Tracking tumour evolution in glioma through liquid biopsies of cerebrospinal fluid

Alexandra M. Miller1,13, Ronak H. Shah2,3,12,13, Elena I. Pentsova1,13, Maryam Pourmaleki4, Samuel Briggs5, Natalie Distefano5, Yoyyun Zheng2,3, Anna Skakodub1, Smrutiben A. Mehta4, Carl Campos4, Wan-Ying Hsieh4, S. Duygu Selekulu3, Lilan Ling3, Fanli Meng3, Xiaohong Jing3, Aliaksandra Samoila6, Tejus A. Bale2, Dana W. Y. Tsui2,3, Christian Grommes1, Agnes Viale3, Mark M. Souweidane3,7, Viviane Tabar2, Cameron W. Brennan3, Anne S. Reiner3, Marc Rosenblum2, Katherine S. Panageas8, Lisa M. DeAngelis1, Robert J. Young9, Michael F. Berger2,3,4,10 & Ingo K. Mellinghoff1,4,11

Diffuse gliomas are the most common malignant brain tumours in adults and include glioblastomas and World Health Organization (WHO) grade II and grade III tumours (sometimes referred to as lower-grade gliomas). Genetic tumour profiling is used to classify disease and guide therapy1,2, but involves brain surgery for tissue collection; repeated tumour biopsies may be necessary for accurate genotyping over the course of the disease3–10. While the detection of circulating tumour DNA (ctDNA) in the blood of patients with primary brain tumours remains challenging11,12, sequencing of ctDNA from the cerebrospinal fluid (CSF) may provide an alternative way to genotype gliomas with lower morbidity and cost13,14. We therefore evaluated the representation of the glioma genome in CSF from 85 patients with gliomas who underwent a lumbar puncture because they showed neurological signs or symptoms. Here we show that tumour-derived DNA was detected in CSF from 42 out of 85 patients (49.4%) and was associated with disease burden and adverse outcome. The genomic landscape of glioma in the CSF included a broad spectrum of genetic alterations and closely resembled the genomes of tumour biopsies. Alterations that occur early during tumorigenesis, such as co-deletion of chromosome arms 1p and 19q (1p/19q co-deletion) and mutations in the metabolic genes isocitrate dehydrogenase 1 (IDH1) or IDH215, were shared in all matched ctDNA-positive CSF–tumour pairs, whereas growth factor receptor signalling pathways showed considerable evolution. The ability to monitor the evolution of the glioma genome through a minimally invasive technique could advance the clinical development and use of genotype-directed therapies for glioma, one of the most aggressive human cancers.

Eighty-five adults with glioma underwent collection of CSF as part of their clinical evaluation for neurological signs or symptoms. Diagnoses included WHO grade IV glioblastoma (46/85; 54%), WHO grade III glioma (26/85; 31%), and WHO grade II glioma (13/85; 15%). The diagnosis of glioma had been established by prior tumour biopsy or resection. All patients had received treatment for glioma before CSF collection, including surgery (85/85; 100%), radiation (84/85; 99%), and at least one systemic tumour-directed chemotherapy (81/85; 95%; Supplementary Table 1). The median duration of disease before CSF collection was 355 days for IDH wild-type glioblastomas (GBMs) (those with wild-type alleles for both IDH1 and IDH2), 473 days for IDH wild-type lower-grade gliomas (LGGs), and 2,077 days for IDH mutant LGGs (Extended Data Fig. 1). Indications for lumbar puncture included signs or symptoms of CNS infection, leptomeningeal tumour spread, or increased intracranial pressure. All samples were analysed using MSK-IMPACT, a custom FDA-authorized next-generation sequencing-based tumour sequencing assay15. We detected at least one tumour-derived genetic alteration in the CSF from 42 out of 85 patients with glioma (49.4%). By contrast, we did not detect any oncogenic variants in CSF from seven individuals with non-malignant neurological conditions (Extended Data Table 1).

In our patients with glioma, several radiographic findings were associated with shedding of tumour DNA into the CSF, including tumour progression (P = 0.0005, Fisher’s exact test), tumour burden (P = 0.0000017, Wilcoxon rank-sum test), and spread of tumour towards the ventricular system or subarachnoid space (P = 0.02, Fisher’s exact test; Table 1, Extended Data Fig. 2). The latter finding is reminiscent of a prior study that collected CSF during surgery, in which tumour DNA was more commonly detected in CSF from patients with tumours abutting a CSF reservoir or cortical surface16. The presence of tumour DNA in the CSF was associated with shorter survival following CSF collection (Extended Data Fig. 3). In a multivariable analysis, the presence of CSF ctDNA remained a statistically significant prognostic factor, even after adjustment for per cent extent of resection at original diagnosis, tumour burden at CSF collection, and IDH status (Extended Data Table 2, Supplementary Table 2). Subjects who had ctDNA in their CSF experienced a fourfold higher risk of death than subjects who did not (P = 0.000024315). We found no significant association between ctDNA-positive CSF and glioma grade, disease duration, or prior therapy. Most patients with ctDNA-positive CSF (35/39, 90%) did not have detectable malignant cells in the CSF as assessed by standard CSF cytopathologic analysis.

Certain genetic alterations occur at the earliest stages of glioma development. These alterations are viewed as ‘truncal’ events during tumour evolution and are used to define prognostically distinct subtypes of LGG16. For example, gliomas with IDH mutation and 1p/19q co-deletion typically also harbour mutations in the telomerase reverse transcriptase (TERT) promoter. By contrast, gliomas with mutations in IDH and TP53 (and without 1p/19q co-deletion) often contain alterations in the gene encoding α-thalassemia/mental-retardation-syndrome-X-linked (ATRX). Both TERT promoter mutations and ATRX gene alterations promote telomerase maintenance, and they are mutually exclusive17. We examined whether these combinations of genetic alterations (‘LGG signatures’) could be detected in the CSF and matched the signature of the original tumour. We sequenced all available tumour biopsies from patients with positive CSF ctDNA (36/42 patients). These included ten LGGs, twenty GBMs, and six tumours with DNA hypermutation in tissue or CSF. In all ten (100%) of the LGG patients without DNA
hypermutation, the combination of genetic alterations that defined the LGG subtype were congruent between the CSF and tumour (Extended Data Fig. 4, patients 1–10). Mutations were also shared between the CSF and tumour in all twenty (100%) patients with GBMs without DNA hypermutation (Extended Data Fig. 4, patients 11–30).

Further analysis of CSF ctDNA revealed a broad spectrum of protein-coding mutations, copy number alterations (CNAs), promoter mutations, and structural rearrangements. The most commonly observed alterations included mutations within the TERT promoter, the protein-coding regions of TP53, and the catalytic domain of IDH1 as well as deletions of CDKN2A and CDKN2B, amplifications of the gene for epidermal growth factor receptor (EGFR), and the in-frame EGFR-variant III deletion (Fig. 1a).

CSF ctDNA from 5 out of 42 patients with glioma (12%; Fig. 1a, asterisks) showed a markedly higher mutation rate and harboured a signature (G:C→A:T transitions) that suggests prior exposure to alkylating agents18–20 (Extended Data Fig. 5). All five patients had received temozolomide before CSF collection. The median mutation rate in CSF ctDNA was 4.90 mutations per Mb. This rate is higher than the reported mutation rate in The Cancer Genome Atlas (TCGA) for the LGG exome (0.7–0.8 per Mb)5 or the GBM exome (2.2 per Mb)21, but not different from the mutation rate in our glioma tumour biopsies previously sequenced with the same gene panel (Fig. 1b, MSK-IMPACT tissue cohort; 4.46 per Mb). This observation may be partially related to the more advanced disease of our patients and the enrichment of frequently mutated genes in MSK-IMPACT. Overall, the frequency of genetic alterations in CSF ctDNA mirrored the distribution of genetic alterations in glioma biopsies taken at the Memorial Sloan Kettering (MSK) Cancer Center (Fig. 1c).

To determine whether mutations detected in the CSF were also present in the plasma, we sequenced plasma cell-free DNA (cfDNA) from 19 patients with gliomas who were positive for CSF ctDNA using a high-sensitivity capture-based next-generation sequencing assay (average raw sequence coverage >18,000×). This assay, while more targeted than MSK-IMPACT, encompassed at least one CSF mutation in each of the 19 patients, and 211 in total. We detected no mutations in plasma cfDNA from 16 out of 19 patients (84%; Supplementary Table 3). A total of 35 mutations were detected in the remaining three patients, including 32 from a single patient with a hypermutated tumour. The average variant allele fraction of these 35 mutations was very low in plasma (0.58%), compared to 23.96% in the CSF (Wilcoxon rank-sum test comparing allele fractions, \( P = 5.821 \times 10^{-11} \); Supplementary Table 4). All three patients with positive plasma ctDNA had radiographic evidence of disseminated disease within the CNS and one of these patients (patient 11) later developed metastatic GBM (with bone and lymph node involvement). These data show that CSF is a more frequent reservoir of tumour DNA than plasma in patients with glioma. It also suggests that tumour DNA is shed from the tumour directly into the CSF, rather than reaching it indirectly through the plasma.

Sequential tumour biopsies in patients with glioma have uncovered considerable evolution of the cancer genome during the course of the disease, with only 33–73% of genetic alterations being shared between
initial and recurrent tumours from the same patient. In our 30 CSF–tumour pairs without DNA hypermutation, the percentage of shared mutations varied considerably across samples (0–100%; Fig. 2a), with a median 81.7% shared mutation rate. In patients with hypermutated tumours, shared mutations were considerably less common (3–49%; Fig. 2b), with a median 19.6% shared mutation rate (Supplementary Table 5). The majority of mutations that were clonal in tissue were also present in CSF, even in patients with DNA hypermutation. By contrast, subclonal mutations were less likely to be shared in CSF (Fig. 2c).

We also examined the concordance between contemporaneously collected CSF replicates and tumour–CSF pairs. Six patients in our study underwent placement of a ventriculo-peritoneal shunt for the treatment of hydrocephalus within three weeks of the initial lumbar puncture. Despite being collected from different anatomical locations within the CNS, the genomic profiles of these CSF replicates were highly concordant (Fig. 2d), even in patients with DNA hypermutation (Extended Data Fig. 5). Five patients in our study underwent a tumour resection within three weeks of CSF collection, and these patients showed near-identical genetic profiles in CSF and tumour tissue (Fig. 2e).

As the interval between tumour and CSF collection increased, we observed greater divergence of the genetic profiles, in particular for genes encoding members of growth factor signalling pathways (Fig. 2f, Extended Data Fig. 7). This discordance followed a pattern of convergent evolution, with later samples showing other genetic alterations within the same gene or related signalling pathway. For example, the initial tumour biopsy from patient 25 harboured a high-level EGFR amplification and EGFR missense mutation, whereas a later CSF sample revealed amplification and mutation of the platelet-derived growth factor receptor alpha gene (PDGFRA) without evidence of the original EGFR alteration (Fig. 2g). Similarly, the initial tumour profile of patient 28 showed an activating mutation in PIK3CA (E545K), whereas a later tumour biopsy showed amplifications of MET and PDGFRA and a subsequent CSF sample retained the MET amplification and acquired a MYC amplification while losing the PDGFRA amplification (Extended Data Fig. 8a). In LGGs, truncal alterations (IDH1, TP53 and ATRX) persisted throughout the disease course, but later samples contained additional mutations in glioma core pathways (Extended Data Fig. 8b), a pattern consistent with branched evolution of the LGG genome.

Our study shows that tumour-derived DNA in CSF from patients with glioma provides a comprehensive and genetically faithful representation of the tumour genome at the time of the CSF collection. The frequency and type of alterations in the CSF closely resembled the genomic landscape of diffuse glioma, as defined in large oncogenic studies, and we observed high concordance between CSF and tumour DNA in patients who underwent both lumbar puncture and neurosurgical tumour resection within a few weeks. Any divergence between CSF and tumour genomes within the same patient followed the pattern of genomic evolution that has been reported in studies of sequential tumour biopsies. This evolution is characterized by the persistence of truncal genetic alterations (IDH1, 1p/19q codeletion, TP53, TERT, ATRX) and by convergent and branched evolution within glioma core pathways, in particular growth factor receptor pathways.
Fig. 2 | CSF ctDNA documents evolution of the glioma genome. 

a, b, Frequency of shared versus tissue-only or CSF-only mutations in matched tumour tissue–CSF sample pairs. a, Results for patients without DNA hypermutation (n = 30). b, Results for patients with DNA hypermutation (n = 6). Insets, aggregate number of mutations for each cohort. Red, shared; blue, tissue-only; teal, CSF-only. c, Fraction of shared versus tissue-only mutations within clonal and subclonal tumour mutations. Top row: results from tumour–CSF pairs without DNA hypermutation (non-HM). Bottom row: tumour–CSF pairs with DNA hypermutation (HM). Shared, red; tissue-only, blue; CSF-only, teal. 

Performing a lumbar puncture in patients with a brain tumour is usually safe and is done routinely for certain brain tumours as part of the staging criteria (for example, CNS lymphoma, medulloblastoma, germ cell tumours). In patients who have a tumour that cannot be approached surgically, a lumbar puncture offers an opportunity to obtain a molecular signature and potentially a definitive diagnosis. In patients in whom tumours recur, a lumbar puncture is a simpler and safer procedure than a second craniotomy for another tumour collected within a three-week period. Tumour samples were collected via biopsy and CSF collections were acquired via lumbar puncture. Heat maps display the variant allele frequencies of all the mutations present in either sample. 

f, Divergence of mutations in growth factor receptor pathways. Shown is the presence (blue) versus absence (white) of selected mutations in matched CSF–tumour pairs (n = 30, non-hypermutated). MSK cancer cohort; n = 533. g, Representative example for convergent evolution. Shown is the disease course of patient 25, who has GBM. MRIs (T2-weighted fluid-attenuated inversion recovery (FLAIR)) are shown from the initial tumour resection (left) and at the time of CSF collection (right). The CSF sample collected at recurrence showed a new PDGFRA amplification and mutation and loss of the previously detected EGFR amplification and EGFR (G719C) mutation (copy number plots shown). Diamond, tumour sample profiled; circle, CSF sample profiled. Heat map variant allelic frequency scale shown.
sample. However, shedding of tumour DNA into the CSF does not appear to be a universal property of diffuse glioma, even in previously treated patients. Our data suggest that the presence of ctDNA in the CSF may be an early indicator of progression in glioma. Further studies are needed to determine when tumour-derived DNA first becomes detectable in the CSF. By demonstrating the accuracy and frequency with which important mutational changes can be detected through CSF profiling, our work provides a framework for such studies.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0882-3.

Received: 7 June 2018; Accepted: 17 December 2018;
Published online 23 January 2019.

1. Louis, D. N. et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature Genet.* **455**, 1061–1068 (2008).
3. Suzuki, H. et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).
4. Johnson, B. E. et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
5. Aihara, K. et al. Genetic and epigenetic stability of oligodendrogliomas at recurrence. *Acta Neuropathol. Commun.* **5**, 18 (2017).
6. Bai, H. et al. Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nat. Genet.* **48**, 59–66 (2016).
7. Kim, H. et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25**, 316–327 (2015).
8. Kim, J. et al. Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell* **28**, 318–328 (2015).
9. Wang, J. et al. Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776 (2016).
10. Lee, J. K. et al. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat. Genet.* **49**, 594–599 (2017).
11. Merker, J. D. et al. Circulating tumor DNA analysis in patients with cancer: American Society of Clinical Oncology and College of American Pathologists joint review. *J. Clin. Oncol.* **36**, 1631–1641 (2018).
12. Bettegowda, C. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
13. Pentsova, E. I. et al. Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. *J. Clin. Oncol.* **34**, 2404–2415 (2016).
14. Wang, Y. et al. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc. Natl Acad. Sci. USA* **112**, 9704–9709 (2015).
15. Cheng, D. T. et al. Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).
16. Brat, D. J. et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
17. Eckel-Passow, J. E. et al. Glioma groups based on IDH, 1p/19q, and TERT promoter mutations in tumors. *N. Engl. J. Med.* **372**, 2499–2508 (2015).
18. Tomita-Mitchell, A. et al. Mismatch repair deficient human cells: spontaneous and MNNG-induced mutational spectra in the HPRT gene. *Mutat. Res.** **450**, 125–138 (2000).
19. Hunter, C. et al. A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res.* **66**, 3987–3991 (2006).
20. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
21. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
22. Parsons, D. W. et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).

**Acknowledgements**

This research was supported by the National Brain Tumor Society Defeat GBM Initiative (I.K.M.), the National Institutes of Health (1 R35 NS105109 01 and P30CA008748), Cycle of Survival (I.K.M.), the American Brain Tumor Association (E.I.P.), the Marie Josée and Henry R. Kravis Center for Molecular Oncology, the MSK Brain Tumor Center, and the MSK Neuro-Oncology Research in Translation (NORTH) Program.

**Reviewer information**

*Nature* thanks A. Bardelli, O. Rueda, M. Taylor and R. Verhaak for their contribution to the peer review of this work.

**Author contributions**

A.M.M., R.H.S., E.I.P., L.M.D., R.J.Y., M.F.B. and I.K.M. conceived and designed the study. A.M.M., R.H.S., E.I.P., R.J.Y., M.F.B. and I.K.M. collected and assembled the data. A.M.M., R.H.S., E.I.P., M.P., S.B., N.D., A. Skakodub, S.D.S., L.L., F.M., X.J., C.G., A.V., M.M.S., V.T., C.W.B., M.R., R.J.Y., M.F.B. and I.K.M. were responsible for provision of the study materials and the patients. A.M.M., R.H.S., E.I.P., Y.Z., A.S.R., K.S.P., R.J.Y., M.F.B. and I.K.M. analysed and interpreted the data. M.P., C.C., S.A.M., A. Samoila and F.M. processed the CSF and blood samples. A.M.M., R.H.S., E.I.P., W.-Y.H., T.A.B., A.V., L.M.D., K.S.P., R.J.Y., M.F.B. and I.K.M. provided administrative, material and technical support. A.M.M., R.H.S., E.I.P., D.W.Y.T., C.G., L.M.D., K.S.P., R.J.Y., M.F.B. and I.K.M. wrote the manuscript. All authors approved the manuscript.

**Competing interests**

E.I.P. reports advisory roles with AstraZeneca, V.T. is a founder of Blue Rock Therapeutics. K.S.P. reports stock ownership in Pfizer. L.M.D. reports advisory roles for Sapience Therapeutics, Togacan, BTG International, Roche, and Syndax. R.J.Y. reports research funding from Agios and advisory roles with Icon plc, NordicNeuroLab, and Puma Biotechnology. M.F.B. reports advisory roles with Roche and research funding from Illumina. I.K.M. reports research funding from General Electric, Amgen, and Lilly; advisory roles with Agios, Puma Biotechnology, and DebioPharm Group; and honoraria from Roche for a presentation.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0882-3.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0882-3.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.F.B. or I.K.M.

**Publisher’s note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

Patients. Our study includes CSF and tumour samples from 85 patients who were treated for glioma at Memorial Sloan Kettering (MSK) Cancer Center from January 2015 to April 2017. All patients had received treatment for glioma before CSF collection, including surgery (85/85; 100%), radiation (84/85; 99%), and at least one systemic tumour-directed chemotherapy (81/85 95%). For 19/85 patients, we isolated ctDNA from plasma within four weeks of CSF collection to compare sequencing results between blood and CSF. We used a previously published sequencing cohort that included 553 MSK glioma samples from 512 patients23,24 to benchmark our CSF findings in patients with glioma against a larger number of glioma tumour biopsies sequenced with the same platform. All patients signed statements of informed consent under protocols approved by the MSK Institutional Review Board.

Magnetic resonance imaging. All patients underwent brain MRIs as indicated by standard of care with standard sequences including axial T1-weighted, T2-weighted FLAIR and contrast T1-weighted images. Brain MRIs before and directly after the initial resection were reviewed by an experienced neuroradiologist without knowledge of the CSF ctDNA results. To calculate the extent of resection, the entire lesion was manually segmented using Aquarius iNtuition (Tera Reon) using axial contrast-enhanced T1-weighted images for enhancing lesions and axial FLAIR images for non-enhancing lesions. The process was repeated for the postoperative scans, with the volumes reported in cm³. The extent of resection was calculated for 63 patients with median 6 days between scans (range, 1–45). The remaining 22 patients were missing preoperative and/or postoperative scans. Brain MRIs that were performed closest to the CSF collection date were also reviewed by a neuroradiologist without knowledge of the CSF ctDNA results. Standard response assessment in neuro-oncology (RANO) criteria were applied to determine tumour burden (SPD) and radiographic progression25. The presence or absence of radiographic signs of tumour spread to subependymal, pial and leptomeningeal sites was also documented. In a binary manner, subependymal spread was determined to be present when the edge of the enhancing lesion extended to the subependymal margin of any ventricle. Similarly, pial spread was determined to be present when the edge of the enhancing lesion extended to any pial surface of the brain, and leptomeningeal spread when the enhancing lesion involved any subarachnoid space (for example, a sulcus or fissure).

CSF collection. CSF collections were performed as part of standard of care. CSF was collected by lumbar puncture in 82/85 (96%) patients. Additional CSF samples were collected during VP shunt placement or intraoperatively.

Isolation of ctDNA. Freshly collected CSF was stored at 4°C until centrifugation (1,400 r.p.m., 5 min, 4°C). CSF supernatants were transferred into cryotubes and stored in 3.5–ml aliquots at −80°C. Immediately before use, CSF (3.5 ml) was thawed in a waterbath at 37°C. The QIAasympom SY SP system (QIAGEN) was used to thawed in a waterbath at 37 °C. The QIAsymphony SP system (QIAGEN) was used to benchmark our CSF findings in patients with glioma against a larger number of glioma tumour biopsies sequenced with the same platform. All patients signed statements of informed consent under protocols approved by the MSK Institutional Review Board.

Comparison of genomic alterations in CSF and tumour. For concordance analysis of matched tumour–CSF pairs (n = 36), mutations were designated as either ‘called’ or ‘present’. Called mutations were independently detected without any prior knowledge using clinically validated filters (total depth ≥ 20, variant allele frequency ≥ 8 and variant allele frequency ≥ 2% for known mutational hotspots; total depth ≥ 20, variant allele depth ≥ 10 and variant allele frequency ≥ 5% for novel mutations). For all mutations called in either the tumour or the CSF, a secondary mutation analysis was performed for all other samples collected from the same patient, in which less stringent criteria were applied to detect the full compilation of SNVs. Mutations were called as present if variant allele depth ≥ 1 and variant allele frequency ≥ 1%. If these criteria were not met a mutation was marked as not present and considered to be limited to the tumour or the CSF. For the matched pair analysis, the tumour used was the closest tumour sequence to the time of CSF collection. In patients with multiple CSF collections, we prioritized CSF samples that met the following criteria: (1) at least one mutation called in CSF; (2) CSF collected through lumbar puncture (that is, rather than during VP-shunt placement or intra-operatively); (3) highest sequence coverage; (4) shortest interval between CSF and tumour collection.

Clonality analysis. We determined total, allele-specific, and integer DNA copy number genome-wide using the FACETS algorithm in all cases26. Allelic imbalance (including tumour-specific loss of heterozygosity) was determined from a change in the zygosity of heterozygous SNPs. We then defined the presence of genome doubling (GD) in samples for which the majority of the genome (>50%) contained multiple copies from the same parent/allele. Gene-level copy number was assigned from spanning segments of integer copy number data in each tumour. Homozygous deletion was determined from regions of total copy number of zero. Integer copy number greater than 5 or 6 in diploid and GD cases, respectively. Partial deletions (with intragenic breakpoints) were called whereas partial amplifications were not.

Statistical analysis. Associations between ctDNA positivity and patient and treatment characteristics were assessed using nonparametric tests by either Wilcoxon rank-sum test or Fisher’s exact test. Wilcoxon rank-sum test was used for the continuous variables and Fischer’s exact test was used for the categorical variables. All statistical tests were two-sided with α = 0.05 for statistical significance. To assess the association between CSF ctDNA positivity and overall survival, we performed a multivariable Cox model which included CSF ctDNA positivity and further adjusted for per cent extent of resection at original diagnosis, tumour burden at CSF collection, and IDH status. Follow-up time was calculated from time of CSF collection until death or last follow-up. Deaths were treated as events in the model and those who were alive at last follow-up were censored.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All genomic results and associated clinical data for all patients in this study are publicly available in the cBioPortal for Cancer Genomics at the following URL: http://www.cbioportal.org/study?id=glioma_msks 2018. The MSK-IMPACT data analysis pipeline can be found here: https://github.com/rshah/IMPACT-Pipeline. The source data for Table 1 and Extended Data Fig. 1 are available in Supplementary Table 1. The source data for the multivariable analysis (Extended Data Table 2) are available in Supplementary Table 2. The source data for the matched pair analysis (Fig. 2a, b) are available in Supplementary Table 5 (separate MS Excel file).

23. AACR Project GENIE Consortium, AACR Project GENIE: powering precision medicine through an international consortium. Cancer Discov. 7, 818–831 (2017).
24. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med. 23, 703–713 (2017).
25. Wen, P. Y. et al. Response assessment in neuro-oncology clinical trials. J. Clin. Oncol. 35, 2439–2449 (2017).
26. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).
27. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
28. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 44, e131 (2016).
Extended Data Fig. 1 | Interval between diagnosis and CSF collection, grouped by glioma subtype. For the comparison between the GBM (IDH WT) and the LGG (IDH WT) groups, P not significant at 0.16; between GBM (IDH WT) and LGG (IDH mutant), $P = 0.0000000689$; between LGG (IDH WT) and LGG (IDH mutant), $P = 0.0054$. Wilcoxon two-sample test was used for two-way comparisons. No adjustments were made for multiple comparisons. The box-plot elements are as follows: all patients ($n = 85$, grey): median, 510; minimum, 62; maximum, 9,122; 25th percentile, 273; 75th percentile, 1,606. GBM (IDH WT) ($n = 44$, red): median, 355; minimum, 62; maximum, 1,606; 25th percentile, 193; 75th percentile, 528. LGG (IDH WT) ($n = 12$, green): median, 473; minimum, 79; maximum, 2,982; 25th percentile, 292; 75th percentile, 1,013. LGG (IDH mutant) ($n = 24$, blue): median, 2,077; minimum, 63; maximum, 7,669; 25th percentile, 1,061; 75th percentile, 4,274. *Five patients were excluded from sub-group analysis owing to unknown IDH status ($n = 3$) or IDH mutant GBM ($n = 2$).
Extended Data Fig. 2 | Glioma growth towards CSF spaces. Representative brain MRI examples (T1 post-contrast) from patients with distinct patterns of tumour spread. Spread of enhancing disease to the pial, subependymal and subarachnoid spaces was used as an imaging surrogate to estimate tumour spread into the CSF, which is otherwise not visible by MRI. a, Enhancing leptomeningeal spread along bilateral cranial nerves VII and VIII (arrows). b, Enhancing pial spread to the surface of the pons (arrows). c, Nodular and curvilinear enhancing subependymal spread along both lateral ventricles (arrows).
Extended Data Fig. 3 | Interval between CSF collection and death for patients with positive (blue) and negative (red) CSF ctDNA.

a, All patients with glioma. Median overall survival (OS): for CSF ctDNA-positive patients, 3.15 months (95% confidence interval, 1.97–4.63); for CSF ctDNA-negative patients, 11.91 months (95% confidence interval, 8.40–30.81). The log-rank P value for comparing the survival experience of all patients with glioma stratified by ctDNA status was 0.0000078675.

b, Patients with GBM (IDH WT). Median overall survival: for CSF ctDNA-positive subjects, 2.04 months (95% confidence interval, 0.98–3.77); for CSF ctDNA-negative subjects, 9.89 months (95% confidence interval, 5.54–12.39). The log-rank P value for comparing the survival experience of patients with GBM (IDH WT) by ctDNA status was 0.000062396 (log-rank test, two-sided).
Extended Data Fig. 4 | Concordance between CSF and tumour in glioma subtype-defining genes. Shown are combinations of genetic alterations (LGG signatures) that are consistently congruent between the CSF and tumour (10/10). This was also the case in glioblastoma (20/20).
Pt # 36

Extended Data Fig. 5 | DNA hypermutation signature in CSF. Disease course for patient 36 with GBM with two tumour resections and one CSF collection. The patient received 14 monthly cycles of temozolomide (TMZ) following the initial tumour resection and postoperative radiation (RT)/TMZ. The initial tumour harboured five mutations, the recurrent tumour 120 mutations and the CSF 132 mutations. MRIs (T1 post-contrast) are shown from the time of diagnosis, first recurrence and second recurrence. The original tumour was in the right parietal lobe and recurrence was in the right frontal lobe. Diamond, tumour samples profiled; circle, CSF sample profiled. Bev, bevacizumab. The bar graph shows the precise number of SNVs that were called by the IMPACT pipeline in the recurrent tumour \( n = 120 \) independent somatic SNVs) and in CSF ctDNA \( n = 132 \) independent somatic SNVs; INDELs were excluded). Bar graphs show the precise number of SNVs for each of the possible trinucleotide combinations.
Extended Data Fig. 6 | Variant allelic frequencies for all SNVs in two independently collected CSF samples from patient 34 with DNA hypermutation. Scatter plot of variant allelic frequencies for all SNVs in two independently collected CSF samples from patient 34. Both CSF replicates harboured more than 200 SNVs. Pearson correlation coefficient ($r^2 = 0.966$) was calculated using a linear regression model in R (following Gist (https://gist.github.com/rhshah/3f4965a80886affb96d847dc2ecf69f5)).
Extended Data Fig. 7 | Divergence of tumour and CSF profiles over time. The histogram (top) depicts the interval (in days) between tumour and CSF collection. The pie charts (bottom) show that the samples that were collected at a very close interval (<3 weeks; red) had a higher percentage of shared mutations (79%) than the samples that were collected at a longer interval (>1,000 days; blue)(29%).
Extended Data Fig. 8  |  Evolution of the glioma genome.  

**a.** Disease course of patient 28 (GBM (IDH WT)) who received treatment with concurrent RT/TMZ, bevacizumab, and a PD-1 inhibitor. The patient underwent three tumour resections and one CSF collection and all four biospecimens were sequenced. The CDK4 amplification was seen in all four samples. Amplifications of PDGFRA/KIT were observed in tumour sample 3, whereas the later CSF sample (4) no longer showed the PDGFRA/KIT amplification. 

**b.** Disease course of patient 7 (IDH mutant anaplastic astrocytoma). The patient underwent four tumour resections and two CSF collections. All six samples were profiled. MRIs (T1 post-contrast) correspond to the time of each tissue resection or CSF recollection. Bottom, heat map shows all mutations across the six samples. Diamond, tumour samples profiled; circle, CSF samples profiled. Heat map indicates the variant allelic frequency of the indicated SNVs.
Extended Data Table 1 | Sequencing of CSF samples collected from individuals with non-malignant neurological conditions (controls)

### a

| Sample # | Disease                      | Reason for lumbar puncture | Coverage |
|----------|------------------------------|-----------------------------|----------|
| Control-1 | Headaches                   | rule out meningitis         | 105x     |
| Control-2 | Cryptococcal meningitis & HIV | infection                   | 1548x    |
| Control-3 | Multiple Sclerosis          | diagnostic work-up          | 3x       |
| Control-4 | Healthy Individual          | n/a                         | 2x       |
| Control-5 | Multiple Sclerosis          | diagnostic work-up          | 2x       |
| Control-6 | Neurofibromatosis Type 2/schwanoma | leg weakness; rule-out inflammatory neuropathy | 4x |
| Control-7 | Meningioma                  | r/o normal pressure hydrocephalus | 5x |

### b

| Sample # | Chr | Pos | Ref | Alt | Gene | cDNA change | A change | Total Depth(DP) | Variant Allele Depth(AD) | Variant Allele Frequency(VAF) | Present in cosmic? |
|----------|-----|-----|-----|-----|------|-------------|----------|-----------------|----------------------------|-------------------------------|--------------------|
| Control-1 | X   | 63409890 | C | T | AMER1 | c.3277G>A  | p.E1093K  | 96  | 39  | 0.40625          | Not present                    |
| Control-1 | 1   | 36908122 | G | A | CSF3R | c.839C>T   | p.A280V   | 83  | 30  | 0.361445783      | Not present                    |
| Control-1 | 2   | 46605868 | G | T | EPAS1 | c.1516G>T  | p.A505S   | 104 | 51  | 0.490384615      | Not present                    |
| Control-1 | 17  | 33428326 | C | T | RAD51D | c.461G>A   | p.R154H   | 88  | 48  | 0.545454545      | Not present                    |
| Control-1 | 8   | 145708723 | G | A | RECOL4 | c.2341C>T  | p.P781S   | 40  | 20  | 0.5              | Not present                    |
| Control-1 | 20  | 31025246 | C | T | ASXL1 | c.2731C>T  | p.P911S   | 1660 | 891 | 0.479032259      | Not present                    |
| Control-2 | 11  | 119155969 | C | T | CBL  | c.1634C>T  | p.P545L   | 1472 | 739 | 0.502038043      | Not present                    |
| Control-2 | 12  | 49437478 | T | C | KMT2D | c.5407A>G  | p.T1803A  | 1591 | 803 | 0.504714016      | Not present                    |
| Control-2 | 3   | 37092086 | G | A | MLH1  | c.2213G>A  | p.G738E   | 1493 | 727 | 0.486939049      | Not present                    |
| Control-2 | 5   | 79950696 | G | T | -    | c.162_170dup | p.A60_A62dup | 1398 | 356 | 0.254649499      | Not present                    |
| Control-2 | 19  | 15302618 | T | C | NOTCH3 | c.740A>G   | p.N247S   | 2107 | 1019 | 0.483626009      | Not present                    |
| Control-2 | 19  | 50095288 | C | T | POLD1 | c.496C>T   | p.R166W   | 1685 | 797 | 0.472997033      | Not present                    |
| Control-2 | 9   | 8499691 | T | G | PTPRD | c.2278A>G  | p.K760Q   | 1566 | 784 | 0.50063857       | Not present                    |
| Control-2 | 3   | 12647740 | T | C | RAF1  | c.640A>G   | p.M214V   | 535  | 255 | 0.47633514       | Not present                    |

---

a. Primary diagnosis, indication for LP, and sequencing coverage. b. List of sequence variants identified in negative control CSF samples. Because matched normal DNA was not available, sequence variants were filtered according to population sequencing databases. The remaining variants (above) may represent rare germline SNPs and have not previously been reported in databases of somatic mutations such as COSMIC.
Extended Data Table 2 | Results from multivariable model for overall survival from the time of CSF collection

| Variable                          | Category | HR     | 95% CI       | p-value       |
|----------------------------------|----------|--------|--------------|---------------|
| % Extent of resection at diagnosis | continuous | 1.005  | 0.997-1.013  | 0.26          |
| Tumor burden at LP/CSF           | continuous | 1.0002 | 1.0001-1.0004 | 0.007        |
| CSF positivity                    | Negative | ref    | --           | --            |
| CSF positivity                    | Positive | 4.16   | 2.15-8.05    | 0.000024315  |
| IDH Status                        | Mutant   | ref    | --           | --            |
| IDH Status                        | WT       | 12.01  | 4.44-32.49   | 0.0000009778250 |

CSF positivity was significantly associated with an increased rate of death from the time of CSF collection in an analysis that also accounted for the following variables: (1) per cent extent of resection at diagnosis; (2) tumour burden at the time of CSF collection; and (3) IDH status. n = 63 biologically independent samples. This number comprises n = 33 samples from CSF ctDNA-positive patients and n = 30 samples from CSF ctDNA-negative patients. The statistical tests used for the analysis were two-sided Wald tests. HR, hazard ratio; CI, confidence interval.
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
- [ ] An 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

Sequence data for all CSF samples were generated in the MSK research sequencing core (IGO) in the Center for Molecular Oncology. Tumor sequence data for comparison were generated in the MSK clinical sequencing laboratory in the Department of Pathology and were obtained from the cbioPortal for Cancer Genomics (http://cbioportal.org/msk-impact).

Data analysis

The MSK-IMPACT data analysis pipeline can be found here: https://github.com/rhshah/IMPACT-Pipeline and is composed entirely of open source software. Multivariable survival analysis was performed using the PHREG procedure in SAS (Cary, NC) to implement Cox regression modeling, a standard statistical modeling technique not unique to SAS.

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

Data Availability: All genomic results and associated clinical data for all patients in this study are publically available in the cBioPortal for Cancer Genomics at the following URL: http://www.cbioportal.org/study?id=glioma_msk_2018

Field-specific reporting

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

- [x] 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

No sample size calculation was performed for this retrospective analysis, however all consecutive samples within 1/12/15-4/5/2017 were obtained. The sample size was sufficient since we were able to detect associations between tumor-derived DNA in CSF with both disease burden and adverse outcome.

Data exclusions

The complete patient cohort consisted of 85 patients on whom all analyses were based with the exception of the multivariable Cox regression model. This model was based on a subset of 63 patients and is the only analysis where exclusion criteria were applied. Due to study design, 22 patients were excluded because the original tumor resection had been performed at an outside hospital and MRIs were not available. To ensure there were no differences in the complete patient cohort (n=85) and the subset on whom the multivariable Cox regression model was based (n=63), we compared the two groups on overall survival experience as well as all covariables in the model: CSF positivity, IDH status, percent extent of resection at diagnosis, and tumor burden at LP/CSF. There were no statistically significant differences between the two groups indicating that the subset of patients on whom the Cox regression model was based was comparable to the complete patient cohort.

Replication

For multiple patients in our study, a contemporaneous tumor biopsy or second CSF collection was available for comparisons with the CSF ctDNA sample. This data is included in Figure 2 and shows the reproducibility of our assay.

Randomization

This was not a setting in which randomization was required since there was no intervention being assessed. Samples from all comers during a specific date range were assessed for signal.

Blinding

Conventional blinding was not used in our design since no intervention was being assessed. However, we note in the Methods section that: "Brain MRIs prior to and directly following the initial resection were reviewed by an experienced neuroradiologist without knowledge of the CSF ctDNA results."

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
  - [ ] Unique biological materials
  - [x] Antibodies
  - [x] Eukaryotic cell lines
  - [x] Palaeontology
  - [x] Animals and other organisms
  - [x] Human research participants

Methods

- n/a Involved in the study
  - [x] ChIP-seq
  - [x] Flow cytometry
  - [x] MRI-based neuroimaging
**Unique biological materials**

Policy information about availability of materials

Obtaining unique materials

The study includes data from tumor samples and cerebro-spinal fluid (CSF) from glioma patients. These unique materials are not available from the authors or from standard commercial sources.

**Human research participants**

Policy information about studies involving human research participants

Population characteristics

Patients were between 22 and 90 years old. The patient population included 54 men and 31 women. 46 patients had glioblastoma (GBM). 39 patients had lower grade glioma (LGG). 84/85 patients had received prior radiation therapy. 78/85 patients had received therapy with alkylating agents as part of their systemic therapy.

Recruitment

All patients in our study were treated and followed at Memorial Sloan Kettering and underwent lumbar puncture for clinical indications. There are no expected biases based on recruitment since all consecutive samples were obtained during a specified time period. There was no self-selection into the study.

**Magnetic resonance imaging**

Experimental design

| Design type               | N/A          |
|---------------------------|--------------|
| Design specifications     | N/A          |
| Behavioral performance measures | N/A          |

Acquisition

| Imaging type(s)         | Structural   |
|-------------------------|--------------|
| Field strength          | 1.5T and 3T  |

Sequence & imaging parameters

MRI included the following parameters: axial contrast-enhanced T1-weighted images with gadolinium contrast (gadobutrol, 0.1 mmol/kg max 14 ml) and field-of-view, 24; slice thickness, 3-5mm; gap, 0; time-to-echo (TE), minum; time-to-repetition (TR), 400-2000; flip angle, auto; matrix, 256x224; and axial fluid attenuated inversion recovery (FLAIR) images with field-of-view, 24; slice thickness, 3-5mm; gap, 0; TE, 120; TR, 8000-9000; flip angle, 90 degrees; time-to-inversion (TI), 2000-2500; matrix, 256x256.

Area of acquisition

whole brain

Diffusion MRI

- Used: [ ]
- Not used: [x]

Preprocessing

| Preprocessing software   | N/A          |
|--------------------------|--------------|
| Normalization            | N/A          |
| Normalization template   | N/A          |
| Noise and artifact removal | N/A         |
| 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: [ ]

Statistic type for inference (See Ekland 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 |