing number of publicly available MRI datasets. Notable examples include [The Cancer Genome Atlas Glioblastoma dataset \(TCGA-GBM\)](#) consisting of 262 subjects and the International Brain Tumor Segmentation (BraTS) challenge dataset consisting of 542 subjects (including 243 preoperative cases from TCGA-GBM). The public availability of these glioma MRI datasets has fostered the growth of numerous emerging AI techniques including automated tumor segmentation, radiogenomics, and MRI-based survival prediction. Despite these advances, existing publicly available glioma MRI datasets have been largely limited to only 4 MRI contrasts (T2, T2/FLAIR, and T1 pre- and post-contrast) and imaging protocols vary significantly in terms of magnetic field strength and acquisition parameters. Here we present the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) dataset. The UCSF-PDGM dataset includes 501 subjects with histopathologically-proven diffuse gliomas who were imaged with a standardized 3 Tesla preoperative brain tumor MRI protocol featuring predominantly 3D imaging, as well as advanced diffusion and perfusion imaging techniques. The dataset also includes isocitrate dehydrogenase (IDH) mutation status for all cases and O[6]-methylguanine-DNA methyltransferase (MGMT) promotor methylation status for World Health Organization (WHO) grade III and IV gliomas. The UCSF-PDGM has been made publicly available in the hopes that researchers around the world will use these data to continue to push the boundaries of AI applications for diffuse gliomas.

## Methods

### Patient Population

Data collection was performed in accordance with relevant guidelines and regulations and was approved by the University of California San Francisco institutional review board with a waiver for consent. The dataset population consisted of 501\* adult patients with histopathologically confirmed grade II-IV diffuse gliomas who underwent preoperative MRI, initial tumor resection, and tumor genetic testing at a single medical center between 2015 and 2021. Patients with any prior history of brain tumor treatment were excluded; however, history of tumor biopsy was not considered an exclusion criterion.

### Genetic Biomarker Testing

All subjects' tumors were tested for IDH mutations by genetic sequencing of tissue acquired during biopsy or resection. All grade III and IV tumors were tested for MGMT methylation status using a methylation sensitive quantitative PCR assay.

### Study participant demographic data

The 501\* cases included in the UCSF-PDGM include 55 (11%) grade II, 42 (9%) grade III, and 403 (80%) grade IV tumors. There was a male predominance for all tumor grades (56%, 60%, and 60%, respectively for grades II-IV). IDH mutations were identified in a majority of grade II (83%) and grade III (67%) tumors and a small minority of grade IV tumors (8%). MGMT promoter hypermethylation was detected in 63% of grade IV gliomas and was not tested for in a majority of lower grade gliomas. 1p/19q codeletion was detected in 20% of grade II tumors and a small minority of grade III (5%) and IV (<1%) tumors. Tabulated details and glossary are available in the Data Access and Detailed Description tabs below.

### Image Acquisition

All preoperative MRI was performed on a 3.0 tesla scanner (Discovery 750, GE Healthcare, Waukesha, Wisconsin, USA) and a dedicated 8-channel head coil (Invivo, Gainesville, Florida, USA). The imaging protocol included 3D T2-weighted, T2/FLAIR-weighted, susceptibility-weighted (SWI), diffusion-weighted (DWI), pre- and post-contrast T1-weighted images, 3D arterial spin labeling (ASL) perfusion images, and 2D 55-direction high angular resolution diffusion imaging (HARDI). Over the study period, two gadolinium-based contrast agents were used: gadobutrol (Gadovist, Bayer, LOC) at a dose of 0.1 mL/kg and gadoterate (Dotarem, Guerbet, Aulnay-sous-Bois, France) at a dose of 0.2 mL/kg.

### Image Pre-Processing

HARDI data were eddy current corrected and processed using the Eddy and DTIFIT modules from FSL 6.0.2 yielding isotropic diffusion weighted images (DWI) and several quantitative diffusivity maps: mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA). Eddy correction was performed with outlier replacement on and topup correction off. DTIFIT was performed with simple least squares regression. Each image contrast was registered and resampled to the 3D space defined by the T2/FLAIR image (1 mm isotropic resolution) using automated non-linear registration (Advanced Normalization Tools). Resampled co-registered data were then skull stripped using a previously described and publicly available deep-learning algorithm: [https://www.github.com/ecalabr/brain\\_mask/](https://www.github.com/ecalabr/brain_mask/).

## Tumor Segmentation

Multicompartment tumor segmentation of study data was undertaken as part of the 2021 BraTS challenge. Briefly, image data first underwent automated segmentation using an ensemble model consisting of prior BraTS challenge winning segmentation algorithms. Images were then manually corrected by trained radiologists and approved by 2 expert reviewers. Segmentation included three major tumor compartments: enhancing tumor, non-enhancing/necrotic tumor, and surrounding FLAIR abnormality (sometimes referred to as edema).

The UCSF-PDGM adds to an existing body of publicly available diffuse glioma MRI datasets that are commonly used in AI research applications. As MRI-based AI research applications continue to grow, new data are needed to foster development of new techniques and increase the generalizability of existing algorithms. The UCSF-PDGM not only significantly increases the total number of publicly available diffuse glioma MRI cases, but also provides a unique contribution in terms of MRI technique. The inclusion of 3D sequences and advanced MRI techniques like ASL and HARDI provides a new opportunity for researchers to explore the potential utility of cutting-edge clinical diagnostics for AI applications. In addition, these advanced imaging techniques may prove useful for radiogenomic studies focused on identification of IDH mutations or MGMT promoter methylation.

The UCSF-PDGM dataset, particularly when combined with existing publicly available datasets, has the potential to fuel the next phase of radiologic AI research on diffuse gliomas. However, the UCSF-PDGM dataset's potential will only be realized if the radiology AI research community takes advantage of this new data resource. We hope that this dataset sparks inspiration in the next generation of AI researchers, and we look forward to the new techniques and discoveries that the UCSF-PDGM will generate.

## Acknowledgements

We would like to acknowledge the individuals and institutions that have provided data for this collection:

- Research was supported by the National Institutes of Health Ruth L. Kirschstein Institutional National Research Service Award under award number T32EB001631 and by the RSNA Research & Education Foundation under grant number RR2011. The content is solely the responsibility of the authors and does not necessarily represent the official views of the RSNA R&E Foundation.

## Data Access

### Data Access

Data Type	Download all or Query/Filter	License
Images and Annotations (NIfTI format , 156 GB)	<a href="#">Download</a> (Download and apply the <a href="#">IBM-Aspera-Connect plugin</a> to your browser to retrieve this faspex package)	CC BY 4.0
Clinical data (CSV)	<a href="#">Download</a>	CC BY 4.0
Clinical metadata glossary (CSV)	<a href="#">Download</a>	CC BY 4.0
bval file (BVAL)	<a href="#">Download</a>	CC BY 4.0
bvec file (BVEC)	<a href="#">Download</a>	CC BY 4.0

## Detailed Description

### Detailed Description

Image Statistics	
Modalities	MR
Number of Patients	495
Number of Studies	501
Number of Files	11,523
Images Size (GB)	156.5 GB

**\*Note: it was discovered that some ID are followup imaging of others, therefore the true number of patients in this Collection is 495 not 501, the pixels have not been changed:**

UCSF-PDGM-0433 was imaged 7 days later as UCSF-PDGM-0315 (version 3 files reflect this change by renaming UCSF-PDGM-0315 to UCSF-PDGM-0433\_FU007d)

UCSF-PDGM-0431 was imaged 1 days later as UCSF-PDGM-0278 (version 3 files reflect this change by renaming UCSF-PDGM-0278 to UCSF-PDGM-0431\_FU001d)

UCSF-PDGM-0396 was imaged 175 days later as UCSF-PDGM-0175 (version 3 files reflect this change by renaming UCSF-PDGM-0175 to UCSF-PDGM-0396\_FU175d)

UCSF-PDGM-0429 was imaged 3 days later as UCSF-PDGM-0138 (version 3 files reflect this change by renaming UCSF-PDGM-0138 to UCSF-PDGM-0429\_FU003d)

UCSF-PDGM-0409 was imaged 1 days later as UCSF-PDGM-0181 (version 3 files reflect this change by renaming UCSF-PDGM-0181 to UCSF-PDGM-0409\_FU001d)

UCSF-PDGM-0391 was imaged 16 days later as UCSF-PDGM-0289 (version 3 files reflect this change by renaming UCSF-PDGM-0289 UCSF-PDGM-0391\_FU016d)

All image data have been "skull stripped", deidentified, pre-processed per the methods section of our abstract, and converted to NIfTI format. We cannot provide original DICOM data, however, these pre-processed files have been prepared to facilitate the type of research that this dataset is intended for. Publicly available deep-learning algorithm for performing this process is here: [https://www.github.com/ecalabr/brain\\_mask/](https://www.github.com/ecalabr/brain_mask/).

Glossary of abbreviations: [UCSF-PDGM-metadata\\_glossary.csv](#)

Term	Represents	Values
ID	DICOM (0010,0020) PatientID	
Sex	DICOM (0010,0040) Patient Sex	M,F
Age at MRI	Age in years at time of MR imaging	
WHO CNS Grade	Grade per the 2021 World Health Organization Classification of Tumors of the Central Nervous System (WHO CNS 2021) ( <a href="https://doi.org/10.1093/neuonc/noab106">https://doi.org/10.1093/neuonc/noab106</a> )	2,3,4



Final pathologic diagnosis (WHO 2021)	Final (integrated) pathologic diagnosis per the 2021 World Health Organization Classification of Tumors of the Central Nervous System (WHO CNS 2021) ( <a href="https://doi.org/10.1093/neuonc/noab106">https://doi.org/10.1093/neuonc/noab106</a> )	<ul style="list-style-type: none"> <li>• Glioblastoma, Isocitrate dehydrogenase (IDH) - wildtype,</li> <li>• Astrocytoma, IDH-mutant,</li> <li>• Astrocytoma, IDH-wildtype,</li> <li>• Oligodendroglioma, IDH-mutant, 1p/19q-codeleted</li> </ul>
MGMT status	O6-methylguanine-DNA methyltransferase status - clinical interpretation of the MGMT index described below.	negative, positive, indeterminate
MGMT index	O6-methylguanine-DNA methyltransferase methylation index (in house method developed by UCSF clinical labs, <a href="https://genomics.ucsf.edu/content/mgmt-promoter-methylation-assay">https://genomics.ucsf.edu/content/mgmt-promoter-methylation-assay</a> ) where numeric values 0-17 indicate the number of promoter methylation sites.	0-17, blank
1p/19q	presence of codeletion of 1p and 19q genes, assayed by fluorescent in-situ hybridization	intact, co-deletion, relative co-deletion, unknown
IDH	isocitrate dehydrogenase mutation subtype characterized with a capture-based targeted next-generation DNA sequencing panel (UCSF500) as described in ( <a href="https://doi.org/10.1093/neuonc/now254">https://doi.org/10.1093/neuonc/now254</a> )	
1-dead 0-alive	Survival status of the patient at last clinical follow up.	
OS	Overall survival in days from initial diagnosis to last clinical follow up.	
EOR	Extent of resection determined by review of operative reports and immediate postoperative imaging	biopsy (only biopsy) Subtotal resection (STR) gross total resection (GTR)
Biopsy prior to imaging	Was a burr-hole biopsy performed prior to imaging	yes, no, blank

Note that The L1/L2/L3 files the eigenvalues. These are direct outputs from FSL DTIFIT as described here: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT/UserGuide#DTIFIT> . We elected not to include the tensor files since they are quite large and are straightforward to derive from the provided 4D DWI data using FSL DTIFIT. The link provided above also includes instructions for how to do this in case the user is not familiar.

#### Citations & Data Usage Policy

### Citations & Data Usage Policy

Users must abide by the [TCIA Data Usage Policy and Restrictions](#). Attribution should include references to the following citations:

#### Data Citation

Calabrese, E., Villanueva-Meyer, J., Rudie, J., Rauschecker, A., Baid, U., Bakas, S., Cha, S., Mongan, J., & Hess, C. (2022). **The University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) (Version 4) [Dataset]**. The Cancer Imaging Archive. DOI: [10.7937/tcia.bdgf-8v37](https://doi.org/10.7937/tcia.bdgf-8v37)

#### Publication Citation

Evan Calabrese, Javier E. Villanueva-Meyer, Jeffrey D. Rudie, Andreas M. Rauschecker, Ujjwal Baid, Spyridon Bakas, Soonmee Cha, John T. Mongan, Christopher P. Hess. (2022) **The UCSF Preoperative Diffuse Glioma MRI (UCSF-PDGM) Dataset**. Radiology: Artificial Intelligence. DOI: <https://doi.org/10.1148/ryai.220058>

#### TCIA Citation

Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: [10.1007/s10278-013-9622-7](https://doi.org/10.1007/s10278-013-9622-7)

### Other Publications Using This Data

The following publications are recommended by the data submitters that may be useful to researchers utilizing this collection:

- Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Cha S. **Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower Grade Gliomas Using 3D Fully Convolutional Neural Networks**. Radiology: Artificial Intelligence. 2021 May 19;e200276.
- Calabrese E, Villanueva-Meyer JE, Cha S. **A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas**. Scientific Reports. 2020 Jul 16;10(1):11852.

TCIA maintains [a list of publications](#) which leverage TCIA data. If you have a manuscript you'd like to add please [contact TCIA's Helpdesk](#).

#### Versions

### Version 4 (Current): Updated 2023/04/07

Data Type	Download all or Query/Filter	License
-----------	------------------------------	---------

Images and Annotations (NIfTI format , 156 GB)	<a href="#">Download</a> (Download and apply the <a href="#">IBM-Aspera-Connect plugin</a> to your browser to retrieve this faspex package)	CC BY 4.0
Clinical data (CSV)	<a href="#">Download</a>	CC BY 4.0
Clinical metadata glossary	<a href="#">Download</a>	CC BY 4.0
bval file	<a href="#">Download</a>	CC BY 4.0
bvec file	<a href="#">Download</a>	CC BY 4.0

Added bvec and bval to data access table. Added metadata glossary from Detailed description to Data Access table

### Version 3: Updated 2023/01/11

Data Type	Download all or Query/Filter	License
Images and Annotations (NIfTI format , 156 GB)	<a href="#">Download</a> (Download and apply the <a href="#">IBM-Aspera-Connect plugin</a> to your browser to retrieve this faspex package)	CC BY 4.0
Clinical data (CSV)	<a href="#">Download</a>	CC BY 4.0

\*Note: it was discovered that some ID are followup imaging of others, therefore **the true number of patients in this Collection is 495**, the pixels have not been changed:

UCSF-PDGM-0433 was imaged 7 days later as UCSF-PDGM-0315 (version 3 files reflect this change by renaming UCSF-PDGM-0315 to UCSF-PDGM-0433\_FU007d)

UCSF-PDGM-0431 was imaged 1 day later as UCSF-PDGM-0278 (version 3 files reflect this change by renaming UCSF-PDGM-0278 to UCSF-PDGM-0431\_FU001d)

UCSF-PDGM-0396 was imaged 175 days later as UCSF-PDGM-0175 (version 3 files reflect this change by renaming UCSF-PDGM-0175 to UCSF-PDGM-0396\_FU175d)

UCSF-PDGM-0429 was imaged 3 days later as UCSF-PDGM-0138 (version 3 files reflect this change by renaming UCSF-PDGM-0138 to UCSF-PDGM-0429\_FU003d)

UCSF-PDGM-0409 was imaged 1 day later as UCSF-PDGM-0181 (version 3 files reflect this change by renaming UCSF-PDGM-0181 to UCSF-PDGM-0409\_FU001d)

UCSF-PDGM-0391 was imaged 16 days later as UCSF-PDGM-0289 (version 3 files reflect this change by renaming UCSF-PDGM-0289 UCSF-PDGM-0391\_FU016d)

## Version 2: Updated 2022/11/30:

Data Type	Download all or Query/Filter	License
Images and Annotations (NIfTI format , 156 GB)	<a href="#">Download</a> (Download and apply the <a href="#">IBM-Aspera-Connect plugin</a> to your browser to retrieve this faspex package)	CC BY 4.0
Clinical data (CSV)	<a href="#">Download</a>	CC BY 4.0

Changes to this version:

1. Fixes to integer rounding errors in \*tumor\_segmentation.nii.gz files by the Collection's investigators.
2. TCIA users have asked for a mapping to the **2021 BraTS data** to prevent any data leak *when both datasets are used*. Updated metadata spreadsheet [attached](#) which includes the BraTS IDs for all relevant cases. This data was confirmed by BraTS Challenge organizers.
3. Correction of name suffix for bias corrected T1 postcontrast images from "T1gad\_bias" to "T1c\_bias" for consistency.

## Version 1: Updated 2022/09/26

Data Type	Download all or Query/Filter	License
Images and Annotations (NIfTI zip, 156 GB)	(deprecated)	
Clinical data (CSV)	<a href="#">Download</a>	CC BY 4.0