



ARTICLE

AI Applied to Cancer

A deep convolutional neural network for segmentation of whole-slide pathology images identifies novel tumour cell-perivascular niche interactions that are associated with poor survival in glioblastoma

Amin Zadeh Shirazi ^{1,2}, Mark D. McDonnell ², Eric Fornaciari³, Narjes Sadat Bagherian⁴, Kaitlin G. Scheer ¹, Michael S. Samuel ^{1,5}, Mahdi Yaghoobi⁶, Rebecca J. Ormsby ⁷, Santosh Poonnoose ^{7,8}, Damon J. Tumes ¹ and Guillermo A. Gomez ¹

BACKGROUND: Glioblastoma is the most aggressive type of brain cancer with high-levels of intra- and inter-tumour heterogeneity that contribute to its rapid growth and invasion within the brain. However, a spatial characterisation of gene signatures and the cell types expressing these in different tumour locations is still lacking.

METHODS: We have used a deep convolutional neural network (DCNN) as a semantic segmentation model to segment seven different tumour regions including leading edge (LE), infiltrating tumour (IT), cellular tumour (CT), cellular tumour microvascular proliferation (CTmvp), cellular tumour pseudopalisading region around necrosis (CTpan), cellular tumour perinecrotic zones (CTpnz) and cellular tumour necrosis (CTne) in digitised glioblastoma histopathological slides from The Cancer Genome Atlas (TCGA). Correlation analysis between segmentation results from tumour images together with matched RNA expression data was performed to identify genetic signatures that are specific to different tumour regions.

RESULTS: We found that spatially resolved gene signatures were strongly correlated with survival in patients with defined genetic mutations. Further in silico cell ontology analysis along with single-cell RNA sequencing data from resected glioblastoma tissue samples showed that these tumour regions had different gene signatures, whose expression was driven by different cell types in the regional tumour microenvironment. Our results further pointed to a key role for interactions between microglia/pericytes/monocytes and tumour cells that occur in the IT and CTmvp regions, which may contribute to poor patient survival.

CONCLUSIONS: This work identified key histopathological features that correlate with patient survival and detected spatially associated genetic signatures that contribute to tumour-stroma interactions and which should be investigated as new targets in glioblastoma. The source codes and datasets used are available in GitHub: https://github.com/amin20/GBM_WSSM.

British Journal of Cancer (2021) 125:337–350; <https://doi.org/10.1038/s41416-021-01394-x>

INTRODUCTION

Glioblastoma is the most frequently diagnosed and aggressive type of brain cancer, accounting for 80% of primary malignant brain tumours of the central nervous system (CNS), and 60% of all malignant brain tumours in adults.¹ There are ~100,000 new cases of glioblastoma diagnosed each year worldwide,^{1,2} with a 1.6-fold higher prevalence in men.³ While rare relative to overall cancer incidence, glioblastoma accounts for 2.5% of total cancer-related deaths.¹

The clinical management of glioblastoma clinical management has not improved in the last 30 years.⁴ First-line therapy for newly

diagnosed glioblastoma is maximal safe resection of the tumour, followed by concurrent chemo-radiation and maintenance chemotherapy.⁵ Despite these aggressive treatments, the disease almost inevitably recurs, in which case there is no standard treatment available.⁶ This lack of progress could be attributed to two possible reasons (i) extensive intra- and inter-tumour heterogeneity and (ii) the highly invasive and infiltrative nature of these tumours, both of which are dependent on the interactions of the tumour cells with the surrounding microenvironment.⁷

Gliomas are graded and categorised according to the World Health Organization (WHO) guidelines based on a combination of

¹Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA, Australia; ²Computational Learning Systems Laboratory, UniSA STEM, University of South Australia, Mawson Lakes, SA, Australia; ³Department of Mathematics of Computation, University of California, Los Angeles (UCLA), CA, USA; ⁴Mashhad University of Medical Sciences, Mashhad, Iran; ⁵Adelaide Medical School, University of Adelaide, Adelaide, SA, Australia; ⁶Electrical and Computer Engineering Department, Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Mashhad, Iran; ⁷Flinders Health and Medical Research Institute, College of Medicine & Public Health, Flinders University, Adelaide, SA, Australia and ⁸Department of Neurosurgery, Flinders Medical Centre, Bedford Park, SA, Australia
Correspondence: Guillermo A. Gomez (Guillermo.Gomez@unisa.edu.au)

These authors contributed equally: Amin Zadeh Shirazi, Mark D. McDonnell, Eric Fornaciari, Guillermo A. Gomez

Received: 23 October 2020 Revised: 16 March 2021 Accepted: 8 April 2021

Published online: 29 April 2021

histologic and molecular features.⁸ Grade IV gliomas correspond to glioblastoma and essential diagnostic features include atypical glial cells, brisk mitotic activity, evidence of microvascular proliferation (MVP) and/or significant necrosis. MVP typically appears as glomeruloid tufts of multi-layered endothelial cells that are mitotically active along with smooth muscle cells or pericytes. Necrosis is a fundamental feature of glioblastoma and, together with the presence of blood vessels, is the strongest predictor of aggressiveness.^{9–11}

Although analysis of gene-expression data (bulk RNA-seq) available from The Cancer Genome Atlas (TCGA) has enabled the successful identification of the molecular signatures/biomarkers associated with the different GBM tumour subtypes, to date, GBM treatment based on this information alone has not resulted in improved patient survival.^{4,12} GBM treatment is further complicated by the fact that glioblastoma tumours also exhibit a high level of cancer cell heterogeneity and plasticity.^{13–19} Indeed, tumour plasticity and the capacity of cancer stem cells to partially and/or reversibly differentiate into different cancer cell populations is believed to be the main cause of resistance to therapy and the development of tumour recurrence.^{13,16,20} Understanding the mechanisms that contribute to glioblastoma plasticity and how it is shaped by the tumour microenvironment (i.e. stem cell niches) has been difficult. This is partly due to the lack of appropriate tools to unravel the spatial and functional interactions that occur between tumour cells and cells in the tumour niche, and to predict which interactions are important regulators of cancer stem cell plasticity and the capacity of cancer cells to infiltrate the surrounding healthy brain tissue.¹²

Recently, a Decision Forests statistical machine learning-driven algorithm ("Mill") was used by the Ivy GAP to identify and segment tumour regions and label the anatomic features in ~12,000 histological images, which were then laser-capture micro-dissected and subjected to bulk RNA-seq.²¹ The Ivy GAP data constituted a significant advance in the field and has allowed the identification of signatures associated with different tumour regions.²¹ However, it has limitations, as the results were highly dependent on the segmentation algorithms that were implemented for the segmentation of the different tumour regions^{22–25} and inherently biased due to the small number ($n=32$) of patient samples analysed.²¹ Furthermore, although multiple sections/regions were processed for data augmentation and the creation of the Ivy GAP database, the small number of patients recruited to the project did not permit the identification of genetic signatures or key demographic characteristics associated with survival and/or particular molecular profiles of the tumours.²¹ Moreover, the gain in insight into spatial resolution by attributing gene signatures to specific tumour regions was not translated into an understanding of the expression of these signatures by specific cell types, as limited cellular ontology analysis was performed on this data, due to the limited number of datasets and bioinformatic tools at the moment of its publication. Consequently, there is currently no available data that permits the identification of the cell populations that are present in each of these tumour anatomic locations and the corresponding transcriptional programs that underpin tumour-stroma interactions that may contribute to poor survival. Gaining insight into these aspects would be critical to progress our understanding on the biology of these very aggressive brain tumours.¹²

Here, we implemented a deep convolutional neural network (DCNN) model for semantic segmentation of histopathological (Haematoxylin-Eosin, H&E) images trained on the Ivy GAP dataset to improve the quality of segmentation of different brain tumour regions (Supplementary Fig. 1A). We sought a pathologist's opinion to advise on the removal of outliers (Supplementary Fig. 1B includes some examples) in the input dataset and for verification of results, that together contribute to improve the accuracy of segmentation of different brain tumour regions. The resulting model was then employed to segment TCGA

histopathological images and correlation analysis between gene expression and tumour region size was used for the identification of different gene signatures associated with different tumour regions as well as the different cell types that express them. We further combined these results with single-cell RNA sequencing (scRNA-seq) data to investigate tumour-stroma interactions that contribute to poor prognosis in glioblastoma.

METHODS

Histology image datasets

We used the Ivy GAP dataset, as the main dataset for training our DCNN semantic segmentation model (GBM_WSSM). This dataset includes 32 patients diagnosed by primary surgery type with a total of 805 whole-slide images (WSIs) along with their corresponding ground truths.²¹ Three distinct datasets (training, validation and test) were created by randomly splitting the data in the ratio of 14:1:1. Twenty eight patients with 687 WSIs were allocated to the training dataset, two patients with 52 WSIs were allocated to the validating dataset and two patients with 48 WSIs were allocated to the testing dataset. The validation set was used to evaluate the performance of our models during the training phase while the test set includes unseen samples for checking the model after training. An additional dataset was downloaded from TCGA²⁶ and includes 640 whole-slides H&E histopathological images from 329 glioblastoma tumours. The GBM_WSSM was applied to this dataset to produce the masks corresponding to all patients. Those masks were then used for further biological investigation as described in the paper.

TCGA RNA-Seq dataset

The following filters were applied to extract RNA glioma datasets from TCGA for correlation analysis. Project: TCGA GBM, Disease Type: gliomas, Vital Status: dead; Data Category, transcriptome profiling; Data Type: RNA-Seq. The resulting cases were further filtered by the availability of matched and successfully segmented whole-slide images from the same tissue portion used for RNA-seq analysis (Supplementary Table 1). FPKM-UQ data from this cohort of patients was used and when available, replicates measurements from the same patient were averaged. ENSG to gene name conversion was performed using the Biomart tool from Ensembl (<https://m.ensembl.org/biomart/martview/e5476740b357b1b968ca8507fd4d853e>) using Homo sapiens (human) genome assembly GRCh38.p13 from Genome Reference Consortium.

Image pre-processing

In the pre-processing stage, all WSIs and GTs from the Ivy GAP dataset were first visually checked by the trained pathologist of our team (N.S.B.) and 18 of 805 were removed as outliers (Supplementary Fig. 1B) because of high-level noise presenting poor segmentation in the masks. All images and masks were then re-sized to four different scales: 4096×4096 , 2048×2048 , 1024×1024 and 512×512 from the original size $18,000 \times 15,000$. Afterwards, different sizes of patches (1024×1024 , 512×512 , 256×256) were systematically extracted (not randomly) from the re-sized images and their masks. Therefore, the extracted patches were the entire datasets obtained from images and used in the training phase. Patches were then normalised by using a pixel-normalisation method into the domain $[0,1]$ by dividing all pixel values by the largest pixel value among all patches i.e. 255. For data augmentation purposes, random crops and vertical flips were applied in the training phase to all models.

Algorithms

For semantic segmentation of the eight different brain tumour regions inside the GBM WSIs, a customised fully convolutional DenseNets for semantic segmentation with 103 layers (based on the tiramisu design)²³ was trained and validated on the normalised

patches. These regions include (Fig. 1a) leading edge (LE), infiltrating tumour (IT), cellular tumour (CT), cellular tumour microvascular proliferation (CTmvp), cellular tumour pseudopalisading region around necrosis (CTpan), cellular tumour perinecrotic zones (CTpnz), cellular tumour necrosis (CTne) and background (BG, i.e. area void of tissue). Spearman correlation analysis (Matlab 2019b, see also Supplementary Materials 1, 2 and 3) between gene expression and brain tumour region size was used to identify genetic signatures that are specific to each tumour region (Supplementary Table 2). *p*-values against the null hypothesis that there is no correlation between these variables were calculated for each Spearman correlation test and are also included in Supplementary Table 2. These *p*-values were not corrected for multiple comparisons²⁷⁻³⁰ since (i) the *p*-value distribution is not the same across all brain tumour regions and is not always uniformly distributed (Supplementary Fig. 2, A and B), and (ii) there are significant correlations within each of the variables (i.e. genes vs genes expression; region size vs region size) used for Spearman calculations (Supplementary Fig. 2, C), which suggest that multiple Spearman tests were not fully independent of each other.

Performance evaluation

We trained 49 different networks and used the random search approach for hyperparameter tuning in each one (Supplementary

Table 3). We used sparse categorical cross-entropy, and RMSprop, as loss function and optimiser method, respectively. The number of epochs in different experiments varied between 30 and 150. For prevention of overfitting, dropout technique and L2 weight regulariser were applied. Experiments' performance comparison was performed based on the evaluator used in the original paper²³ and by using pixel-wise classification accuracy, where each predicted pixel in the mask produced by a DCNN model is compared with the corresponding pixel in the original ground truth.

Hardware and software

We used python as the programming language, Keras,³¹ a high-level neural networks API with the Tensorflow platform,³² and trained all networks using four NVIDIA 1080Ti GPUs. PRISM and Matlab2019b were used for Pearson's and Spearman correlation analysis (codes provided Supplementary Materials 1, 2 and 3). Survival plots were generated in Matlab using MatSurv function,³³ Cell ontology analysis was performed using CellKb (<https://www.CellKb.com>), Clustergrams were generated using Morpheus (<https://software.broadinstitute.org/morpheus>) and Matlab2018 (Supplementary Materials 1). Euclidean distance was used for clustering of variables. Venn diagrams were drawn using InteractiVenn.³⁵ Gene ontology analysis was performed using

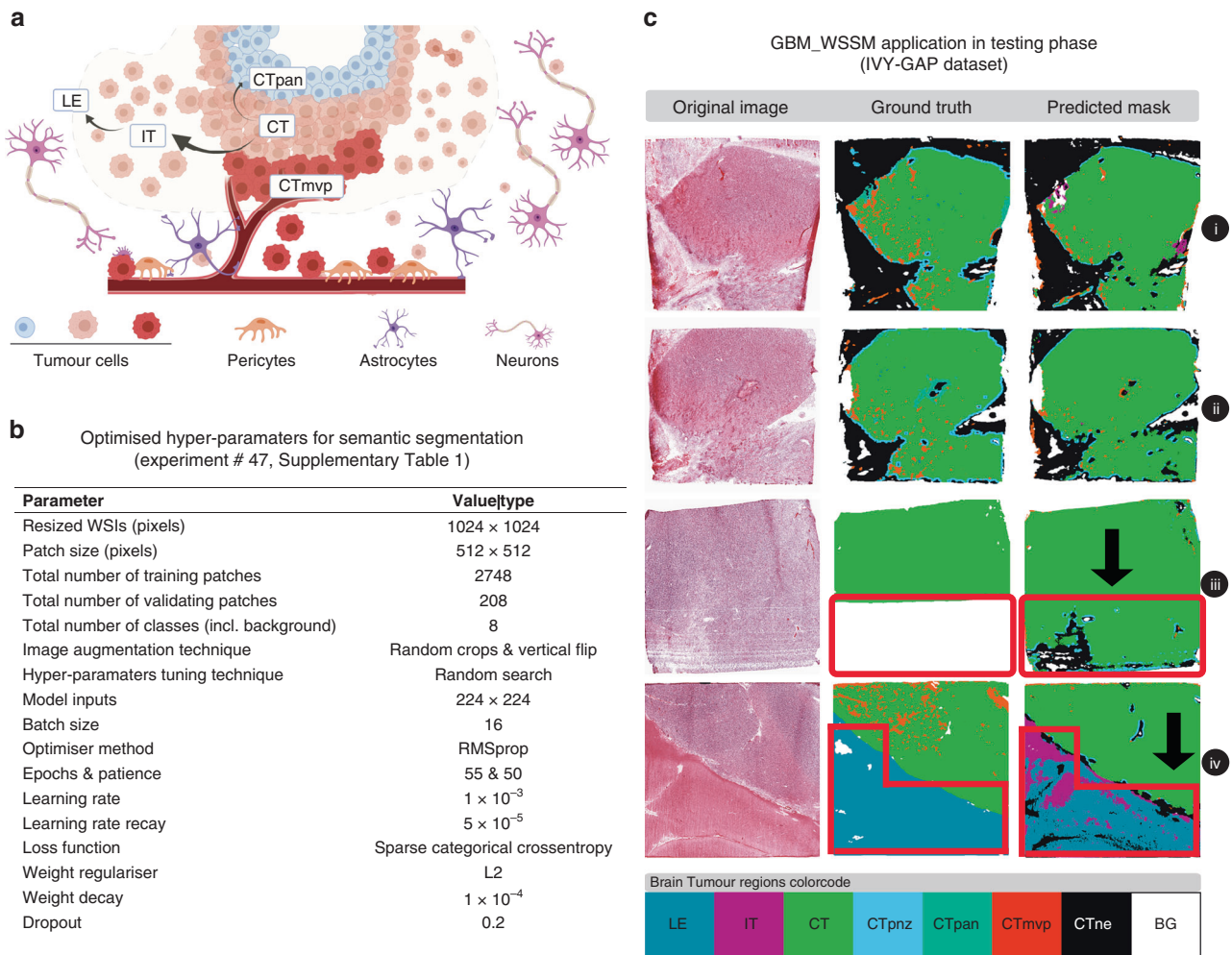


Fig. 1 The training and testing phases of the GBM_WSSM model on the IVY-GAP dataset. **a** Schematic representation of different brain tumour regions in glioblastoma. **b** Optimal experiment details for GBM WSI semantic segmentation model (GBM_WSSM). **c** Representative images from segmentation results of the proposed model in the testing phase (GBM_WSSM segmentation superiority has been shown by black arrows).

Gorilla³⁶ and network graphs generated using Cytoscape.³⁷ Scientific Illustrations were created with BioRender.com and all Figures were compiled using Adobe Illustrator 2020.

scRNA-seq experiments and data analysis

Preparation of samples, library preparation and bioinformatic pipelines for the generation of curated scRNA-seq data from three resected glioblastoma tissue samples was described before.³⁸ Cell quality control and clustering were done using Seurat version 3.0³⁹ as we described before³⁸ and analysis of paracrine ligand–receptor pairs between clusters were done using SingleCellSignalR.⁴⁰

We use average expression per cluster of differential expressed genes in this data for analysis of cell ontology and hierarchical clustering analysis.

RESULTS

A DCNN model improves semantic segmentation accuracy of whole-slides histopathological images of glioblastoma. Large labelled datasets are essential parts of the training phase of the deep learning models with supervised learning. Sometimes, it can be seen that the labels (ground truths) for training deep learning models have been generated by non-experts or automated methods and hence, the level of noise in such labels is typically higher than the labels annotated by experts. However, recent studies confirm that DCNNs are extremely robust to handle the high level of noisy labels in supervised learning approaches since well-designed DCNNs applied to sufficiently large and diverse dataset do not memorise the data and they learn the dominant patterns shared among all samples.⁴¹

The Ivy GAP portal (<http://glioblastoma.alleninstitute.org/>) is a freely accessible online database that contains GBM slides and their corresponding ground truths annotated using a statistical machine learning method. Although this dataset is very helpful for further research, we identified variable segmentation accuracy in some masks by manual inspection.

Because of the wide use of the Ivy GAP histopathological images database within the brain cancer research community, we decided to use the GBM slides and their corresponding labels from the Ivy GAP to train a deep convolutional neural network (DCNN) architecture and obtain a semantic segmentation model of glioblastoma histopathological images. To increase the accuracy of the semantic segmentation model, we took the advantage of DCNN models in addressing noisy labels⁴¹ and introduced the pathologist opinion after the training phase of each experiment (Supplementary Fig. 1, Supplementary Table 3). Then, the best model was applied to TCGA as a larger database containing more GBM patients to segment its GBM slides.

Our results indicated that WSI and mask re-sizing from its original size 18,000 × 15,000 to 1024 × 1024 and extracted patches with the size of 512 × 512 led to the highest segmentation accuracy, reaching ~70% determined by pixel by pixel comparison (model 47 in Supplementary Table 3, and Fig. 1b). However, when compared to the original H&E image, the model accuracy was better than the ground truth (as can be seen in Fig. 1c). In this figure, it can be seen that the segmentation accuracy in our model is better than that achieved in the original GTs, particularly in the third and fourth examples shown by the black arrows and the red polygons. Based on our pathologist's advice, our implemented DCNN model in segmentation of GBM slides has the following advantages (in comparison with the original GTs produced by a statistical machine learning model):

- The DCNN model can segment unclear regions in the original slide (Fig. 1c, iii)
- The DCNN model is able to properly detect the regions corresponding to the Infiltrating Tumour (specified by the

purple colour) and Cellular Tumour Necrosis (specified by the black colour) regions without over-segmenting the Leading-Edge region (Fig. 1c, iv)

- The DCNN model can tackle the problem of over-segmentation of the Cellular Tumour Microvascular Proliferation region (specified by the orange colour in the masks, Fig. 1c, i, ii, and iv)
- The DCNN model can tackle the problem of over-segmentation of the Cellular Tumour (specified by the green colour) region (in almost the majority of the masks produced).

We reported the results using the global accuracy metric across all eight classes²³:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

This equation calculates the ratio of the correctly classified pixels (TP + TN) with regard to the total pixels (TP + TN + FP + FN). We have used the accuracy metric as an evaluator to compare between our 49 different DCNN experiments during the training phase, hyper-parameters tuning, and to select the best models. For this purpose, we applied the Human_In_The_Loop (HITL) approach, which is a setting in a loop where an expert can insert prior knowledge into an AI machine to enhance its output.⁴² In our case, the HITL approach entailed our pathologist supervising/evaluating the results (segmented masks) obtained by the experiments with the better performance (i.e. higher accuracy, see also Supplementary Fig. 1). Subsequently, if the results were not satisfactory (based on the pathologist's opinion), even if the accuracy values achieved by the experiment was good, the experiments were repeated by adjusting the hyper-parameters to eventually achieve a better segmentation result among all regions than the previous model producing the original GTs. Furthermore, while we have eight distinct regions/classes, we do not see a significant class imbalance in the original GTs, particularly in the main parts i.e. Cellular Tumour, Infiltrating Tumour, Leading Edge and Necrosis. The Background (BG) region also covers a small portion of each GT and the ~70% segmentation accuracy has been achieved among all eight classes. For simplicity, we call our best model "GBM_WSSM".

Different brain tumour regions are associated with distinct GBM mutation profiles that are indicative for poor prognosis

We then applied the GBM_WSSM to the GBM TCGA dataset of whole-slide H&E stained histopathological images. This dataset includes 640 whole-slide images from 329 glioblastomas. All the WSIs were re-sized and fed to the GBM_WSSM as new inputs and their corresponding masks were produced (Fig. 2a), which were also evaluated by our team pathologist for segmentation accuracy. To produce a numerical dataset, the area of the masks for different tumour regions were quantified across all 329 samples analysed. For tumour region quantification, we used the averaged pixel counting approach. In this approach, for each patient, the same pixels related to one region are counted across all masks of that patient and then, the counted value is averaged over the number of slices. This was performed in order to capture the total amount of tumour region across different sections and have a better representation of the volume (measured as total number of pixels) that each of these occupies in the portion of tissue analysed at the same time that normalises the total number of pixels detected for each region by the number of sections available for each patient, as, in many instances, for each patient, there are the different number of available sections. Figure 2b shows the quantified regions for some of the patients. The Grubbs' test⁴³ was then applied on this numerical dataset and 27 outliers were removed in order to obtain a normally distributed survival data, which is required for the analysis of the relationship between

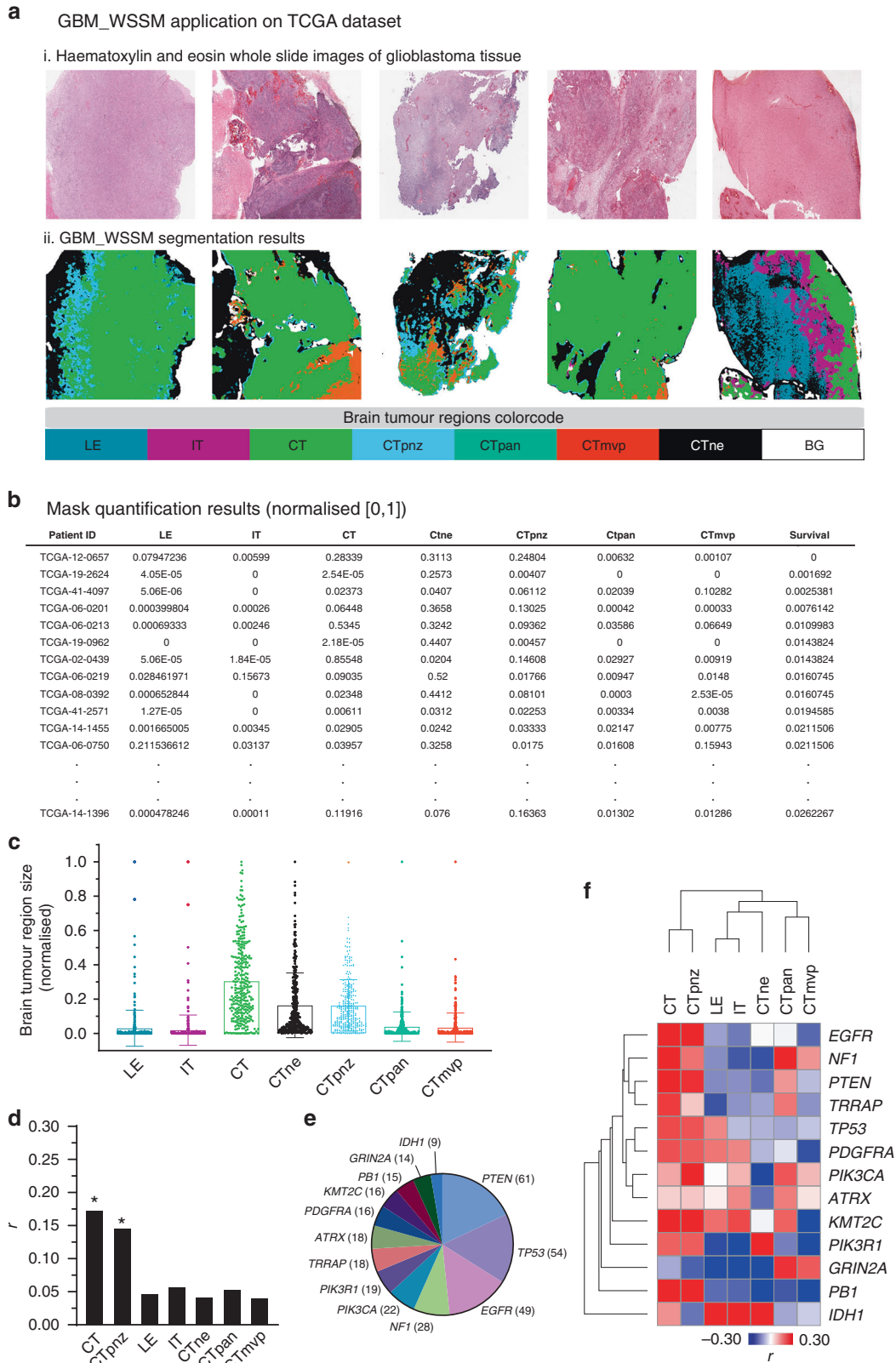


Fig. 2 The application of the GBM_WSSM model to the TCGA dataset. **a** Representative results from GBM_WSSM application on the TCGA samples. (i) WSIs from TCGA. (ii) Corresponding masks produced by GBM_WSSM. **b** Measurement of different tumour regions (normalised) and patient survival in each TCGA glioblastoma case. **c** Dot plot graph of the results shown in **b**. **d** Pearson correlation values between survival rates and normalised areas corresponding to each tumour region (asterisk denotes $p < 0.05$). Comparable statistically significant results were obtained in the presence or absence of outliers using Spearman correlation analysis, which is less sensitive to the presence of outliers. **e** Distribution of glioblastoma patients harbouring specific oncogenic mutations in TCGA. **f** Heatmap plot of Pearson correlation values between the size of the tumour regions and survival rates.

segmentation results and patient survival through the calculation of correlation coefficients^{44,45} (Supplementary Table 4).

The results of tumour region size quantification revealed that across the set of TCGA glioblastoma tumours, large areas of WSI images contained brain tumour regions corresponding to the cellular tumour mass including the CT, CTne and CTpnz regions (Fig. 2c, Supplementary Fig. 3, Supplementary Table 4).

This is expected as these regions are the primary focus of the neurosurgery and are thus more likely to appear in resected samples. Other more peripheral regions exhibit higher variability, likely due to differences in surgical procedures and the ability to delimit the tumour boundaries during the operation. With this in mind, we believe our segmentation results agreed with what would be expected from the analysis of data derived from samples collected in multiple centre/institutions, where practices and experimental protocols might not be standardised across the sites and site-to-site differences are likely to be present.

We have previously shown that DCNN models can be trained and applied to TCGA H&E histopathological images to predict brain cancer patient survival⁴⁶ (see also review in ref. ⁴⁷). To analyse the relationship between segmentation results and patient survival, we next performed Pearson correlation analysis between the area of different brain tumour regions and patient survival across (i) all TCGA glioblastomas, and (ii) brain tumours harbouring the 13 most frequent mutations (PTEN, TP53, EGFR, NF1, PIK3CA, PI3KR1, TRAPP, ATRX, PDGFRA, KMT2C, PB1, GRIN2A, IDH1; see Supplementary Fig. 3 for tumour region size distribution across different patient's cohorts). When we analysed all patients in the TCGA GBM cohort, we found that survival rates are positively correlated with the area CT and CTpnz but poorly correlated with other regions (Fig. 2d). The relatively low correlation values could be due to the heterogeneous nature of the data (i.e. pooled across all patients). To determine if this was the case, we next stratified our analysis by patients harbouring specific oncogenic mutations (Fig. 2e). Mutation-specific stratification resulted in higher correlation values between tumour region size and survival rates and also revealed new additional cases where region size was negatively correlated with survival (Fig. 2f). In brain tumours harbouring mutations in the PIK3R1 and PB1 genes, the presence of LE, IT and CTmvp regions negatively correlated with patient survival, likely reflecting the highly infiltrative nature of tumours bearing these histologies.^{48–50} Moreover, in patients harbouring mutations in the PDGFRA, KMT2C or EGFR genes, we observed a negative correlation between the presence of highly vascularised tumour regions and patient survival, again reflecting the more aggressive phenotype of this type of tumour. Interestingly, we also note that in patients harbouring mutations in the NF1, PTEN, TRAPP, TP53 and PIK3CA, there is a strong positive correlation between survival and central tumour regions (i.e. CT and CTpnz) but little or no correlation with the other infiltrative regions LE, IT and CTmvp). This probably reflects non-invasive and therefore more benign tumours. Overall, these results highlight the varying influence of gene mutations on histopathological features and reveal tumour features associated with prognosis. These data also highlight critical differences that significantly influence the biology of these tumours and their capacity to escape complete resection during surgery as well as to develop therapy resistance.

Gene signatures of specific brain tumour regions are associated with distinct biological processes in tumour biology and are indicative of patient survival

RNA-seq data from GBM patients in TCGA enabled us to identify gene signatures (markers) associated with each of the different regions. We reason was that genes specific to one region would show higher expression in tumour samples where the area for that region (measured as the total pixel count) is bigger. For this, we used Spearman correlation analysis results between gene

expression (>35,000 genes, whose expression was detected in at least 25 patients) and brain tumour area for different brain tumour regions. The results indicated that for different regions, there are both positively and negatively correlated genes (Fig. 3a, Supplementary Table 2). Of these, we focused only on positively ($\rho > 0.1$) and significantly ($p < 0.05$) correlated genes (Fig. 3b, Supplementary Table 2). Overall, we found that the different brain regions have different numbers of genes that behave as “markers” according to our definitions, with IT and CTmvp regions exhibiting the largest number of marker genes (174 and 373 genes, respectively, Fig. 3c, Supplementary Table 5). Importantly, our analysis validated many of the genes previously annotated as markers in the Ivy GAP Database (Fig. 3d). However, the number of genes identified in the Ivy GAP project was considerably larger than the number that we identified. This is likely due to the fact that not all markers identified in Ivy GAP have a prognostic value as well as the lowest accuracy of Ivy GAP segmentation results, which coupled to analysis of the fewer number of patients, could also lead to a significant increase in the number of false positives.

Next, we analysed the relationship of the region-specific gene signatures with patient survival, which was possible due to the large number of patients whose data are available in the TCGA glioma database ($n = 129$). We found that most of the genes in the CT and CTpnz signatures positively correlated with survival, which agreed with our previous observation in cohorts of patients with a specific set of mutations (Fig. 2d, 3c). We also found that for LE, CTne and CTpan, an approximately equal number of genes positively and negatively correlated with patient survival (Fig. 3c). Interestingly, we found that in IT and CTmvp, the regions with more specific markers, most genes negatively correlated with survival (Fig. 3c). This is consistent with tumour infiltration into the surrounding brain tissue being unfavourable for complete surgical resection and with the capacity for infiltrating tumours to rapidly develop therapy resistance, which together contributes to poor patient survival.

These results revealed different gene signatures for different regions that have different correlations with patient survival. In order to gain insight on these, we then focus on IT and CTmvp gene markers whose expression is negatively correlated with patient survival (IT^{neg} and CTmvp^{neg}) as well as CT and CTpnz gene markers whose expression is positively correlated with patient survival (CT^{pos} and CTpnz^{pos}). Gene ontology (GO) biological processes analysis allowed us to investigate the cellular processes associated with these gene signatures (Fig. 3e, f and Supplementary Fig. 4, Supplementary Table 6). For the IT^{neg} signature, we identified several GO processes, including cell proliferation in midbrain and regeneration, possibly reflecting the stem-cell-like properties of tumour cells in those regions.^{18,19,51} Processes involving ‘glial cell migration’, ‘locomotion’, ‘cell contractility’, ‘cell movement’ and ‘lymphocyte migration’ were also identified to be associated with the gene signatures specific to this region. This highlights the migratory and invasive potential of cancer cells⁵² as well as the motility of microglia and lymphocytes, suggesting that in the IT region these different cell types within the microenvironment may be cooperating to enhance the invasive capacity of glioblastoma cells.⁵³ Interestingly, several GO processes that belong to stress cell responses were also identified in this region (i.e. response to: ‘cycloheximide’, ‘organic’ and ‘inorganic’ substances; ‘glucocorticoids’, and to ‘chemicals’), again reflecting properties of the microenvironment to which tumour and inflammatory cells are exposed.¹⁶

GO biological processes analysis revealed a more complex microenvironment in the CTmvp^{neg} signature relative to IT^{neg}. Important GO processes associated with this signature corresponded to ‘immune’ and ‘immune-inflammatory responses’, ‘cytokine signalling’ and ‘regulation of leukocyte proliferation’, indicating a fine regulation of immune responses in the tumour microenvironment in CTmvp regions. Other key GO processes

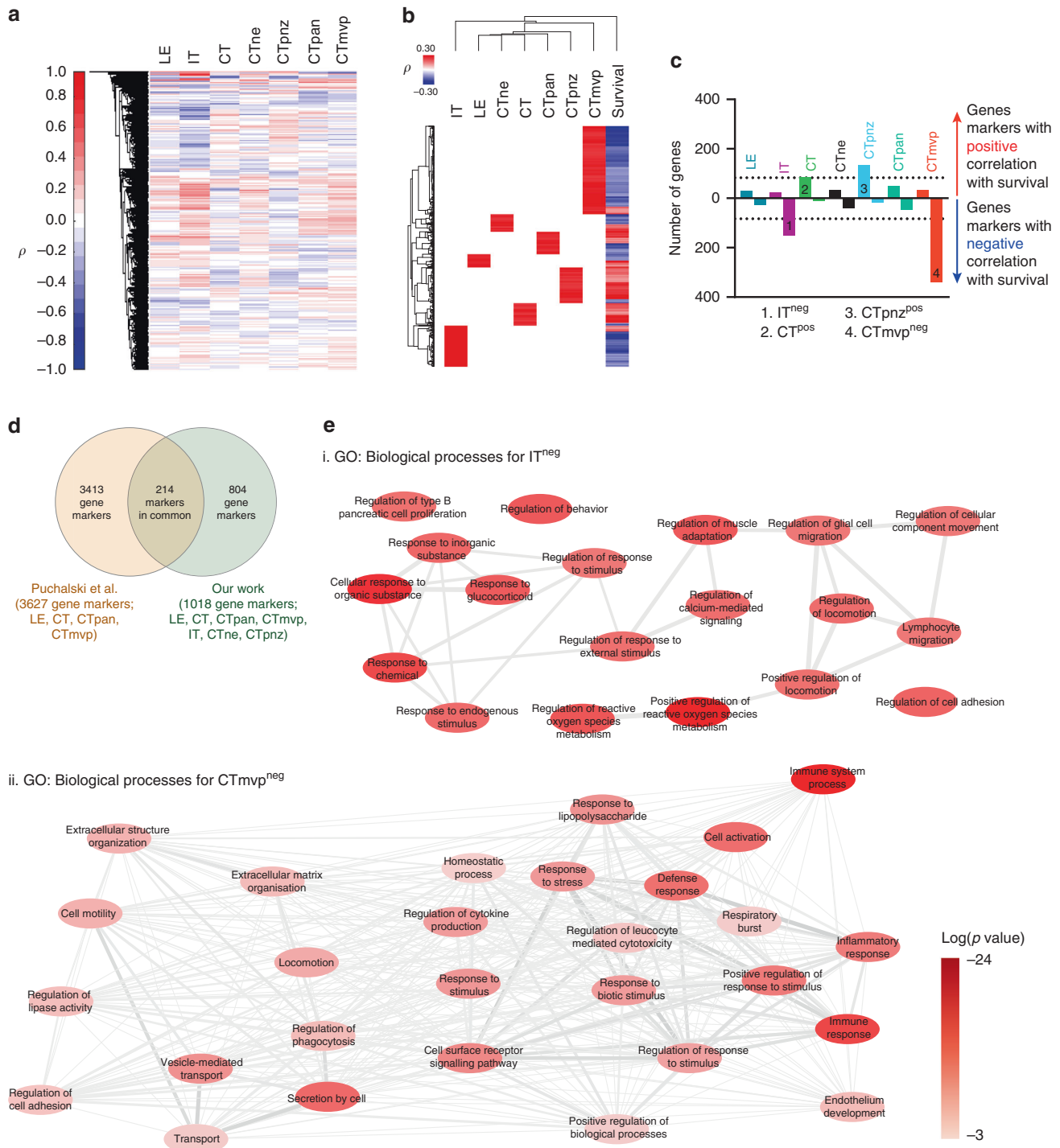


Fig. 3 Gene expression analysis in the TCGA segmented tumour regions. **a** Hierarchical cluster analysis (columns only) of Spearman correlation coefficients (ρ) between gene expression and tumour region area measurements for the cohort of TCGA patients in Supplementary Table 1. This cohort was obtained from TCGA cases that have both RNA-seq and WSI data available (See Supplementary Table 4 for Spearman and p -values numerical results). **b** Columns for IT, LE, CTne, CT, CTpna, CTpznz and CTmvp are same as in A but the data was filtered for genes that have $\rho > 0.1$ and this correlation is significantly different from the null hypothesis of no correlation ($p < 0.05$). The 'survival' column shows Spearman correlation coefficient values between patient survival and gene expression for all the genes analysed. **c** Bar graph showing the number of markers for each tumour region whose expression is positively and negatively correlated with patient survival (See Supplementary Table 5 for the list of genes for each condition). **d** Venn diagram showing the comparison of gene markers for each tumour region identified in this work and previously by Puchalski et al.²¹ **e, f** Summary of GO: Biological Processes analysis for IT^{neg} and CTmvp^{neg} gene signatures (see also Supplementary Table 6 for process description, p -values and genes associated to each process). The network was simplified by highlighting only GO: Biological Processes with $\log(p\text{-value}) < -3.5$ and the nodes (i.e. GO: Biological Processes) colour coded (red levels) to highlight the more significant processes associated to each signature. Colour bar indicates different p -values for each node.

identified in this region included 'secretion by cell', 'vesicle-mediated transport', 'collagen metabolism', 'cell motility', 'locomotion', 'extracellular matrix (ECM) organisation'. These are related to the regulation of ECM production and the interaction of that ECM with cells and is consistent with the capacity of tumour cells to interact with the ECM surrounding the vasculature and infiltrate healthy brain tissue along the microvasculature.⁵⁴

In contrast to the IT^{neg} and $CTmpv^{neg}$ signatures, the CT^{pos} and $CTpnz^{pos}$ gene signatures showed a strong connection to those processes related to DNA, RNA and protein synthesis through different metabolic pathways, that align overall with higher macromolecule synthesis rates required by highly proliferative tumour cells (Supplementary Fig. 4). Taken together, these results establish the power of using AI to segment tumour regions in many patient-derived samples and the use of this information for the identification of gene signatures characteristic of each region.

Gene signatures in different brain regions predict the presence of specific cell types in the tumour microenvironment

The diverse GO processes associated with the different gene signatures motivated us to investigate the identity of the cells that express these signatures in the different tumour regions. Accordingly, we investigated the cell types that are more likely to express these gene signatures in each tumour region using CellKb application software, which uses publicly available scRNA-seq data to predict the cells types that are likely to express a list of differentially expressed genes ("match score"³⁴). This match score is calculated for each cell type and correspond to the sum of rank-based scores calculated for overlapping genes between the query and the cell type. The match score thus accounts for the gene rank, the difference in gene ranks and the total number of significant genes in the cell type. Using CellKb, we obtained a list of cell types and for each cell type an associated match score and a list of genes specific for that cell type (Supplementary Table 7).

In this list, a cell type may be predicted more than once, since multiple databases (scRNA-seq experiments) were used. Thus, to uniquely assign a cell type to our gene signature, we weighted the analysis for not only using the cell type match score but also taking into account the number of times a particular cell type was predicted using this analysis (Fig. 4ai–iv, Supplementary Table 8). The results revealed that IT markers that are negatively correlated with survival (IT^{neg}) are predicted to be highly expressed by microglial cells (CL:0000129) and monocytes (CL:0000576). This is consistent with the GO processes discussed above but highlighted a key non-cell autonomous role of the tumour microenvironment (microglia) in the poor prognosis of glioblastoma. Cells in $CTmpv^{neg}$ compartment also exhibited monocyte/microglia characteristics. Moreover, this compartment also included cells expressing genes that are specific for macrophages (CL:0000235), endothelial cells (CL:0000115), dendritic cells (CL:0001056), cerebral cortex endothelial cells (CCEC, CL:0001056). Fibroblasts are probably representative of pericytes, which coat blood vessels and contribute to the architecture and function of the blood–brain barrier. The results thus indicate that the genes characteristic of these different regions, that are associated with poor patient prognosis, are more likely expressed by cells of the tumour microenvironment than the cancer cells.

Finally, we performed cell ontology analysis on the CT^{pos} and $CTpnz^{pos}$ regions. These regions are expected to be enriched in tumour cells, which was supported by our analysis revealing that neoplastic cells are more likely to express the gene signature corresponding to each region. Interestingly, T cells were only associated with the $CTpnz$, a potential inflammatory region that might chemoattract T cells (Fig. 4a). These data also indicate that this region may be particularly important for targeting with checkpoint inhibitor therapy.⁵⁵

CellKb analysis also incorporates scRNA-seq data from a large database of scRNA-seq studies from a range of tissues and disease

states that do not necessarily correspond to glioblastoma tumour tissue. We therefore sought to validate our results using scRNA-seq data that we obtained from three glioblastoma patients.³⁸ Our previous analysis using cell markers for different cell types in the literature allowed us to define the different cell populations in the scRNA-seq UMAP plots. The tumours analysed were principally composed of glioma-associated macrophages (including microglia and macrophages), lymphocytes, endothelial cells, pericytes and tumour cells.³⁸ In order to map the gene signatures identified by semantic segmentation to this data, we first re-analysed our scRNA-seq data using CellKb (Supplementary Table 9). Using this approach, we were able to assign a cell ontology type to each of the different cell clusters in the data derived from each individual patient (Fig. 4b, Supplementary Table 10), which correlated well with our previous manual cell assignment.³⁸ Once individual clusters were analysed, we then calculated the average expression of each gene in the gene signatures of the CT^{pos} , IT^{neg} and $CTmpv^{neg}$ regions and performed hierarchical clustering to determine which cell types within the brain tumour expresses each of the genes in the different gene signatures that we identified for each region (Fig. 4c, Supplementary Figs. 5 and 6, Supplementary Tables 11–13).

Consistent with our results examining gene expression in publicly available scRNA-seq datasets, we found that the gene signature associated with CT tumour regions is predominantly associated with neoplastic tumour cells across the three patient-derived samples analysed (Supplementary Figs. 5 and 6). Furthermore, a group of genes within this signature was identified to be specifically associated with T, B and NK cells. However, no significant expression of these genes was observed in the microglia/macrophages lineage, which contrasted with what we observed in the corresponding results of the IT and $CTmpv$ regions (Fig. 4c). In these regions, a considerable number of genes of each signature were expressed by cell types including macrophages, microglia and monocytes. A larger number of genes within $CTmpv$ gene signature were also expressed in these cell types compared to that in the IT gene signature. We also observed that pericytes and endothelial cells express a significant number of genes that match the $CTmpv$ signature and a smaller number of genes in the IT signature. Overall, these results reinforce the notion that different brain tumour exhibit different microenvironment properties and cellular compositions that may contribute to prognosis in glioblastoma.

Gene signatures in different brain regions may mediate tumour-microenvironment interactions

Our cell ontology results indicate that different cell types in the tumour microenvironment contribute to poor patient prognosis. However, the molecular events underpinning this effect are unknown. A possibility is that the gene signatures mediate specific types of cell–cell interactions between tumour cells and the surrounding microenvironment (Fig. 5a). To investigate this, we performed a bioinformatic analysis to explore paracrine ligand–receptor interactions between the previously defined cell clusters in our scRNA-seq data by focusing on ligand–receptor pairs where the tumour cells express the receptor (Fig. 5a). For this, we used the recently developed approach to identify paracrine interactions since it includes one of the most comprehensive datasets for this type of analysis.⁴⁰ After running this bioinformatic pipeline and filtering the results for (i) receptors identified in the IT or $CTmpv$ gene signatures, and;—(ii) expression of the ligand–receptor pair in all three patient-derived scRNA-seq datasets; we identified the reticulin 4 (RTN4): Gap junction beta-2 protein (GJB2) paracrine ligand–receptor pair as a potential mediator of cell–cell interactions between tumour cells and the surrounding microenvironment⁵⁶ (Fig. 5b, Supplementary Table 14). The *RTN4* gene (also known as *NOGO*) encodes three isoforms (isoforms A–C) with Nogo-A enriched in the central

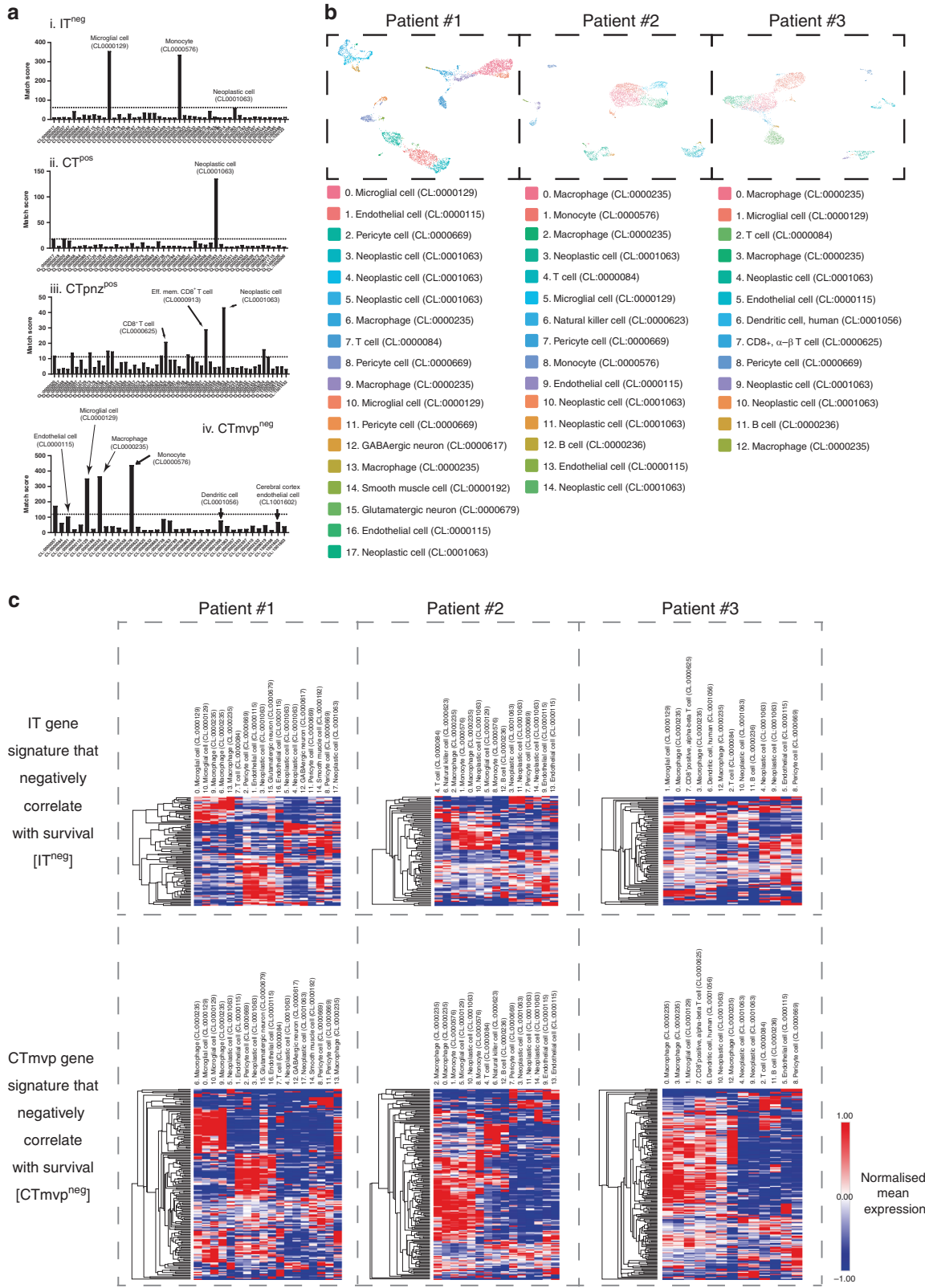


Fig. 4 Cell ontology analysis in GBM tumour regions. **a** (i–iv) Cell ontology analysis for the IT^{neg} , CT^{pos} , $CTpzn^{pos}$ and $CTmvp^{neg}$ gene signatures. The dotted line corresponds to the standard deviation calculated from the average match score across all cell types. **b** Cell type assignment to each of cell clusters from scRNA-seq data derived from three glioblastoma resected tissue samples. **c** Cluster analysis of average gene expression for genes in the IT^{neg} and $CTmvp^{neg}$ signatures in different cell types identified in our scRNA-seq data (See Supplementary Fig. 6 for a high-resolution image that also includes the corresponding gene names).

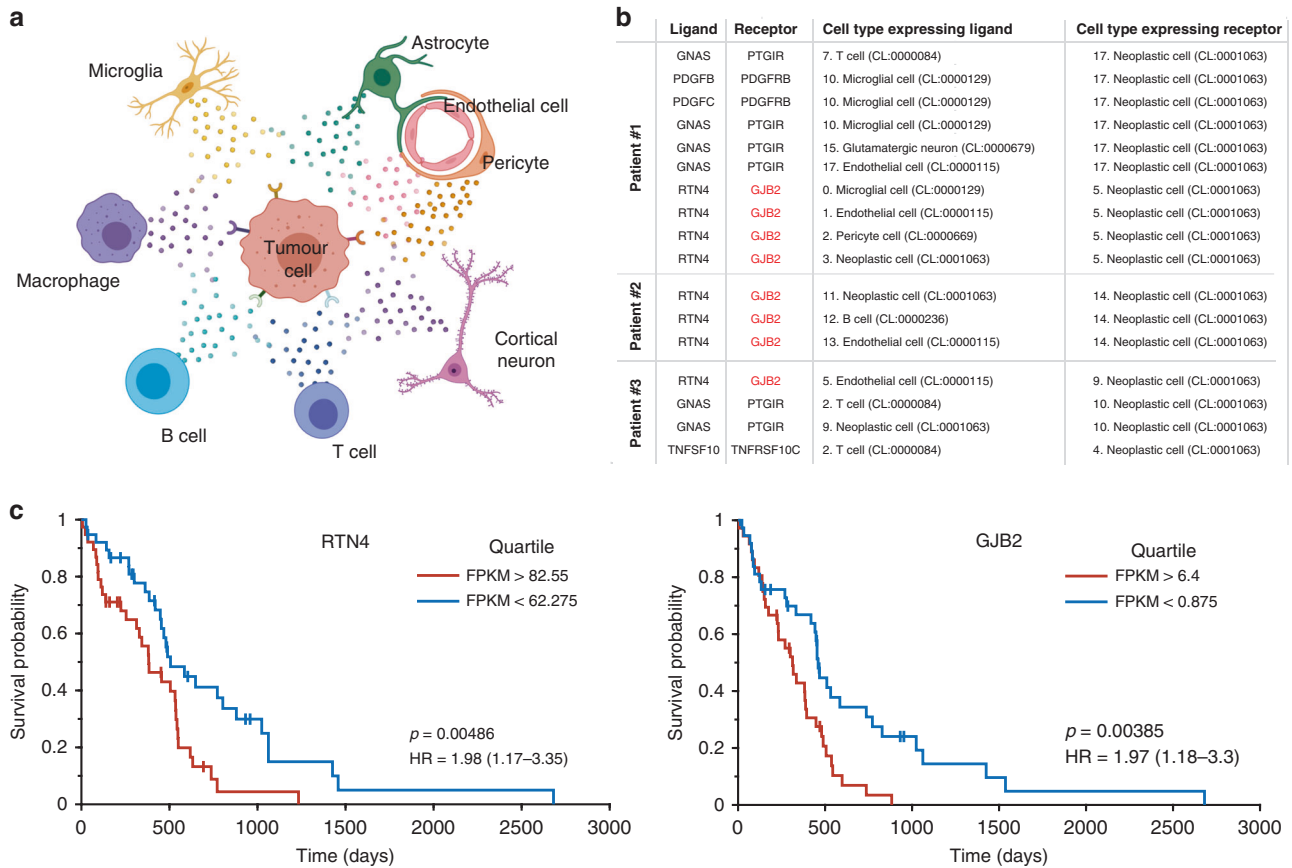


Fig. 5 GBM tumour-microenvironment interactions and survival analysis. **a** Schematic representation of different tumour-stroma interactions in glioblastoma. **b** Summary of potential ligand–receptor interactions identified in our dataset. Please note that GNAS (a $G\alpha$ subunit) and PTGIR (a G protein-coupled receptor) interacts on the cytoplasmic side of the plasma membrane. Supplementary Table 14 provides a full list of ligand–receptor pairs identified in this work. **c** Survival plots of TCGA GBM data for patients with low and high expression of GJB2 and RTN4 reveal their association with a poor prognosis in glioblastoma.

nervous system (CNS). NOGO isoforms contain two transmembrane domains and are localised through the exocytic pathway, plasma membrane and cell–cell junctions where it also interacts with cadherin 5 (Cdh5⁵⁷). NOGO proteins have been implicated in glioma cell invasion,⁵⁸ tumour angiogenesis⁵⁹ and regulation of blood vessel homeostasis.^{60,61} In contrast, gap junction protein 2B (GJB2, also known as Connexin26) is a structural component of gap junctions^{62–67} and is expressed at the plasma membranes of cells. GJB2 has previously been identified as a gene associated with prognostic value in brain^{68,69} and pancreatic cancers^{70,71} and contribute to tumorigenesis in breast cancer.⁷² Survival plots of TCGA GBM data for patients with low and high expression of GJB2 and RTN4 reveal their association with a poor prognosis in glioblastoma (Fig. 5c).

DISCUSSION

Here, we used DCNN for semantic segmentation of different brain tumour regions on TCGA GBM datasets that contain matched histopathological images, patient demographics and gene-expression data. DCNN is a robust tool that can handle the high-level noise in ground truths⁴¹ derived from semantic segmentation results using machine learning models, that has low-level accuracy.²³

Using results acquired using DCNN, we have identified specific markers for each of these tumour regions and evaluated their prognostic value in glioblastoma. Moreover, we combined this information with scRNA-seq experiments to determine the cohort of different cell types within these regions that express the genes

associated with these signatures as well as protein–protein interactions that mediate tumour-stroma relationships that can serve as potential targets for glioblastoma.

Several studies have been undertaken to characterise gene signatures relating to the tumour microenvironment.^{21,73–75} However, it has not been possible until now to investigate the spatial organisation and possible roles of non-cancer cell types that populate the microenvironment⁷⁶ and pinpoint their occurrence within the context of different brain tumour regions. Amongst, the most notable is the pioneering work by the Allen Institute that lead to the generation of the Ivy GAP database. By using laser microdissection and bulk RNA sequencing, this effort has generated spatial information about the different tumour regions within glioblastomas and associated genetic signatures with them. However, a limitation of this database, is that gene-signature results were derived from laser microdissection of a relatively small number of independent glioblastoma tissue samples ($n = 32$). Although our gene signatures exhibited some overlap with the gene signatures in the reported in the Ivy GAP database, our gene signatures had fewer genes. This is probably a consequence of the thresholds we applied for the correlation coefficient and the p -value, to define the different gene signatures and the fact that Ivy GAP tumour regions markers were defined by differences in expression between regions, which is distinct from the correlation versus no correlation approach between gene expression and brain tumour regions taken in this study. Interestingly, our analysis permitted the identification of a large number of genes that did not overlap with the results from Ivy GAP database. This reflects again the power of the combination of

AI for segmentation and the possibility of analysing large datasets to be able to perform robust correlation analysis for the identification of gene signatures.

The results of multiple laboratories performing scRNA-seq experiments using cellular suspensions derived from resected tumour tissue^{13–17,19,20,38,77} as well as bioinformatics approaches designed to extract microenvironment features from bulk RNA-seq data in TCGA⁷⁸ have significantly advanced the characterisation of the genetic signatures and cellular composition of the tumour microenvironment in glioblastoma. These previous studies have provided significant insight into the tumour cell heterogeneity in glioblastoma as well as the cellular composition of the tumour microenvironment.⁷⁹ Moreover, these experiments have not only provided a better understanding of tumour plasticity but also the transcriptional program changes that occur in different cell types in the tumour microenvironment.⁷⁷ However, it has not been possible until now to investigate the spatial organisation of non-cancer cell types that populate the microenvironment and pinpoint their occurrence within the context of different brain tumour regions. In a recent study, Rajapakse et al. used computational modelling to define cancer cell states and allocate active and dormant cancer cell types within these different tumour regions.⁸⁰ In contrast, in this work, we combined publicly available scRNA-seq data (using the CellKb platform) and our own experiments to determine the gene signatures expressed by different cell types in the microenvironment and how these relate to different tumour compartments. Although spatial transcriptomics^{81–83} and multiplex immunofluorescence imaging^{74,75} approaches can offer similar outcomes to that presented here, these techniques are still prohibitively expensive and time-consuming. We expect these results to facilitate the spatial analysis of tumour biopsies and promote the adoption of complementary techniques such as spatial transcriptomics^{81–83} and highly multiplexed immunofluorescence imaging.^{74,75} Although expensive and time-consuming, these approaches enable high-resolution detection of RNA or protein in tissue sections, which facilitates precise spatial analysis of the tumour microenvironment and measurements of distances (and possible interactions) between different cells of interest when combined with single-cell RNA sequencing analysis.⁸⁴

scRNA-seq data have also proved to be a powerful tool for assessing potential protein–protein interactions that can control tumour cell behaviour.^{40,85} Using these methods, we investigated membrane receptors expressed on tumour cells that can interact with secreted and/or membrane proteins, and determined whether these ligand–receptor interactions are part of the gene signatures that we have identified. Interestingly, we found the GJB2 (also known as connexin 26): RTN4 (also known as NOGO) pair in the CTmvp signature, with both genes being indicative of poor prognosis in glioblastoma.

Although the role of connexins has been primarily associated with its ability to function as hemichannels forming a direct transmembrane communication pathway between neighbouring cells, recent studies of clinical samples suggested connexins can contribute to cancer progression through multiple pathways, namely (1) gap junction intercellular communication, (2) C-terminal tail-mediated signalling and (3) cell–cell adhesion during gap junction formation.⁸⁶ In our data, the expression of GJB2 (and another receptor for RTN4, RTN4R) was restricted mostly to cancer cells, whereas the expression of RTN4 was more widely distributed between cells in the tumour microenvironment including microglial cells, endothelial cells, pericytes and B cells. GJB2 is a tetraspan membrane protein with two extracellular domains and is expressed in the plasma membrane. GJB2 has previously been identified as a gene with prognostic significance in brain,^{68,69} pancreatic^{70,71} and breast cancer.⁷² RTN4 has two long (32 aa) transmembrane domains and the region between these domains (NOGO-66) has been shown to mediate protein interaction with

cells expressing RTN4 receptors.^{87,88} The long transmembrane domains have been also postulated to confer distinct topological organisations of this protein, particularly exposing the N-terminus (the more divergent domain between RTN proteins) extracellularly.

In terms of downstream signalling, RTN4 localises both at the endoplasmic reticulum and at the plasma membrane. In the endoplasmic reticulum, it is known to interact with and suppress the anti-apoptotic functions of Bcl2 and BclXL,⁸⁹ suggesting a role as a tumour suppressor. However, more recent data in prostate cancer has shown that RTN4 regulates cell fate and its over-expression led to cell cycle arrest and senescence.⁹⁰ Besides, RTN4 binding to RTN4R, the more established receptor RTN4, contributes to ROCK and STAT3 activation in the cell expressing RTN4R,⁸⁸ which can contribute to glioblastoma tumour cells invasion (through ROCK⁹¹) and chronic microglia activation (through ROCK and STAT3^{92,93}). It is still unclear how RTN4's interaction with GJB2 and RTN4R contributes to poor prognosis in glioblastoma, but redistributing RTN4 from the ER to the plasma membrane and preventing it to interact with other RTN4 intracellular binding partners could be a possible therapeutic strategy that needs further investigation.

In summary, our analysis of the TCGA GBM database using segmentation of histopathological images provides new opportunities for the characterisation of the tumour microenvironment in glioblastoma within distinctly spatial regions within these very heterogeneous brain tumours. These results have led to the identification of novel prognosis-associated region-specific gene signatures and targets for treating glioblastoma and are now available to the rest of the brain cancer research community.

ACKNOWLEDGEMENTS

We thank the South Australia Neurological Tumour Bank (SANTB) for providing access to patient-derived tissue samples and our laboratory colleagues for their continuous support and fellowship. scRNA-seq experiments were performed at the Australian Cancer Research Foundation (ACRF) Cancer Genomics and Cancer Discovery Accelerator facilities, established with the generous support of the Australian Cancer Research Foundation. We also thanks to Dr. Lisa M. Ebert, Prof. Angel Lopez and Prof. Stuart M. Pitson at the Centre for Cancer Biology for valuable feedback discussions during the preparation of this work.

AUTHOR CONTRIBUTIONS

A.Z.S. and G.A.G. conceived the project with input from M.Y., M.S.S. and D.T. The project was supervised by G.A.G. and M.D.M. A.Z.S., M.D.M. and E.F. analysed the Ivy GAP data and implemented the DCNN semantic segmentation model. G.A.G. and D.T. performed bioinformatics analysis. A.Z.S., G.A.G. and M.D.M. analysed the TCGA samples using the trained DCNN model. K.G.S. has prepared data from Ivy GAP Dataset for training, validation and test of the DCNN model. N.S.B. inspected all histological images in the pre-processing phase and results obtained by the models. R.J.O. and S.P. contributed to bio-specimen collections. A.Z.S. and G.A.G. wrote the manuscript and illustrated the figures. All authors contributed to manuscript editing. All authors have read and approved the manuscript.

ADDITIONAL INFORMATION

Ethics approval and consent to participate The study was performed in accordance with the Declaration of Helsinki and approved by the Southern Adelaide Clinical Human Research Ethics Committee (#286.10, HREC/17/RAH/358 and HREC/18/SAC/16). Brain tumour tissue was obtained from the SANTB. Informed consent obtained from study participants was provided in writing. Consent forms and ethical approval to enable the collection of brain tumour tissues was provided by the Southern Adelaide Clinical (SAC) Human Research Ethics Committee (#286.10, HREC/17/RAH/358 and HREC/18/SAC/16). Brain tumour tissues were obtained from patients undergoing biopsy or resection of their brain tumour at Flinders Medical Centre in Adelaide, South Australia, Australia. SANTB specimens were de-identified before use in this study and all specimens were linked with non-identifiable, comprehensive clinical information.

Consent to publish Images presented in the manuscript correspond to data derived using de-identified patient-derived samples and there are no details on individuals reported within the manuscript.

Data availability Computational source codes along with the datasets used for image segmentation and RNA-seq analysis are provided in the Supplementary Information and have been deposited at GitHub (https://github.com/amin20/GBM_WSSM).

Competing interests The authors declare no competing interests.

Funding information This work was supported by grants from the National Health and Medical Research Council of Australia [grant numbers 1067405, 1123816 to G.A.G.]; the Cure Brain Cancer Foundation [to G.A.G., R.J.O., and S.P.]; the University of South Australia and the MAWA Foundation [to G.A.G.]; the Neurosurgical Research Foundation [to R.J.O., S.P. and G.A.G.]; the Cancer Council SA Beat Cancer Project [to G.A.G.]; and the Australian Research Council [FT160100366 to G.A.G.]. A.Z.S. is supported by an Australian Government Research Training Program (RTP) Scholarship. The funders had no role in study design, data collection and analysis, interpretation of data, the decision to publish or preparation of the manuscript.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41416-021-01394-x>.

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

REFERENCES

- Hanif, F., Muzaffar, K., Perveen, K., Malhi, S. M. & Simjee, Sh. U. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pac. J. Cancer Prev.* **18**, 3–9 (2017).
- Xu, H., Chen, J., Xu, H. & Qin, Z. Geographic variations in the incidence of glioblastoma and prognostic factors predictive of overall survival in US adults from 2004–2013. *Front. Aging Neurosci.* **9**, 352 (2017).
- Tamimi, A. F. & Juweid, M. in *Glioblastoma* (ed. De Vleeschouwer S), (Exon, 2017). <https://doi.org/10.15586/codon.glioblastoma.2017.ch8>.
- Australian Institute of Health and Welfare. *Brain and other central nervous system cancers*. <https://www.aihw.gov.au/reports/cancer/brain-other-central-nervous-system-cancers/contents/table-of-contents> (2017).
- Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
- Lieberman, F. Glioblastoma update: molecular biology, diagnosis, treatment, response assessment, and translational clinical trials. *F1000Res.* **6**, 1892 (2017).
- Perrin, S. L., Samuel, M. S., Koszyca, B., Brown, M. P., Ebert, L. M., Oksdath, M. et al. Glioblastoma heterogeneity and the tumour microenvironment: implications for preclinical research and development of new treatments. *Biochem. Soc. Trans.* **47**, 625–638 (2019).
- Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
- Hambardzumyan, D. & Bergers, G. Glioblastoma: Defining tumor niches. *Trends Cancer* **1**, 252–265 (2015).
- Olar, A. & Aldape, K. D. Using the molecular classification of glioblastoma to inform personalized treatment. *J. Pathol.* **232**, 165–177 (2014).
- Vigneswaran, K., Neill, S. & Hadjipanayis, C. G. Beyond the World Health Organization grading of infiltrating gliomas: advances in the molecular genetics of glioma classification. *Ann. Transl. Med.* **3**, 95 (2015).
- Aldape, K., Brindle, K. M., Chesler, L., Chopra, R., Gajjar, A., Gilbert, M. R. et al. Challenges to curing primary brain tumours. *Nat. Rev. Clin. Oncol.* **16**, 509–520 (2019).
- Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P. et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).
- Muller, S., Kohanbash, G., Liu, S. J., Alvarado, B., Carrera, D., Bhaduri, A. et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* **18**, 234 (2017).
- Muller, S., Liu, S. J., Di Lullo, E., Malatesta, M., Pollen, A. A., Nowakowski, T. J. et al. Single-cell sequencing maps gene expression to mutational phylogenies in PDGF- and EGF-driven gliomas. *Mol. Syst. Biol.* **12**, 889 (2016).
- Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849 e821 (2019).
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
- Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–NF110 (2010).
- Yuan, J., Levitin, H. M., Frattini, V., Bush, E. C., Boyett, D. M., Samanamud, J. et al. Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med.* **10**, 57 (2018).
- Dirkse, A., Golebiewska, A., Buder, T., Nazarov, P. V., Muller, A., Poovathingal, S. et al. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat. Commun.* **10**, 1787 (2019).
- Puchalski, R. B., Shah, N., Miller, J., Dalley, R., Nomura, S. R., Yoon, J.-G. et al. An anatomic transcriptional atlas of human glioblastoma. *Science* **360**, 660–663 (2018).
- Bueno, G., Fernandez-Carrobles, M. M., Gonzalez-Lopez, L. & Deniz, O. Glomerulosclerosis identification in whole slide images using semantic segmentation. *Comput. Methods Prog. Biomed.* **184**, 105273 (2020).
- Jégou, S., Drozdal, M., Vazquez, D., Romero, A. & Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. in *Proc. IEEE conference on computer vision and pattern recognition workshops* (2017).
- Mejbri, S., Franchet, C., Reshma, I. A., Mothe, J., Brousset, P. & Faure, E. Deep Analysis of CNN Settings for New Cancer whole-slide Histological Images Segmentation: the Case of Small Training Sets. in *6th International conference on Biomedicine (BIOIMAGING 2019)*. (2019).
- Ronneberger, O., Fischer, P. & Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28 (2015).
- Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
- Lovric, M. in *Springer reference International encyclopedia of statistical science*. 1st edn. (Springer, International Encyclopedia of Statistical Science, 2011).
- Ményhárt, O., Weltz, B. & Györfy, B. MultipleTesting.com: a tool for life science researchers for multiple hypothesis testing correction. *bioRxiv* <https://doi.org/10.1101/2021.01.11.426197> (2021).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
- Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q -value. *Ann. Stat.* **31**, 2013–2035 (2003).
- Chollet, F. *keras*. <https://keras.io/>. (2015).
- Girija, S. S. *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. <https://www.tensorflow.org/> (2016).
- Creed, J. H., Gerke, T. A. & Berglund, A. E. MatSurv: Survival analysis and visualization in MATLAB. *J. Open Source Softw.* <https://doi.org/10.21105/joss.01830> (2020).
- Patil, A. CellKb Immune: a manually curated database of mammalian immune marker gene sets optimized for rapid cell type identification. *bioRxiv* <https://doi.org/10.1101/2020.12.01.389890> (2020).
- Heberle, H., Meirles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).
- Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Ebert, L., Yu, W., Gargett, T., Toubia, J., Kollis, P., Tea, M. et al. Endothelial, pericyte and tumor cell expression in glioblastoma identifies fibroblast activation protein (FAP) as an excellent target for immunotherapy. *Clin. Transl. Immunol.* <https://onlinelibrary.wiley.com/doi/full/10.1002/cti2.1191> (2020).
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. 3rd et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
- Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M. & Colinge, J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
- Karimi, D., Dou, H., Warfield, S. K. & Gholipour, A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. *Med. Image Anal.* **65**, 101759 (2020).

42. Hu, B.-G. & Dong, W.-M. A design of human-like robust AI machines in object identification. Preprint at *arXiv* <https://arxiv.org/abs/2101.02327> (2021).
43. Couderc, N. GRUBBS: Stata module to perform Grubbs' test for outliers. <https://ideas.repec.org/c/boc/bocode/s456803.html> (2007).
44. de Winter, J. C. F., Gosling, S. D. & Potter, J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol. Methods* **21**, 273–290 (2016).
45. Xu, W., Hou, Y., Hung, Y. S. & Zou, Y. A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Process.* **93**, 261–276 (2013).
46. Shirazi, A. Z., Fornaciari, E., Bagherian, N. S., Ebert, L. M., Koszyca, B. & Gomez, G. A. DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images. *Med. Biol. Eng. Comput.* **58**, 1031–1045 (2020).
47. Zadeh Shirazi, A., Fornaciari, E., McDonnell, M. D., Yaghoobi, M., Cevallos, Y., Tello-Oquendo, L. et al. The application of deep convolutional neural networks to brain cancer images: a survey. *J. Personalized Med.* **10**, 224 (2020).
48. Xiao, X., Tang, C., Xiao, S., Fu, C. & Yu, P. Enhancement of proliferation and invasion by MicroRNA-590-5p via targeting PBRM1 in clear cell renal carcinoma cells. *Oncol. Res.* **20**, 537–544 (2013).
49. Tanaka, S., Batchelor, T. T., Iafraite, A. J., Dias-Santagata, D., Borger, D. R., Elisen, L. W. et al. PIK3CA activating mutations are associated with more disseminated disease at presentation and earlier recurrence in glioblastoma. *Acta Neuropathol. Commun.* **7**, 66 (2019).
50. Weber, G. L., Parat, M. O., Binder, Z. A., Gallia, G. L. & Riggins, G. J. Abrogation of PIK3CA or PIK3R1 reduces proliferation, migration, and invasion in glioblastoma multiforme cells. *Oncotarget* **2**, 833–849 (2011).
51. Zhang, L. H., Yin, Y. H., Chen, H. Z., Feng, S. Y., Liu, J. L., Chen, L. et al. TRIM24 promotes stemness and invasiveness of glioblastoma cells via activating SOX2 expression. *Neuro Oncol* <https://doi.org/10.1093/neuonc/noaa138> (2020).
52. Mair, D. B., Ames, H. M. & Li, R. Mechanisms of invasion and motility of high-grade gliomas in the brain. *Mol. Biol. Cell* **29**, 2509–2515 (2018).
53. Juliano, J., Gil, O., Hawkins-Daarud, A., Noticewala, S., Rockne, R. C., Gallaher, J. et al. Comparative dynamics of microglial and glioma cell motility at the infiltrative margin of brain tumours. *J. R. Soc. Interface* <https://doi.org/10.1098/rsif.2017.0582> (2018).
54. Pacioni, S., D'Alessandris, Q. G., Buccarelli, M., Boe, A., Martini, M., Larocca, L. M. et al. Brain invasion along perivascular spaces by glioma cells: relationship with blood-brain barrier. *Cancers* <https://doi.org/10.3390/cancers12010018> (2019).
55. Brahm, C. G., van Linde, M. E., Enting, R. H., Schuur, M., Otten, R. H. J., Heymans, M. W. et al. The Current status of immune checkpoint inhibitors in neuro-oncology: a systematic review. *Cancers (Basel)* <https://doi.org/10.3390/cancers12030586> (2020).
56. Xiao, Z. A., Xie, D. H., Hu, P., Xia, K., Cai, F. & Pan, Q. [Functional interaction of the C-terminal of Nogo protein with connexin 26 and the expression of Nogo's mRNA in the murine inner ear]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **23**, 492–496 (2006).
57. Di Lorenzo, A., Manes, T. D., Davalos, A., Wright, P. L. & Sessa, W. C. Endothelial reticulon-4B (Nogo-B) regulates ICAM-1-mediated leukocyte transmigration and acute inflammation. *Blood* **117**, 2284–2295 (2011).
58. Jin, S. G., Ryu, H. H., Li, S. Y., Li, C. H., Lim, S. H., Jang, W. Y. et al. Nogo-A inhibits the migration and invasion of human malignant glioma U87MG cells. *Oncol. Rep.* **35**, 3395–3402 (2016).
59. Lugano, R., Ramachandran, M. & Dimberg, A. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cell Mol. Life Sci.* **77**, 1745–1770 (2020).
60. Walchli, T., Pernet, V., Weinmann, O., Shiu, J. Y., Guzik-Kornacka, A., Decrey, G. et al. Nogo-A is a negative regulator of CNS angiogenesis. *Proc. Natl Acad. Sci. USA* **110**, E1943–E1952 (2013).
61. Walchli, T., Ulmann-Schuler, A., Hintermuller, C., Meyer, E., Stampanoni, M., Carmeliet, P. et al. Nogo-A regulates vascular network architecture in the postnatal brain. *J. Cereb. Blood Flow. Metab.* **37**, 614–631 (2017).
62. Oshima, A., Tani, K., Hiroaki, Y., Fujiyoshi, Y. & Sosinsky, G. E. Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule. *Proc. Natl Acad. Sci. USA* **104**, 10034–10039 (2007).
63. Bennett, B. C., Purdy, M. D., Baker, K. A., Acharya, C., McIntire, W. E., Stevens, R. C. et al. An electrostatic mechanism for Ca(2+)-mediated regulation of gap junction channels. *Nat. Commun.* **7**, 8770 (2016).
64. Bicego, M., Beltramello, M., Melchionda, S., Carella, M., Piazza, V., Zelante, L. et al. Pathogenetic role of the deafness-related M34T mutation of Cx26. *Hum. Mol. Genet.* **15**, 2569–Cx2587 (2006).
65. Choi, S. Y., Park, H. J., Lee, K. Y., Dinh, E. H., Chang, Q., Ahmad, S. et al. Different functional consequences of two missense mutations in the GJB2 gene associated with non-syndromic hearing loss. *Hum. Mutat.* **30**, E716–E727 (2009).
66. Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y. et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature* **458**, 597–602 (2009).
67. Oshima, A., Tani, K., Toloue, M. M., Hiroaki, Y., Smock, A., Inukai, S. et al. Asymmetric configurations and N-terminal rearrangements in connexin26 gap junction channels. *J. Mol. Biol.* **405**, 724–735 (2011).
68. Berezovsky, A. D., Poisson, L. M., Cherba, D., Webb, C. P., Transou, A. D., Lemke, N. W. et al. Sox2 promotes malignancy in glioblastoma by regulating plasticity and astrocytic differentiation. *Neoplasia* **16**, 193–206 (2014). 206 e119-125.
69. Yu, S. C., Xiao, H. L., Jiang, X. F., Wang, Q. L., Li, Y., Yang, X. J. et al. Connexin 43 reverses malignant phenotypes of glioma stem cells by modulating E-cadherin. *Stem Cells* **30**, 108–120 (2012).
70. Sun, D., Jin, H., Zhang, J. & Tan, X. Integrated whole genome microarray analysis and immunohistochemical assay identifies COL11A1, GJB2 and CTRL as predictive biomarkers for pancreatic cancer. *Cancer Cell Int.* **18**, 174 (2018).
71. Zhu, T., Gao, Y. F., Chen, Y. X., Wang, Z. B., Yin, J. Y., Mao, X. Y. et al. Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. *Oncotarget* **8**, 21281–21289 (2017).
72. Shettar, A., Damineni, S., Mukherjee, G. & Kondaiah, P. Gap junction beta2 expression is negatively associated with the estrogen receptor status in breast cancer tissues and is a regulator of breast tumorigenesis. *Oncol. Rep.* **40**, 3645–3653 (2018).
73. Berens, M. E., Sood, A., Barnholtz-Sloan, J. S., Graf, J. F., Cho, S., Kim, S. et al. Multiscale, multimodal analysis of tumor heterogeneity in IDH1 mutant vs wild-type diffuse gliomas. *PLoS ONE* **14**, e0219724 (2019).
74. Bernstock, J. D., Vicario, N., Rong, L., Valdes, P. A., Choi, B. D., Chen, J. A. et al. A novel in situ multiplex immunofluorescence panel for the assessment of tumor immunopathology and response to virotherapy in pediatric glioblastoma reveals a role for checkpoint protein inhibition. *Oncimmunology* **8**, e1678921 (2019).
75. Cieremans, D., Kim, J. Y., Valencia, A., Santos, J., Bordeaux, J., Tran, T. et al. Predictive evaluation of quantitative spatial profiling of the tumor microenvironment by multiplex immunofluorescence in recurrent glioblastoma treated with PD-1 inhibitors. *J. Clin. Oncol.* **38**, e14524–e14524 (2020).
76. Sadeghi Rad, H., Monkman, J., Warkiani, M. E., Ladwa, R., O'Byrne, K., Rezaei, N. et al. Understanding the tumor microenvironment for effective immunotherapy. *Med. Res. Rev.* <https://doi.org/10.1002/med.21765> (2020).
77. Couturier, C. P., Ayyadhury, S., Le, P. U., Nadaf, J., Monlong, J., Riva, G. et al. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* **11**, 3406 (2020).
78. Jia, D., Li, S., Li, D., Xue, H., Yang, D. & Liu, Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)* **10**, 592–605 (2018).
79. Pombo Antunes, A. R., Scheyltjens, I., Duerinck, J., Neyns, B., Movahedi, K., Van Ginderachter, J. A. Understanding the glioblastoma immune microenvironment as basis for the development of new immunotherapeutic strategies. *elife* <https://doi.org/10.7554/eLife.52176> (2020).
80. Rajapakse, V. N., Herrada, S. & Lavi, O. Phenotype stability under dynamic brain-tumor environment stimuli maps glioblastoma progression in patients. *Sci. Adv.* **6**, eaaz4125 (2020).
81. Ravi, V. M., Neidert, N., Will, P., Joseph, K., Maier, J. P., Kückelhaus, J. et al. Lineage and spatial mapping of glioblastoma-associated immunity. *bioRxiv* <https://doi.org/10.1101/2020.06.01.121467> (2020).
82. Wang, Y., Mashock, M., Tong, Z., Mu, X., Chen, H., Zhou, X. et al. Changing technologies of RNA sequencing and their applications in clinical oncology. *Front. Oncol.* **10**, 447 (2020).
83. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
84. Zanotelli, V. R., Leutenegger, M., Lun, X. K., Georgi, F., de Souza, N. & Bodenmiller, B. A quantitative analysis of the interplay of environment, neighborhood, and cell state in 3D spheroids. *Mol. Syst. Biol.* **16**, e9798 (2020).
85. Krieger, T. G., Tirier, S. M., Park, J., Jechow, K., Eisemann, T., Peterziel, H. et al. Modeling glioblastoma invasion using human brain organoids and single-cell transcriptomics. *Neuro Oncol.* **22**, 1138–1149 (2020).
86. Wu, J. I. & Wang, L. H. Emerging roles of gap junction proteins connexins in cancer metastasis, chemoresistance and clinical application. *J. Biomed. Sci.* **26**, 8 (2019).
87. Yang, Y. S. & Strittmatter, S. M. The reticulons: a family of proteins with diverse functions. *Genome Biol.* **8**, 234 (2007).
88. Pradhan, L. K. & Das, S. K. The regulatory role of reticulons in neurodegeneration: insights underpinning therapeutic potential for neurodegenerative diseases. *Cell Mol. Neurobiol.* <https://doi.org/10.1007/s10071-020-00893-4> (2020).
89. Tagami, S., Eguchi, Y., Kinoshita, M., Takeda, M. & Tsujimoto, Y. A novel protein, RTN-X5, interacts with both Bcl-XL and Bcl-2 on endoplasmic reticulum and reduces their anti-apoptotic activity. *Oncogene* **19**, 5736–5746 (2000).

90. Zhao, H., Su, W., Zhu, C., Zeng, T., Yang, S., Wu, W. et al. Cell fate regulation by reticulon-4 in human prostate cancers. *J. Cell Physiol.* **234**, 10372–10385 (2019).
91. Barnes, J. M., Kaushik, S., Bainer, R. O., Sa, J. K., Woods, E. C., Kai, F. et al. A tension-mediated glyocalyx-integrin feedback loop promotes mesenchymal-like glioblastoma. *Nat. Cell Biol.* **20**, 1203–1214 (2018).
92. Yan, J., Zhou, X., Guo, J. J., Mao, L., Wang, Y. J., Sun, J. et al. Nogo-66 inhibits adhesion and migration of microglia via GTPase Rho pathway in vitro. *J. Neurochem.* **120**, 721–731 (2012).
93. Fang, Y., Yan, J., Li, C., Zhou, X., Yao, L., Pang, T. et al. The Nogo/Nogo receptor (NgR) signal is involved in neuroinflammation through the regulation of microglial inflammatory activation. *J. Biol. Chem.* **290**, 28901–28914 (2015).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021