

1 **Deep learning-based identification of necrosis and microvascular proliferation in adult diffuse**

2 **gliomas from whole-slide images**

3

4 Yumeng Guo^{1,#}, Hanli Huang^{1,#}, Xing Liu², Wanjing Zou², Fufang Qiu^{1,3}, Yunqing Liu¹, Ruichao

5 Chai^{1,2}, Tao Jiang^{2*}, Jiguang Wang^{1,4*}

6

7 ¹Division of Life Science, Department of Chemical and Biological Engineering, State Key Laboratory

8 of Molecular Neuroscience, The Hong Kong University of Science and Technology, Hong Kong SAR,

9 China

10 ²Department of Neuropathology, Beijing Neurosurgical Institute, Beijing Tiantan Hospital of Capital

11 Medical University, Beijing 100070, China

12 ³Department of Neurosurgery, Huashan Hospital, Shanghai Medical College, Fudan University,

13 Shanghai

14 ⁴SIAT-HKUST Joint Laboratory of Cell Evolution and Digital Health, HKUST Shenzhen-Hong Kong

15 Collaborative Innovation Research Institute, Futian, Shenzhen, China

16

17 #Co-first authors.

18 *Corresponding authors.

19 **Abstract**

20 For adult diffuse gliomas (ADGs), most grading can be achieved through molecular subtyping,
 21 retaining only two key histopathological features for high-grade glioma (HGG): necrosis (NEC) and
 22 microvascular proliferation (MVP). We developed a deep learning (DL) framework to automatically
 23 identify and characterize these features. We trained patch-level models to detect and quantify NEC and
 24 MVP using a dataset that employed active learning, incorporating patches from 621 whole-slide images
 25 (WSIs) from the Chinese Glioma Genome Atlas (CGGA). Utilizing trained patch-level models, we
 26 effectively integrated the predicted outcomes and positions of individual patches within WSIs from The
 27 Cancer Genome Atlas (TCGA) cohort to form datasets. Subsequently, we introduced a patient-level
 28 model, named PLNet (Probability Localization Network), which was trained on these datasets to
 29 facilitate patient diagnosis. We also explored the subtypes of NEC and MVP based on the features
 30 extracted from patch-level models with clustering process applied on all positive patches. The patient-
 31 level models demonstrated exceptional performance, achieving an AUC of 0.9968, 0.9995 and AUPRC
 32 of 0.9788, 0.9860 for NEC and MVP, respectively. Compared to pathological reports, our patient-level
 33 models achieved the accuracy of 88.05% for NEC and 90.20% for MVP, along with a sensitivity of
 34 73.68% and 77%. When sensitivity was set at 80%, the accuracy for NEC reached 79.28% and for
 35 MVP reached 77.55%. DL models enabled more efficient and accurate histopathological image
 36 analysis which will aid traditional glioma diagnosis. Clustering-based analyses utilizing features
 37 extracted from patch-level models could further investigate the subtypes of NEC and MVP.

38

39 **Keywords**

adult diffuse gliomas, necrosis, microvascular proliferation, histopathologic whole-slide images, deep learning

42

43 **Introduction**

44 Glioma is the most common histological type of primary Central Nervous System (CNS) cancer, with
45 an annual incidence of 5-6 cases per 100,000 individuals, and adult diffuse gliomas account for
46 approximately 70% of all glioma cases¹. According to the 2021 World Health Organization (WHO)
47 classification of CNS tumors², genetic mutations and DNA methylation have been recommended as
48 key molecular parameters to be considered for guiding clinical diagnosis, tumor grading, and
49 therapeutic interventions. Meanwhile, histological features are also critically important during the
50 diagnostic process. Especially the presence of necrosis (NEC) and/or microvascular proliferation (MVP)
51 is considered as hallmarks of HGG in adult diffuse gliomas, which is more aggressive and indicates a
52 poor prognosis³.

53

54 Histopathology analyses are routine assessments conducted by pathologists to microscopically examine
55 a biopsy or surgical specimen that is processed and fixed onto glass slides⁴. Histopathology studies
56 provide a gold standard for diagnosis, disease treatment, and prognostic evaluation. However, it has
57 been demonstrated that for many diagnostic tasks, human assessments are highly subjective, and the
58 reproducibility among different pathologists is less than optimal^{5, 6}. For example, in the diagnosis of
59 specific types of breast cancer, the pathological consistency can be as low as 48%⁷, a scenario also
60 observed in prostate cancer⁸. Additionally, the examination of histopathological sections is a highly
61 complicated and time-consuming task. Typically, it demands highly experienced doctors to

meticulously examine gigapixel-sized whole-slide images (WSIs). This field necessitates more
reproducible, precise, and efficient diagnostics for histopathological analysis.

The current application of deep learning in pathology analysis primarily focuses on automatic diagnosis
tasks, such as tumor detection, classification and grading, image segmentation, etc. In recent years,
numerous exciting breakthroughs have been achieved in this field. Earlier in 2016, the International
Symposium on Biomedical Imaging (ISBI) organized a grand challenge to evaluate computational
systems for the automatic detection of metastatic breast cancer in WSI of sentinel lymph node biopsies.
During this challenge, a patch-level classification deep learning model achieved comparable
performance to that of pathologists, showcasing the potential of deep learning to significantly enhance
the accuracy of pathological diagnoses⁹. In 2019, a study implemented U-Net models (a classical CNN
algorithm) for the automated multiclass segmentation of PAS-stained nephrectomy samples and
transplant biopsies¹⁰. Previous studies have indicated that deep learning algorithms can also be used to
learn the histological features from WSIs and forecast the prognosis for glioma. Mehmet and Daniel
proposed a deep learning-based classification pipeline for grading glioma patients on the TCGA cohort,
achieving 96% accuracy for GBM versus LGG classification and 71% accuracy for differentiating
LGG WHO grade II and grade III¹¹. Recently, several interesting studies have advanced the field by
demonstrating that deep learning algorithms can combine histological images with genetic information
to predict the clinical outcome of diseases, such as overall survival time and time to progression^{12, 13, 14},
¹⁵. For example, Cooper et al. reported a computational approach that integrated histology images and
genomic biomarkers information into a single unified network to predict the survival of glioma

83 patients¹⁶. In 2023, Hollon et al. developed DeepGlioma, which is trained on a multimodal dataset
 84 comprising stimulated Raman histology (SRH) and genomic data, to streamline the molecular
 85 diagnosis of diffuse gliomas¹⁷. Wang et al. presented a CNN-based integrated diagnosis model that was
 86 capable of automatically classifying adult-type diffuse gliomas according to the 2021 WHO standard
 87 from annotated-free WSIs¹⁸. In 2024, Hoang et al. developed Deep lEarning from histoPathoLOgy and
 88 methYlation (DEPLOY), a deep learning model that classifies CNS tumors to ten major categories
 89 from histopathology¹⁹. Mahootiha et al. developed a multimodal deep learning framework using
 90 preoperative MRI and several EHR-derived clinical variables to improve postoperative recurrence
 91 prediction and risk-stratification for pediatric low-grade gliomas²⁰. Jin et al. employed deep learning
 92 technologies to classify 5 major types of histopathological glioma images²¹, which greatly inspired us.
 93 They achieved 86.5% patch-level accuracy and 87.5% patient-level accuracy on the independent
 94 testing dataset. While the results were promising, they only focused on the histopathological glioma
 95 classification not the grading part which is more important for diagnosis and treatment. Our research
 96 significantly improved the automation of grading system for glioma and achieved better results at both
 97 patch and patient levels.

98
 99 In this study, we developed a framework based on deep learning for the quantitatively detecting
 100 necrosis and microvascular proliferation from WSIs in glioma patients, aiming to provide objective
 101 high accuracy prediction to aid diagnosis in clinical practice (**Fig. 1, 2a, 4c**). We utilized a dataset from
 102 Beijing Tiantan Hospital to train the patch-level models namely NEC detector and MVP detector. This
 103 dataset encompassed over 600 newly collected WSIs from gliomas ranging from WHO grade 2 to

grade 4 to ensure a comprehensive representation of glioma tumor grades. Initially, we trained deep learning models for necrosis and microvascular proliferation patch detection and quantification. Subsequently, we analyzed 1,699 WSIs from an independent cohort to train a patient-level model, named Probability Localization Network (PLNet). This model can be used for computer-aided patient diagnosis by integrating the patch-level predictors and spatial information features. Performance of this model was then evaluated using patient label that extracted from TCGA pathology reports. Finally, all patch-level and patient-level features were extracted from the models for further association analysis with clinical and genomic data (**Fig. 1**), highlighting the clinical significance for diagnosis and treatment.

113

114 **Materials and Methods**

115 **Data collection**

This study was based on data from the Chinese Glioma Genome Atlas (CGGA : 1,968 WSIs) cohort from Beijing Tiantan Hospital and publicly available cohort The Cancer Genome Atlas (TCGA: 1,705 WSIs). The image samples include low-grade glioma (LGG) and high-grade glioma (HGG) which are all digitized hematoxylin and eosin (H&E) stained from formalin-fixed, paraffin-embedded (FFPE) histology slides, referred to Whole-slide images (WSIs). All the image samples with section folds, large areas of pen markings, and poor staining were removed under the review of pathologists. The pathology reports were downloaded from TCGA. The clinical and molecular features of 1,122 TCGA patients were collected from cBioPortal, Mu et al.³ and Ceccarelli et al.²² (**SMa. 1**).

124

125 **WSI and patch preprocessing**

126 Whole-slide images (WSIs) are considerably too large, exceeding 70,000 pixels in each dimension, to
 127 be directly used as input for training deep learning models. These circumstances introduce specific
 128 complexities within the domain of image analysis. To balance the size of images appropriate for
 129 histological patterns detection and training speed, we tiled WSIs into patches with the size of $2,048 \times$
 130 $2,048$ pixels with a 50% overlap, resulting in 3,000 to 4,000 patches per WSI, and discarded patches
 131 containing small tissue area at $40 \times$ magnification. Meanwhile, for WSIs with a maximum
 132 magnification of $20\times$, the tiled patch size will be set to $1,024 \times 1,024$ pixels. These patch sizes are
 133 appropriate for most WSIs, enabling pathologists to identify crucial diagnostic features. Here we
 134 employed the Gaussian smoothing filter and Otsu's method to differentiate between the tissue area and
 135 background area in thumbnail images that compressed from the WSIs, resulting in the generation of
 136 binary segmentation masks for each WSI (**Fig. S1a**). A stain normalization algorithm was subsequently
 137 applied to the tiled patches to mitigate variations in the staining process. A standard patch was chosen
 138 to normalize all other patches to the same staining level (**Fig. S1b**).

139

140 **Patch-level datasets amplification and annotation**

141 The annotated dataset is a scarce resource in the field of medical image deep learning. An active
 142 learning pipeline was established (**Fig. 2a**) to obtain two datasets for training, validating, and testing
 143 the NEC and MVP patch-level models, respectively. In this pipeline, the initial dataset is manually
 144 labeled by expert pathologists and consists of 150 positive patches and 150 negative patches. Based on
 145 the established dataset, three initial deep learning models were trained and utilized to predict NEC or

146 MVP on a patch pool containing all available patches from the CGGA dataset. With the assistance of
147 pathologists, we were able to randomly verify the predicted results. The incorrectly predicted patches
148 were relabeled and reintegrated into the original dataset. Subsequently, a new model was trained based
149 on the updated dataset. After multiple rounds of active learning, we obtained the NEC dataset
150 comprising 3,552 positive and 8,369 negative patches, as well as the MVP dataset containing 3,139
151 positive and 11,488 negative patches. Finally, all patches were meticulously reviewed by three
152 pathologists to verify the accuracy of the labels.

153

154 **Patch-level and patient-level models**

155 In this study, we used different pre-trained CNN architectures for the necrosis and microvascular
156 proliferation to better accomplish the patch-level classification tasks, including DenseNet201²³,
157 EfficientNetB2²⁴, and ResNet50²⁵. The reason for selectin these three typical models is described in
158 detail in **SMe. 1**.

159

160 We also proposed a deep learning model, named Probability Localization Network (PLNet) to treat the
161 heatmap as an image to do patient-level classification. The PLNet is a CNN model which is inspired by
162 the skip connections and designed based on the heatmap data characteristics (50% overlap patch tiling)
163 with 3×3 kernel size for the first layer feature extraction (**Fig. S4d**).

164

165 **Patch-level models training, validation and evaluation**

166 We utilized 80% of patches in the dataset for training and allocated 20% for validation. The networks
 167 parameters were initialized with the best set achieved on the ImageNet dataset. The training details of
 168 the models were provided in the **SMe. 2**. After completing the training of the classification models, we
 169 assessed their performance on an independent dataset. 14 WSIs, which were not involved in data
 170 amplification and model training, were tiled into patches and labeled by 3 pathologists. Any patches
 171 with disagreements among the pathologists or that lacked consensus on a single patch were excluded.
 172 Ultimately, the necrosis independent test dataset comprised 777 positive and 16,941 negative patches,
 173 while the microvascular proliferation independent test dataset comprised 485 positive and 14,885
 174 negative patches. All models were evaluated by area under the receiver operating characteristic curve
 175 (AUC) scores and area under the precision-recall curve (AUPRC) scores. The Grad-CAM²⁶
 176 visualization tool was employed to further verify if the models accurately predicted necrosis and
 177 microvascular proliferation based on precise areas. Additionally, for further evaluation, the open-
 178 sourced tool QuPath was utilized to visualize the predicted outcomes of all the patches within the WSIs.

179

180 **Generating probability localization maps and extracting labels for patient-level NEC and MVP** 181 **models training and validation**

182 After predicting all the patches within WSIs, we assembled the patch-level predicted results and
 183 coordinates to construct the probability localization map for subsequent analysis (**Fig. S4a**). The
 184 detailed steps were as follows: First, the positive probabilities, predicted for NEC or MVP, of all
 185 patches within a single WSI were compiled. Second, the positive probabilities of all patches were

186 assembled based on their coordinates, which were tiled from the WSI. For instance, in the tiling of a
 187 WSI into patches, each patch was considered equivalent to a pixel, thereby converting the WSI into a
 188 matrix. Third, because of the assembled matrices with varying sizes for each WSI, which posed
 189 challenges in learning the probability distribution map. While a direct approach to address this issue
 190 would involve resizing all maps to a uniform size, the resizing process could alter the shape, size, and
 191 other characteristics of the NEC or MVP area. To preserve these characteristics, a straightforward
 192 method was implemented, matching the matrix to the center of a template with dimensions of $224 \times$
 193 224 pixels, while setting the blank area to 0 positive probability. The selection of 224×224 pixels was
 194 based on the statistics of all matrix sizes from CGGA and TCGA, ensuring compatibility with all
 195 statuses. Ultimately, the uniform-size probability distribution maps were produced.

196
 197 To obtain the patient-level labels for WSIs, we proposed a pipeline to employ optical character
 198 recognition (OCR) techniques to extract searchable text documents from pathology reports in PDF
 199 format downloaded from TCGA website. Additionally, we incorporated a fuzzy keyword search for
 200 necrosis and/or microvascular proliferation related diagnoses to acquire patient-level labels for the
 201 WSIs (**Fig. S4b**). For NEC, 95 positive and 156 negative patients, and for MVP, 100 positive and 145
 202 negative patients were included in this processing. The probability localization map integrated
 203 previously was used to train and valiate our proposed PLNet (Probability Localization Network) using
 204 5-fold cross-validation (**SMe. 3**).

205

206 NEC and MVP pattern clustering with differential expression and gene set enrichment analyses

207

208 Necrosis and microvascular proliferation are two of the most potent predictors of poor prognosis in

209 adult diffuse gliomas. Beyond their roles as mere morphological markers, these structures are

210 undeniably pivotal in driving the accelerated growth properties that define the transition from low

211 grade to high grade. To further explore the characteristics of NEC and MVP, as well as patients with

212 NEC and/or MVP, the mini-batch K-Means method was employed to cluster 11,923 NEC positive

213 patches and 19,297 MVP positive patches tiled from 1,699 TCGA WSIs. One patch was represented by

214 the combined 2,272-dimensional feature vectors extracted from the 3 CNN patch-level models

215 previously trained (DenseNet201: 896, EfficientNetB4: 352, ResNet50: 1,024). Principle component

216 analysis (PCA) was subsequently employed to reduce the feature vector dimension to 100 for

217 subsequent clustering. All these patches were grouped into 2 clusters to differentiate between

218 pseudopalisading necrosis (PAN) enriched patients and other necrosis enriched patients or between

219 glomeruloid microvascular proliferation (GMP) enriched patients and other MVP enriched patients

220 using the Mini-Batch K-Means method. Additionally, UMAP²⁷ was utilized for visualizing the

221 clustering outcomes. All patients were categorized into groups based on the enrichment of specific

222 patches within their WSIs (**SMe. 4**).

223

224 We conducted differential expression with volcano plot and gene set enrichment analysis (GSEA)^{28, 29}

225 with the normalized RNA-seq expression data acquired for the TCGA glioma dataset, which were

226 selected samples from two groups based on the PAN clustering result: PAN enriched patients (n=42)

227 and other necrosis enriched patients (n=70), and GMP clustering result: GMP enriched patients (n=56)

228 and other MVP enriched patients (n=71) (Details in **SMe. 5**). Only Hallmark gene sets^{30, 31} with a p-
229 value < 0.05 were selected as the functionally enriched biological states or processes.

230

231 **Results**

232 **Patch-level models classification performance**

233 Commencing with a modest dataset comprising 838 patches from 20 WSIs, we separately trained 3
234 distinct common CNN architectures, which were DenseNet201²³, EfficientNetB2²⁴, and ResNet50²⁵,
235 for the NEC and MVP patch-level classification tasks. Subsequently, after updating the training dataset
236 with models and pathologists, we retrained the 3 CNN models. AUC and AUPRC of all models were
237 evaluated on the independent test dataset, comprising 17,718 NEC and 15,370 MVP patches from 14
238 WSIs that were excluded from training dataset. The annotations of all patches in the training and test
239 datasets were independently reviewed by three neuropathologists (**Fig. 2a**). We evaluated the
240 improvement in model performance as the training dataset size increased during active learning (**Fig.**
241 **2b**). The voting results represented the median value of the 3 models from each dataset. The results
242 indicated that model performance improved with the inclusion of more WSIs in the training dataset,
243 ultimately reaching the limitation of label annotation accuracy. Finally, two well-annotated datasets
244 were established for training the NEC and MVP patch-level models. These datasets include 3,552
245 positive patches and 8,369 negative patches for NEC, and 3,139 positive patches and 11,488 negative
246 patches for MVP. AUC and AUPRC results indicated that models exhibited high performance (Table
247 in **Fig. 2c**), notably achieving an AUC of 0.9968 and AUPRC of 0.9788 for NEC models' voting
248 results, and an AUC of 0.9995 and AUPRC of 0.9860 for MVP models' voting results.

249

250 Quantitative analysis in WSIs and visual explanations in patches

251 Previously, conducting quantitative analysis of histological features was excessively time-consuming
 252 for pathologists during diagnosis. To tackle this issue, we shifted our focus to whole-slide images and
 253 investigated whether CNN models could be leveraged to quantitatively analyze histological features
 254 from WSIs, like the percentage of NEC and MVP positive patches in one WSI. For this purpose, we
 255 selected a WSI from the independent test set to demonstrate the quality of quantitative analysis of NEC
 256 and MVP compared to the annotation made by pathologists, as illustrated in **Fig. 3a** and **Fig. S2b, c**. To
 257 evaluate the predicted performance of the CNN models, we tasked pathologists with independently
 258 classifying all patches from the WSIs through visual inspection. The results revealed that the CNN
 259 model could accurately detect and quantify NEC and MVP with predicted performance compared to
 260 pathologists. The overall NEC and MVP positive patches identified by the CNN models highly
 261 overlaps with the positive patches labeled by pathologists (voting results: **Fig. S2a**, separate results:
 262 **Fig. S3**). In this WSI, the precisions for NEC and MVP patches were 1 and 0.9726, respectively, while
 263 the sensitivities were 0.9375 and 0.9342. Furthermore, the model accurately calculated the percentage
 264 of NEC and MVP positive patches, demonstrating results comparable to those of the pathologists as a
 265 quantification metric for patient diagnosis.

266

267 Given the satisfactory predicted performance of the classification models, it is essential to delve deeper
 268 into how the deep learning model arrived at its predictions and ascertain whether the model genuinely
 269 learned the meaningful histopathological features. To address these questions, we carried out gradient-

270 weighted class activation map analysis (Grad-CAM) on the CNN models. 4 NEC and 4 MVP positive
271 patches were randomly selected for these assessments. As depicted in **Fig. 3b, c**, the localization of
272 specific features that CNN models focused on during prediction aligned well with the localization of
273 NEC and MVP identified by the pathologists. These alignments indicate that the CNN models trained
274 on our datasets successfully learned meaningful histological features and utilized the information for
275 prediction.

276

277 **Patient-level models for computer-aided diagnosis**

278 We conducted analyses utilizing the generated probability localization maps and extracted labels for
279 patient level. The caption content of all the pathology reports generated by our pipeline is illustrated in
280 **Fig. S4c**. Each patch was assigned a predicted probability for NEC or MVP and positioned based on its
281 coordinates within the WSI. Subsequently, a heatmap of probability localization would be generated.
282 We proposed a deep learning model, named PLNet (Probability Localization Network, **Fig. S4d**), to
283 treat the heatmap as an image for patient-level classification. The datasets contained gliomas ranging
284 from WHO grade II to grade IV in TCGA cohort (Table in **Fig. 4a**). The patient-level predicted labels
285 were directly derived from the patch-level outcomes, where the labels were determined based on
286 whether the whole-slide image included predicted positive NEC or MVP patches. Utilizing pathology
287 reports as the ground truth, accuracy (NEC: 81.27%, MVP: 87.76%) and sensitivity (NEC: 55.79%,
288 MVP: 70%) were calculated for TCGA WSIs (**Fig. 4b**). The pipeline for the patient-level aggregation
289 models is illustrated in **Fig. 4c**. Accuracy and sensitivity were determined by setting different
290 probability thresholds for TCGA WSIs, using pathology reports as ground truth. Additionally, they

291 were evaluated through 5-fold cross-validation across all the data. The results demonstrated that our
292 PLNet model achieved the highest accuracy of 88.05% for NEC and 90.20% for MVP, along with a
293 sensitivity of 73.68% for NEC and 77% for MVP. When the sensitivity was set at 80%, the accuracy
294 for NEC reached 79.28% and for MVP reached 77.55% (**Fig. 4d**). Compared to **Fig. 4b**, both accuracy
295 and sensitivity were enhanced for NEC and MVP. The results from TCGA datasets also indicated that
296 our deep learning model, trained on CGGA datasets, possessed the capability to generalize effectively
297 in identifying and quantifying NEC and MVP across independent cohorts.

298

299 **Survival analysis for outcomes of patient-level models**

300 We further investigated whether our patient-level model could be utilized for prognosis prediction.
301 Following the diagnosis guideline in conjunction with the NEC and MVP predicted results and
302 pathology reports, these 878 patients with WSIs from 1,122 patients were categorized into four groups:
303 Oligodendroglioma (IDH-mutant and 1p/19q-codeleted, WHO Grade 2/3), Astrocytoma (IDH-mutant,
304 1p/19q-non-codeleted and CDKN2A/B wildtype, WHO Grade 2/3), Astrocytoma (IDH-mutant,
305 1p/19q-non-codeleted and CDKN2A/B HD, WHO Grade 4), Glioblastoma (IDH wild type, WHO
306 Grade 4). Using our patient-level model, we conducted several analyses to verify its practicality and
307 clinical potential. We performed a correlation analysis between the true labels from pathology reports
308 and the model predicted scores (**Fig. 5a, b**). For NEC, when the model predicted score was less than or
309 equal to 0.5 (n=173), the survival probabilities of patients were similar ($P=0.161$) to those with a true
310 label of 0 (n=153). When the model predicted score was greater than 0.5 (n=74), the survival
311 probabilities of patients were similar ($P=0.818$) to those with a true label of 1 (n=94). For MVP, when

the model predicted score was less than or equal to 0.5 (n=165), the survival probabilities of patients were similar ($P=0.097$) to those with a true label of 0 (n=142). When the model predicted score was greater than 0.5 (n=77), the survival probabilities of patients were similar ($P=0.772$) to those with a true label of 1 (n=100). Additionally, our model could effectively classify patients into two groups with distinct survival predictions ($P=1.1596E-13$ for NEC and $P=3.7353E-12$ for MVP between predicted scores ≤ 0.5 and > 0.5). Subsequently, we utilized various model predicted scores as thresholds to further divide the samples into two groups: one group with scores less than or equal to the threshold and the other group with scores greater than the threshold (**Fig. 5c**). We calculated the survival differences between these two groups and determined the corresponding p-values. For NEC samples with a true label of 0, we identified four thresholds (0.34, 0.35, 0.36, 0.37) where their $-\log_{10}$ p-values were greater than $-\log_{10} 0.05$. The highest $-\log_{10}$ p-value was observed at 0.34, where 22 samples with model-predicted scores less than or equal to 0.34 had better survival (**Fig. 5d**, $P=0.031$, $-\log_{10} P=1.508$) compared to 131 samples with scores greater than 0.34. We further extended our analysis to include all TCGA data with WSI and observed that our model can effectively differentiate between survival outcomes (**Fig. 5e, g**). For type A, IDHmut-non-codel, when our model predicted value is less than or equal to 0.66 or greater than 0.66 for NEC, or when our model predicted value is less than or equal to 0.33 or greater than 0.3 for MVP, the group with lower values exhibits superior survival compared to the group with higher values (**Fig. 5f, h**, $P=0.001$ for NEC and $P<0.001$ for MVP). Based on this, for A, IDHmut-non-codel CDKN2A/B wildtype records, we further subtyped the samples with more significant NEC and/or MVP characteristics. These records, derived from 186 CDKN2A/B wildtype samples across all A, IDHmut-non-codel samples in TCGA, have the potential to be a higher grade

(Fig. S5b). Similar results were also observed in type O, IDHmut-codel (Fig. S5a, c). We updated the Grade (WHO 2021) based on our model as shown in Supplementary Materials.

Clustering-based analyses with feature extraction

Uniform Manifold Approximation and Projection (UMAP)²⁷ demonstrated pseudopalisading necrosis (PAN) patches enriched in one cluster of all NEC patches, and glomeruloid microvascular proliferation (GMP) patches also enriched in one cluster of all MVP patches, which were manually checked by pathologists (Fig. 6a). Patients enriched with PAN or GMP were differentiated from all TCGA patients based on the quantity of PAN or GMP patches present in each WSI. Kaplan-Meier analysis was utilized to assess whether the PAN and GMP enriched patients could be effectively discriminate from other NEC and MVP enriched patients, respectively. The results revealed that the PAN (n=128) and GMP (n=178) enriched patients exhibited poorer survival outcomes compared to other NEC (n=120) and MVP (n=135) enriched patients, respectively, with p-values of 1.04E-04 for the NEC group and 1.30E-03 for the MVP group (Fig. 6b). We extended our analyses by conducting differential expression analysis with volcano plot (Fig. 6c), and gene set enrichment analysis (GSEA) (Fig. S6a-c) for both the PAN enriched patients compared with the other NEC enriched patients, as well as for the GMP enriched patients compared with the other MVP enriched patients. In the volcano plots (Fig. 6c) comparing the PAN enriched patients with other NEC enriched patients, several genes exhibit significant upregulation in the PAN enriched patient group, including FEZF1, SEPTIN14, SCNN1B, and IRX5. Conversely, notable downregulated genes such as PCDHGA11, KRT14, CCL19, and NMUR2 are observed in the PAN enriched patient group. Comparing the GMP enriched patients with

354 other MVP enriched patients, several genes exhibit significant upregulation in the GMP enriched
355 patient group, including CDK4, TSPAN31, CYP27B1, and PRDM13. Conversely, notable
356 downregulated genes such as SELL, KRT6A, SMCP, and CCL21 are observed in the GMP enriched
357 patient group. These upregulated genes suggest potential pathways or mechanisms that are more active
358 in the PAN or GMB enriched patient group compared to the other NEC or MVP enriched patient group.
359 The downregulation of these genes may indicate suppressed pathways or functions in the PAN or GMB
360 enriched patient group, potentially highlighting areas of biological significance or therapeutic interest.
361 The gene set enrichment analysis (GSEA) based on Hallmark gene sets revealed the significantly
362 altered pathways (**Fig. S6a, b**). For p-value less than 0.05, the 7 enriched pathways in the PAN
363 enriched patient group are primarily associated with cellular proliferation related, including E2F targets,
364 G2M checkpoint, MYC targets V1, and Mitotic spindle, metabolic related like Glycolysis, pathway
365 related like Hypoxia, and signaling related like MTORC1 signaling (**Fig. S6a**). The enrichment of these
366 pathways in the PAN enriched patient group suggests potential alterations in cell cycle regulation,
367 metabolism, cellular response to oxygen levels, and signaling pathways related to growth and
368 proliferation. This molecular profile may indicate specific biological processes, cellular adaptations, or
369 dysregulations that are characteristic of this group being studied. Meanwhile, the 16 enriched pathways
370 in the GMP enriched patient group are primarily associated with cellular proliferation related, including
371 E2F targets, G2M checkpoint, MYC targets V1, V2, and Mitotic spindle, metabolic related like
372 Glycolysis, and pathway related such as Hypoxia and Unfolded protein response, development related
373 such as Angiogenesis, Epithelial mesenchymal transition, Spermatogenesis, and Pancreas beta cell,
374 DNA damage related like DNA repair, immune related like Interferon alpha response, and signaling

375 related such as MTORC1 signaling, and TNFA signaling via NFKB (**Fig. S6b**). The presence of these
 376 pathways in the GMP enriched patient group suggests a complex interplay of cellular processes related
 377 to cell cycle regulation, metabolism, stress responses, angiogenesis, epithelial-mesenchymal transition,
 378 specialized cell functions, DNA integrity maintenance, immune responses, and signaling pathways.
 379 This intricate network of pathways may collectively contribute to the unique molecular profile and
 380 potentially the underlying biology of this group being studied. Previous studies investigated that PAN
 381 in glioblastoma was associated with hypoxia and GMP in glioblastoma was associated with
 382 angiogenesis^{32, 33, 34, 35, 36}, which were consistent with our findings. The GSEA results also indicated
 383 that the highest normalized enrichment scores were observed in the E2F targets pathway for both PAN
 384 and GMP enriched patients (**Fig. S6c**). Furthermore, we conducted more detailed analyses and obtained
 385 more sufficient information. Ultimately, we combined all the NEC and MVP predicted results together
 386 for generating diagnosis reports (**Fig. S7a**). Additionally, we utilized the patch-level predictions to
 387 generate the patient-level heatmap for online demonstration (**Fig. S7b**).

388

389 Discussion

390 In this study, we proposed an active learning pipeline to annotate patch-level datasets and train patch-
 391 level models for classifying and quantifying NEC and MVP. Subsequently, we proposed a patient-level
 392 CNN classification model named PLNet (Probability Localization Network), which was integrated
 393 with 3 patch-level CNN models to identify and quantify tumor NEC and MVP from histopathological
 394 glioma patients. The high predicted performance of the patch-level and patient-level models indicate
 395 that AI-based automatic histopathological diagnosis has the potential to be applied in clinical practice,

396 achieving performance comparable to that of pathologists. The models can improve the efficiency and
397 consistency of the diagnosis process by assisting pathologists in repetitive detection tasks within
398 whole-slide images (WSIs). Moreover, they have the potential to adjust previously uncertain patient
399 grades to updated grades, thereby facilitating more precise diagnoses. We further clustered NEC and
400 MVP positive patches into different subtypes using high-level features extracted by patch-level CNN
401 models to characterize patients enriched with PAN and other NEC, and with GMP and other MVP. The
402 results suggest that deeper information correlated with tumor progression stage and patients' genomic
403 features can be dug out by our model. This information can be ignored in clinical diagnosis and is
404 difficult to obtain through traditional pathology methods. It would be a starting point to develop an AI-
405 based diagnosis and treatment suggestion system that combines digital pathology and genomics, which
406 can really provide precision medicine specific to each patient.

407
408 We have evaluated our models on different datasets and a large number of patients to enhance their
409 generalization ability. However, the diversity of WSIs introduced by variations in tissue collection
410 methods, fixation technologies, levels of hematoxylin and eosin (H&E) staining, different digital slide
411 scanners, various data sources and collecting time makes it challenging to apply the models to every
412 dataset. Our approach demonstrates a promising pipeline from dataset construction, training, evaluating,
413 for further analysis on WSIs. Hopefully, it is feasible to adapt this pipeline to many other diseases.

414
415 We utilized formalin-fixed, paraffin-embedded (FFPE) WSIs in this study as they are commonly
416 preferred in pathology diagnosis practices. Another type of WSI is based on frozen tissue, which can be

417 rapidly produced during surgery to aid the surgeon in determination whether the tumor margins are
 418 clean. Most of the frozen tissue WSIs will exhibit tissue damaged, displaying a Swiss-cheese-like
 419 appearance, but can still provide pathological features for diagnosis. Several previous studies have
 420 explored using both frozen tissue and FFPE WSIs as inputs for deep learning models in tumor
 421 diagnosis^{12, 37, 38}.

422
 423 Necrosis and microvascular proliferation (MVP) are two primary pathological features in glioma
 424 diagnosis. Additional features such as mitosis activity^{39, 40, 41}, nuclei atypia, and nuclei density^{42, 43, 44}
 425 are also crucial for a comprehensive diagnosis. We are utilizing the approach presented in this paper for
 426 the automatic detection and quantification of all these features. With all features automatically detected
 427 by the models, we can integrate them to generate a comprehensive diagnostic report that includes all
 428 quantified features, genomic indications, treatment suggestions, and potential drug choices. This holds
 429 promise for the future of precision medicine era.

430

431 **Acknowledgments**

432 The work described in this paper was supported by a grant from the Research Grants Council of the
 433 Hong Kong Special Administrative Region, China (Project No. R6003-22). This work was also
 434 partially supported by ITC grant (MHP/004/19, ITCPD/17-9) and a seed fund of the Big Data for Bio-
 435 Intelligence Laboratory (Z0428) from The Hong Kong University of Science and Technology. J. Wang
 436 is also supported by Padma Harilela Professorship.

437

438 **Author Contributions**

439 J. Wang and T. Jiang conceived the project and designed the experiments. X. Liu and T. Jiang obtained
440 the patient samples and scanned the WSIs. W. Zou, X. Liu and F. Qiu labeled the necrosis and
441 microvascular proliferation dataset. Y. Guo and H. Huang performed the model training and testing and
442 the dataset correction. Y. Liu and Y. Guo perform survival analysis. Y. Guo performed the differential
443 expression, and gene set enrichment analyses. Y. Guo, H. Huang, Y. Liu and J.W. wrote the manuscript
444 with input from all authors.

445

446 **Declaration of interests**

447 The authors declare no competing interests.

448

449 **Data sharing**

450 CGGA data, including WSIs (not yet), tumor types and demographics, are available at
451 <http://www.cgga.org.cn/>. Registration and acceptance of terms and data use agreements are required for
452 access. The human LGG and GBM diagnostic slide images were derived from the TCGA Research
453 Network at <https://portal.gdc.cancer.gov/>. The code is available for academic research purposes via
454 GitHub at <https://github.com/WangLabHKUST/NMADG>.

455

456 **References**

- 457 1. Jiang, T., Nam, D.H., Ram, Z., Poon, W.S., Wang, J., Boldbaatar, D., Mao, Y., Ma, W., Mao, Q.,
458 You, Y. and Jiang, C. Clinical practice guidelines for the management of adult diffuse gliomas.

459 *Cancer letters*, 499, 60-72 (2021).

460 2. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C.,

461 Ng, H.K., Pfister, S.M., Reifenberger, G. and Soffiatti, R., The 2021 WHO classification of tumors

462 of the central nervous system: a summary. *Neuro-oncology*, 23(8), 1231-1251 (2021).

463 3. Mu, Q., Chai, R., Pang, B., Yang, Y., Liu, H., Zhao, Z., Bao, Z., Song, D., Zhu, Z., Yan, M. and

464 Jiang, B. Identifying predictors of glioma evolution from longitudinal sequencing. *Science*

465 *translational medicine*, 15(716), p.eadh4181 (2023).

466 4. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M. and Yener, B.

467 Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2, 147-171

468 (2009).

469 5. Mittler, M.A., Walters, B.C. and Stopa, E.G. Observer reliability in histological grading of

470 astrocytoma stereotactic biopsies. *Journal of neurosurgery*, 85(6), 1091-1094 (1996).

471 6. Meyer, J.S., Alvarez, C., Milikowski, C., Olson, N., Russo, I., Russo, J., Glass, A., Zehnbauer,

472 B.A., Lister, K. and Parwaresch, R. Breast carcinoma malignancy grading by Bloom–Richardson

473 system vs proliferation index: reproducibility of grade and advantages of proliferation index.

474 *Modern pathology*, 18(8), 1067-1078 (2005).

475 7. Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N., Nelson, H.D.,

476 Pepe, M.S., Allison, K.H., Schnitt, S.J. and O’Malley, F.P. Diagnostic concordance among

477 pathologists interpreting breast biopsy specimens. *Jama*, 313(11), 1122-1132 (2015).

478 8. Van der Kwast, T.H., Roobol, M.J., Wildhagen, M.F., Martikainen, P.M., Määtänen, L., Pihl, C.G.,

479 Santonja, C., Bubendorf, L., Neetens, I., Di Lollo, S. and Hoedemaeker, R.F. Consistency of

480 prostate cancer grading results in screened populations across Europe. *BJU international*, 92, 88-
481 91 (2003).

482 9. Jing, J., Wan, T., Cao, J. and Qin, Z. An improved hybrid active contour model for nuclear
483 segmentation on breast cancer histopathology. In *2016 IEEE 13th International Symposium on*
484 *Biomedical Imaging (ISBI)*, 1155-1158 (IEEE, 2016).

485 10. Hermesen, M., de Bel, T., Den Boer, M., Steenberg, E.J., Kers, J., Florquin, S., Roelofs, J.J.,
486 Stegall, M.D., Alexander, M.P., Smith, B.H. and Smeets, B. Deep learning-based histopathologic
487 assessment of kidney tissue. *Journal of the American Society of Nephrology*, 30(10), 1968-1979
488 (2019).

489 11. Ertosun, M.G. and Rubin, D.L. Automated grading of gliomas using deep learning in digital
490 pathology images: a modular approach with ensemble of convolutional neural networks. In *AMIA*
491 *annual symposium proceedings*, Vol. 2015, 1899 (2015).

492 12. Wulczyn, E., Steiner, D.F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., Mermel, C.H.,
493 Chen, P.H.C., Liu, Y. and Stumpe, M.C. Deep learning-based survival prediction for multiple
494 cancer types using histopathology images. *PloS one*, 15(6), e0233678 (2020).

495 13. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A.,
496 Valous, N.A., Ferber, D. and Jansen, L. Predicting survival from colorectal cancer histology slides
497 using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1), e1002730 (2019).

498 14. Yu, K.H., Zhang, C., Berry, G.J., Altman, R.B., Ré, C., Rubin, D.L. and Snyder, M. Predicting
499 non-small cell lung cancer prognosis by fully automated microscopic pathology image features.
500 *Nature communications*, 7(1), 12474 (2016).

501 15. Zhu, X., Yao, J., Zhu, F. and Huang, J. Wsisa: Making survival prediction from whole slide
502 histopathological images. In *Proceedings of the 30th IEEE conference on computer vision and*
503 *pattern recognition*, Vol. 2017, 7234-7242 (2017).

504 16. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega,
505 J.E., Brat, D.J. and Cooper, L.A. Predicting cancer outcomes from histology and genomics using
506 convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13), E2970-E2979
507 (2018).

508 17. Hollon, T., Jiang, C., Chowdury, A., Nasir-Moin, M., Kondepudi, A., Aabedi, A., Adapa, A., Al-
509 Holou, W., Heth, J., Sagher, O. and Lowenstein, P. Artificial-intelligence-based molecular
510 classification of diffuse gliomas using rapid, label-free optical imaging. *Nature Medicine*, 29(4),
511 828-832 (2023).

512 18. Wang, W., Zhao, Y., Teng, L., Yan, J., Guo, Y., Qiu, Y., Ji, Y., Yu, B., Pei, D., Duan, W. and
513 Wang, M., Neuropathologist-level integrated classification of adult-type diffuse gliomas using
514 deep learning from whole-slide pathological images. *Nature Communications*, 14(1), p.6359
515 (2023).

516 19. Hoang, D.T., Shulman, E.D., Turakulov, R., Abdullaev, Z., Singh, O., Campagnolo, E.M.,
517 Lalchungnunga, H., Stone, E.A., Nasrallah, M.P., Ruppin, E. and Aldape, K., Prediction of DNA
518 methylation-based tumor types from histopathology in central nervous system tumors with deep
519 learning. *Nature Medicine*, pp.1-10 (2024).

520 20. Mahootiha, M., Tak, D., Ye, Z., Zapaishchykova, A., Likitlersuang, J., Climent Pardo, J.C., Boyd,
521 A., Vajapeyam, S., Chopra, R., Prabhu, S.P. and Liu, K.X. Multimodal deep learning improves

522 recurrence risk prediction in pediatric low-grade gliomas. *Neuro-oncology*, p.noae173 (2024).

523 21. Jin, L., Shi, F., Chun, Q., Chen, H., Ma, Y., Wu, S., Hameed, N.F., Mei, C., Lu, J., Zhang, J. and

524 Aibaidula, A. Artificial intelligence neuropathologist for glioma classification using deep learning

525 on hematoxylin and eosin stained slide images and molecular markers. *Neuro-oncology*, 23(1),

526 pp.44-52 (2021).

527 22. Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova,

528 O., Newton, Y., Radenbaugh, A., Pagnotta, S.M. and Anjum, S. Molecular profiling reveals

529 biologically discrete subsets and pathways of progression in diffuse glioma. *Cell*, 164(3), 550-563

530 (2016).

531 23. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. Densely connected convolutional

532 networks. In *Proceedings of the 30th IEEE conference on computer vision and pattern*

533 *recognition*, Vol. 2017, 4700-4708 (2017).

534 24. Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In

535 *International conference on machine learning*, 6105-6114 (PMLR, 2019).

536 25. He, K., Zhang, X., Ren, S. and Sun, J. Deep residual learning for image recognition. In

537 *Proceedings of the IEEE conference on computer vision and pattern recognition*, Vol. 2016, 770-

538 778 (2016).

539 26. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., Grad-CAM:

540 visual explanations from deep networks via gradient-based localization. *International journal of*

541 *computer vision*, 128, pp.336-359 (2020).

542 27. McInnes, L., Healy, J. and Melville, J. Umap: Uniform manifold approximation and projection for

543 dimension reduction. *arXiv preprint*, arXiv:1802.03426 (2018).

544 28. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich,
545 A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. Gene set enrichment analysis: a
546 knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the*
547 *National Academy of Sciences*, 102(43), 15545-15550 (2005).

548 29. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S.,
549 Jenkins, S.L., Jagodnik, K.M., Lachmann, A. and McDermott, M.G. Enrichr: a comprehensive
550 gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), W90-W97
551 (2016).

552 30. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P.
553 Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12), 1739-1740 (2011).

554 31. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. The
555 molecular signatures database hallmark gene set collection. *Cell systems*, 1(6), 417-425 (2015).

556 32. Brat, D.J., Castellano-Sanchez, A.A., Hunter, S.B., Pecot, M., Cohen, C., Hammond, E.H., Devi,
557 S.N., Kaur, B. and Van Meir, E.G. Pseudopalisades in glioblastoma are hypoxic, express
558 extracellular matrix proteases, and are formed by an actively migrating cell population. *Cancer*
559 *research*, 64(3), 920-927 (2004).

560 33. Rong, Y., Durden, D.L., Van Meir, E.G. and Brat, D.J. 'Pseudopalisading' necrosis in
561 glioblastoma: a familiar morphologic feature that links vascular pathology, hypoxia, and
562 angiogenesis. *Journal of Neuropathology & Experimental Neurology*, 65(6), 529-539 (2006).

563 34. Straume, O., Chappuis, P.O., Salvesen, H.B., Halvorsen, O.J., Haukaas, S.A., Goffin, J.R., Bégin,

564 L.R., Foulkes, W.D. and Akslen, L.A. Prognostic importance of glomeruloid microvascular
565 proliferation indicates an aggressive angiogenic phenotype in human cancers. *Cancer Research*,
566 62(23), 6808-6811 (2002).

567 35. Brat, D.J. and Van Meir, E.G. Glomeruloid microvascular proliferation orchestrated by
568 VPF/VEGF: a new world of angiogenesis research. *The American journal of pathology*, 158(3),
569 789-796 (2001).

570 36. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A.,
571 Lankervich, M., Bertagnolli, D., Bickley, K. and Boe, A.F. An anatomic transcriptional atlas of
572 human glioblastoma. *Science*, 360(6389), 660-663 (2018).

573 37. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyő, D., Moreira, A.L.,
574 Razavian, N. and Tsirogas, A. Classification and mutation prediction from non-small cell lung
575 cancer histopathology images using deep learning. *Nature medicine*, 24(10), 1559-1567 (2018).

576 38. Jang, H.J., Song, I.H. and Lee, S.H. Generalizability of deep learning system for the pathologic
577 diagnosis of various cancers. *Applied Sciences*, 11(2), 808 (2021).

578 39. Nateghi, R., Danyali, H. and Helfroush, M.S. A deep learning approach for mitosis detection:
579 application in tumor proliferation prediction from whole slide images. *Artificial intelligence in*
580 *medicine*, 114, 102048 (2021).

581 40. Mahmood, T., Arsalan, M., Owais, M., Lee, M.B. and Park, K.R. Artificial intelligence-based
582 mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs.
583 *Journal of clinical medicine*, 9(3), 749 (2020).

584 41. Sohail, A., Khan, A., Wahab, N., Zameer, A. and Khan, S. A multi-phase deep CNN based mitosis

585 detection framework for breast cancer histopathological images. *Scientific Reports*, 11(1), 6215
586 (2021).

587 42. Wang, P., Hu, X., Li, Y., Liu, Q. and Zhu, X. Automatic cell nuclei segmentation and
588 classification of breast cancer histopathology images. *Signal Processing*, 122, 1-13 (2016).

589 43. Abdolhoseini, M., Kluge, M.G., Walker, F.R. and Johnson, S.J. Segmentation of heavily clustered
590 nuclei from histopathological images. *Scientific reports*, 9(1), 4551 (2019).

591 44. Prasath, V.S., Fukuma, K., Aronow, B.J. and Kawanaka, H. November. Cell nuclei segmentation
592 in glioma histopathology images with color decomposition based active contours. *In 2015 IEEE*
593 *International Conference on Bioinformatics and Biomedicine (BIBM)*, 1734-1736 (IEEE 2015).
594

595 **Figure legends**

596

597 **Figure 1 | Computational and analytical pipeline of the study.** (a) We initially applied WSI
598 preprocessing to obtain tiled and stained patches and then developed deep learning models to detect
599 and quantify histopathological features, including necrosis (NEC) and microvascular proliferation
600 (MVP), at the patch level with labeled data. (b) We then integrated the patch-level results to aid
601 patient-level diagnosis and grading based on an AI-based aggregation model. (c) Finally, we conducted
602 several analyses, including survival analysis, differential expression analysis, and gene set enrichment
603 analysis, to investigate the clinical implications of our computational pathology models.

604

605 **Figure 2 | Patch level classification of necrosis (NEC) and microvascular proliferation (MVP).** (a)
606 Diagram summarizing the model training pipeline. An active learning strategy was applied to label
607 patches based on whether NECs and/or MVPs were recognized in a particular patch. After active
608 learning, the final training set included 3,552 NEC-positive, 8,369 NEC-negative patches, 3,139 MVP-
609 positive, and 11,488 MVP-negative patches. These patches were extracted from 621 glioma WSIs. Our
610 deep learning models were trained in the above training sets and then tested in an independent dataset,
611 which was extracted from another 14 WSIs (777 NEC-positive, 16,941 NEC-negative, 485 MVP-
612 positive, and 14,885 MVP-negative patches) and manually labelled by three pathologists. (b) Using the
613 training set, we first trained three deep learning models using the widely used network frameworks, *i.e.*,
614 DenseNet, EfficientNet and ResNet, and then concatenated results from these three models to generate
615 the fourth results based on voting. Performance of the four results on test set were compared. Using

necrosis-recognition model as an example, we demonstrated that the performance of patch-level models improved as the size of training dataset increased from patches of 20 WSIs to patches of 621 WSIs, until all model results converged. AUPRC: Area Under the Precision-Recall Curve. (c) Performance of various model architectures under the final training set (patches derived from 621 WSIs). The best performance was highlighted in bold. For both NEC and MVP, the best AUC and AUPRC were achieved with the voting results of three models (NEC: AUC of 0.9968 and AUPRC of 0.9788, MVP: AUC of 0.9995 and AUPRC of 0.9860). AUC: Area Under the Receiver Operating Characteristic Curve.

Figure 3 | Performance of the patch-level deep learning models. (a) A WSI in the independent test set was reviewed to demonstrated the model performance (left panel). In the middle panel, the dark blue (or red) areas highlighted the NEC (or MVP)-positive patches classified by pathologists through visual inspection. The right panel included the results of the voting models with light blue (or red) areas highlighting the NEC (or MVP)-positive patches by the deep learning models. (b-c) Grad-CAM analysis of key pathology features. NEC (b) and MVP (c) were manually highlighted by pathologists (blue area in the top panels), while Grad-CAM analyses of the deep learning model provided the class-discriminative visualization of the corresponding features shown in the heatmap in the bottom panels.

Figure 4 | Patient-level models of NEC and MVP. (a) NEC and MVP annotations extracted from TCGA pathology reports. (b) Accuracy and sensitivity of patch-level models to predict TCGA pathology reports. The patient-level prediction label derived from the patch-level results, whether the

WSI has positive NEC or MVP patches as the patient-level label. Using pathology reports as ground truth, accuracy and sensitivity were respectively calculated for TCGA WSIs (NEC: accuracy of 81.27% and sensitivity of 55.79%, MVP: accuracy of 87.76% and sensitivity of 70%). (c) Pipeline for the patient-level aggregation models. First, we utilized our trained patch-level models to predict the tiled patches from TCGA WSIs. Each patch was assigned a predicted probability for NEC or MVP and positioned according to its coordinates within the WSI. Subsequently, a probability localization map would be generated. We then proposed a deep learning model called PLNet (Probability Localization Network) to treat the map as an image for patient-level classification. (d) NEC and MVP AUC results of 5-fold cross-validation for TCGA cohort of patients based on our proposed PLNet. The results demonstrated that our PLNet model achieved the highest accuracy of 88.05% for NEC and 90.20% for MVP, along with a sensitivity of 73.68% for NEC and 77% for MVP. When the sensitivity was set at 80%, the accuracy for NEC reached 79.28% and for MVP reached 77.55%. Compared to (b), both accuracy and sensitivity were enhanced for NEC and MVP.

Figure 5 | Survival analysis using the patient-level model. (a)-(d) are based on data with reports, while (e)-(h) are based on all TCGA data. (a) Survival difference analysis for NEC between the true label and the model-predicted score. (b) Survival difference analysis for MVP between the true label and the model-predicted score. (c) By using different model-predicted scores as thresholds, we divide the samples into two groups. The survival difference between these two groups is calculated, and the corresponding p-value is determined. (d) Survival curves based on the most significant p-value from (c). This demonstrates that our model-predicted scores can further subtype the true labels. (e) In

different molecular subtypes, the p-values of survival curves using different probability prediction thresholds of the NEC model. O: Oligodendroglioma, A: Astrocytoma, GBM: Glioblastoma. (f) Survival curves based on the most significant p-value for A, IDHmut-non-codel from (e). (g) In different molecular subtypes, the p-values of survival curves using different probability prediction thresholds of the MVP model. (h) Survival curves based on the most significant p-value for A, IDHmut-non-codel from (g).

Figure 6 | Clustering of NEC and MVP patches and their subtype analysis. (a) UMAP projection of K-Means clustering of all 11,923 NEC and 19,297 MVP concatenated 2,272-dimensional feature vectors (DenseNet201: 896, EfficientNetB4: 352, ResNet50: 1,024) of positive patches of TCGA WSI dataset. 2 typical PAN (pseudopalisading necrosis) patches, and 2 other NEC patches were evaluated by pathologists. PAN patches were enriched in dark-blue cluster, and other NEC patches were enriched in light-blue cluster. 2 typical GMP (glomeruloid microvascular proliferation), and 2 other MVP patches were evaluated by pathologists. Other MVP patches were enriched in dark-red cluster, and GMP patches were enriched in light-red cluster. (b) Survival analysis for PAN (n=128) vs. Other NEC (n=120) and GMP (n=178) vs. Other MVP (n=135) enriched patients. Both NEC and MVP groups demonstrated significant survival differences. (c) Volcano plots depicting differentially expressed genes in PAN and GMP enriched patients.

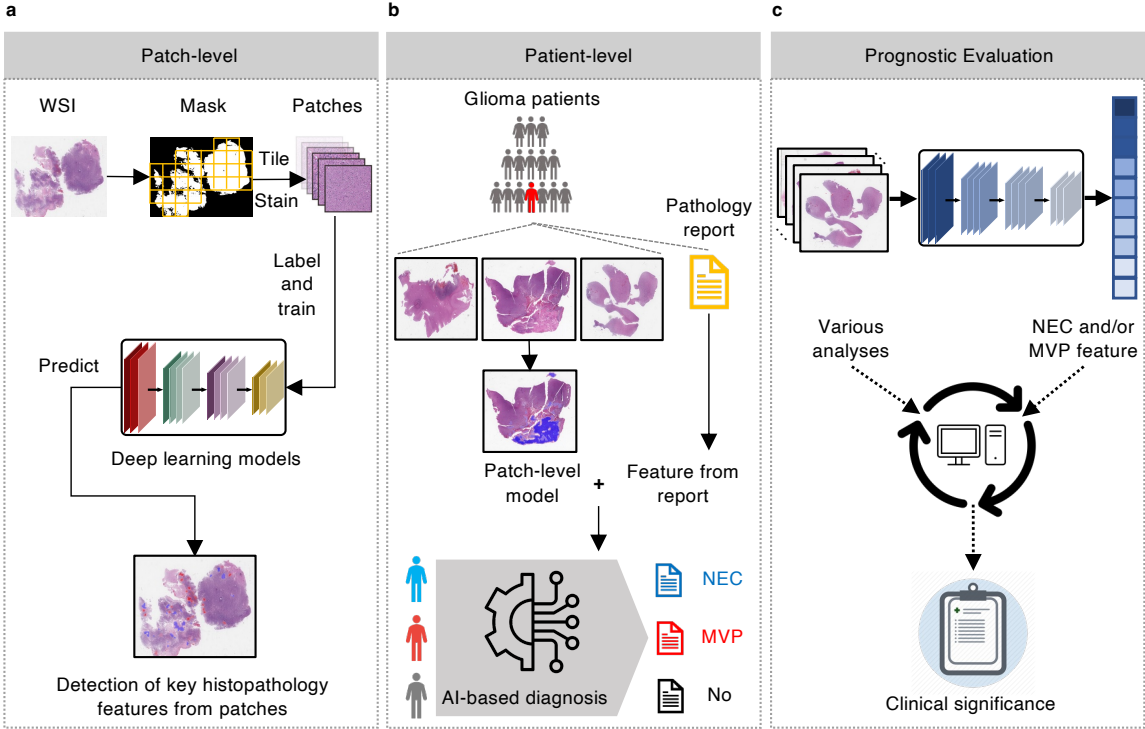


Figure 1

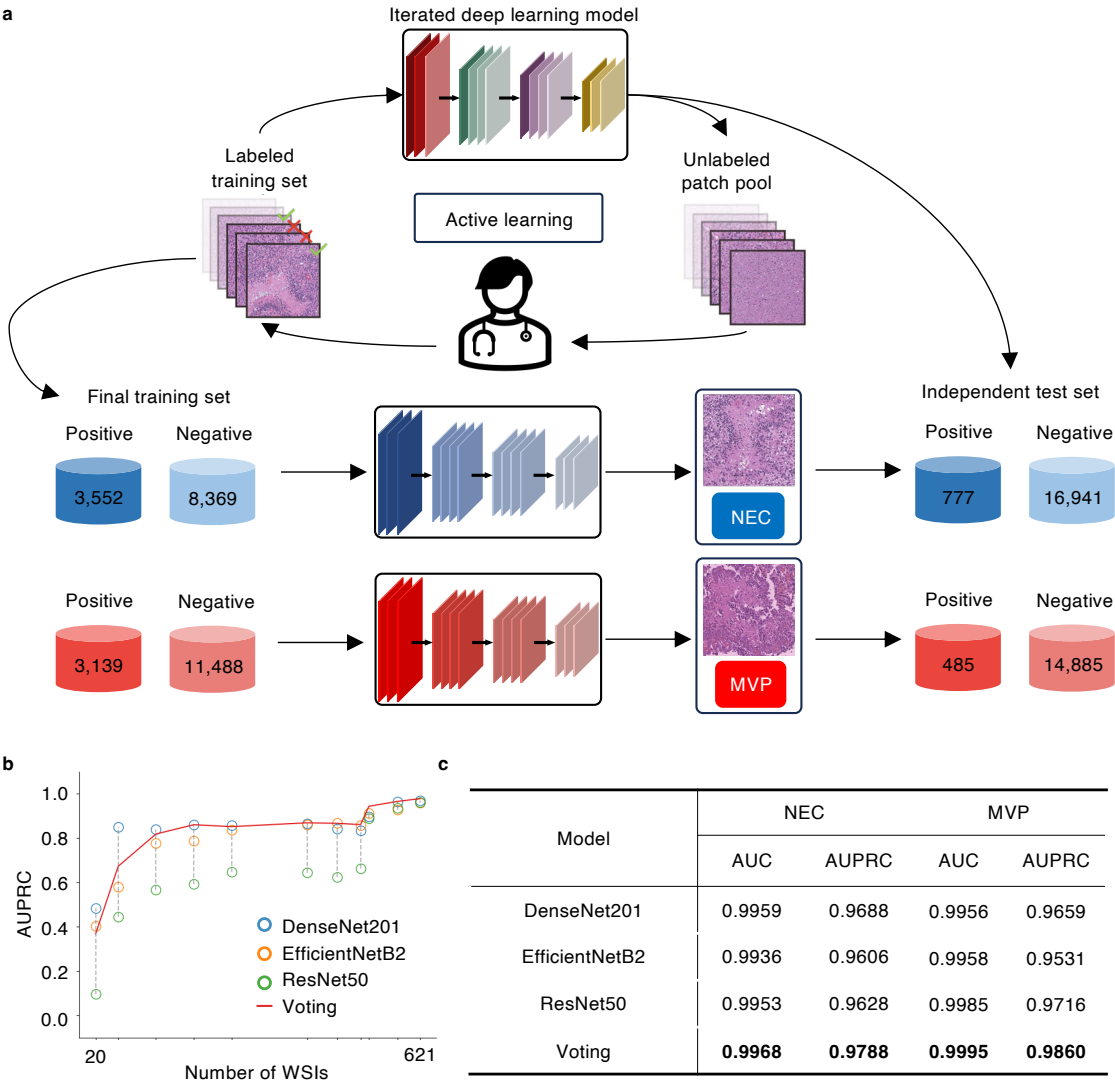
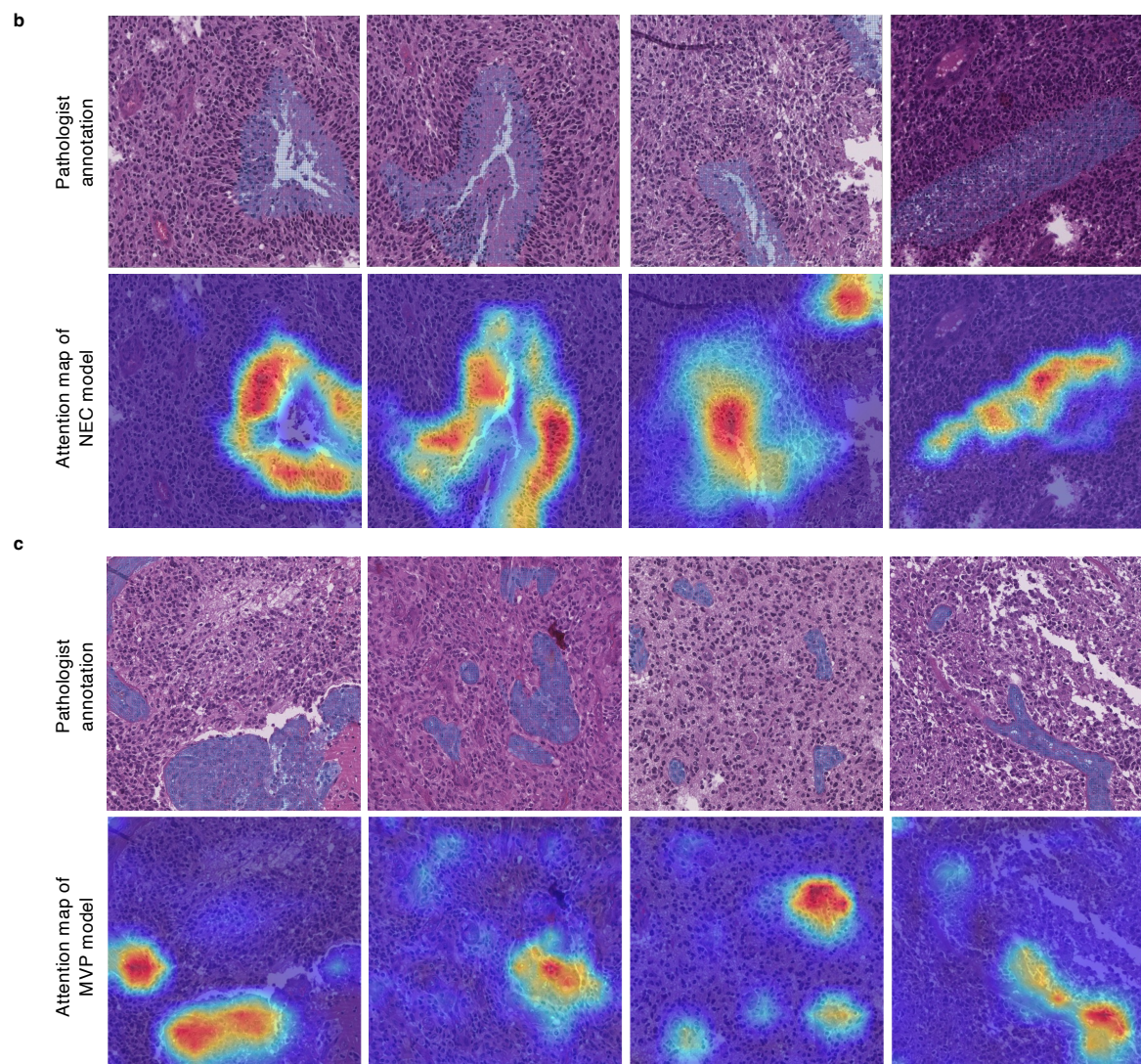


Figure 2



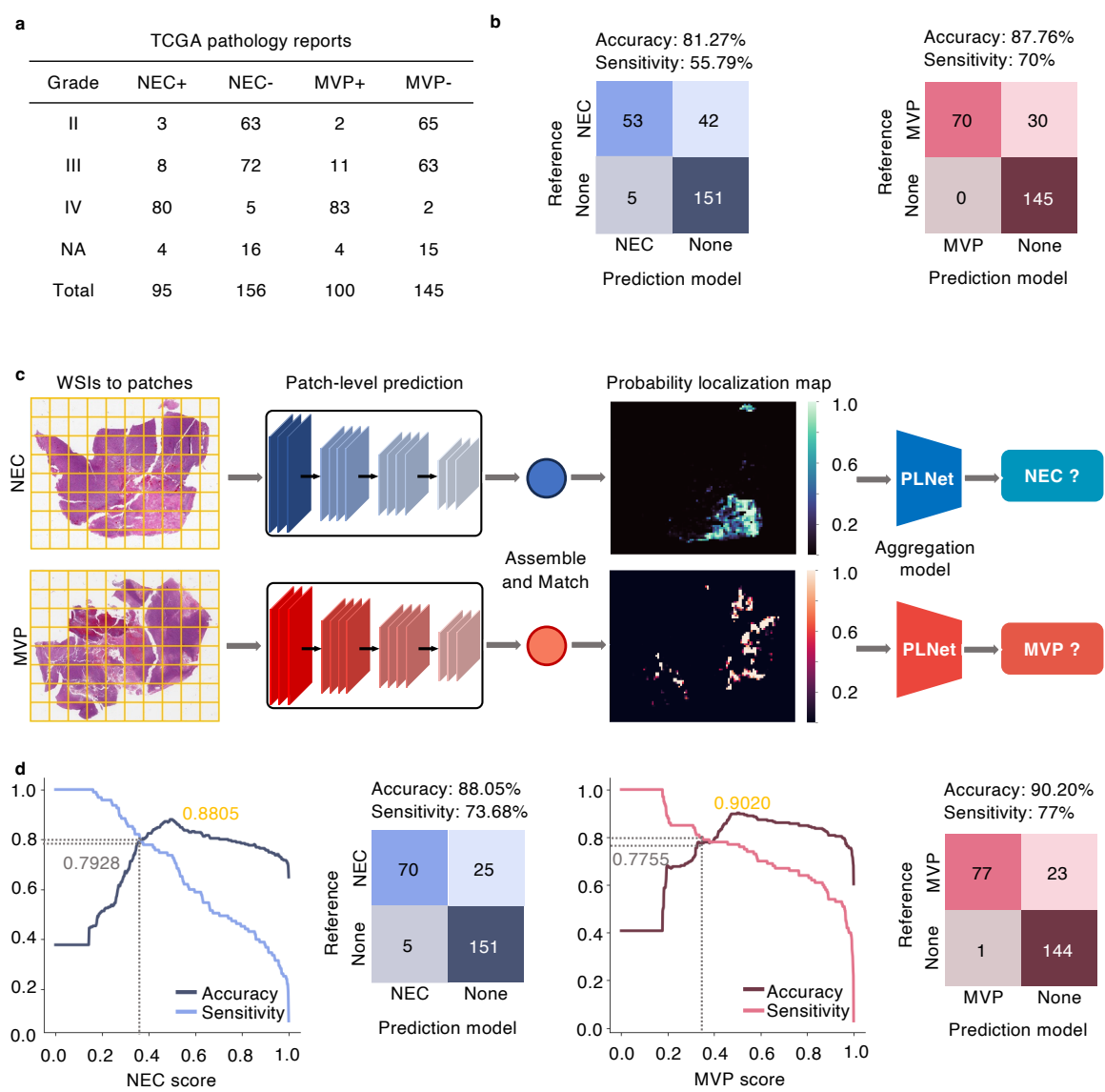


Figure 4

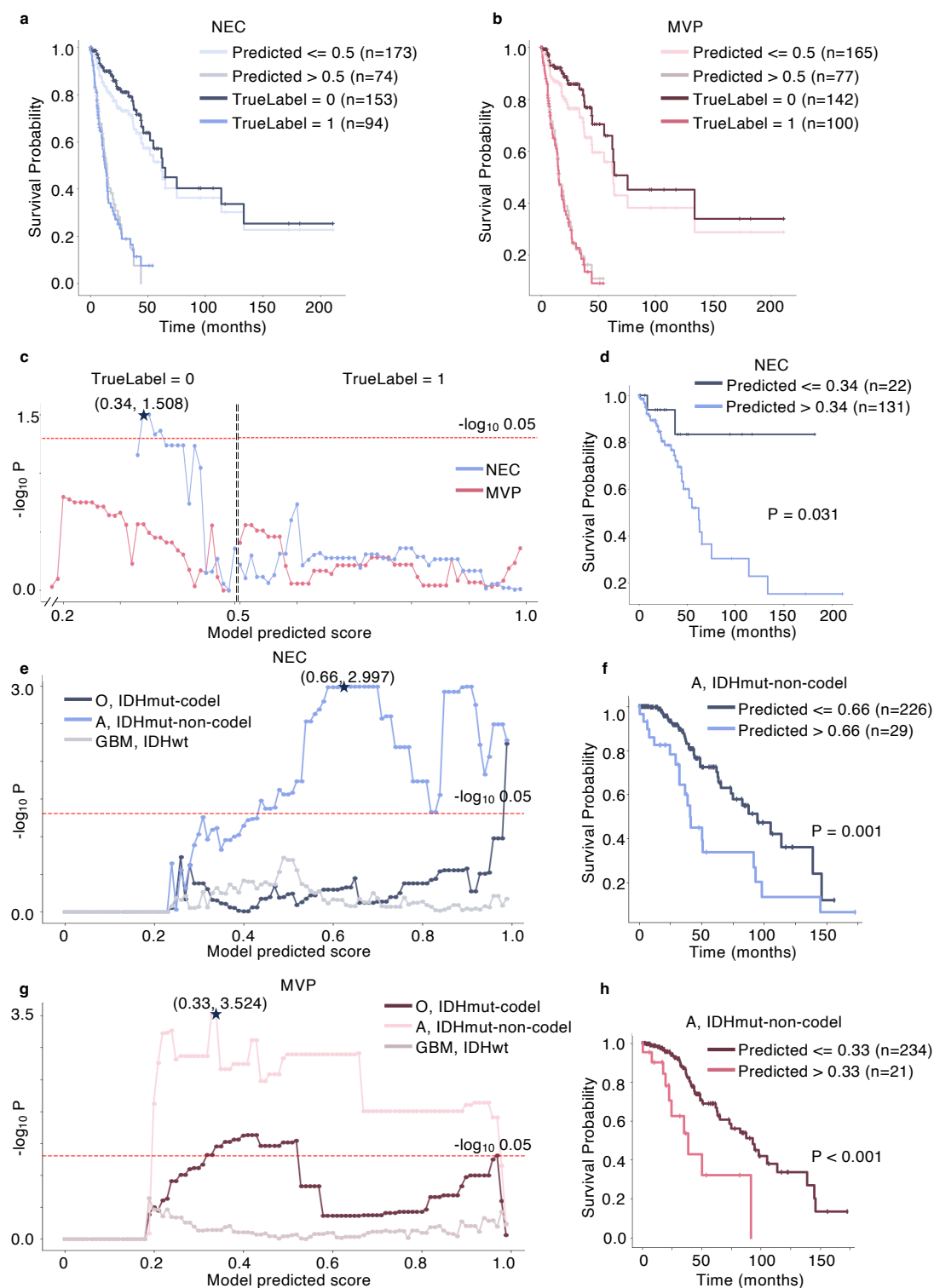


Figure 5

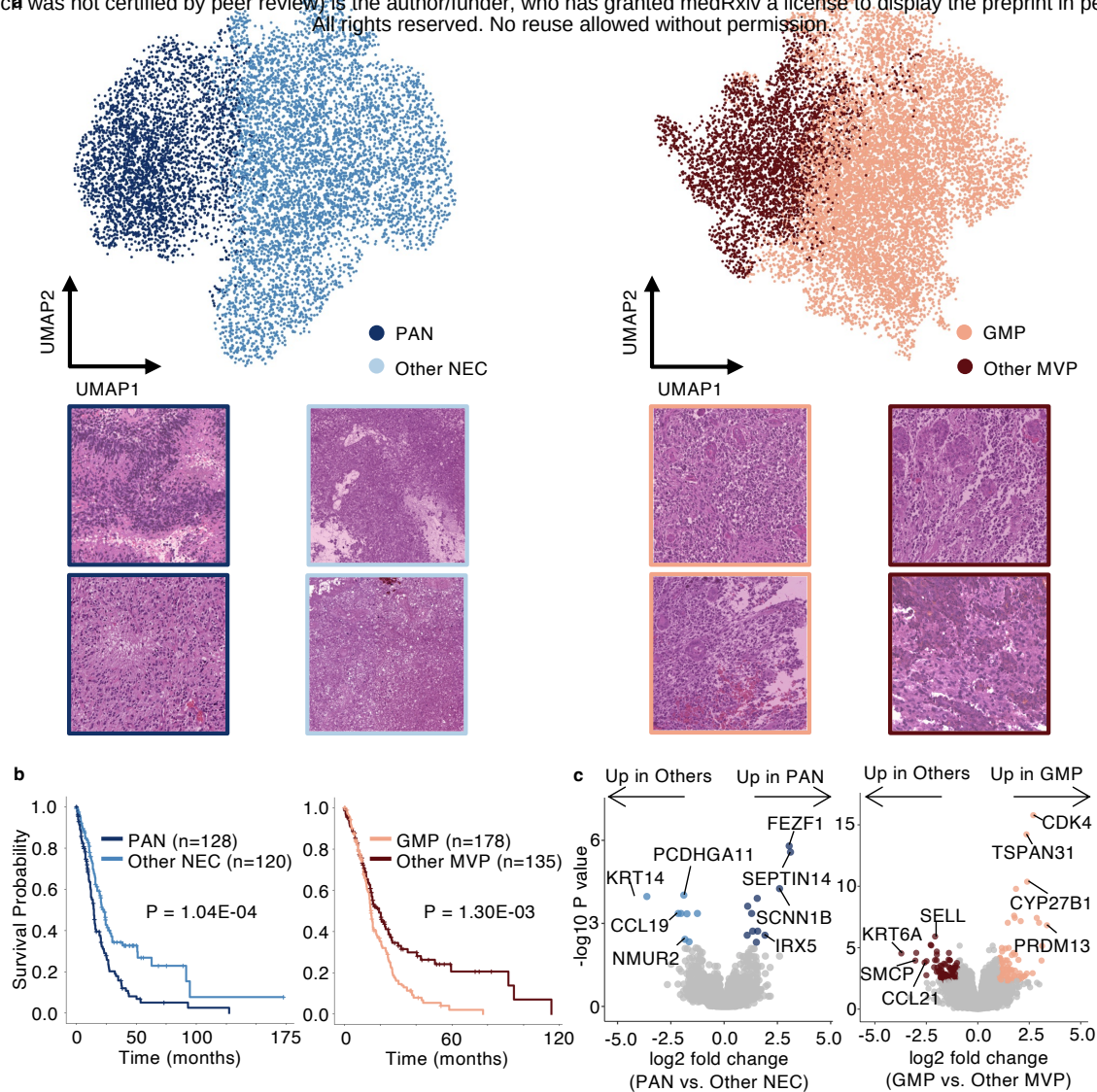


Figure 6