nature research

Corresponding author(s): Timothy F. Cloughesy, Robert M. Prins

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Data was collected with Microsoft Excel or R on a secure server.
Data analysis	Statistical analysis was performed with SAS version 9.2, GraphPad Prism version 6.01, R version 3.4.3 or higher, HISAT2 version 2.1.0, CIBERSORT version 1.06. HALO image analysis platform version 2.1. R packages: ConsensusClusterPlus version 1.46.0, corrplot version 0.84, destiny version 2.12.0, flowCore version 1.48.0, FlowSOM version 1.14.0, ggplot2 version 3.1.0, glmnet version 2.0-16, GSVA version 1.30.0, limma version 3.38.2, lme4 version 1.1-19, mice version 3.3.0, multcomp version 1.4-8, premessa version 0.1.8, RColorBrewer version 1.1-2, readxl version 1.1.0, survival version 2.43-1, survminer version 0.4.3, Rtsne version 0.15.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA sequencing data is available in the Gene Expression Omnibus (GEO) under the accession code GSE121810. Source data for Figure 2B and Extended Data Figures 3, 4 and 5 are provided with the paper. The remainder of data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Based on our preliminary data, the mean of tumor infiltrating lymphocyte density was estimated to be 0.4 T cells per nucleated cell (standard deviation = 0.5) in the control group. Fifteen patients per group was deemed sufficient to achieve 85% power to detect an increase of 0.5 in tumor infiltrating lymphocyte density comparing the neoadjuvant group against the adjuvant-only group at an alpha of 0.05 (one sided) using a two-sample t test.
Data exclusions	Two patients were replaced according to the study protocol based on insufficient histological evidence of glioblastoma. Three other patients withdrew consent prior to receiving study intervention, but all were included in the intention-to-treat efficacy analysis.
Replication	For elastic net Cox regression, the value of lambda was tuned using 5-fold cross-validation; this was repeated by utilizing different starting seeds to ensure that identified nonzero coefficients were reproducible. For multiple imputation of missing variables, different starting seeds were also utilized and checked using density plots. Multiplex staining was performed in one standardized run per patient. For each staining run, two sequential slides were used as duplicates for each patient. Computational analysis was performed two or more times per sample. Mass cytometry, T cell receptor sequencing, Nanostring and bulk tumor RNA sequencing were performed once due to clinical sample availability.
Randomization	Subjects were randomized at enrollment into either the neoadjuvant group or the adjuvant-only group.
Blinding	Neither investigators nor patients were blinded to allocated treatment arm in this multi-institutional randomized study. As an investigator initiated pilot study, there was inadequate funding available for blinding. However, at no time before or during the study did the investigators or the patients consider an advantage to either arm as the study was principally designed to require a control for the immune monitoring performed in the perioperative and post operative period. Clinical equipoise was presumed maintained as both treatment arms received adjuvant study drug post operatively. The study evaluations were not analyzed until 4 months after the last patient was registered . Efficacy results were shared with the clinical investigators in June 2018 which is the time of the clinical data lock.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

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

Antibodies

Antibodies used

Antibodies used for mass cytometry are as follows:
Target (clone) Vendor Catalog number
CD45RA (HI100) - 169Tm Fluidigm 3169008B
CD69 (FN50) - 144Nd Fluidigm 3144018B
CD4 (RPA-T4) - 145Nd Fluidigm 3145001B
CD25 (2A3) - 143Nd (custom)
CD19 (HIB19) - 142Nd Fluidigm 3142001B
CD274 (29E.2A3) - 148Nd Fluidigm 3148017B
CD14 (M5E2) - 151Eu Fluidigm 3151009B
CD279 (EH12.2H7) - 155Gd Fluidigm 3155009B
CD33 (WM53) - 158Gd Fluidigm 3158001B
CD8a (RPA-T8) - 146Nd Fluidigm 3146001B
CD15 (W6D3) - 164Dy Fluidigm 3164001B
CD16 (3G8) - 165Ho Fluidigm 3165001B
CD3 (UCHT1) - 154Sm Fluidigm 3154003B
CD11b (ICRF44) - 209Bi Fluidigm 3209003B
HLA-DR (L243) - 174Yb Fluidigm 3174001B
CD127 (A019D5) - 176Yb Fluidigm 3176004B
CD38 (HIT2) - 172Yb Fluidigm 3172007B
CD206 (15-2) - 168Er Fluidigm 3168008B
TIM-3 (F38-2E2) - 153Eu Fluidigm 3153008B
CD223 (LAG-3) - 150Nd (custom)
CD152 (14D3) - 161Dy Fluidigm 3161004B
CD11c (Bu15) - 159Tb Fluidigm 3159001B
CD27 (O323) - 167Er Fluidigm 3167002B
CD56 (NCAM 16.2) - 163Dy Fluidigm 3163007B

Validation

According to the manufacturer's website, each lot of conjugated antibodies is quality control tested by mass cytometry analysis of stained cells using the appropriate positive and negative cell staining and/or activation controls. Specifically, CD45RA (HI100) was validated by Fluidigm on human PBMCs; CD56 (NCAM 16.2) was validated by Fluidigm on human PBMCs - according to the manufacturer's website, human PBMCs were incubated for 6 hours in media alone or with PMA and ionomycin in the presence of monensin and brefeldin A. Cells were then fixed, permeabilized and stained with 154Sm-anti-CD45 (HI30) and 144Nd-anti-CD69 (FN50). Anti CD4-145Nd (RPA-T4), CD19 (HIB19), CD274 (29E.2A3), CD14 (M5E2), CD279 (EH12.2H7), CD33 (WM53), CD8a (RPA-T8), CD15 (W6D3), CD16 (3G8), CD3 (UCHT1), CD11b (ICRF44), HLA-DR (L243), CD127 (A019D5), CD38 (HIT2), CD206 (15-2), TIM-3 (F38-2E2), CD152 (14D3), CD11c (Bu15), CD27 (O323) were also validated by the manufacturer on human PBMCs; Custom conjugated antibodies CD25 (2A3) and CD223 (LAG-3) were validated on human PBMCs at the UCLA Jonsson Comprehensive Cancer Center Flow Cytometry Core.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Patients were aged ≥18 years, male and female, of all ethnicities, with recurrent World Health Organization grade IV malignant glioma that were candidates for surgical debulking. Key eligibility criteria included Karnofsky performance status ≥70, previous first line therapy with at least radiotherapy, first or second relapse with with unequivocal evidence of tumor progression, adequate organ function and absence of previous anti-angiogenic or anti-vascular endothelial growth factor agents.

Recruitment

Participants were recruited by site-specific co-investigators at their respective institutions based on protocol eligibility criteria and verified by the study coordinator. Self-selection bias may be present and may affect survival; however, patients randomized to the control arm of this study exhibited overall survival similar to that of other patients with recurrent GM.

Magnetic resonance imaging

Experimental design

Design type

N/A - The current study did not involve "functional" MRI, only anatomic MRI for visualizing and quantifying treatment response.

Design specifications

N/A - The current study involved acquisition of structural (anatomic) MRI at screening, prior to surgery, after surgery, and every treatment cycle until tumor progression, plus subsequent "off treatment" MRI scans until patient death.

Behavioral performance measures

N/A

Acquisition

Imaging type(s): Structural/Anatomic

Field strength: 1.5T & 3T

Sequence & imaging parameters: Parameter matched pre- and post-contrast (gadopentate dimegumine, 0.1 mmol/kg BW) T1-weighted images were acquired using either a 2D fast spin-echo or 3D gradient echo (MPRAGE, SPGR, or IR-SPGR) sequence (repetition time (msec)/echo time (msec)/ inversion time (msec) = 400–3209/3.6– 21.9/0–1238; slice thickness = 1–6.5 mm; intersection gap = 0–2.5 mm; number of averages = 1-2; matrix size = 176–512 x 256–512; and field of view = 24–25.6 cm). 2D T2-weighted fast spin-echo and fluid-attenuated inversion-recovery (FLAIR) images were also acquired but not used in the current study. All on trial scans were compliant with the consensus recommendations for the international standardized brain tumor imaging protocol (Ellingson et al., Neuro Oncol 2015; 17(9): 1188-98.)

Area of acquisition: Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: Linear registration was performed between all images (T2, FLAIR, pre-contrast T1, post-contrast T1) to post-contrast T1-weighted images at screening using a 12–degree-of-freedom linear transformation and a correlation coefficient cost function in FSL (FLIRT; FMRIB Soft- ware Library, Oxford, England; http:// www.fmrib.ox.ac.uk/fsl/).

Estimates of tumor volume were performed using contrast-enhanced T1-weighted digital subtraction maps to exclude areas of post-surgical blood products or other sources of T1 shortening. T1 subtraction maps were created by first performing linear registration as described above. Next, Gaussian normalization of image intensity was performed for both nonenhanced and contrast enhanced T1-weighted images using custom c-code courtesy of the National Institutes of Health Magnetoencaphalography Core Facility (3dNormalize; NIMH MEG Core, Bethesda, MD; kurage.nimh.nih.gov/ meglab/Med/3dNormalize), which normalizes image intensity by dividing each voxel by the standard deviation of the image intensity from the whole brain [SNor(x,y,z) = S (x,y,z)/stdWB], where S is raw image signal intensity, Nor is normalized, x,y,z are voxel coordinates, and stdWB is whole brain standard deviation. Next, voxel-by-voxel subtraction between normalized nonenhanced and contrast-enhanced T1-weighted images was performed using the Analysis of Functional NeuroImages software package (AFNI; 3dcalc; https://afni.nimh.nih.gov/). Image voxels with a positive (greater than zero) before-to-after change in normalized contrast enhancement signal intensity (ie, voxels increasing in MR signal after contrast agent administration) within T2-weighted FLAIR hyperintense regions were isolated to create the final T1 subtraction maps in order to exclude large vessels and other hyperintense regions outside the primary tumor area. Estimates of tumor volume included areas of contrast enhancement on T1 subtraction maps. Initial segmentation was performed automatically and final segmented volumes were edited by by an experienced independent observer with more than 10 years of experience to exclude large vessels and any obvious non-tumor regions.

Normalization: Intensity normalization was performed using custom c-code courtesy of the National Institutes of Health Magnetoencaphalography Core Facility (3dNormalize; NIMH MEG Core, Bethesda, MD; kurage.nimh.nih.gov/ meglab/Med/3dNormalize), which normalizes image intensity by dividing each voxel by the standard deviation of the image intensity from the whole brain [SNor(x,y,z) = S (x,y,z)/stdWB], where S is raw image signal intensity, Nor is normalized, x,y,z are voxel coordinates, and stdWB is whole brain standard deviation.

Normalization template: Images were not normalized/registered to a standard template space. All images were registered to the patient-specific screening MRI exam as mentioned above.

Noise and artifact removal: No noise or artifact removal was performed.

Volume censoring: N/A

Statistical modeling & inference

Model type and settings: N/A

Effect(s) tested: N/A

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s): Areas of contrast enhancing tumor burden

Statistic type for inference (See Eklund et al. 2016): N/A

Correction: N/A

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

Given the relatively low number of patients and large number of possible parameters with high correlation, we used an elastic net regularized regression for variable selection. Both clinical (including IDH mutation status, age, sex, MGMT methylation status, number of prior relapses, Karnofsky performance status, steroid dose in milligrams at time of registration) and laboratory data (including mass cytometry cluster percentages, T cell receptor overlap, tumor T cell density, expanded tumor-associated T cell clones, interferon-γ-related signature scores, presence or absence of inducible PD-L1 expression on multiplex immunofluorescence) were considered potential covariates. We used 5-fold cross-validation to obtain the value of λ that gave the minimum mean cross-validated error and determined the corresponding coefficients for each covariate. Variables with nonzero coefficients were then checked for collinearity and subsequently fitted into a Cox proportional hazards model, forcing in age and sex. Peripheral blood T cell receptor clonality and tumor infiltrating lymphocyte fractions were standardized before fitting into the Cox regression due to their wide ranges. As the tumor infiltrating T cell fraction was a measurement taken at the time of surgery, we ran both a two-sample t-test ($P = 0.52$, $t = 0.65$, $df = 27.6$) and Wilcoxon rank sum test ($P = 0.41$, $W = 133$) to ensure that there was no significant imbalance between the two groups at alpha = 0.05 that could potentially introduce confounding into the model. We performed a Cox-Snell residual plot, which did not suggest any lack-of-fit for the multivariate Cox model.