Foundation models for fast, label-free detection of glioma infiltration

https://doi.org/10.1038/s41586-024-08169-3

Received: 7 March 2024

Accepted: 8 October 2024

Published online: 13 November 2024

Open access

Check for updates

Akhil Kondepudi[1,2], Melike Pekmezci[3], Xinhai Hou[1,2], Katie Scotford[4], Cheng Jiang[1,2], Akshay Rao[1], Edward S. Harake[1], Asadur Chowdury[1], Wajd Al-Holou[5], Lin Wang[1], Aditya Pandey[5], Pedro R. Lowenstein[5], Maria G. Castro[5], Lisa Irina Koerner[6], Thomas Roetzer-Pejrimovsky[7,13], Georg Widhalm[6], Sandra Camelo-Piragua[8], Misha Movahed-Ezazi[9], Daniel A. Orringer[10], Honglak Lee[11], Christian Freudiger[12], Mitchel Berger[4], Shawn Hervey-Jumper[4]✉ & Todd Hollon[1,5]✉

A critical challenge in glioma treatment is detecting tumour infiltration during surgery to achieve safe maximal resection[1–3]. Unfortunately, safely resectable residual tumour is found in the majority of patients with glioma after surgery, causing early recurrence and decreased survival[4–6]. Here we present FastGlioma, a visual foundation model for fast (<10 s) and accurate detection of glioma infiltration in fresh, unprocessed surgical tissue. FastGlioma was pretrained using large-scale self-supervision (around 4 million images) on rapid, label-free optical microscopy, and fine-tuned to output a normalized score that indicates the degree of tumour infiltration within whole-slide optical images. In a prospective, multicentre, international testing cohort of patients with diffuse glioma ($n = 220$), FastGlioma was able to detect and quantify the degree of tumour infiltration with an average area under the receiver operating characteristic curve of 92.1 ± 0.9%. FastGlioma outperformed image-guided and fluorescence-guided adjuncts for detecting tumour infiltration during surgery by a wide margin in a head-to-head, prospective study ($n = 129$). The performance of FastGlioma remained high across diverse patient demographics, medical centres and diffuse glioma molecular subtypes as defined by the World Health Organization. FastGlioma shows zero-shot generalization to other adult and paediatric brain tumour diagnoses, demonstrating the potential for our foundation model to be used as a general-purpose adjunct for guiding brain tumour surgeries. These findings represent the transformative potential of medical foundation models to unlock the role of artificial intelligence in the care of patients with cancer.

The importance of detecting tumour infiltration within surgical specimens during an operation has been recognized for over a century[7]. Despite our best efforts to deliver precision healthcare to patients with cancer, residual tumour after surgery is a major public health problem in the United States and globally[8]. For solid cancers and brain tumours, residual tumour results in worse quality of life, decreased patient survival and increased burden on healthcare systems[2,3,9,10]. Rates of residual tumour have not improved over the past two decades, and corrective surgical procedures and post-surgical treatment have an estimated total cost of more than US$1 billion annually in the United States[9,11].

Here we present FastGlioma, an open-source, artificial intelligence (AI)-based diagnostic system for detecting brain tumour infiltration in fresh, unprocessed, unlabelled surgical tissue at the patient's bedside (an interactive demo is available at https://fastglioma.mlins.org). Conventional microscopy analysis with haematoxylin and eosin (H&E)-stained tissue during surgery is slow, resource intensive and reliant on a shrinking pathology workforce[12]. FastGlioma solves these limitations by combining rapid, user-friendly, bedside optical microscopy and visual foundation models trained on a diverse dataset of over 11,000 surgical specimens and 4 million unique microscopy fields of view. Foundation models, such as the GPT-4 and DALL-E3 models, are AI models that are trained on massive, diverse datasets and can be adapted to a wide range of downstream tasks[13–16]. Foundation models for medical AI have the potential to solve challenging

[1]Machine Learning in Neurosurgery Laboratory, Department of Neurosurgery, University of Michigan, Ann Arbor, MI, USA. [2]Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [3]Department of Pathology, University of California, San Francisco, San Francisco, CA, USA. [4]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA. [5]Department of Neurosurgery, University of Michigan, Ann Arbor, MI, USA. [6]Department of Neurosurgery, Medical University Vienna, Vienna, Austria. [7]Comprehensive Center for Clinical Neurosciences and Mental Health, Medical University of Vienna, Vienna, Austria. [8]Department of Pathology, University of Michigan, Ann Arbor, MI, USA. [9]Department of Pathology, Rutgers University, New Brunswick, NJ, USA. [10]Department of Neurosurgery, New York University, New York, NY, USA. [11]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. [12]Invenio Imaging, Inc., Santa Clara, CA, USA. [13]Present address: Division of Neuropathology and Neurochemistry, Department of Neurology, Medical University Vienna, Vienna, Austria. ✉e-mail: Shawn.Hervey-Jumper@ucsf.edu; tocho@med.umich.edu

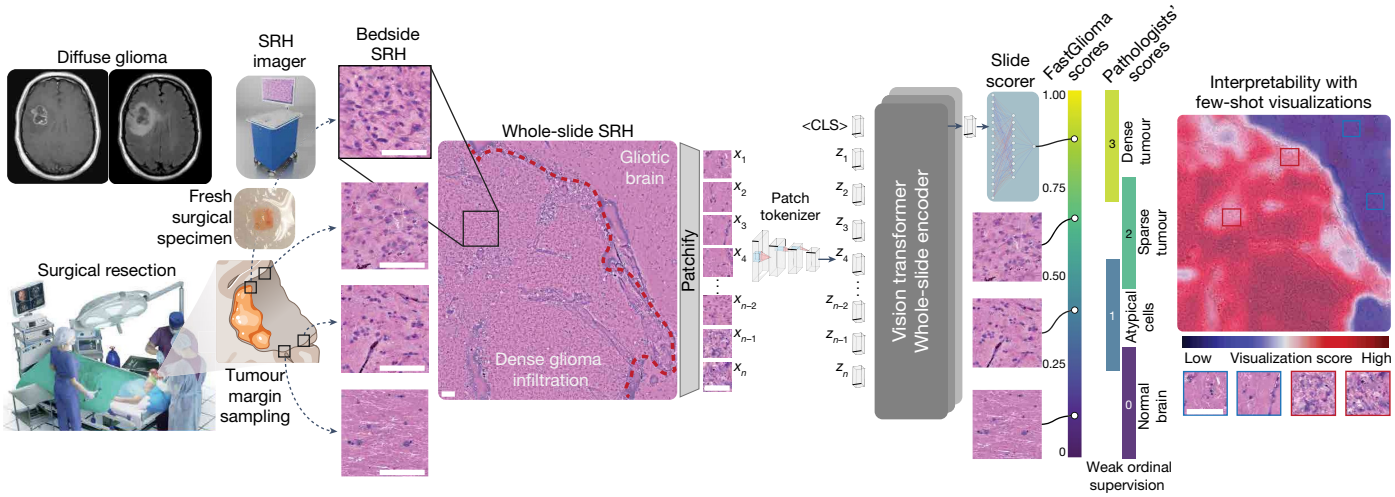


Fig. 1 | FastGlioma workflow. A patient with a suspected diffuse glioma undergoes surgical resection. During tumour resection, the surgeon samples tissue from the surgical margin. The portable SRH imaging system acquires microscopy images in the operating room, performed by a single technician using simple touchscreen instructions. A freshly excised surgical specimen is loaded directly into a custom microscope slide and inserted into the SRH imager without the need for tissue processing. Additional details on image acquisition are provided in Extended Data Fig. 1. SRH images can be virtually stained using an H&E-like colour scheme for clinician review as shown above[20]. A whole-slide SRH image is divided into patches and each patch undergoes a feedforward pass through a patch tokenizer (Extended Data Fig. 3a). The patch tokens, plus an appended classification token <CLS>, are then input into a whole-slide SRH encoder that is a vision transformer. The patch tokenizer and whole-slide encoder are pretrained as a visual foundation model using large-scale self-supervision (Extended Data Fig. 3b). For tumour-infiltration scoring, a slide scorer model is fine-tuned to output a normalized continuous score between 0 and 1 that predicts the degree of tumour infiltration within the whole-slide image that corresponds to a four-tier whole-slide ordinal infiltration scale as defined by expert neuropathologists (Extended Data Figs. 2 and 4). Ordinal labels are weak because they apply to the slide level only. Despite the weak labels, FastGlioma provides regional interpretability by identifying areas within whole-slide SRH images with a high probability of tumour infiltration. Scale bars, 100 µm. Tumour resection and microscopy slide images were adapted from ref. 4. The operating room graphic was adapted from ref. 48. The neural network architectures were adapted from https://alexlenail.me/NN-SVG/.

clinical tasks by requiring little to no task-specific annotations to achieve good generalization performance and clinically meaningful results[17–19]. FastGlioma combines visual foundation model training and efficient fine-tuning strategies to generalize across patient demographics, healthcare systems and WHO brain tumour diagnoses with minimal supervised training. As a general-purpose surgical adjunct, FastGlioma provides physicians with real-time, accurate and clinically actionable diagnostic information within seconds of tissue biopsy that can improve the surgical care of patients with brain tumours, thereby increasing patient quality of life and overall survival.

Visual foundation model training

The FastGlioma workflow starts during the surgical resection of a brain tumour (Fig. 1). Fresh tissue specimens are sampled at the surgical margins of a resection cavity to detect microscopic tumour infiltration. Specimens are imaged at the patient's bedside using stimulated Raman histology (SRH)—a rapid, label-free, sub-micrometre-resolution optical imaging method[20,21]. A major advantage of SRH over other intraoperative imaging methods is that image contrast is generated from the intrinsic biochemical properties of the specimen and does not rely on stains, dyes or labels. SRH images can be acquired in either full-resolution mode (~100 s) or fast mode (~10 s, lower resolution) depending on the clinician's preference and the clinical context (Extended Data Fig. 1). Whole-slide images range from 1 mm^2 to 10 mm^2 and are divided into smaller, non-overlapping fields of view, or patches, for model input. The FastGlioma foundation model training dataset was acquired from 13 medical centres and includes imaging data from over 3,000 patients, spanning the diagnostic spectrum of central nervous system tumours and human cancers (Extended Data Fig. 2). We developed a two-stage self-supervised learning method specifically designed for training vision transformer architectures on whole-slide microscopy images[22]. First, patch features are extracted using a patch tokenizer trained using hierarchical self-supervised learning[23] (Extended Data Fig. 3a,b). Second, whole-slide image features are learned by generating two views of the same whole-slide image by randomly splitting, cropping and masking the patch tokens. Both views then undergo a feedforward pass through a vision transformer, with the patches as input tokens, and a whole-slide self-supervised objective is minimized[24] (Extended Data Fig. 3c). Using this two-stage strategy to train a vision foundation model on the full SRH dataset, we obtained high-quality whole-slide representations and state-of-the-art performance on a previously benchmarked multiclass brain tumour diagnosis task[25]. The classification task includes diagnosing the most common brain tumour types, including diffuse lower-grade and malignant gliomas with variable amounts of tumour infiltration. The high-quality patch-level and slide-level self-supervised training results in minimal trade-off between imaging speed/resolution versus model performance, with fast SRH reaching a mean class accuracy of 88.0 ± 2.1% versus full-resolution SRH at 90.2 ± 3.0% (Extended Data Fig. 3c). The proposed vision foundation model training strategy enables clinicians to seamlessly analyse multiple specimens throughout the tumour resection without sacrificing model performance.

Fine-tuning for infiltration scoring

Fine-tuning can improve task-specific foundation model performance[26]; however, a major disadvantage of fine-tuning is the need for new large and annotated datasets[16]. Biomedical datasets are challenging to obtain, making efficient fine-tuning strategies essential to ensure safe and effective medical AI performance. To adapt our SRH foundation model for tumour infiltration detection and scoring, we developed a data-efficient ordinal representation learning

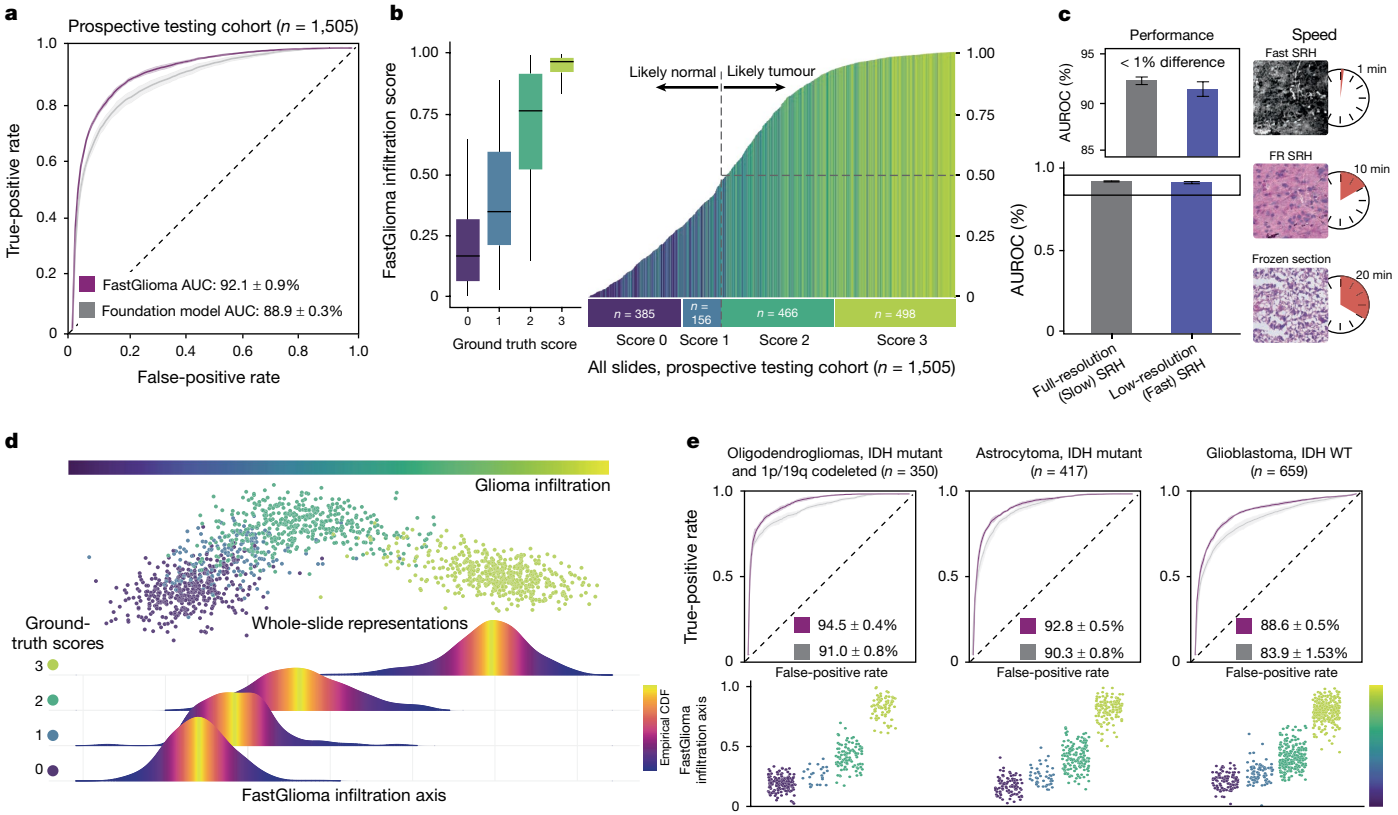


Fig. 2 | Prospective clinical testing results. a, The prediction results for the full prospective, international, multicentre testing cohort of patients with diffuse glioma ($n = 220$) are shown. ROC curves (mean ± s.d.) show the average performance for predicting four levels of tumour infiltration. Subgroup analysis is shown in Extended Data Fig. 6. SRH foundation model pretraining showed strong prediction performance without fine-tuning. When fine-tuning with ordinal metric learning, FastGlioma had a 3.2% increase in overall performance. FastGlioma outperforms models trained using standard supervised training (84.7 ± 1.1% mAUC) as shown in Supplementary Table 4. **b**, Box and whisker plots, shown in the standardized quartile format, of FastGlioma infiltration scores by ground-truth value. Scores had strong correlation with ground-truth ordinal scores ($\rho = 0.77$, 95% confidence interval = 0.74–0.78). Individual scores are shown in a histogram and correspond to AUROC values in **a**. **c**, FastGlioma performance on full-resolution (FR) versus low-resolution SRH images. Data are mean ± s.d. FastGlioma allows for a 10× increase in imaging speed with minimal performance trade-off. **d**, Whole-slide SRH representations are plotted on a linear discriminant axis. FastGlioma-learned representations rank whole-slide SRH images on a near-linear tumour-infiltration axis. **e**, Subgroup analysis by WHO adult-type diffuse glioma subtypes (ROC curves are plotted as mean ± s.d.). FastGlioma performs well across all three adult-type diffuse gliomas. Importantly, FastGlioma performs well on lower-grade gliomas in which tumour infiltration and tissue cellularity can be low (Extended Data Fig. 7). Low-grade and lower tumour infiltration are major challenges for other surgical adjuncts, such as fluorescence-guided surgery. WT, wild type.

method called ordinal metric learning that was implemented using an existing, previously annotated SRH tumour-infiltration dataset[4]. The tumour-infiltration dataset is 100× smaller than the foundation SRH dataset (around 100 versus approximately 10,000 images) and was annotated by three expert neuropathologists (lead pathologist, M.P.). The pathologists ranked the degree of tumour infiltration within each SRH image on a consensus four-tier scale: (0) normal brain tissue/no tumour; (1) atypical cells/possible tumour but not definitive; (2) sparse tumour infiltration; (3) dense tumour infiltration. Each specimen then underwent H&E staining and tumour-marker-specific immunohistochemistry, such as isocitrate dehydrogenase-1/2 (IDH) and p53, to confirm tumour-infiltration scores[4]. Using this tumour-infiltration dataset, ordinal metric learning fine-tunes the SRH foundation model by maximizing the latent distance, or metric, between whole-slide SRH images with different degrees of tumour infiltration (Extended Data Fig. 5a,b). Moreover, the increased efficiency of ordinal metric learning stems from enforcing that the model ranks images based on their tumour infiltration by performing a pairwise comparison between all images in a training mini-batch. The model then implicitly learns similar representations for whole-slide SRH images with the same degree of tumour infiltration and the representations are appropriately ordered in the whole-slide embedding space (Extended Data Fig. 5c). The fine-tuned model uses a linear slide scoring layer to output a single scalar value between 0 and 1 that indicates the degree of tumour infiltration within a whole-slide SRH image, providing clinically actionable information for each specimen within seconds of biopsy. In addition to assessing surgical margins, FastGlioma can identify specimens with dense tumour infiltration early in the surgical resection to obtain high-yield diagnostic tissue for intraoperative and final pathologic diagnosis. We demonstrate that ordinal metric learning outperforms other state-of-the-art ordinal regression methods, especially when training data are limited, achieving a mean area under the receiver operator characteristic curve (mAUROC) of 88.7 ± 1.6% on hold-out testing using the SRH infiltration dataset.

Prospective testing of FastGlioma

We tested the fine-tuned FastGlioma model in a multicentre, prospective cohort of patients with diffuse glioma to evaluate how the model generalizes across different continents, medical centres, patient demographics and WHO diffuse glioma molecular subgroups. Model testing was designed as a single-arm, non-inferiority, diagnostic clinical trial with a minimum sample size of 565 SRH specimens for both

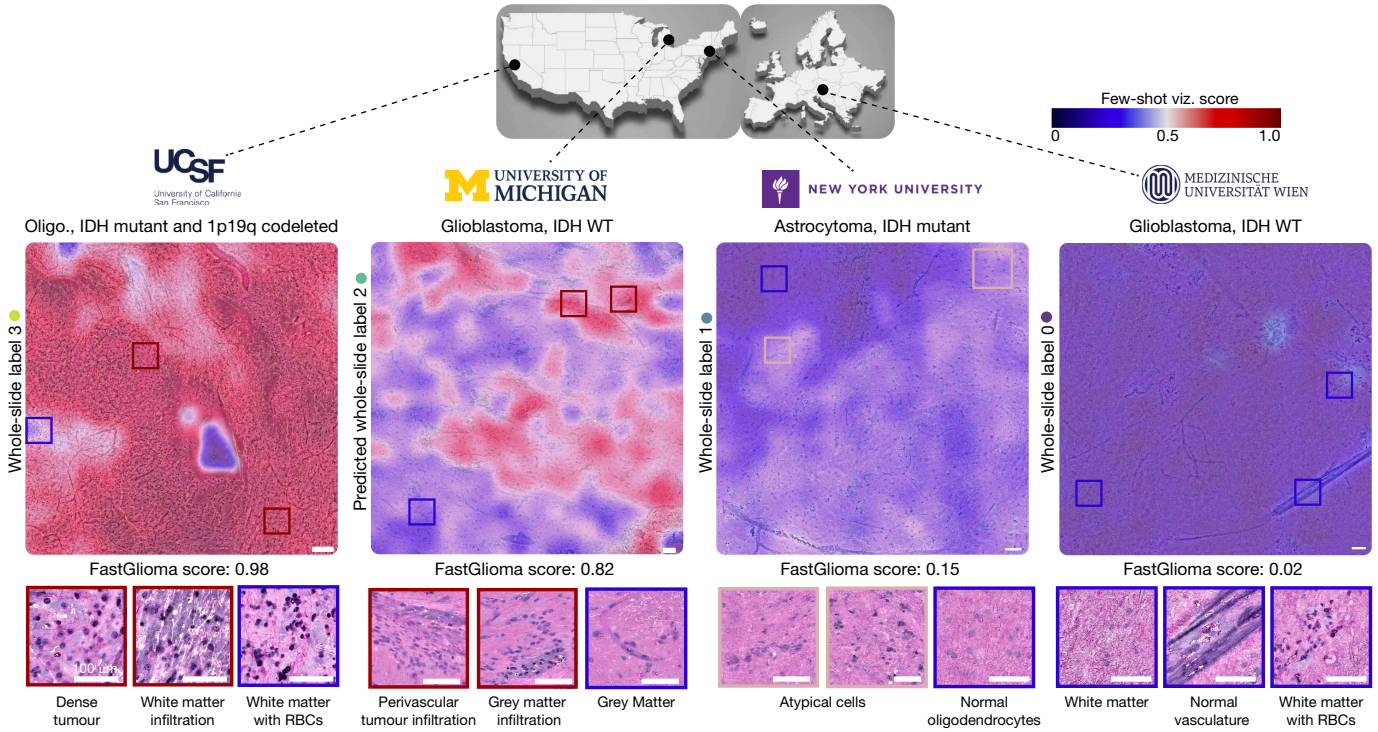


Fig. 3 | General and interpretable FastGlioma predictions. Diffuse glioma specimens from the four study medical centres are shown. Specimens span the full diagnostic spectrum of WHO adult-type diffuse gliomas. Few-shot visualizations highlight regions of tumour infiltration contained within surgical specimens using a support set of expert-physician-selected SRH patch keys to match the underlying SRH query features using FastGlioma. As a visual foundation model, FastGlioma requires few examples (around 10, University of Michigan images only) to achieve interpretable tumour-infiltration visualizations that generalize across medical centres and degrees of tumour infiltration. FastGlioma predictions and visualizations are consistent across diverse histological features, including dense tumour, infiltrated normal brain parenchyma, tumour-associated immune cell infiltration and atypical cells. FastGlioma differentiates between increased cellularity due to tumour cell infiltration versus non-neoplastic cells, such as red blood cells (RBCs). Additional detailed few-shot visualizations are provided in Extended Data Fig. 8. Scale bars, 100 µm. The maps were created using Vecteezy.

IDH-mutant and IDH-wild-type diffuse gliomas (1,130 total specimens). All specimen processing and annotation was performed using the validated four-tiered tumour-infiltration rating system described above. Three tertiary medical centres across the United States and Europe were included as testing recruitment sites: University of California San Francisco (UCSF), New York University (NYU) and Medical University of Vienna (MUV). Patients were recruited as a consecutive cohort of adult patients (aged ≥18 years) with diffuse gliomas who underwent tumour resection. A total of 220 patients were included (Extended Data Fig. 2), resulting in 767 IDH-mutant specimens and 659 IDH-wild-type specimens. FastGlioma achieved a mean AUROC of 92.1 ± 0.9% for differentiating the four degrees of diffuse glioma infiltration (Fig. 2a). Normalized infiltration scores were strongly correlated with ground-truth ordinal labels, with a correlation coefficient of $\rho = 0.77$ (95% confidence interval = 0.74–0.78, $P = 0.00$) (Fig. 2b). Importantly, visual foundation model pretraining allowed FastGlioma to generalize to the fast, low-resolution images acquired at 10× imaging speed without a clinically significant reduction in prediction performance (<1% decrease) (Fig. 2c). SRH foundation model fine-tuning with ordinal metric learning resulted in the FastGlioma representing whole-slide SRH images on a linear infiltration axis that stratifies SRH images according to their ground-truth infiltration label (Fig. 2d). Model performance and infiltration scores were consistent across patient demographics, including sex, age and race. mAUROC remained high across medical centres: UCSF (92.1 ± 0.2%), MUV (88.6 ± 0.13%) and NYU (92.9 ± 0.1%) (Extended Data Fig. 6a,b). While other surgical adjuncts can be limited to a single brain tumour diagnosis, FastGlioma was intentionally designed as a general-purpose model for all diffuse gliomas and degrees of tumour infiltration. FastGlioma maintained accurate tumour-infiltration scores despite the significant cytological and histoarchitectural differences related to tumour grade, molecular genetics, treatment effect or WHO subtypes (Fig. 2e and Extended Data Fig. 6c–e). Lastly, FastGlioma outperforms cellularity/segmentation-based methods for tumour detection by a large margin with a >10% mAUROC performance increase, especially in regions of lower cellularity or tumour densities (Extended Data Fig. 7).

Interpretability and zero-shot results

Interpretable visualizations that highlight regions of tumour infiltration within whole-slide images are essential to ensure safe, reliable and trustworthy predictions[27]. We developed a visualization strategy that takes advantage of FastGlioma's self-supervised training and generalization performance, called few-shot visualizations. For a query SRH region or patch, few-shot visualizations use a small support set of physician-selected SRH patch examples as keys to assess feature similarity (around ten images). The support set includes a diverse selection of diffuse gliomas and normal brain parenchyma SRH patch examples. FastGlioma few-shot visualizations compare the query cosine similarity with tumour keys and the dissimilarity with normal keys to generate tumour-infiltration heat maps. Few-shot visualizations are a flexible foundation-model-based interpretability framework that can accommodate any support set selected by pathologists, clinicians or investigators. Adding or changing support set examples does not require model retraining. Figure 3 shows interpretable few-shot FastGlioma visualizations from the prospective testing medical centres. The heat maps demonstrate slide-level segmentation of tumour-infiltrated regions.

Visualization quality generalizes across medical centres, degrees of tumour infiltration and molecular subgroups (Extended Data Fig. 8). Similar to FastGlioma tumour-infiltration scoring, few-shot visualizations reliably identify regions of tumour infiltration despite variations in the underlying histological features. FastGlioma learned invariant tumour-infiltration representations across the molecular subgroups of diffuse gliomas (Extended Data Fig. 8a,b). Importantly, FastGlioma demonstrates medical foundation model properties with zero-shot generalization. FastGlioma achieves accurate tumour-infiltration detection and few-shot visualization for non-glioma brain tumour diagnoses, including metastatic brain tumours, primary central nervous system lymphomas, embryonal tumours and meningiomas (Extended Data Fig. 9). These results demonstrate the advantage of visual foundation models for medical AI applications and the potential to generalize to other human cancers without requiring extensive model retraining or fine-tuning.

FastGlioma as a surgical adjunct

Finally, we evaluated the feasibility and safety of FastGlioma as a surgical adjunct by simulating an interventional clinical trial in which surgical resections are guided by FastGlioma predictions. FastGlioma predictions (experimental arm) were compared in a head-to-head, prospective comparison against standard-of-care intraoperative surgical adjuncts (control arm): image-guided surgery with magnetic resonance imaging (MRI)-based neuronavigation and fluorescence-guided surgery with 5-aminolevulinic acid (5-ALA). Both adjuncts have been studied in clinical trials that demonstrated improved extent of resection and decreased postoperative residual tumour volume[28,29]. In total, 129 patients with diffuse glioma, a subset of the above prospective cohort, were included, resulting in 624 surgical specimens with matched FastGlioma predictions and radiographic features (contrast enhancement/FLAIR) and/or 5-ALA status for each surgical specimen according to previously published and validated protocols[4,30] (Supplementary Table 3). We evaluated both study arms on the classification task of differentiating surgical specimens with the ground truth label of normal brain (score 0) versus dense tumour infiltration (score 3). Errors on this task are clinical high-risk errors because they represent actionable and decisive predictions: normal brain predictions signal to stop resection, dense tumour signals to continue resection if otherwise safe. FastGlioma outperformed both image-guided and fluorescence-guided methods for detecting tumour infiltration by a wide margin (Fig. 4a). FastGlioma achieved an AUROC of 98.1% compared to 76.3% for FLAIR positivity, 71.8% for contrast enhancement and 89.0% for 5-ALA fluorescence. A major challenge in diffuse glioma surgery is interpreting FLAIR positivity, which can indicate tumour infiltration or cerebral oedema, or both. FastGlioma was able to correctly differentiate tumour infiltration and cerebral oedema in FLAIR-positive regions, with an AUROC of 98.7% (ref. 31) (Supplementary Fig. 2). Next, we analysed these results for each study patient to identify the number of patients that had one or more high-risk tumour miss errors, or false-negative predictions, for both study arms. Tumour miss errors place patients at high risk of dense, safely resectable residual tumour left within the resection cavity after surgery. Only 3.8% (5 out of 129) of patients in the FastGlioma arm had at least 1 high-risk tumour miss compared with 24.0% (31 out of 129) in the surgical adjuncts arm (Fig. 4b). Patients who undergo diffuse glioma resections guided using current standard-of-care surgical adjuncts alone are potentially at 6.3× increased relative risk of dense, safely resectable residual tumour after surgery compared to including FastGlioma as a surgical adjunct. FastGlioma powered by real-time, label-free, optical imaging with SRH overcomes the inherent limitations of existing adjuncts, such as insufficient tumour fluorescence, non-specific radiographic features and brain shift, enabling more precise and effective surgical care of patients with diffuse glioma[30,32,33] (Extended Data Fig. 10).

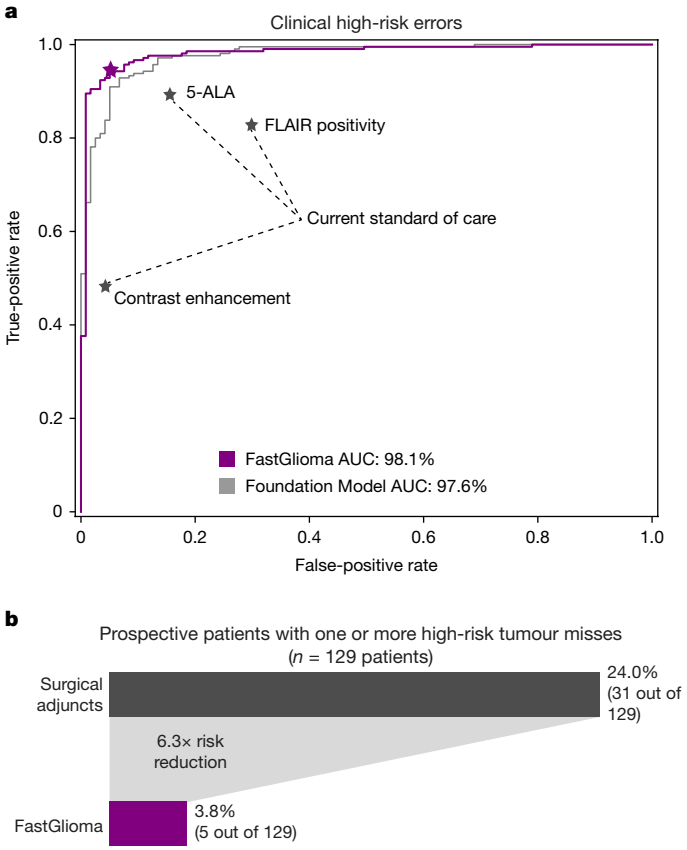


Fig. 4 | Simulated clinical trial results. a, The results of the prospective simulated clinical trial comparing FastGlioma with current standard-of-care surgical adjuncts. Both FastGlioma and SRH foundation models outperform image-guided and fluorescence-guided surgical methods by a wide margin for differentiating normal/gliotic brain (score 0) from dense tumour infiltration (score 3). The performance boost results from FastGlioma predicting directly on high-resolution, label-free SRH images without requiring indirect labels of infiltration, such as 5-ALA fluorescence or contrast enhancement. **b**, Graphical summary of the patient-level false-negative rates in the prospective cohort. To ensure fair and clinical representative comparison, false-negative rates were calculated for the optimal surgical adjunct given the patient's diffuse glioma subtypes: FLAIR for IDH-mutant diffuse gliomas and contrast enhancement and 5-ALA was used for IDH-wild-type tumours. False-negative errors on the score 0 versus score 3 classification task represent high-risk tumour misses and place the patient at the highest risk of residual tumour after surgery resection. The FastGlioma study arm had a 3.8% false-negative/tumour miss rate compared with 24.0% in the standard-of-care surgical adjuncts arm. The results indicate a potential 6.3× decrease in the relative risk of residual tumour within resection cavities by using FastGlioma to guide tumour resections. Detailed classification results are provided in Extended Data Fig. 10.

Discussion

Here we present FastGlioma, an open-source medical foundation model for fast, label-free detection of diffuse glioma infiltration during surgery. Within seconds of tissue sampling, FastGlioma accurately predicts the degree of tumour infiltration in fresh, unprocessed surgical specimens from both IDH-wild-type glioblastomas and IDH-mutant diffuse gliomas. FastGlioma outperforms standard-of-care image-guided and fluorescence-guided intraoperative methods for tumour-infiltration detection by a substantial margin. FastGlioma can be generalized to other paediatric and adult brain tumour diagnoses, demonstrating its potential as a general-purpose medical foundation model for guiding brain tumour surgery.

FastGlioma has the potential for immediate clinical impact on improving the comprehensive management of patients with diffuse glioma. The prognostic importance of the extent of resection and residual tumour burden has been reproducibly demonstrated in multiple independent international clinical studies[2,3,6,34]. FastGlioma represents an innovative departure from existing surgical adjuncts by rapidly identifying tumour infiltration at microscopic resolution using AI, nearly eliminating the potential risk of residual dense tumour within glioma resection cavities. AI-based computer vision and visual foundation models can minimize reliance on radiographic features, contrast enhancement or extrinsic fluorescent labels to optimize the extent of resection and achieve safe maximal tumour removal.

Worldwide, over 18 million people will be diagnosed with cancer annually[35]. The importance of delivering safe, timely and affordable cancer surgery has been identified as a major area of improvement in the global fight against cancer[8]. Global cancer initiatives have recommended incorporating new technologies, including advanced imaging modalities and AI, into cancer surgery[36]. SRH and FastGlioma are accessible and affordable surgical adjuncts for cancer surgery. Future studies will focus on applying a similar FastGlioma workflow to other human cancers, including lung, prostate, head and neck, and breast cancer[37,38].

As part of the growing field of AI-based diagnostics[39–41], FastGlioma is a rapid and scalable alternative to conventional wet-lab methods. Conventional intraoperative pathology using smear preparation and frozen sectioning is limited in diagnostic and clinical value. Histological diagnosis[20,42], tumour recurrence[43], molecular classification[44] and tumour-infiltration detection are now possible through an integrated bedside SRH-AI platform. SRH visual foundation models can enable computational staining for histological diagnoses, molecular markers and tumour infiltration within fresh, unprocessed cancer specimens.

The SRH visual foundation model can facilitate generalization of FastGlioma to other human cancers. Foundation models will have a central role in the safe and effective deployment of AI in healthcare and clinical medicine[17,45,46]. Large and diverse medical datasets combined with high-quality self-supervised training results in better performance and less bias across a wide range of complex clinical and diagnostic tasks[18,47]. Here we demonstrate how visual foundation models can facilitate tumour detection across patient demographics, medical centres, molecular genetics and brain tumour diagnoses. In conclusion, FastGlioma represents the transformative potential of medical foundation models to unlock the role of artificial intelligence in the care of patients with cancer.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08169-3.

1. Karschnia, P. et al. Prognostic validation of a new classification system for extent of resection in glioblastoma: a report of the RANO resect group. *Neuro. Oncol.* **25**, 940–954 (2023).
2. Sanai, N., Polley, M.-Y., McDermott, M. W., Parsa, A. T. & Berger, M. S. An extent of resection threshold for newly diagnosed glioblastomas. *J. Neurosurg.* **115**, 3–8 (2011).
3. Smith, J. S. et al. Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas. *J. Clin. Oncol.* **26**, 1338–1345 (2008).
4. Pekmezci, M. et al. Detection of glioma infiltration at the tumor margin using quantitative stimulated Raman scattering histology. *Sci. Rep.* **11**, 12162 (2021).
5. Petrecca, K., Guiot, M.-C., Panet-Raymond, V. & Souhami, L. Failure pattern following complete resection plus radiotherapy and temozolomide is at the resection margin in patients with glioblastoma. *J. Neurooncol.* **111**, 19–23 (2013).
6. Hervey-Jumper, S. L. et al. Interactive effects of molecular, therapeutic, and patient factors on outcome of diffuse low-grade glioma. *J. Clin. Oncol.* **41**, 2029–2042 (2023).
7. Gal, A. A. & Cagle, P. T. The 100-year anniversary of the description of the frozen section procedure. *JAMA* **294**, 3135–3137 (2005).
8. Sullivan, R. et al. Global cancer surgery: delivering safe, affordable, and timely cancer surgery. *Lancet Oncol.* **16**, 1193–1224 (2015).
9. Orosco, R. K. et al. Positive surgical margins in the 10 most common solid cancers. *Sci. Rep.* **8**, 5686 (2018).
10. Brown, P. D. et al. A prospective study of quality of life in adults with newly diagnosed high-grade gliomas: the impact of the extent of resection on quality of life and survival. *Neurosurgery* **57**, 495–504 (2005).
11. *The Big Question: Precision Surgical Interventions* https://arpa-h.gov/research-and-funding/programs/psi (ARPA-H, 2024).
12. Robboy, S. J. et al. Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch. Pathol. Lab. Med.* **137**, 1723–1732 (2013).
13. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at arxiv.org/abs/2108.07258 (2021).
14. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M. Hierarchical text-conditional image generation with CLIP latents. Preprint at arxiv.org/abs/2204.06125 (2022).
15. OpenAI et al. GPT-4 technical report. Preprint at arxiv.org/abs/2303.08774 (2023).
16. Brown, T. B. et al. Language models are few-shot learners. In *Proc. 34th Conference on Neural Information Processing Systems* (eds. Larochelle, H. et al.) 1877–1901 (Curran, 2020).
17. Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
18. Azizi, S. et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nat. Biomed. Eng.* **7**, 756–779 (2023).
19. Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med.* **30**, 850–862 (2024).
20. Orringer, D. A. et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat. Biomed. Eng.* **1**, 0027 (2017).
21. Freudiger, C. W. et al. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* **322**, 1857–1861 (2008).
22. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. Preprint at arxiv.org/abs/2010.11929 (2020).
23. Jiang, C. et al. Hierarchical discriminative learning improves visual representations of biomedical microscopy. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2023**, 19798–19808 (2023).
24. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *Proc. 37th International Conference on Machine Learning* (eds Iii, H. D. & Singh, A.) 119, 1597–1607 (PMLR, 2020).
25. Jiang, C. et al. OpenSRH: optimizing brain tumor surgery using intraoperative stimulated Raman histology. *Adv. Neural Inf. Process. Syst.* **35**, 28502–28516 (2022).
26. Ouyang, L. et al. Training language models to follow instructions with human feedback. In *Proc. 36th Conference on Neural Information Processing Systems* (eds. Kovejo, S. et al.) 27730–27744 (Curran, 2022).
27. Wiens, J. et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.* **25**, 1337–1340 (2019).
28. Stummer, W. et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol.* **7**, 392–401 (2006).
29. Senft, C. et al. Intraoperative MRI guidance and extent of resection in glioma surgery: a randomised, controlled trial. *Lancet Oncol.* **12**, 997–1003 (2011).
30. Lau, D. et al. A prospective phase II clinical trial of 5-aminolevulinic acid to assess the correlation of intraoperative fluorescence intensity and degree of histologic cellularity during resection of high-grade gliomas. *J. Neurosurg.* **124**, 1300–1309 (2016).
31. Broggi, G. et al. Histologic definition of enhancing core and FLAIR hyperintensity region of glioblastoma, IDH-wild type: a clinico-pathologic study on a single-institution series. *Brain Sci.* **13**, 248 (2023).
32. Gerard, I. J. et al. Brain shift in neuronavigation of brain tumors: a review. *Med. Image Anal.* **35**, 403–420 (2017).
33. Wen, P. Y. et al. RANO 2.0: update to the response assessment in neuro-oncology criteria for high- and low-grade gliomas in adults. *J. Clin. Oncol.* **41**, 5187–5199 (2023).
34. Jakola, A. S. et al. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *JAMA* **308**, 1881–1888 (2012).
35. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
36. Are, C. et al. Global cancer surgery: pragmatic solutions to improve cancer surgery outcomes worldwide. *Lancet Oncol.* **24**, e472–e518 (2023).
37. Mannas, M. P. et al. Stimulated Raman histology, a novel method to allow for rapid pathologic examination of unprocessed, fresh prostate biopsies. *Prostate* **83**, 1060–1067 (2023).
38. Hoesli, R. C., Orringer, D. A., McHugh, J. B. & Spector, M. E. Coherent Raman scattering microscopy for evaluation of head and neck carcinoma. *Otolaryngol. Head Neck Surg.* **157**, 448–453 (2017).
39. Tiu, E. et al. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat. Biomed. Eng.* **6**, 1399–1406 (2022).
40. Zhang, K. et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **182**, 1360 (2020).
41. He, B. et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature* **616**, 520–524 (2023).
42. Hollon, T. C. et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.* **26**, 52–58 (2020).
43. Hollon, T. C. et al. Rapid, label-free detection of diffuse glioma recurrence using intraoperative stimulated Raman histology and deep neural networks. *Neuro. Oncol.* https://doi.org/10.1093/neuonc/noaa162 (2020).

44. Hollon, T. et al. Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. *Nat. Med.* **29**, 828–832 (2023).
45. Jiang, L. Y. et al. Health system-scale language models are all-purpose prediction engines. *Nature* **619**, 357–362 (2023).
46. Singhal, K. et al. Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).
47. Mittermaier, M., Raza, M. M. & Kvedar, J. C. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit. Med.* **6**, 113 (2023).
48. Ozyoruk, K. B. et al. A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. *Nat. Biomed. Eng.* **6**, 1407–1419 (2022).

Methods

Study design

We had three main objectives for this study: (1) train a vision transformer model on a large and diverse dataset of SRH images using self-supervision to develop the first SRH visual foundation model; (2) fine-tune the visual foundation models to develop FastGlioma for detection and quantification of diffuse glioma infiltration in fresh, unprocessed surgical specimens; (3) validate FastGlioma in a prospective, multicentre, international cohort of patients with diffuse glioma and compare results to current surgical adjuncts. We adopted the common working definition of a foundation model: (1) any machine learning model that is (2) trained on a large and diverse dataset using (3) self-supervision at scale and (4) can be adapted to a wide range of downstream tasks[13]. We also added to this definition (5) evidence of zero-shot generalization to new, unseen data. Foundation modelling had not been previously investigated in studies on the clinical applications of SRH and we focused on tumour infiltration as the most clinically important and ubiquitous problem in cancer surgery as the major downstream tasks. We aimed to design FastGlioma to detect microscopic tumour infiltration for all diffuse glioma molecular subtypes. A major data-centric contribution of this work was developing a multicentre, international, label-free SRH tumour-infiltration dataset annotated by expert neuropathologists. Preliminary results demonstrated the feasibility of generating this complex biomedical dataset[4]. Moreover, previous studies that combine SRH and AI were done such that the same imaging dataset was used both for human interpretation and AI model training[42,44]. Here we aimed to push the limits of AI-based computer vision performance in lower image resolution/faster image acquisition regimes at 10 times the speed of conventional SRH imaging. Finally, we aimed to demonstrate the feasibility of using FastGlioma as a surgical adjunct and compare tumour detection performance with existing image-guided and fluorescence-guided surgical adjuncts.

SRH imaging

All of the images in our study were acquired using intraoperative fibre-laser-based stimulated Raman scattering microscopy[21,49]. The NIO Imaging System (Invenio Imaging) was used for all training and testing data collection. We have provided a detailed description of the imager and laser configuration in previous studies[20,49]. In brief, a pump beam at 790 nm and a Stokes beam with a tuneable range from 1,015 nm to 1,050 nm was used to stimulate the surgical specimens. The settings allow for access to the Raman shift spectral range between 2,800 cm^{-1} and 3,130 cm^{-1}. Images were acquired as 1,000 pixel-width strips with an imaging speed of 0.4 Mpx per strip. In normal imaging mode, each strip row is acquired independently in a left–right manner using a custom beam-scanning[20,21]. Two image channels are acquired sequentially at 2,845 cm^{-1} (CH2 channel) and 2,930 cm^{-1} (CH3 channel) Raman wavenumber shifts. A stimulated Raman signal at 2,845 cm^{-1} represents the CH2 symmetric stretching mode of lipid-rich structures, such as myelinated axons. A second Raman peak at 2,930 cm^{-1} corresponds to protein- and nucleic acid-rich regions such as the cell nucleus and collagen. As all SRH strips are acquired through standard horizontal line scanning[20,21,49], low-resolution SRH images can be generated by directly downsampling SRH strip rows by a downsampling factor, such as 1/2, 1/4, 1/8 and so on. Halving the line sampling factor corresponds to a 2× imaging time savings. In fast imaging mode, single-channel images with a user-specified downsampling factor are acquired. The whole-slide SRH images are then split into 300 × 300 pixel patches without overlap using a sliding raster window over the full image. All models are trained using 16-bit, raw, greyscale SRH images. For the purposes of the study, SRH images were acquired as two-channel images (2,845 cm^{-1}, 2,930 cm^{-1}) for pathologist's review to determine ground truth tumour-infiltration labels.

SRH dataset

Clinical SRH imaging began at the University of Michigan (UM) on 1 June 2015 following Institutional Review Board approval (HUM00083059). All patients with a suspected brain tumour were recruited for intraoperative SRH imaging in a prospective manner. The inclusion criteria were as follows: patients who were undergoing surgery for (1) suspected central nervous system tumour and/or (2) epilepsy, (3) subject or durable power of attorney was able to provide consent and (4) preoperative assessment that additional tumour specimens would be available in addition to what is required for clinical pathologic diagnosis. Exclusion criteria were (1) grossly inadequate tissue, (2) insufficient diagnostic tissue (for example, haemorrhagic, necrotic) or (3) imaging malfunction. A similar imaging protocol was implemented at 12 other medical centres with clinical SRH imaging deployed in their operating rooms. A total of 2,799 patients, 11,462 whole-slide SRH images and approximately 4 million unique 300 × 300 pixel SRH patches were included for SRH foundation model training. Dataset statistics and diagnostic information are provided in Extended Data Fig. 2.

SRH foundation model training

SRH foundation models consist of two modular components trained using self-supervision: the patch tokenizer and the whole-slide encoder.

Patch tokenizer training with hierarchical discrimination. In standard vision transformers, converting small, fixed-size image patches, such as 8 × 8 or 16 × 16 pixel patches, into tokens can be done by flattening. This tokenization strategy is not feasible due to the size of whole-slide SRH images (>6,000 × 6,000 pixels). We therefore developed a data-driven patch tokenization method that leverages the inherent patient-slide-patch hierarchy of SRH images to define a hierarchical discriminative learning task[23]. We previously demonstrated that hierarchical discrimination, called HiDisc, outperforms instance discrimination methods for biomedical microscopy computer vision tasks. HiDisc uses self-supervised contrastive learning such that positive image patches are defined based on a shared ancestry in the patient-slide-patient data hierarchy. The HiDisc loss is a summation of three losses, each of which corresponds to instance discrimination at a level of the patch-slide-patient hierarchy. We define the HiDisc loss at the level ℓ to be:

$$L_{\text{HiDisc}}^{\ell} = \sum_{i \in \mathcal{I}} \frac{-1}{|\mathcal{P}_{\ell}(i)|} \sum_{p \in \mathcal{P}_{\ell}(i)} log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a \in \mathcal{A}(i)} \exp(z_i \cdot z_a/\tau)}, \quad (1)$$

where $\ell \in \{\text{Patch, Slide, Patient}\}$ is the level of discrimination, and I is the set of all images in the minibatch. $\mathcal{A}_{\ell}(i)$ is the set of all images in I except for the anchor image i,

$$\mathcal{A}_{\ell}(i) = \mathcal{I} \setminus \{i\}, \quad (2)$$

and $\mathcal{P}_{\ell}(i)$ is a set of images that are positive pairs of i at the ℓ level,

$$\mathcal{P}_{\ell}(i) = \left\{p \in \mathcal{A}_{\ell}(i) : \text{ancestry}_{\ell}(p) = \text{ancestry}_{\ell}(i)\right\}, \quad (3)$$

where $\text{ancestry}_{\ell}(\cdot)$ is the ℓ-level ancestry for the anchor patch. For example, patches x_i and x_j from the same patient would have the same patient ancestry, that is, $\text{ancestry}_{\text{Patient}}(x_i) = \text{ancestry}_{\text{Patient}}(x_j)$. The component HiDisc losses calculate the same overall contrastive objective with positive pairs at different levels in the hierarchy. Finally, the complete HiDisc loss is the sum of the patch-, slide- and patient-level losses defined above:

$$\mathcal{L}_{\text{HiDisc}} = \sum_{\ell \in \{\text{Patch,Slide,Patient}\}} \lambda_{\ell} L_{\text{HiDisc}}^{\ell}, \quad (4)$$

where λ_ℓ is a weighting hyperparameter for level ℓ in the total loss. As HiDisc is a self-supervised representation learning method, we used the full SRH dataset as shown in Extended Data Fig. 2a. We found that HiDisc patch tokenization improved classification performance compared with ImageNet transfer learning (Extended Data Fig. 3c).

Patch encoding was accomplished using the ResNet-34 architecture as the backbone feature extractor and a one-layer multilayer perceptron to project the embedding to 128-dimensional latent space for HiDisc self-supervised training[50]. We performed ablation studies over the batch size, learning rate and loss hyperparameters to optimize performance on the SRH7 dataset. The encoder was trained using a batch size of 512 and an AdamW optimizer with a learning rate of 0.001 on a cosine decay schedule with warmup for the first 10% of training iterations for a total of 100,000 iterations on the foundation SRH dataset. To train using the HiDisc loss, the mini-batches were constructed by first selecting 64 patients, followed by sampling two slides per patient, two patches per slide and finally applying two random augmentations per patch, yielding 512 patches. The patch, slide and patient losses were weighted equally, and the temperature was set to 0.7. All of the patch experiments were performed using mixed-precision and data parallelism on four NVIDIA A40 GPUs, taking up to 3 days. We performed additional ablation experiments with open-source foundational patch encoders to assess the quality of HiDisc feature learning compared with other pretrained models[19] (Supplementary Table 4).

Whole-slide encoder. A major contribution of this work was developing an efficient and effective method for whole-slide self-supervised training with vision transformer architectures. The major advantage of vision transformers for whole-slide inference in computational pathology and optical imaging is their ability to handle large and variably sized images. The whole-slide self-supervised learning strategy is a Siamese architecture that requires two random transformations of the same whole-slide image. The slide-level transformation strategy is as follows. First, the whole slide is split into two mutually exclusive patch sets (splitting). Next, two random spatial crops are selected from the whole-slide image (cropping). Finally, 10–80% of patches from a crop are dropped (masking). This strategy is ideally suited for vision transformers because it allows for variable sized inputs and random dropping of patch tokens/spatial regions within a whole-slide image. After generating two transformed views, we then minimize a variance–invariance–covariance (VICReg) self-supervised objective function[51]. VICReg is well suited for whole-slide encoding because it is computationally efficient, does not require negative examples and maintains high expressivity by avoiding dimensional collapse[52].

The whole-slide transformer consists of 2 hidden layers with dimension 512, with 4 attention heads per layer. The output of the transformer is distilled into a <CLS> token, with seven additional register tokens employed to stabilize training[53]. Positional information is learned concurrently in a Fourier feature positional embedding generator network[54]. The Fourier feature and MLP hidden dimension of this network are 96 and 36, respectively. For self-supervision purposes, a one-layer MLP was trained to project the embedding to 128-dimensional latent space. The VICReg objective was used for whole-slide self-supervised training, with the coefficients being 10, 10 and 1 for the variance, invariance and covariance losses, respectively. Pretraining was done with an effective batch size of 256 and the AdamW optimizer with a learning rate of 3×10^{-4} for 100 epochs on a single NVIDIA Titan V100 GPU. Checkpoints were saved every 10 epochs, with the optimal one selected using slide-level metrics on the histological brain tumour diagnosis task with a hold-out validation set. A schematic of SRH foundation model training is shown in Extended Data Fig. 3. Detailed model training configurations, including batch size, learning rate and other hyperparameters, are shown in Supplementary Table 6 and are available at GitHub (https://github.com/MLNeurosurg/fastglioma).

SRH foundation model evaluation

Validation of the foundation model was performed on a multiclass SRH brain tumour diagnostic task. This dataset consists of 3,560 whole-slide images from 896 patients (852,000 total patches). Diagnostic classes are normal brain, high-grade glioma (HGG), low-grade glioma (LGG), meningioma, pituitary adenoma, schwannoma and metastatic tumour. In all previous benchmarking studies, training required supervised filtering of nondiagnostic patches and patch-level average pooling for whole-slide inference, which is known to degrade performance[42–44]. Here we demonstrate that high-quality self-supervised patch and whole-slide representation learning with vision transformers bypasses the need for preprocessing, filtering or patch-level voting/averaging. We used nearest-neighbour classification (k-NN) for SRH foundation model evaluation. First, we generated whole-slide representations for both the training and testing data. Next, the k-NN classifier was used to match each slide in the testing dataset to the k most similar representations in the training dataset as determined by their cosine similarity. We set $k = 10$ in our experiments for all models to ensure consistent results. This enables us to determine a class prediction for each slide in the testing dataset. We then calculate the mean class accuracy (MCA) and mean average precision (mAP) for the seven-class task for slide metrics (Extended Data Fig. 4). Whole-slide representations were visualized using t-distributed stochastic neighbour embedding (t-SNE) to qualitatively assess slide representations with respect to tumour classes. Embeddings for k-NN and subsequent evaluations were generated on a single NVIDIA Titan V100 GPU.

Fine-tuning with ordinal representation learning

Our SRH foundation models were specifically developed to adapt to downstream diagnostic tasks for clinical decision support. Here we aimed to fine-tune the foundation model for the detection and quantification of tumour infiltration using intraoperative SRH imaging. While diffuse glioma infiltration is a continuous random variable, the majority of previous work modelled glioma infiltration as an ordinal variable[4,55], such that expert pathologists score the degree of tumour infiltration on a discrete, ordered scale. We fine-tuned the foundation model using the glioma-infiltration dataset from a previous study[4]. The dataset consists of 161 surgical specimens imaged using SRH from 35 patients. The degree of tumour infiltration in each SRH image was scored on a scale from 0 to 3 by three independent expert neuropathologists, where 0 is no tumour present; 1 is mildly cellular tissue either due to reactive gliosis or with scattered atypical cells, without definitive tumour; 2 is tumour present but in mild/sparse density; and 3 is moderate to severe density of tumour cells. This dataset is approximately 100 times smaller than the SRH foundation model training dataset and approximately 10 times smaller than the calculated sample size for model testing. Owing to this extreme data sparsity for fine-tuning, we developed a general, data-efficient, few-shot ordinal representation learning method called ordinal metric learning. Ordinal metric learning aims to minimize the feature distance, or metric, between images with the same ordinal rank. Moreover, it implicitly learns to order images based on their ordinal label by performing a pairwise comparison between all images in a mini-batch. Ordinal metric learning accomplishes this by applying a binary cross entropy objective on the distance $d_{i,j} = s_i - s_j$ between scores for all possible pairs of images in a mini-batch to enforce the image with the higher label is assigned a higher score. The following loss equation accomplishes this:

$$\mathcal{L}_{\text{OrdinalMetric}} = \sum_{i \in \mathcal{I}} \left\{ \frac{1}{|\mathcal{B}(i)|} \sum_{b \in \mathcal{B}(i)} \text{BCE}\left(d_{i,b}, \mathbf{1}_{i,b}\right) \right\}, \tag{5}$$

where

$$\text{BCE}(x, y) = y \cdot \log\sigma(x) + (1 - y) \cdot \log(1 - \sigma(x)), \tag{6}$$

and

$$\mathbf{1}_{x,y} = \begin{cases} 1 & \text{if } l_x > l_y \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

$\mathcal{I}$ is the set of all images in the minibatch. $\mathcal{B}(i)$ is the set of all images in $\mathcal{I}$ except for those with the same label as the anchor image i, $\mathcal{L}(i)$, denoted as,

$$\mathcal{B}(i) = \mathcal{I} \setminus \mathcal{L}(i), \quad (8)$$

A schematic of ordinal metric learning is shown in Extended Data Fig. 4.

Ordinal metric learning was used to train the FastGlioma model and included fine-tuning the slide encoder and a one layer linear slide scorer. Tumour-infiltration labels were balanced by whole-slide oversampling of the minority classes during training. The model was trained with a batch size of 16 and adjusted learning rate of 1.875×10^{-5} for 100 epochs. The best checkpoint was selected using a hold-out validation set. To evaluate the SRH foundation model, a standard linear evaluation protocol was followed with only the slide scorer being trained. Our linear evaluation protocol is similar to other self-supervised visual representation learning methods, such as SimCLR[56] or DINO[57], where the visual feature extractor is frozen and a final classification/regression layer is trained.

Tumour-infiltration scoring metrics

To evaluate the performance of FastGlioma in distinguishing various levels of diffuse glioma infiltration, we employ two key metrics: mAUROC and mean absolute error (MAE). The MAE is calculated by passing the FastGlioma whole-slide logit through a sigmoid activation function to rescale between 0 to 1. Similarly, the ground truth labels, which range from 0 to 3, are also normalized to 0 to 1. We then compute the MAE by measuring the average absolute difference between the rescaled logits and the normalized labels. The mAUROC provides a straightforward metric to assess FastGlioma's ability to discern between different degrees of tumour infiltration. mAUROC is calculated as the average of the AUROC for three binary classification tasks: 0 versus 123, 01 versus 23 and 012 versus 3. This metric reflects the ordinal label distribution and emphasizes the clinical diagnostic task.

Prospective testing of FastGlioma

Our prospective FastGlioma clinical testing included a primary and secondary end point. The primary end point was to validate FastGlioma's ability to reproducibly and accurately detect tumour infiltration within SRH images across patient populations, demographics, medical centres and diffuse glioma subgroups. The secondary end point was to compare the performance of FastGlioma with the standard-of-care methods for intraoperative tumour-infiltration detection currently in use for brain tumour surgery.

Primary testing end point: SRH-based tumour-infiltration detection. Our primary study end point was to achieve a diagnostic performance for detecting diffuse glioma infiltration in SRH images on par with previous SRH classification tasks, such as intraoperative tissue diagnosis and molecular classification[42,44]. We designed the primary testing using the same principles as a single-arm, non-inferiority diagnostic clinical trial[42,44]. To obtain a minimum sample size estimate, we used previous studies that combined SRH and AI to classify normal brain versus any tumour tissue. Previously reported accuracy values range from 89.3 to 95.8% with an average value of 93.2% (±3.6%)[42–44,55]. We used this average value to define the expected performance, the equivalence/non-inferiority limit was set to 5%, the alpha value to 2% and the power to 90%, resulting in a sample size value of 565 SRH images obtained from surgical margins. We aimed to achieve this sample size for both IDH-wild-type and IDH-mutant diffuse gliomas to ensure generalizability and reproducibility across diffuse glioma molecular subtypes as defined by the WHO. The calculation resulted in a final minimum sample size of 1,130 surgical specimens. Prospective patient recruitment was continued until minimum sample sizes were reached in both IDH-mutant and IDH-wild-type cohorts. All sample size calculations were performed using the epiR package (v.2.0.46) in R (v.3.6.3). Ground-truth SRH tumour-infiltration labels were provided by the on-site study pathologists (M.P., M.M.-E., T.R.-P.). All pathologists were provided with written and video instructions for SRH tumour-infiltration scoring using the four-tiered system by our primary study pathologist (M.P.).

Secondary testing end point: FastGlioma comparison with image- and fluorescence-based surgical adjuncts for tumour-infiltration detection. Our secondary study end point was to compare the FastGlioma intraoperative workflow (experimental arm) with the two most common surgical adjuncts for identifying tumour infiltration intraoperatively (control arm) in a simulated prospective surgical trial. 'Simulated' terminology is used because FastGlioma is not approved by the Food and Drug Administration or European Medicines Agency to guide treatment decisions, such as extent of tumour resections. However, we aimed to demonstrate the feasibility and safety of using FastGlioma to guide resections by predicting on surgical specimens sampled at the resection margin of patients with diffuse glioma. FastGlioma predictions in this setting produce the actionable information needed to guide resection and simulates the clinical setting that FastGlioma would be deployed. FastGlioma was compared in a head-to-head prospective comparison study to (1) image-guided surgery with MRI-based neuronavigation and (2) fluorescence-guided surgery with 5-ALA. Both methods have been shown to improve the extent of resection in randomized controlled trials[28,29]. In general, neuronavigation and 5-ALA fluorescence can indicate the presence of tumour infiltration but, in contrast to FastGlioma, do not quantify the degree of infiltration. For the purposes of this study and others[30], tumour detection using neuronavigation or 5-ALA was treated as a binary indicator, for example, yes/no contrast enhancement, yes/no 5-ALA fluorescence. To perform a fair comparison between FastGlioma and the surgical adjuncts, we designed this secondary end point to differentiate normal brain tissue (score 0) versus dense tumour (score 3). We focused specifically on this task because errors are clinical high-risk errors and these tumour-infiltration scores are actionable and decisive: score 0 means stop resection, score 3 means continue resection if otherwise safe. Moreover, this strategy avoids biasing performance results in favour of FastGlioma, which provides a continuous score that can differentiate degrees of tumour infiltration. We aimed to show that FastGlioma was non-inferior to both neuronavigation and 5-ALA fluorescence for detecting tumour within surgical specimens collected at the margin of resection cavities during surgical resection. Details of generating the matched SRH/MRI/5-ALA specimen dataset as a subset of the primary testing endpoint data are described below.

Prospective testing dataset

Three medical centres acted as external FastGlioma testing sites: UCSF, NYU and MUV. Each medical centre prospectively enrolled patients for testing. Inclusion criteria were: (1) patient age, ≥18 years; (2) a suspected diffuse glioma on preoperative radiographic imaging; and (3) planned brain tumour resection. Exclusion criteria included: (1) aborted tumour resection; (2) non-glioma final pathology; and (3) SRH imager malfunction. We aimed to accurately simulate the clinical setting that FastGlioma would be implemented for surgical interventions. Study neurosurgeons were therefore instructed to sample surgical margins at their discretion to identify microscopic tumour infiltration within the tumour resection cavity. We aimed to provide as minimal instruction as possible to account for surgeon/user variability

during FastGlioma testing. After intraoperative SRH imaging, surgical specimens were removed from the premade microscope slide and preserved in formalin for downstream tissue processing (Extended Data Fig. 1). Each SRH image was scored postoperatively by an onsite, board-certified neuropathologist with dedicated training and expertise in intraoperative SRH imaging. Our central neuropathologist (M.P.) provided verbal and video instructions for tumour-infiltration scoring. We used the previously developed and validated protocol for 0–3 tumour-infiltration scoring[4]. For the primary testing end point that is evaluated at the image level, SRH tumour-infiltration scores provided by the neuropathologists were used as the ground truth. For the secondary testing end point that is evaluated at the specimen level, neuronavigation coordinates, radiographic features (that is, contrast enhancement, FLAIR positive) and 5-ALA fluorescence status were recorded in real-time by a study technician for each specimen. Secondary end-point testing was completed at UCSF by a dedicated study technician (K.S.) and central neuropathologist (M.P.) to standardize all matched data collection. To optimize for the secondary testing endpoint, annotated data from UM, NYU and MUV were used to fine-tune FastGlioma. After intraoperative SRH imaging, the specimen was extracted from the premade microscopy slide and sent for downstream whole-slide/specimen analysis using H&E/immunohistochemistry testing as previously detailed[4]. Specimen-level ground truth tumour-infiltration scores were determined based on whole-slide analysis. This strategy allows for an unbiased comparison between all three surgical adjuncts.

FastGlioma versus cellularity-based tumour-infiltration scoring

The cellularity within the SRH whole-slide images was calculated to be the average number of cells per 300 × 300 pixel SRH patch. The number of cells was determined using an SRH single-cell segmentation model training using full supervision. Specifically, a Mask R-CNN model with a ResNet-50 backbone pre-trained on the Microsoft COCO dataset was fine-tuned on 1,000 annotated SRH patches of normal brain and 6 different brain tumour diagnoses[58]. The final model predictions were filtered with a non-maximal suppression algorithm to remove overlapping cell bounding boxes with >20% area and predictions with less than 80% confidence. Correlation between cellularity and FastGlioma tumour-infiltration score was calculated using Pearson's correlation coefficient. To evaluate whether cellularity can be used to detect diffuse glioma infiltration, the surrogate tumour-infiltration score for a whole slide was calculated using the cellularity value. This was then used to calculate the mAUROC across the three different tumour-infiltration tasks to compare with FastGlioma infiltration scores as shown in Extended Data Fig. 7.

Few-shot visualizations and model interpretability

We aimed to develop a whole-slide visualization method that can accurately and flexibly identify regions of tumour infiltration within SRH images to improve model interpretability. Studies on vision transformers have generally relied on plotting self-attention coefficients to generate data visualizations[57]. Unfortunately, this strategy does not guarantee uniformly high attention coefficients on foreground/tumour infiltrated regions and is known to produce spurious high attention coefficients in background regions[53]. We therefore developed a few-shot visualization strategy that uses a curated support set of expert physician-selected SRH patches, or keys, that include diverse examples of normal brain parenchyma and diffuse glioma subtypes. This strategy takes advantage of the representational power of the self-supervised patch tokenizer to identify similar SRH features within any given whole-slide field-of-view. Specifically, for any SRH patch query, x_q, within a whole-slide SRH image, we calculate the dot product between the tokenized query patch z_q and a support set of tokenized keys, S. We first determine whether the query patch is foreground/diagnostic by determining if the maximal dot product across the support set exceeds a threshold, ϕ. If not, then the patch is classified as background. If the query dot product exceeds ϕ for any patch in the support set, we then assign it a few-shot visualization score, s_q. This is defined as the difference between the maximum dot product from the tumour exemplars in the support subset, S_{tumour}, and the maximum dot product from the normal exemplars, S_{normal}:

$$\mathbf{s_q} = \begin{cases} s_q' & \text{if } \max\limits_{\forall p \in S} \text{sim}(z_q, z_p) > \phi \\ s_{\min} & \text{otherwise} \end{cases}, \tag{9}$$

where

$$s_q' = \max_{\forall t \in S_{\text{tumour}}} \text{sim}(z_t, z_q) - \max_{\forall n \in S_{\text{normal}}} \text{sim}(z_n, z_q), \tag{10}$$

and sim is the cosine similarity, $\text{sim}(x, y) = \frac{x^T y}{\|x\|\|y\|}$ and ϕ was 0.5 for our visualizations. This visualization strategy has the advantage of leveraging both the feature similarity between tumour patches and the dissimilarity between tumour and normal patches. If a patch has a high similarity to any of the tumour exemplars and high dissimilarity with the normal exemplars, then $s_q > 0$, and vice versa. Empirically, 10 or less patch exemplars per subset can yield high-quality and interpretable visualizations using FastGlioma. Moreover, this strategy demonstrates good zero-shot generalization to non-glioma brain tumour diagnoses without needing to add tumour-specific examples (for example, meningioma or medulloblastoma exemplars) to the support set, as shown in Extended Data Fig. 10.

Computational hardware and software

SRH images were processed using an Intel Core i76700K Skylake Quad-Core 4.0 central processing unit with our custom Python-based (v.3.9) mlins-package. We used the pydicom package (v.2.3.1) to process the SRH images from the NIO Imaging System. All archived postprocessed image patches were saved as 16-bit TIFF images and handled using the tifffile package (v.2020.10.1). All models were trained using the University of Michigan Advanced Research Computing (ARC) Armis2 high-performance computing cluster. Visual patch and whole-slide encoders were trained on NVIDIA A40 and Titan V100 graphical processing units (GPUs), respectively. Evaluations were performed on NVIDIA Titan V100 GPUs. All custom code for training and inference can be found in our open-source FastGlioma repository. Our models were implemented in PyTorch Lightning (v.1.8.4). We used the ImageNet pretrained ResNet-34 model from torchvision (v.0.14.0). Scikit-learn (v.1.4.1) was used to compute performance metrics on model predictions at both training and inference. Additional dependencies and specifications can be found at our GitHub page (https://github.com/MLNeurosurg/fastglioma).

Ethics and inclusion statement

Our research was approved by the University of Michigan institutional review board (HUM00083059) and the methods were carried out in accordance with the institutional review board's guidelines, regulations and policies. All human participants who met the inclusion criteria as stated above were included in the study.

Reporting summary

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

Data availability

The FastGlioma model parameters will be made publicly available for investigational use only under a Creative Commons Attribution Non Commercial Share Alike 4.0 license through HuggingFace

(https://huggingface.co/mlinslab/fastglioma). Institutional Review Board approval was obtained from all of the participating institutions for SRH imaging and data collection. Restrictions apply to the availability of raw patient imaging or genetic data, which were used with institutional permission through IRB approval for the current study, and are therefore not publicly available. All data sharing between medical centres is regulated through data use agreements with the study authors. A similar data sharing protocol may be established for interested investigators. Public access to an open-source repository of SRH images can be found at OpenSRH (https://opensrh.mlins.org/). Please contact the corresponding authors for any requests for data sharing. All requests will be evaluated based on institutional and departmental policies to determine whether the data requested is subject to intellectual property or patient privacy obligations. Data can be shared only for non-commercial academic purposes and will require a formal material transfer agreement.

Code availability

All code was implemented in Python (v.3.9) using PyTorch Lightning (v.1.8.4) as the primary machine learning framework. The following packages were used for data analysis: pydicom (v.2.3.1), tifffile (v.2020.10.1), PyTorch (v.1.13.0), torchvision (v.0.14.0), pandas (v.1.5.3), NumPy (v.1.24.4), matplotlib (v.3.6.3), opencv-python (v.3.4.18.65) and scikit-learn (v.1.4.1). For data visualization and scientific plotting, we used R (v.3.5.2) packages ggplot2 (v.3.3.5), dplyr (v.2.1.1), and the tidyverse (v.1.3.1). All code and scripts to reproduce the main experiments of this paper are available at GitHub (https://github.com/MLNeurosurg/fastglioma) under an MIT license.

49. Freudiger, C. W. et al. Stimulated Raman scattering microscopy with a robust fibre laser source. *Nat. Photon.* **8**, 153–159 (2014).
50. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2015).
51. Bardes, A., Ponce, J. & LeCun, Y. Variance-invariance-covariance regularization for self-supervised learning. In *Proc. ICLR 2022* https://openreview.net/forum?id=xm6YD62D1Ub (2022).
52. Jing, L., Vincent, P., LeCun, Y. & Tian, Y. Understanding dimensional collapse in contrastive self-supervised learning. Preprint at arxiv.org/abs/2110.09348 (2021).
53. Darcet, T., Oquab, M., Mairal, J. & Bojanowski, P. Vision transformers need registers. Preprint at arxiv.org/abs/2309.16588 (2023).
54. Li, Y., Si, S., Li, G., Hsieh, C.-J. & Bengio, S. Learnable Fourier features for multi-dimensional spatial positional encoding. In *Proc. 35th Conference on Neural Information Processing Systems* (eds. Beygelzimer, A, et al.) https://openreview.net/forum?id=YXjhRGvqfFN (2021).
55. Ji, M. et al. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. *Sci. Transl. Med.* **7**, 309ra163 (2015).
56. Chen, R. J. et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 16123–16134 (IEEE, 2022).
57. Caron, M. et al. Emerging properties in self-supervised vision transformers. In *Proc. 2021 ICCV* 9630–9640 (IEEE, 2021).
58. He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 386–397 (2020).
59. Zhai, X., Kolesnikov, A., Houlsby, N. & Beyer, L. Scaling vision transformers. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1204–1213 (IEEE, 2021).
60. Goyal, P. et al. Vision models are more robust and fair when pretrained on uncurated images without supervision. Preprint at arxiv.org/abs/2202.08360 (2022).
61. Cao, W., Mirjalili, V. & Raschka, S. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognit. Lett.* **140**, 325–331 (2020).
62. Paplhám, J. & Franc, V. A call to reflect on evaluation practices for age estimation: comparative analysis of the state-of-the-art and a unified benchmark. In *Proc. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2024).
63. Hollon, T. C. et al. Rapid intraoperative diagnosis of pediatric brain tumors using stimulated raman histology. *Cancer Res.* **78**, 278–289 (2018).
64. Oquab, M. et al. DINOv2: learning robust visual features without supervision. Preprint at arxiv.org/abs/2304.07193 (2023).
65. Eidel, O. et al. Tumor infiltration in enhancing and non-enhancing parts of glioblastoma: a correlation with histopathology. *PLoS ONE* **12**, e0169292 (2017).
66. Kubben, P. L. et al. Correlation between contrast enhancement on intraoperative magnetic resonance imaging and histopathology in glioblastoma. *Surg. Neurol. Int.* **3**, 158 (2012).
67. Roberts, D. W. et al. Coregistered fluorescence-enhanced tumor resection of malignant glioma: relationships between δ-aminolevulinic acid-induced protoporphyrin IX fluorescence, magnetic resonance imaging enhancement, and neuropathological parameters. *J. Neurosurg.* **114**, 595–603 (2011).
68. Zhao, S. et al. Intraoperative fluorescence-guided resection of high-grade malignant gliomas using 5-aminolevulinic acid-induced porphyrins: a systematic review and meta-analysis of prospective studies. *PLoS ONE* **8**, e63682 (2013).

Acknowledgements We thank K. Eddy and A. Marshall for providing technical support. This work was supported by the following National Institute of Health (NIH) funding sources: K12NS080223 (T.H.), R01CA226527 (D.A.O.), F31NS135973 (C.J.), and T32GM007863 (A.K.). This work was supported by the Chan Zuckerberg Foundation (CZI) Advancing Imaging Through Collaborative Project grant, Cook Family Brain Tumour Research Fund (T.H.), Mark Trauner Brain Research Fund, Zenkel Family Foundation (T.H.), Ian's Friends Foundation (T.H.), UM Precision Health Investigators Awards grant program (T.H.), Translational AI Award from the UM Department of Neurosurgery (T.H.), UM Stanley and Judith Frankel Institute for Heart and Brain Health Innovative Multidisciplinary Research Pilot Award (T.H.) and UM Research Scouts program (T.H.). This research was also supported, in part, through computational resources and services provided by Advanced Research Computing, a division of Information and Technology Services at the University of Michigan.

Author contributions A.K., M.P., X.H., K.S., C.J., C.F., M.B., S.H.-J. and T.H. contributed to the conceptualization, study design and analysis of results. A.K., M.P., X.H., K.S., C.J., A.R., E.S.H., A.C. and T.H. contributed to the experimentation, acquisition, analysis and interpretation of data. A.K., M.P., X.H., K.S., C.J. and T.H. contributed to generating the figures and tables for the manuscript. K.S., W.A.-H., L.W., A.P., L.I.K., G.W., M.B., S.H.-J. and T.H. contributed to obtaining tissue for SRH imaging. M.P., T.R.-P., S.C.-P. and M.M.-E. provided pathologic evaluation of tissue. All of the authors were involved in the editing, analysis and review of all data and manuscript versions.

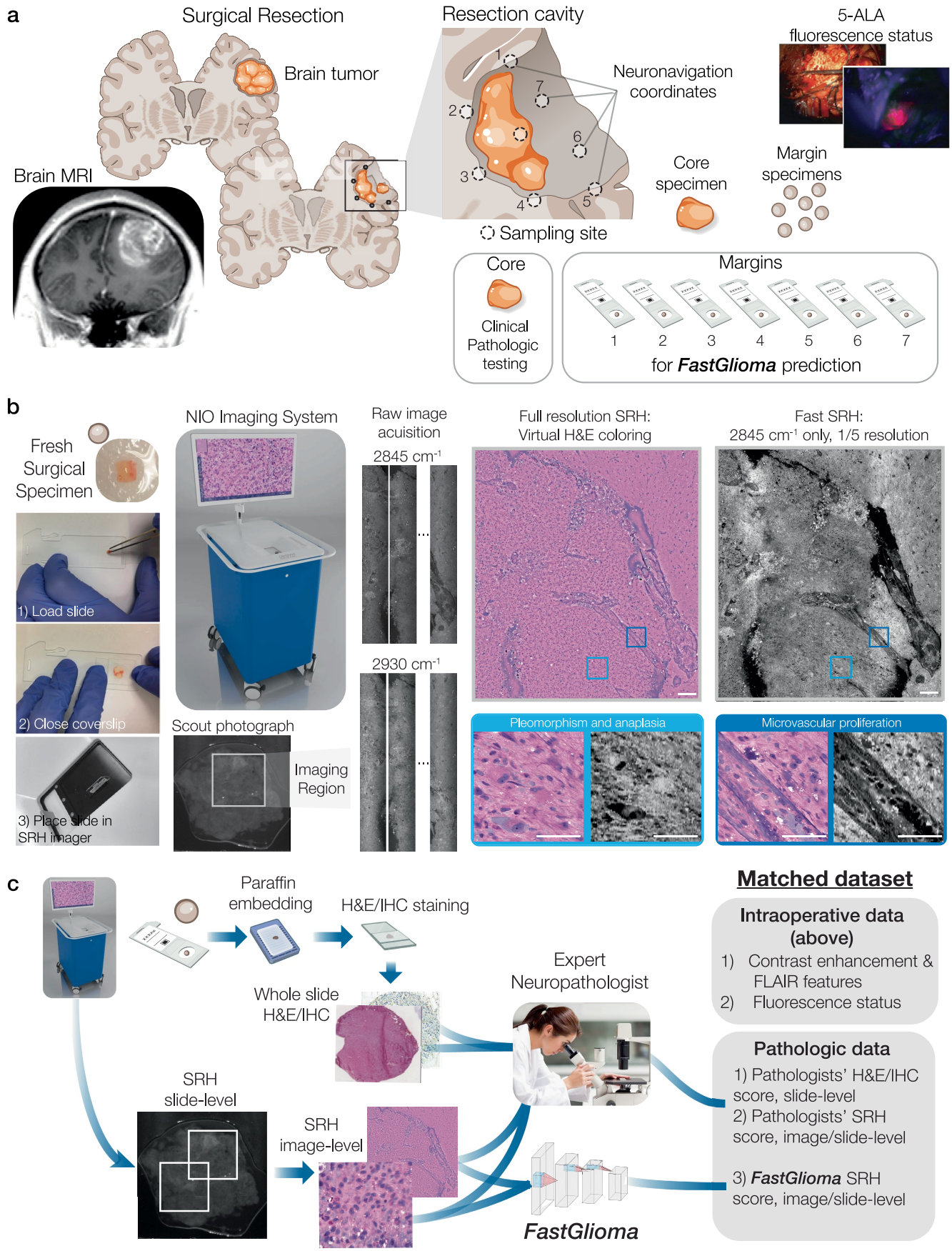
Competing interests C.F., D.A.O. and T.H. are shareholders in Invenio Imaging. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08169-3.

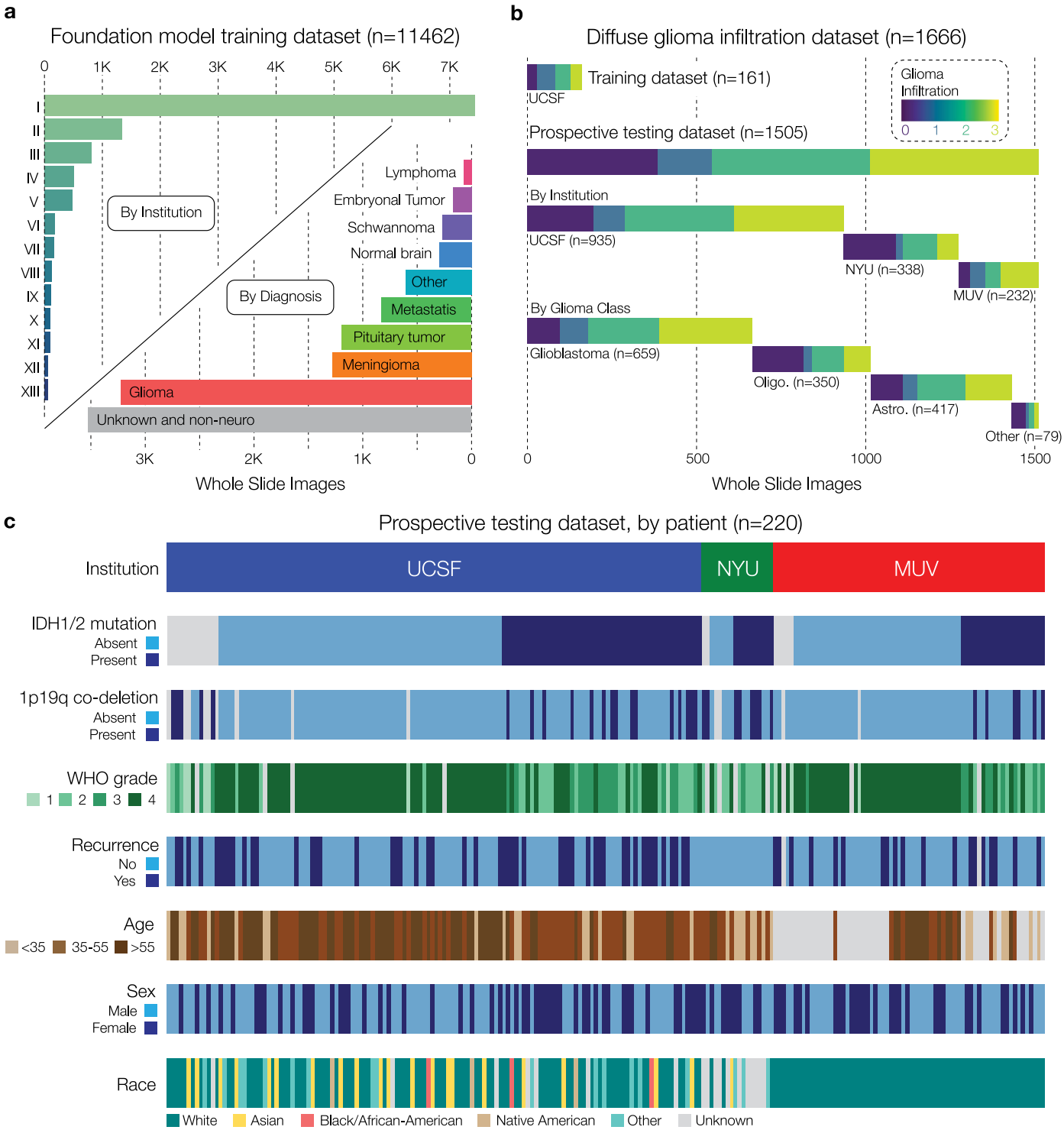
Correspondence and requests for materials should be addressed to Shawn Hervey-Jumper or Todd Hollon.

Peer review information *Nature* thanks Faisal Mahmood and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

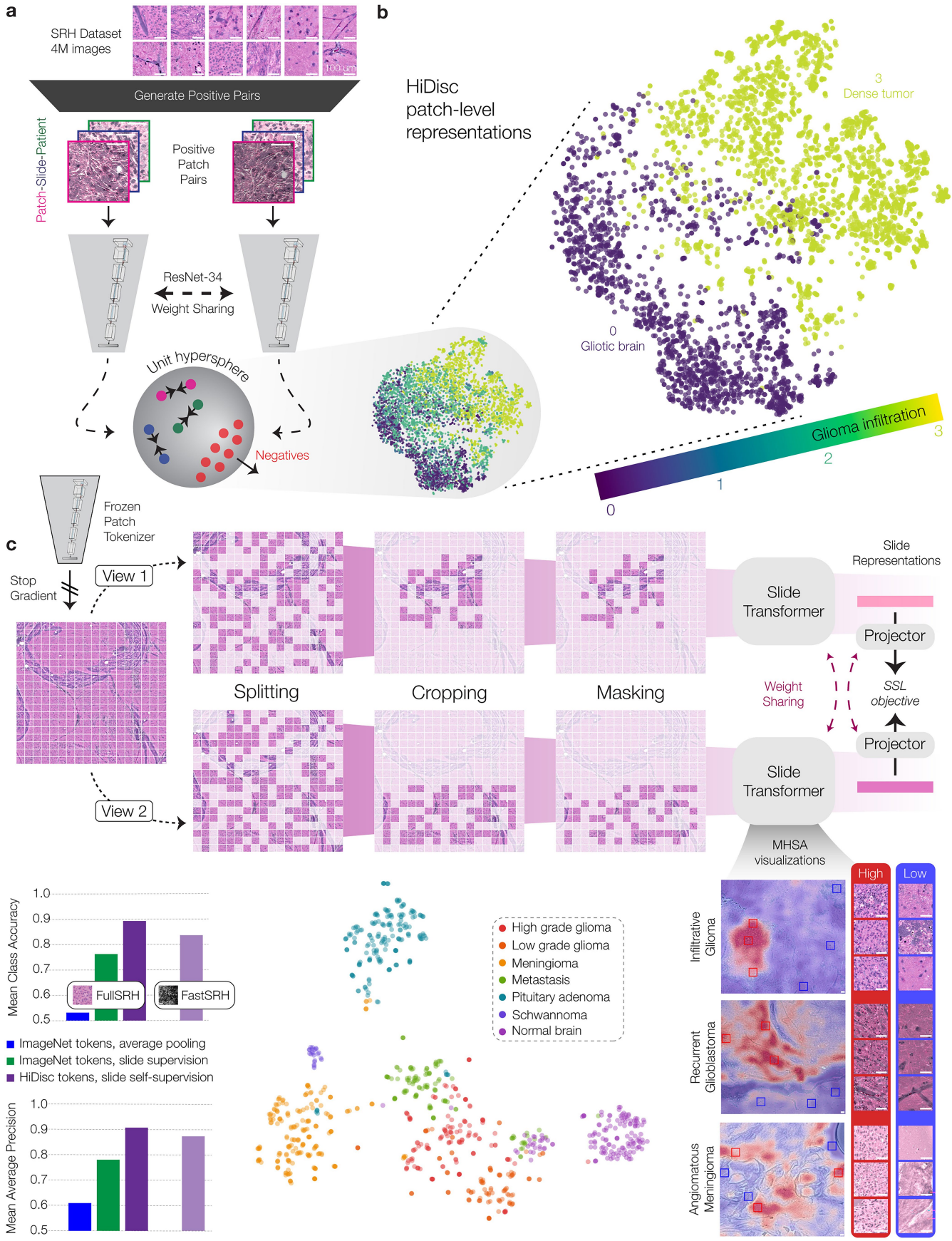


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Extended FastGlioma workflow and dataset generation. a, FastGlioma is intended for patients with a suspected diffuse glioma on preoperative imaging studies. Surgical specimens are sampled from the patient resection cavity. During sampling, neuronavigation coordinates and 5-ALA fluorescence status are recorded by a laboratory technician in real time. Core specimens are sent for clinical pathologic testing, including histologic and molecular classification with SRH[44], and the margin specimens are sent for FastGlioma prediction. Surgeons were encouraged to sample within the resection cavity according to their clinical judgement and context. **b**, Each surgical specimen is loaded into a premade microscope slide and imaged at the patient's bedside using the NIO Imaging System (Invenio Imaging, Inc., Santa Clara, CA). A scout photograph of the specimen is taken and the imaging field of view selected by the user using simple touchscreen instructions. SRH imaging does *not* require a skilled imaging technician. SRH images are acquired sequentially as strips at two Raman shifts, 2845 cm^{-1} and 2930 cm^{-1}. The size and number of strips to be acquired is set by the operator who defines the desired image size. Images can be acquired in either two-channel, full resolution mode or one-channel (2845 cm^{-1} only), low resolution fast mode to decrease imaging time by a factor of 10. **c**, Following SRH imaging, surgical specimens were removed from the microscope, formalin-fixed, and paraffin-embedded for H&E and immunohistochemistry (IHC) staining. Similar to previous protocols, tumour-specific IHC labels were used, if available, to determine the degree of tumour infiltration based on molecular genetics[4]. A neuropathologist from each institution then scores the degree of tumour infiltration within the SRH images and the H&E/IHC-stained slides on a 0–3 scale. Slide-level predictions are generated from image-level FastGlioma predictions by taking the average image-level prediction. Scale bars, 100 µm. Tumour resection and microscopy slide images were adapted from ref. 4. The neural network architectures were adapted from https://alexlenail.me/NN-SVG/.

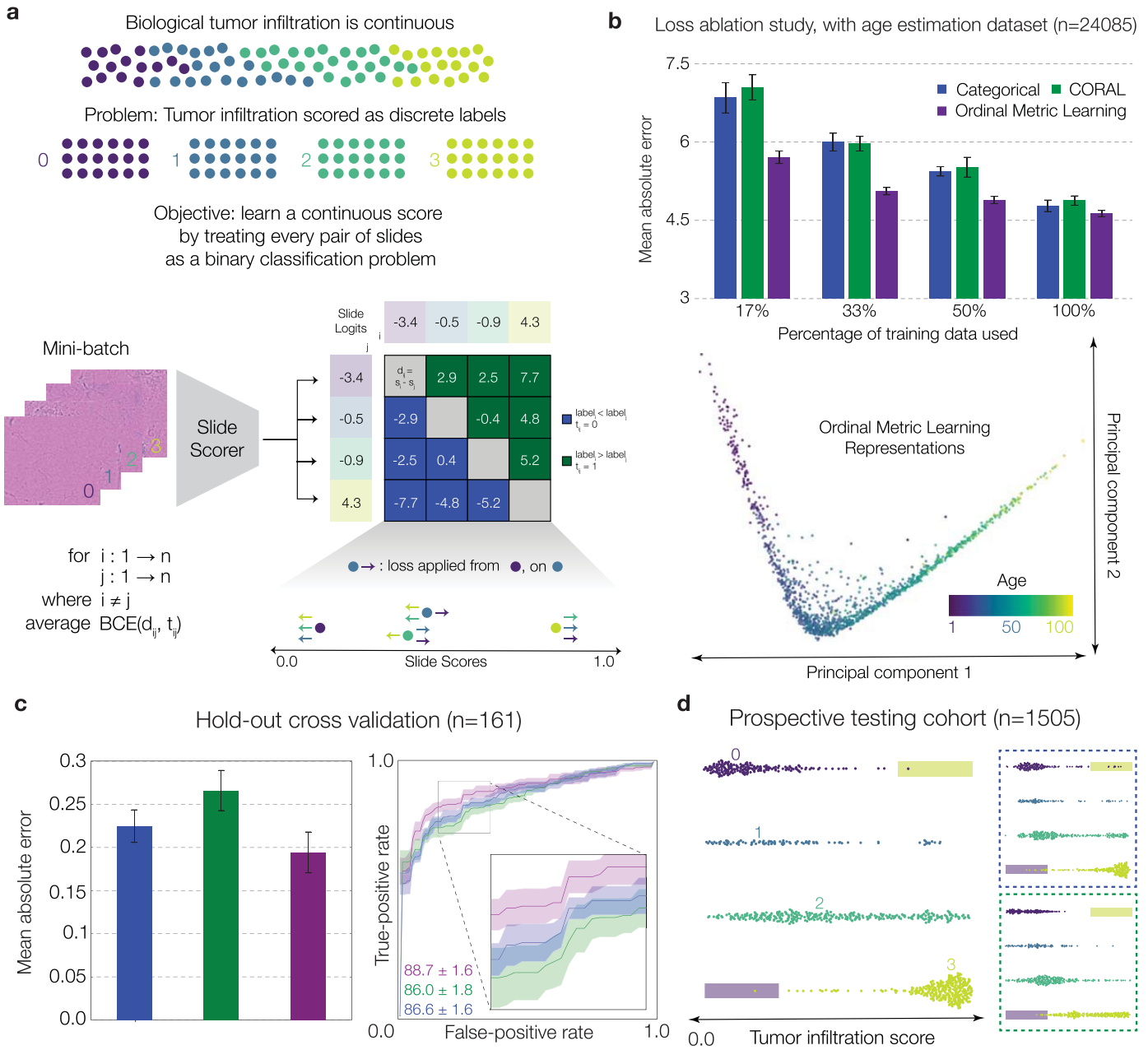


Extended Data Fig. 2 | SRH foundation model and FastGlioma training datasets. a, The foundation model training dataset consisted of over 11 K whole slide SRH images from 13 medical centres across the United States and Europe. Similar to other foundation model training datasets, such as Google's JFT-3B[59] or Meta's SEER[60] datasets, our SRH dataset was generated 'in the wild' at the discretion of the medical centres and treating physicians. The dataset consisted of a diverse set of central nervous system tumours, but also contained specimens from other organ systems, including head and neck tumours, breast, prostate, and lung surgical specimens. **b**, The diffuse glioma infiltration dataset consisted of two parts. The training dataset was generated entirely from a previously published study evaluating tumour margin specimens with SRH[4]. No additional annotation or data collection was performed for this study. The multicenter diffuse glioma testing dataset was generated from three tertiary medical centres with brain tumour programs. Details of data collection and annotation can be found in the Methods section. **c**, Details regarding the patient demographics, WHO glioma subtypes, recurrence status, tumour grade, and institution for the prospective testing dataset are shown.

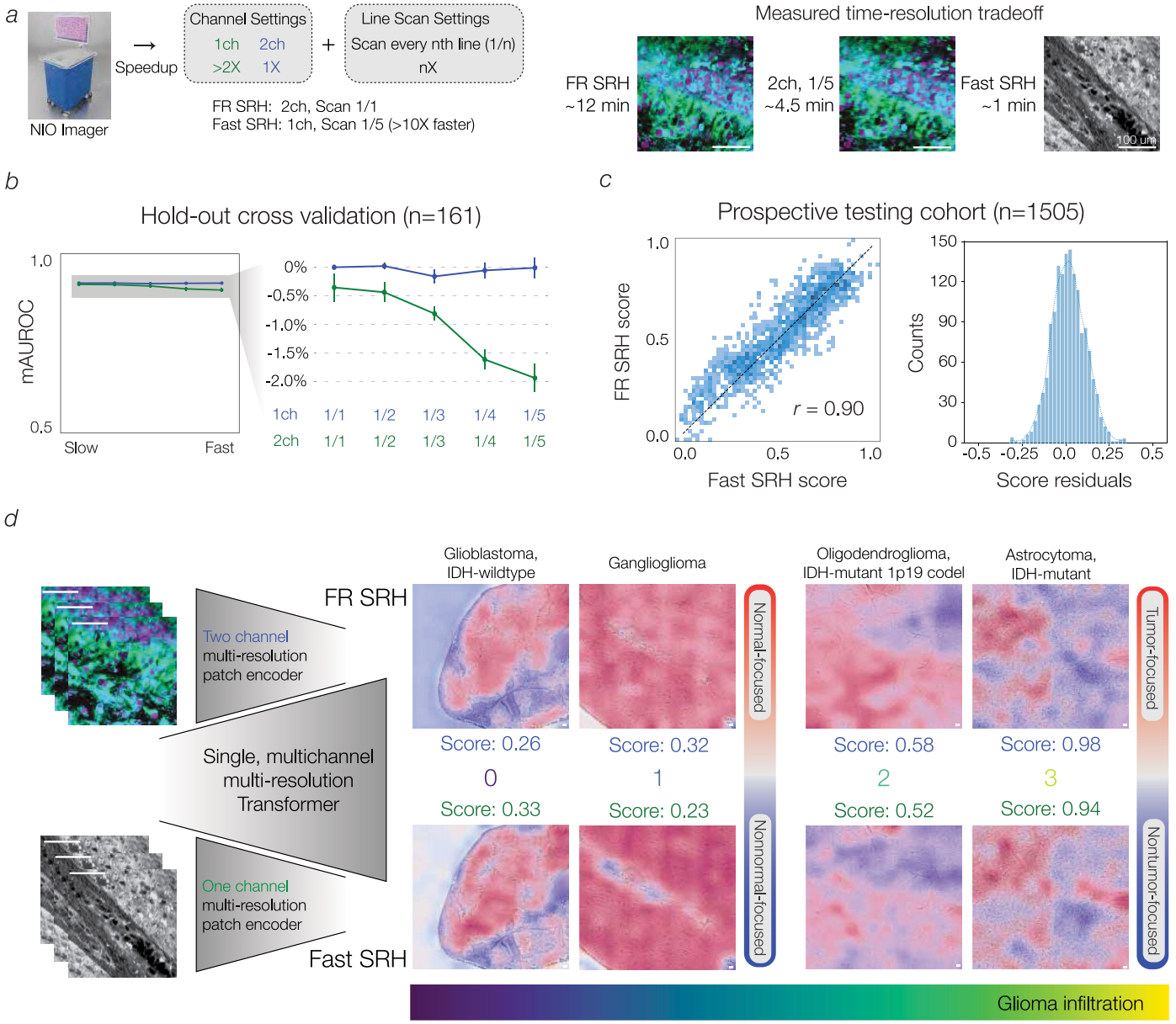


Extended Data Fig. 3 | See next page for caption.

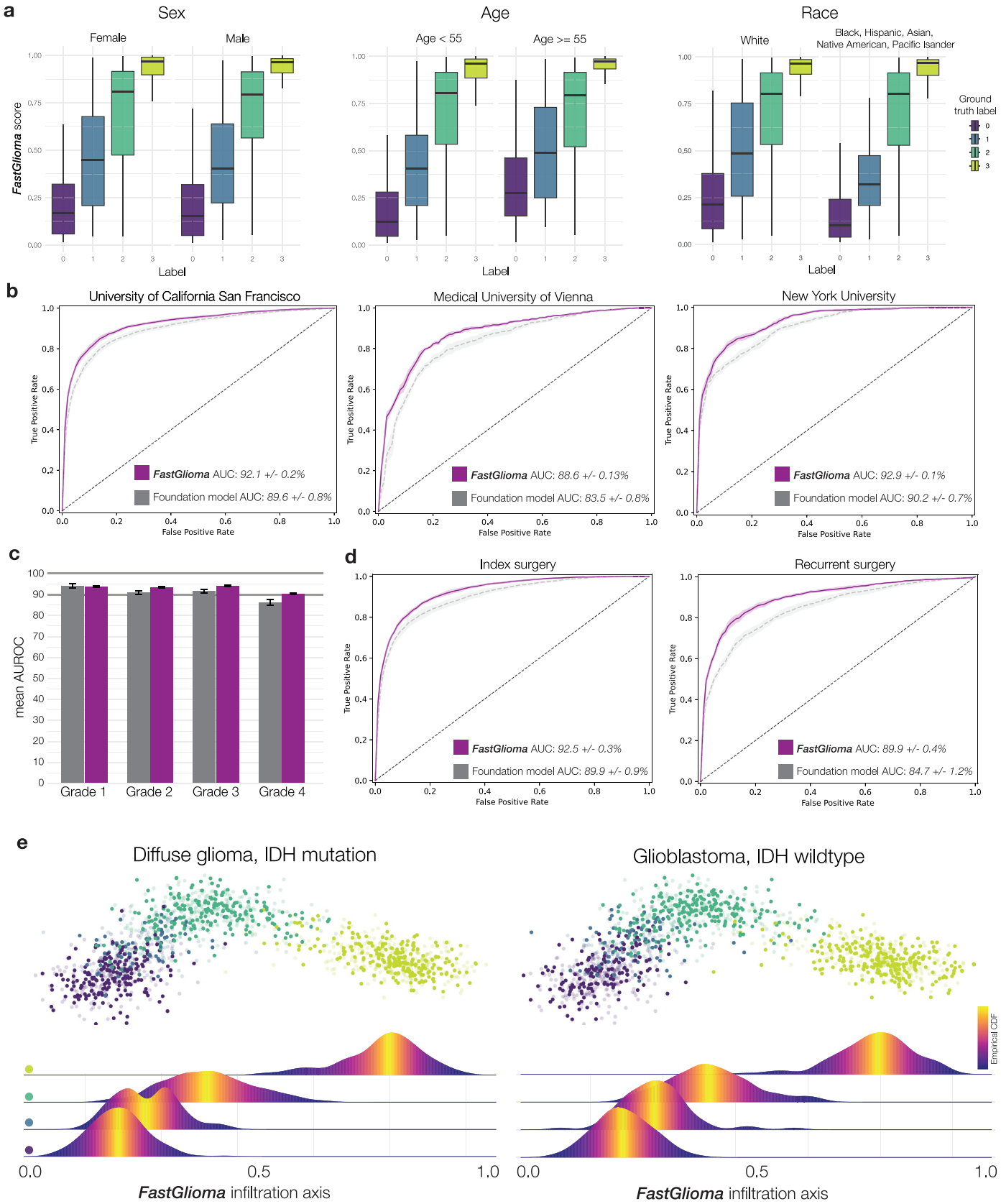
Extended Data Fig. 3 | Self-supervised foundation model training. **a**, The FastGlioma patch tokenizer is trained via self-supervised hierarchical contrastive learning on over 4 million 300×300 pixel SRH patches. Positive patch pairs are generated using the patch-slide-patient hierarchy of clinical SRH images. Patches undergo a feedforward pass through a ResNet-34 model and a contrastive learning objective is minimized[23]. **b**, t-SNE patch representations are plotted and coloured according to the ground truth whole slide label. Patch-level self-supervised training results in representations that span a tumour infiltration axis without the need for patch-level supervision. **c**, Whole slide self-supervision is accomplished by generating two views of a whole slide SRH image via random splitting, cropping, and masking of the patch tokens. The random sequence of patches in each view then undergo a feedforward pass through a slide transformer. The slide-level representations are then projected into a lower-dimensional projection space where a self-supervised objective is minimized (VICReg[51]). Our whole slide self-supervised training strategy was benchmarked using several internal SRH classification metrics, including multiclass brain tumour classification. Mean class accuracy of ablation studies are plotted. Whole slide SRH representations are plotted via t-SNE and show tumour diagnosis discrimination. Trained slide transformer produces interpretable multi-headed self-attention maps that differentiate tumour infiltration from normal brain and nondiagnostic regions. Scale bars, 100 µm. The neural network architectures were adapted from https://alexlenail.me/NN-SVG/.



Extended Data Fig. 4 | Ordinal metric learning for efficient ordinal representation learning. a, A major design challenge for FastGlioma was enforcing the model output to be a semantically meaningful continuous tumour infiltration score while being trained using discrete, ordinal labels. This design choice was warranted because biologic tumour infiltration is continuous given a surgical specimen or microscopy image. Each whole slide SRH image undergoes a feedforward pass through the slide encoder and slide scorer, which outputs a single scalar value. We then compute a pairwise distance matrix between each instance in the minibatch using the scores. Finally, a binary cross entropy (BCE) loss is computed between the pair's dot product, $x_i \cdot x_y$, and the label t_{ij} associated with the pair, defined as 1 if label_i is greater than label_j, and 0 otherwise. All BCE losses are summed and minimized. Intuitively, this representation learning strategy corresponds to ordering each whole slide according to the amount of accumulated 'force' generated by the other examples within the minibatch. For example, in the figure above, whole slides with label 3 will be maximally pushed to the right by all other examples. Whole slides with label 2 will be pushed to the right by two examples (0 s and 1 s) and to left by one example (3 s). **b**, We performed an ablation study on ordinal metric learning using the well studied and benchmarked face age estimation dataset (n = 24085)[61,62]. We show that ordinal metric learning outperforms other state-of-the-art ordinal regression methods and is especially effective when annotated data is sparse (plotted as mean ± s.d.). Ordinal metric learning produces a near linear subspace that orders each face image according to their respective ages. **c**, Hold-out cross validation studies were performed on our training dataset (n = 161) that showed improved ordinal regression performance on tumour infiltration scoring than other methods (ROC curves plotted as mean ± s.d.). **d**, Tumour infiltration scores are shown for ordinal metric learning versus other baseline strategies.

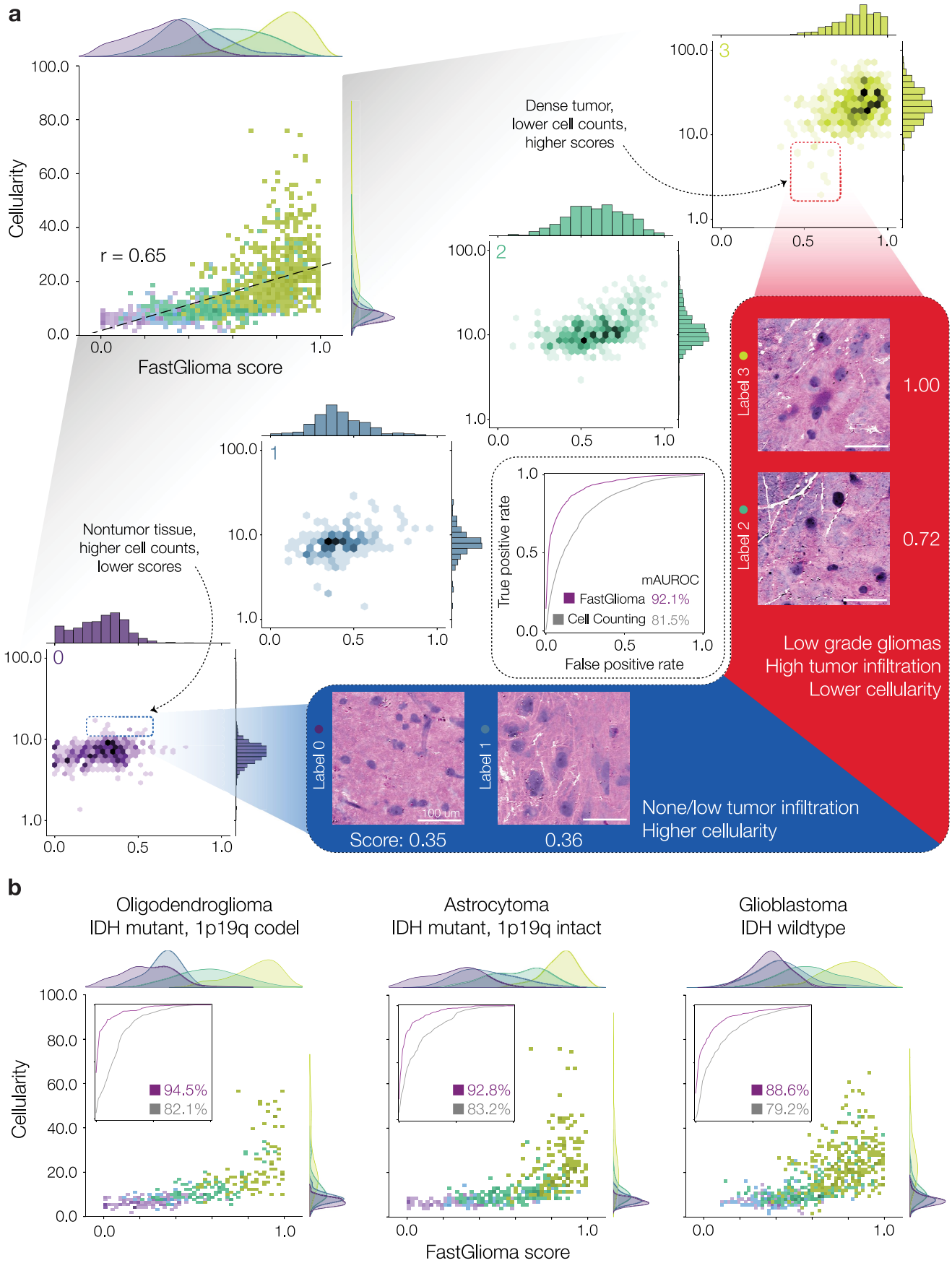


Extended Data Fig. 5 | Speed versus performance trade-off ablation studies. a, The NIO Imaging System can speed imaging time by acquiring only a single-channel image or by decreasing line scanning density. Fast SRH is a single channel greyscale image acquired at Raman wavenumber 2845 cm^{-1} and scanning every 5th line, which corresponds to approximately 10X speedup in imaging time. Examples of full resolution and low resolution images are shown. **b**, We performed hold-out cross validation ablation studies (n = 161) to examine the trade-off between image resolution/speed and model performance (plotted as mean ± s.d.). We found less than 2% decrease in mAUROC with the 10X speed-up of Fast SRH. **c**, We examined how tumour infiltration scores are related using both full resolution and fast SRH imaging. Scores are strongly correlated with a Pearson correlation coefficient, *r*, of 0.90 and score residual (S_{FR}-S_{Fast}) standard deviation of 0.102. **d**, A demonstration of the similarity between the whole slide encoder self-attention coefficients for full resolution versus fast SRH image. Despite significant pixel-level differences between the two imaging domains, self-attention and tumour infiltration scores on the whole slides are similar. Note in the Fast SRH row above, virtual H&E colorscheme is shown for visualization purposes only. The true underlying image is the low-resolution Fast SRH image. Scale bars, 100 µm.



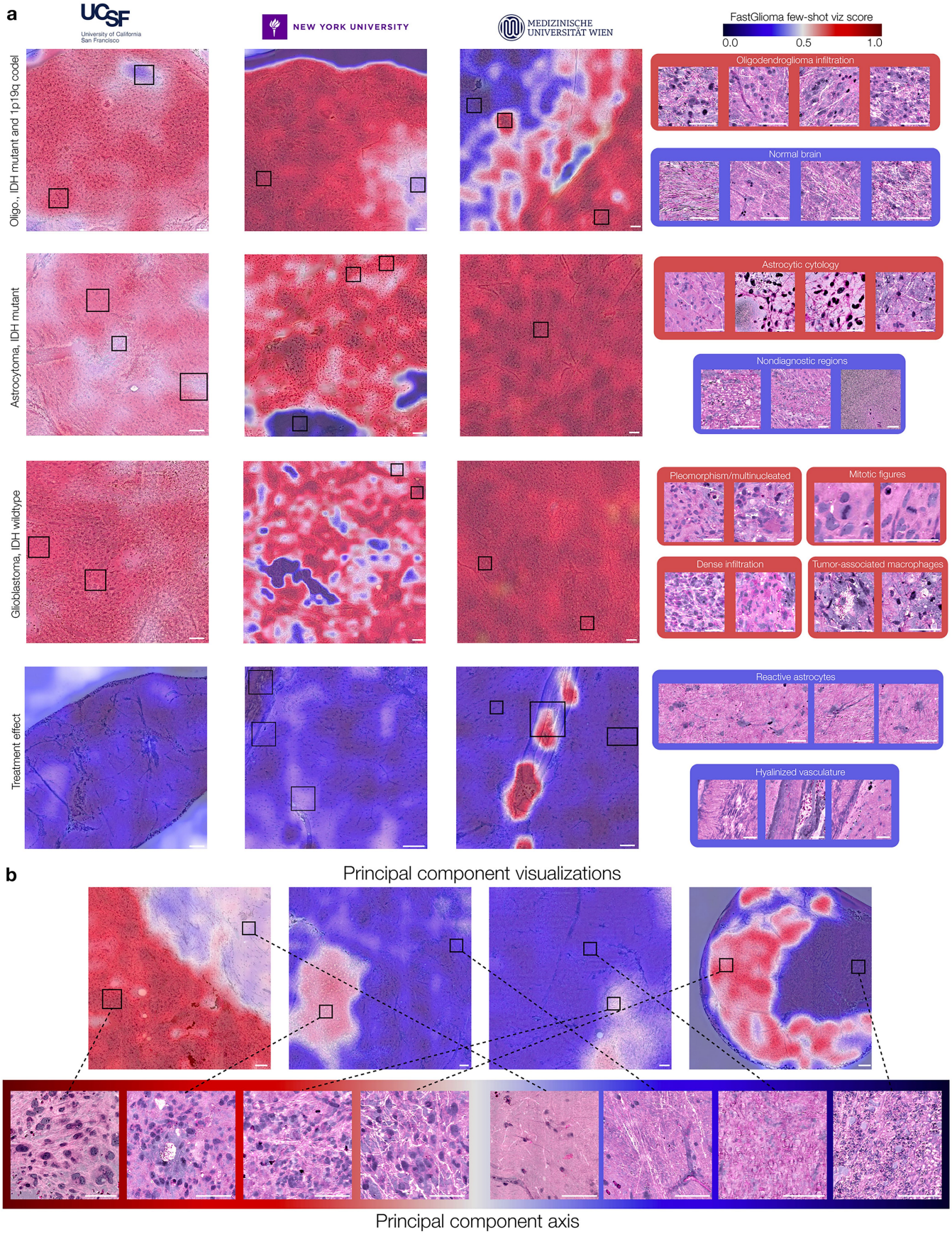
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Extended prospective testing results with subgroup analysis. a, Subgroup analysis by patient demographics. FastGlioma scores by sex, age, and race are shown using box and whisker plots in the standardized quartile format. **b**, Subgroup analysis by medical centre. mAUROC value is plotted for both FastGlioma and the SRH foundation model (plotted as mean ± s.d.). Performance remained high across the three external testing medical centres. **c**, Mean AUROC values are plotted by tumour grade. Importantly, similar performance is seen across all diffuse glioma grades, indicating that FastGlioma scores are not reliant on specific histologic features found within diffuse glioma grades. Lower grade cytologic features and low tumour infiltration were preliminary concerns about the performance of FastGlioma that did not bear out in prospective testing. Mean AUROC values and standard deviations are shown. **d**, Subgroup analysis by index/first surgery versus surgery for recurrent tumour (ROC curves plotted as mean ± s.d.). A total of 442 surgical specimens were sampled from patients with recurrent tumours. FastGlioma maintained good performance for detecting tumour infiltration in recurrent tumours, which is known to be challenging for intraoperative frozen sectioning[42,43]. **e**, Whole slide SRH representations are shown for IDH mutant and IDH wildtype diffuse gliomas. Points are coloured by their ground truth tumour infiltration scores. Despite differences in molecular features, the representation of whole slide SRH images is similarly distributed according to tumour infiltration. These findings demonstrate that FastGlioma is invariant to molecular markers, tumour subtypes, and histologic features for defining the degree of tumour infiltration within SRH images. This is contrasted with 5-ALA fluorescence which is only approved for use in glioblastomas, IDH-wildtype.



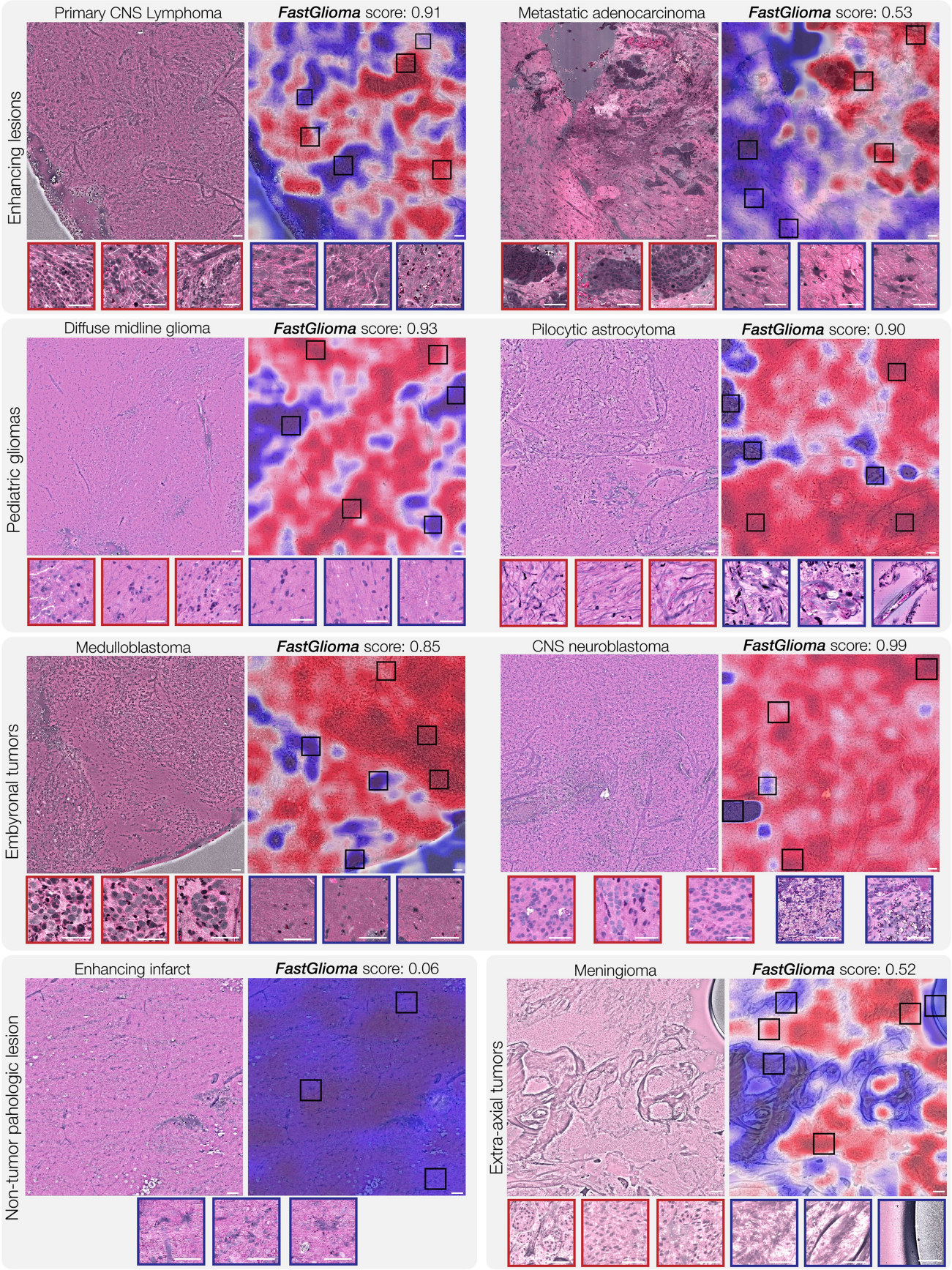
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Cytologic analysis of FastGlioma tumour infiltration scores. a, Increased cellularity, or cellular density, is a cytologic feature of diffuse gliomas and is correlated with the degree of tumour infiltration[55,63]. However, cellularity alone is a poor predictor for tumour infiltration *across* tumour grade, diffuse glioma molecular subgroups, and degree of infiltration. Cellularity as a predictor performs progressively worse as degree of tumour infiltration decreases and with lower tumour grades[4]. We evaluate the relationship between cellularity and FastGlioma scores. Cellularity and FastGlioma scores have a Pearson correlation coefficient of 0.65. Consistent with previous results, we identify surgical specimens without tumour infiltration (0 s) that have cellular densities comparable to specimens with a low to moderate degree of tumour infiltration (1 s and 2 s). Additionally, some dense tumour specimens (3 s) were found to have relatively low cellularity (<10). Examples of both of these scenarios are shown. The kernel density estimates (KDE) of the FastGlioma scores demonstrate the expected ordinal distribution across the ground truth labels; cellularity KDEs are poorly distributed according to ground truth labels, especially for 0–2 scores. FastGlioma has a clinically significant increase in mAUROC compared to cellularity-based predictions. **b**, Cellularity versus FastGlioma scores are plotted by molecular subgroup. Despite differences in cytologic features (anaplasia, pleomorphism, nuclear-cytoplasmic ratios) across molecular subgroups, FastGlioma tumour infiltration scoring is consistently accurate. Diffuse glioma molecular subgroups show a similar correlation between cellularity and score values. FastGlioma effectively uses the full tumour infiltration score range to quantify the degree of tumour infiltration across all molecular subgroups. FastGlioma outperforms cellularity-based metrics for predicting degree of tumour infiltration. Scale bars, 100 µm.



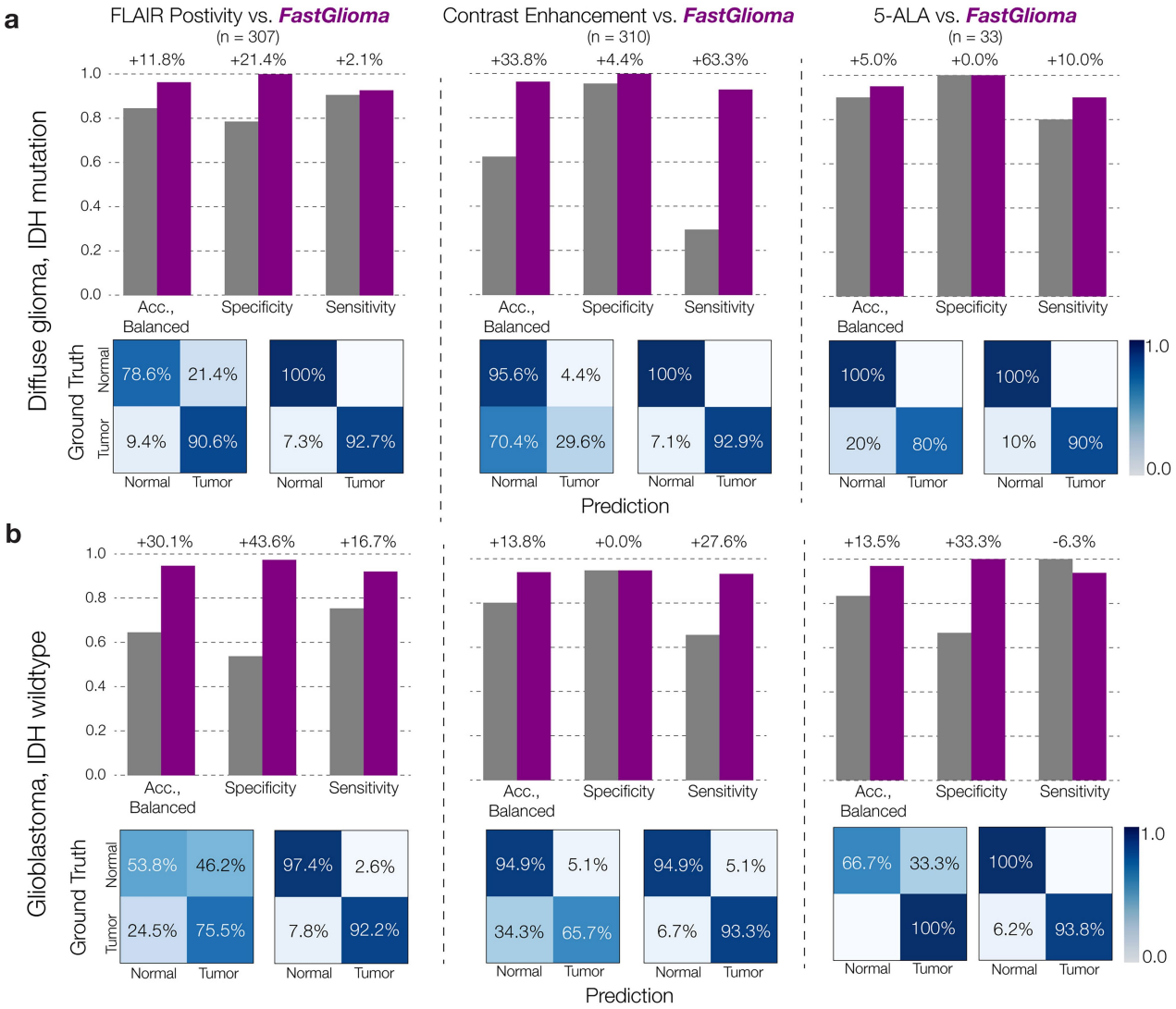
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Few-shot visualizations of diffuse glioma infiltration. **a**, A panel of whole slide SRH images across diffuse glioma molecular subgroups are shown. We have selected cases across the prospective testing medical centres. Few-shot visualizations identify tumour infiltrated regions in molecular oligodendrogliomas, astrocytomas, and glioblastoma. Right, we show insets from each whole slide image to demonstrate the diversity of underlying histologic and cytologic features found within diffuse glioma subgroups. We also show several SRH images with 0 ground truth scores from patients diagnosed with glioblastomas who previously underwent surgery and chemoradiation. SRH images show evidence of treatment effect, including reactive astrocytes and hyalinized blood vessels. Despite these non-neoplastic pathologic findings, FastGlioma gives low tumour infiltration scores and does not identify regions of dense tumour infiltration. However, the limitations of few-shot visualization are shown here with erroneous tumour matching shown in some regions of pathologic hyalinized vasculature. **b**, Similar to previous studies on visual foundation models[64], we used principal component analysis (PCA) on patch features extracted by FastGlioma. PCA provides an unsupervised visualization strategy to better elucidate the learned patch features. We observe that tumour infiltrated regions are matched between SRH images despite changes in degree of tumour infiltration, molecular subtype, and histologic features. Detailed comparisons between FastGlioma scores, few-shot visualizations, and SRH images can be found at fastglioma.mlins.org. Scale bars, 100 μm.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Zero-shot predictions and visualizations. We evaluated FastGlioma's zero-shot generalization to non-adult-type diffuse glioma brain tumour diagnoses. FastGlioma fine-tuning was restricted to adult-type diffuse gliomas; here we show FastGlioma predictions on non-adult-type diffuse gliomas (zero-shot). Moreover, few-shot visualization heatmaps are shown for non-glioma brain tumours using *diffuse glioma SRH keys*. The figure shows examples of brain tumours from several broad categories, such as ring-enhancing lesions, paediatric gliomas, embryonal tumours, non-tumour lesions, and extraaxial tumours. CNS lymphomas are infiltrative brain tumours and can be challenging to differentiate from glioblastomas intraoperatively. SRH images show classic perivascular arrangement (angiocentricity) with tumour cells forming layers around the blood vessels. Residual microscopic metastatic tumour is the major cause of tumour recurrence after surgery. FastGlioma identified regions of microscopic residual metastatic tumour. Tumour infiltration from paediatric gliomas, such as diffuse midline gliomas and pilocytic astrocytomas, is detected by FastGlioma. Paediatric diffuse midline gliomas can have a spectrum of histologic morphologies that can differ in appearance compared to glioblastomas. Rosenthal fibres (black lobules), present in pilocytic astrocytomas but rare in diffuse gliomas, are not a source of error for FastGlioma. Acellular and nondiagnostic regions are not segmented as regions of tumour infiltration. Embryonal tumours have distinctive 'small round blue cell' cytologic features that FastGlioma correctly identifies as tumour infiltration. FastGlioma provides real-time confirmation of the *absence* of tumour infiltration during surgery, such as infarcts. Meningiomas are known to invade adjacent dura, which is the major source of tumour recurrence. FastGlioma can identify meningioma infiltration within normal dura sampled at the dural margin. These results provide evidence for the potential of FastGlioma to generalize *beyond brain tumours, such as breast, lung, and prostate cancer*. Scale bars, 100 µm.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Comparison of FastGlioma and surgical adjuncts for diffuse glioma resection. a, Classification performance for distinguishing between normal brain and dense tumour by FastGlioma versus surgical adjuncts on diffuse gliomas, IDH-mutation. FLAIR positivity in IDH-mutant gliomas is considered a radiographic marker of tumour infiltration and is used to define extent of resection[33]. In our matched cohort, FastGlioma balanced accuracy is +11.8% compared to FLAIR positivity. FLAIR had relatively lower specificity due to a higher number of false positives. Non-specific causes of FLAIR positivity, such as vasogenic oedema, result in decreased accuracy. **b**, Classification performance of FastGlioma versus surgical adjuncts on glioblastomas, IDH-wildtype. Patients with complete removal of contrast enhancing regions have improved progression-free and overall survival[2,6]. In our matched cohort, FastGlioma balanced accuracy is +13.8% compared to contrast enhancement for IDH-wildtype glioblastomas. Tumour infiltration is known to extend beyond regions of contrast enhancement and dense, viable tumour is identified in non-enhancing regions[6,55]. Our findings of poor sensitivity are consistent with previous radiopathologic correlation studies[65,66]. 5-ALA fluorescence is used in glioblastoma surgery to guide surgical resection[28]. Previous studies have reported a wide range of sensitivity and specificity values of 5-ALA fluorescence as an indicator of tumour infiltration depending on tumour types, tumour grade, and recurrence status[30,67,68]. In our matched cohort, FastGlioma achieved a + 13.5% increase in balanced accuracy compared to 5-ALA. **c**, Illustrative examples of matched surgical specimens with FastGlioma predictions, neuronavigation coordinates, radiographic features, and 5-ALA status. The lower left shows concordance between FLAIR positivity and FastGlioma prediction with dense tumour within the specimen. Centre shows a specimen outside the contrast enhancing rim of a glioblastoma with white matter tumour infiltration identified by FastGlioma. Lower right shows an example of 5-ALA fluorescent positive tissue. The patient had a recurrent glioblastoma with associated treatment effect and reactive astrocytes.

nature portfolio

Corresponding author(s): Todd Hollon, MD

Last updated by author(s): Mar 7, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The software and code used in this study for development of the FastGlioma model is publicly available at https://github.com/MLNeurosurg/fastglioma. Proprietary software used in the NIO Imaging System (Invenio Imaging, Inc) was used for optical imaging.
Data analysis	All code was implemented in Python (version 3.9) using PyTorch Lightning (1.8.4) as the primary machine learning framework. The following packages were used for data analysis: pydicom (2.3.1), tifffile (2020.10.10), PyTorch (1.13.0), torchvision (0.14.0), pandas (1.5.3), NumPy (1.23.5), matplotlib (3.6.3), opencv-python (3.7.0), and scikit-learn (1.4.1). For data visualization and scientific plotting, we used R (3.5.2) packages ggplot2 (3.3.5), dplyr (2.1.1), and the tidyverse (1.3.1). All code and scripts to reproduce the experiments of this paper are available on GitHub at https://github.com/MLNeurosurg/fastglioma under an MIT license.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The FastGlioma model parameters will be made publicly available for investigational use only under a MIT license through HuggingFace (https://huggingface.co/mlinslab/fastglioma). Institutional Review Board approval was obtained from all participating institutions for SRH imaging and data collection. Restrictions apply to the availability of raw patient imaging or genetic data, which were used with institutional permission through IRB approval for the current study, and are thus not publicly available. All data sharing between medical centers is regulated through data use agreements with the study authors. A similar data sharing protocol may be established for interested investigators. Public access to an open-source repository of SRH images can be found at OpenSRH (https://opensrh.mlins.org/). Please contact the corresponding authors (T.H., S.H.J.) for any requests for data sharing. All requests will be evaluated based on institutional and departmental policies to determine whether the data requested is subject to intellectual property or patient privacy obligations. Data can only be shared for non-commercial academic purposes and will require a formal material transfer agreement.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	All reported findings apply to patients of any sex or gender. Subgroup analysis was performed based on sex and gender. See Extended Data Figure 6.
Reporting on race, ethnicity, or other socially relevant groupings	All reported findings apply to patients of any race, ethnicity, and other socially relevant groupings. Subgroup analysis was performed based on race, ethnicity, and other social relevant groupings. See Extended Data Figure 6.
Population characteristics	Prospective testing dataset characteristics: Sex: Males=128, Female=92 Age: < 35 year=32, 35-55 years=74, > 55 years=75 Race: White=157, Non-white=45 Recurrence status: Primary=148, Recurrent=70, unknown=2 WHO grade: I=10, II=38, III=31, IV=130, unknown=11 IDH status: Wildtype=119, Mutant=81, unknown=20
Recruitment	Patients were recruited by clinical staff at the time of initial presentation and/or immediately preoperatively. Patients with suspected diffuse glioma were identified as potential study candidates preoperatively. Patients were recruited consecutively and based on clinical presentation without systematic screening or exclusion. All patients were treated according the standard-of-care based on the local hospital system. Self-selection bias was of minimal concern because FastGlioma testing was non-interventional and did not effect patient treatment. Inclusion criteria for SRH imaging and FastGlioma testing: (1) patient age ≥ 18 years old, (2) a suspected diffuse glioma on preoperative radiographic imaging, and (3) planned brain tumor resection. Exclusion criteria included: (1) aborted tumor resection, (2) non-glioma final pathology, and (3) SRH imager malfunction.
Ethics oversight	University of Michigan Institutional Review Board (HUM00083059)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We elected to perform prospective, international, multicenter clinical testing of FastGlioma in order to adhere to the rigorous standards of responsible machine learning in healthcare. Our prospective clinical testing was designed using the same principles as a non-inferiority diagnostic clinical trial. We set the expected accuracy for both the control and experimental arms to be 93.2%, the equivalence limit was set to 5%, power to 90%, and alpha to 2%, resulting in a sample size value of 565 SRH images from surgical margins. We aimed to achieve this sample size for both IDH-wildtype and IDH-mutant diffuse gliomas, resulting in a final minimum sample size of 1130 surgical specimens. All sample size calculations were performed using the epiR package

(version 2.0.46) in R (version 3.6.3).

Data exclusions	Exclusion criteria for SRH imaging was (1) insufficient diagnostic tissue as determined by surgeon or pathologist, (2) grossly inadequate tissue (e.g. hemorrhagic, necrotic, fibrous, liquid, etc.), and (3) SRH imager malfunction.
Replication	We performed a leave-K-out cross-validation, such that our methods were replicated across different subsets, or folds, of the training dataset. We were able to replicate FastGlioma performance results across each fold, which included a minimum of 5 held-out evaluation sets. We also replicated our predictions results across three external, independent, international medical centers.
Randomization	Randomization was not relevant for this study. Study design was as a single-arm, non-interventional, diagnostic, clinical trial. Direct one-to-one comparison was made between FastGlioma versus image-guided and fluorescence-guided surgery. Matched comparison between FastGlioma predictions of surgical specimens and radiographic/fluorescent features necessitated evaluation for both modalities. Surgical specimens imaged in the operating room were selected by the clinician and the remainder of the clinical brain tumor specimen was sent for final pathologic analysis.
Blinding	Surgical procedures could not be blinded to the surgeon. Pathologists evaluated the SRH and H&E/IHC specimens independently without patient information. Final FastGlioma inference was performed without PHI and tumor infiltration labels. There were no sources of data leakage during model training or inference.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Our prospective study was a non-interventional, diagnostic study so does not meet criteria for clinical trial registration.
Study protocol	Prospective study protocol can be found at this link: https://www.dropbox.com/scl/fi/rz5x8g4ybnqo7rtj4nh5v/Prospective-Diagnostic-Testing-of-FastGlioma-for-Intraoperative-Tumor-Infiltration-Detection.docx?rlkey=6yrcuj7slucbjua55ikr2hauv&dl=0
Data collection	Three medical centers acted as external FastGlioma testing sites: University of California San Francisco, New York University, and Medical University of Vienna. Each center collected prospective data over the following intervals: UCSF: 03/2022-02/2023, NYU: 06/2020-03/2023, MUV: 11/2020-07/2023, Each medical center prospectively enrolled patients for testing. Inclusion criteria were: (1) patient age greater or equal to 18 years old, (2) a suspected diffuse glioma on preoperative radiographic imaging, and (3) planned brain tumor resection. Exclusion criteria included: (1) aborted tumor resection, (2) non-glioma final pathology, and (3) SRH imager malfunction. We aimed to accurately simulate the clinical setting that FastGlioma would be implemented for surgical interventions.
Outcomes	Primary testing endpoint was SRH-based tumor infiltation detection with FastGlioma. Second testing endpoint was was to compare the FastGlioma intraoperative workflow (experimental arm) with the two most common surgical adjuncts for identifying tumor infiltration intraoperatively (control arm) in a simulated prospective surgical trial. 'Simulated' terminology is used because FastGlioma is not approved by the Food and Drug Administration or European Medicines Agency to guide treatment decisions, such as extent of tumor resections.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*