Published in final edited form as:

Nat Med. 2020 July ; 26(7): 1044–1047. doi:10.1038/s41591-020-0932-2.

Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes

Farshad Nassiri^{1,2,8}, Ankur Chakravarthy^{3,8}, Shengrui Feng^{3,4,8}, Shu Yi Shen³, Romina Nejad¹, Jeffrey A. Zuccato^{1,2}, Mathew R. Voisin^{1,2}, Vikas Patil¹, Craig Horbinski^{5,6}, Kenneth Aldape^{1,7}, Gelareh Zadeh^{1,2,9}, Daniel D. De Carvalho^{3,4,9}

¹MacFeeters Hamilton Neuro-Oncology Program, Princess Margaret Cancer Centre, University Health Network and University of Toronto, Toronto, Ontario, Canada.

²Division of Neurosurgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada.

³Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada.

⁴Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada.

⁵Department of Pathology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.

⁶Department of Neurological Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.

⁷Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA.

⁸These authors contributed equally: Farshad Nassiri, Ankur Chakravarthy, Shengrui Feng.

⁹These authors jointly supervised this work: Gelareh Zadeh, Daniel D. De Carvalho.

Correspondence and requests for materials should be addressed to G.Z. or D.D.D.C., gelareh.zadeh@uhn.ca; daniel.decarvalho@uhnresearch.ca.

Author contributions

F.N., A.C., S.F., G.Z. and D.D.D.C. conceived and designed the study. F.N., R.N., G.Z., K.A. and C.H. were responsible for clinical care and collected all the biomaterials. F.N., S.F., S.Y.S., J.A.Z. and M.R.V. carried out the laboratory testing. F.N. and R.N. collected and collated the clinical data. F.N., A.C., S.Y.S. and V.P. contributed to the data processing. G.Z. and D.D.D.C. supervised all aspects of the study. F.N., A.C., G.Z. and D.D.D.C. contributed to initial data interpretation and wrote the first draft. All authors contributed to final data interpretation and critical revision of the manuscript, and approved the final version of the manuscript.

Code availability

R markdowns of the code used to generate the results in this paper are available in a Zenodo archive at <https://doi.org/10.5281/zenodo.3715312>

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-0932-2>.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-0932-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-0932-2>.

Competing interests

D.D.D.C., S.Y.S. and A.C. are listed as inventors on patents filed that are related to this method. D.D.D.C. received research funding from Pfizer and Nektar therapeutics not related to this project.

Abstract

Definitive diagnosis of intracranial tumors relies on tissue specimens obtained by invasive surgery. Noninvasive diagnostic approaches provide an opportunity to avoid surgery and mitigate unnecessary risk to patients. In the present study, we show that DNA-methylation profiles from plasma reveal highly specific signatures to detect and accurately discriminate common primary intracranial tumors that share cell-of-origin lineages and can be challenging to distinguish using standard-of-care imaging.

A major challenge in management of intracranial pathologies is accurate diagnosis of lesions identified on imaging that have a broad differential diagnosis ranging from indolent low-grade tumors to aggressive cancers. Current practice necessitates invasive neurosurgery to obtain tissue for diagnosis and molecular subtyping, which introduces risk and neurological morbidity, as well as anxiety for the patient. Reliable noninvasive strategies to diagnose and subtype tumors would be transformative for patient care, either by providing diagnostic information preoperatively that can improve neurosurgical planning or by avoiding the need for highly invasive procedures altogether. Sampling of circulating tumor cell-free DNA (ctDNA) from biological fluids of patients, such as blood, provides an opportunity to noninvasively establish definitive diagnosis (that is, liquid biopsy). However, genomic analysis of ctDNA is restricted in sensitivity primarily due to its limited abundance in plasma^{1,2}, and the limited number of diagnostic alterations that distinguish normal from cancerous ctDNA³.

We have previously shown that DNA-methylation alterations can reliably detect extracranial cancers with distinct cells of origin in plasma despite low ctDNA abundance⁴. In the present study, we sought to determine whether our approach, which recovers and profiles methylated DNA fragments from plasma, called cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP-seq)^{4,5}, could be used to detect and discriminate common intracranial tumors noninvasively, some of which share similar cell-of-origin lineages.

To determine whether cfMeDIP-seq could detect circulating free DNA (cfDNA) from intracranial tumors, we initially tested diffuse glioma as a model brain tumor type and generated cfMeDIP-seq profiles on 60 independent patient plasma samples. Data from these samples were merged with previously published cfMeDIP-seq data to yield a dataset of 447 samples across 9 classes (see Extended Data Fig. 1a)⁴. We harnessed machine-learning approaches to compile all cfMeDIP profiles (see Extended Data Fig. 1b), splitting our cohort into 50 different training and test sets, with 80% of the data in a training set and the remainder in the corresponding test set. Binomial random forest (RF) classifiers (gliomas versus all other cancers) were trained using the top 300 differentially methylated regions (DMRs) for gliomas versus each other class, and model performance was tested on the held-out test sets. A median of eight CpGs contributed to each DMR for the glioma-versus-other RF classifier, and all of the included DMRs contained at least one CpG site with a median CpG enrichment score of 3.02 (95% confidence interval (CI) = 2.89–3.15), in line with our previous results⁴. We observed high sensitivity and discriminative capacity for our models to classify gliomas among other cancerous and healthy patients (area under the curve (AUC) =

0.99, 95% CI = 0.96–1.00; Fig. 1a), with similar performance in IDH mutant (AUC = 0.992; 95% CI = 0.949–1.000) and wild-type (AUC = 0.993; 95% CI = 0.966–1.000) gliomas (Fig. 1b) as well as in lower-grade (AUC = 0.999, 95% CI = 0.990–1.000) and high-grade (AUC = 0.983, 95% CI = 0.903–1.000) gliomas (Fig. 1c). Visualization of the plasma-derived features computed on training sets (see Extended Data Fig. 1c,d) revealed that glioma samples cluster together, with a set of these DMRs being hypermethylated in most glioma samples compared with other tumors and healthy controls, further demonstrating the utility of plasma methylomes to distinguish patients with gliomas from patients with systemic cancers and healthy controls.

We observed that plasma cfMeDIP-seq signals correlated well with corresponding tumor tissue DNA-methylation values overall (Pearson's $r = 0.37$, two-tailed $P < 2.2 \times 10^{-16}$; Fig. 1d), suggesting that the signatures we identified to detect gliomas were in fact primarily derived from glioma tumor DNA. We formally tested whether tumor-derived signals alone could drive effective classification of gliomas by restricting our dataset to windows that are unmethylated in healthy plasma⁶, and repeating the machine-learning workflow described in Extended Fig. 1b for an unbiased estimation of classification performance. We observed remarkably high accuracy, even after excluding nonmalignant contributors to plasma methylation (AUC = 0.982, 95% CI = 0.93–1.00; Fig. 1e), with similar correlation between plasma and tumor methylation signatures (Pearson's $r = 0.42$, two-tailed $P < 2.2 \times 10^{-16}$; see Extended Data Fig. 1e), confirming that ctDNA alone captured with cfMeDIP-seq is sufficient to drive effective tumor classification.

To further demonstrate that the circulating methylome signature used to classify gliomas was specific to glioma tumor cells, we defined a glioma-specific methylation signature using orthogonal datasets by intersecting windows that are typically unmethylated in healthy plasma⁶, with windows containing differentially methylated CpG sites in glioma cell lines versus 1,028 cell lines from 33 other cancers⁷. We observed that the median signal of the windows hypermethylated in glioma cell lines, which are devoid of contribution from the tumor microenvironment, was significantly enriched in cfMeDIP-seq profiles of glioma patients compared with other cancer patients and health controls (see Extended Data Fig. 1f; $P = 1.767 \times 10^{-12}$, Wilcoxon's rank-sum test). We also observed that the patterns of clustering between gliomas and other sample classes were retained even after intersecting our glioma plasma-derived features with the orthogonally derived, glioma-specific methylation signature (Fig. 1f,g). Moreover, the hypermethylated subset of this signature was markedly enriched in glioma plasma samples versus other plasma samples (Fig. 1h; $P < 2.2 \times 10^{-16}$, Wilcoxon's rank-sum test). Taken together, these findings suggested that cfMeDIP-seq can adequately detect glioma ctDNA to drive classification.

Given that we could capture methylome profiles of glioma ctDNA in plasma, and that tumor-derived methylome signatures have been shown to reliably discriminate intracranial tumors⁸, we next sought to determine whether cfMeDIP-seq profiles of plasma could be used to further distinguish common primary intracranial tumors. For this, we generated cfMeDIP-seq profiles on an additional 161 samples that are typically included in the differential diagnosis of solitary extra-axial tumors (meningiomas, hemangiopericytomas) and intra-axial tumors (low-grade glial-neuronal tumors, IDH mutant gliomas, IDH wild-

type gliomas and brain metastases from systemic cancers; see Extended Data Fig. 2a). This cohort was split into 50 balanced training (80%) and test (20%) sets, and the top 300 DMRs for each class compared with other classes were selected in the training sets to train a series of one versus another, regularized, generalized linear models (see Extended Data Fig. 2b). Model evaluation tested on held-out samples demonstrated reliable discrimination of common extra-axial tumors (AUC of meningioma versus others = 0.89, 95% CI = 0.80–0.97; AUC of hemangiopericytoma versus others = 0.95, 95% CI = 0.73–1.00), as well as intra-axial tumors, ranging from low-grade indolent glial–neuronal tumors (AUC = 0.93, 95% CI = 0.80–1.00) to diffuse intra-axial gliomas with distinct molecular composition (AUC of IDH mutant glioma versus others = 0.82, 95% CI = 0.66–0.98; AUC of IDH wild-type glioma versus others = 0.71, 95% CI = 0.53–0.90; Fig. 2a). Visualizing the top differentially methylated regions used for model generation also demonstrated separability of clinically relevant brain tumor types (Fig. 2b). These results demonstrate the potential for cfMeDIP-seq profiles to not only detect ctDNA, but to discriminate tumors with similar cell-of-origin lineages, which can be challenging to distinguish using standard-of-care magnetic resonance imaging (MRI) (Fig. 2c).

In the present study, we present the first application of our plasma-based liquid biopsy approach to detect and distinguish clinically relevant tumors of the central nervous system. We demonstrate that our approach, which does not rely on information obtained from a tumor tissue biopsy, can be used to distinguish gliomas from extracranial cancer types that may metastasize to the brain and healthy controls, and can accurately discriminate different primary brain tumors that may otherwise be indistinguishable using standard-of-care MRI. The evidence we present for the detection of plasma ctDNA could have practice-changing implications for the diagnosis and management of intracranial lesions. When complemented with standard-of-care clinical work-up, our liquid biopsy approach, which has a turn-around time of less than 1 week, could reduce the need for invasive procedures to establish diagnosis when cytoreduction is not necessary (for example, low-grade glial–neuronal tumors) or when risks of invasive surgery outweigh benefits (for example, multifocal diffuse gliomas). Furthermore, for tumors that benefit from cytoreduction, knowledge of diagnosis before surgery can help improve neurosurgical planning and safety, for example in the case of hemangiopericytomas that are almost indistinguishable on preoperative MRI from meningiomas, but pose a risk for excessive blood loss if the surgeon is not prepared. Our study is limited in that we compared regional cfMeDIP-seq signals with corresponding array-based methylation at single CpG loci. Direct comparisons using similar platforms as well as prospective independent validation remain warranted.

Methods

Patients and plasma samples.

Patient tumor samples and clinical data for the present study were obtained from the University Health Network, Toronto and the Feinberg School of Medicine, Northwestern University, Chicago, with protocols approved by the institutional research ethics boards. Sample selection was determined based on availability of stored plasma samples and clinical data. A total of 220 patient samples (70 IDH mutant gliomas, 52 IDH wild-type gliomas,

60 meningiomas, 9 hemangiopericytomas, 14 low-grade glial–neuronal tumors (that is, dysembryoplastic neuroectodermal tumors, gangliogliomas and pilocytic astrocytomas) and 15 brain metastases of unknown primary cancer on presentation) were selected for the present study. The clinical data for glioma samples profiled are detailed in Supplementary Tables 1 and 2.

To determine IDH mutation status, Sanger sequencing for *IDH1* was performed on all tumors. Sequencing for *IDH2* mutations was performed on samples that were *IDH1* wild-type. For patients with matching tumor DNA-methylation data, IDH status was also inferred from methylation data as an additional check using the online brain tumor classifier⁸.

Plasma processing and cfMeDIP-seq.

Peripheral blood (20 ml) was drawn for each patient at the time of surgery and plasma was separated and collected (1–2 ml in acid citrate dextrose vacutainers) within 30 min by centrifugation (2,500*g* for 15 min at 20 °C) and cryopreserved immediately after collection.

Processing and sequencing of plasma samples were performed blinded to pathology and molecular subtypes. The ctDNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) and quantified through Qubit (Thermo Fisher Scientific). For each sample, between 1 and 10 ng ctDNA was subjected to library preparation using Kapa HyperPrep Kit (Kapa Biosystems), followed by the cfMeDIP protocol using Diagenode MagMeDIP kit (catalog no. C02010021) with modifications that have been previously detailed^{4,5}. All the final libraries were submitted for BioAnalyzer analysis and sequenced at our institution (Princess Margaret Genomics Centre). Processing and sequencing of samples were performed in two runs. First, 60 gliomas were processed as previously outlined^{4,5} and sequenced on an Illumina HiSeq 2500, 1 × 50 bp reads, multiplexed as 7 samples per lane. Of these, one sample did not meet the quality control metrics that we detailed previously⁴. The median sequencing depth of this cohort was similar to our previously published samples and therefore these tumors were analyzed together. The second run of samples included a total of 161 samples (63 gliomas, 60 meningiomas, 9 hemangiopericytomas, 15 brain metastases of unknown primary and 14 low-grade glial–neuronal tumors, which included 6 gangliogliomas, 5 dysembryoplastic neuroectodermal tumors and 3 pilocytic astrocytomas). Processing of these samples was performed similar to samples in the first run, with the exceptions that: (1) solid-phase reversible immobilization beads were used for size selection in place of traditional agarose gel size-selection procedures (as has previously been outlined⁵); (2) samples were sequenced pair-ended with greater depth of coverage (median 59 million reads per sample); and (3) samples were sequenced using a NovaSeq instrument. To avoid confounding biological results with technical or sequencing differences, we analyzed these 161 samples separately from the original glioma cohort and previously published samples. The sequencing metrics for all samples are provided in Supplementary Table 3.

Tumor DNA-methylation profiles.

To correlate cfMeDIP-seq signals with tumor methylation values we generated genome-wide DNA-methylation profiles on 112 glioma samples, of which 65 also had cfMeDIP-seq

data available. Between 250 and 500 ng DNA was extracted from 50 mg of frozen tissue, bisulfite converted (Zymo EZ DNA methylation Kit, Zymo 190 Research) and profiled using the Illumina HumanBeadChip 850K array. Beta values of CpG sites that encompassed windows used in the classification of gliomas versus systemic cancers and healthy controls were used to correlate tumor methylation signals (delta–beta of primary tumor versus healthy plasma from EPIC array) with cfMeDIP-seq signals (fold-change of normalized read counts in gliomas versus healthy plasma). Correlation was determined using Pearson's correlation.

Cell-line DNA-methylation profiles.

To generate an orthogonal tumor-specific signature, we harnessed data from two previously published datasets^{6,7}. First, regions that are likely to be unmethylated in normal plasma were computed by multiplying cell-type specific methylation and the abundance of that cell type in healthy plasma, followed by summation to yield total methylation in the plasma and filtering of probes with a total beta value <10% in the plasma⁶. We then intersected this with a glioma cell-line signature that was derived by identifying all differentially methylated probes in glioma cell lines versus all other classes⁷ at a median delta–beta of 0.3 and false discovery rate <0.01. All microarray data were normalized using ssNoob normalization.

Statistical analyses.

A two-tailed $P < 0.05$ corrected for multiple comparisons using a false discovery rate when relevant was considered statistically significant. Nonparametric statistics (Wilcoxon's rank-sum test) were used to compare cfMeDIP-seq signals across tumor types. Pearson's correlation was used to test for associations between cfMeDIP-seq signals, at a given 300-bp window, with corresponding averaged CpG values from array-based tumor methylation data. Details about the generation and evaluation of classifiers using machine-learning algorithms are detailed under Machine-learning approaches for classification and subtyping.

Machine-learning approaches for classification and subtyping.

Data from cfMeDIP-seq profiles were reduced to 300-bp genomic windows mapping to CpG islands, shores, shelves and FANTOM5 enhancers (regulatory features) as previously described for computational efficiency⁴.

For all analyses, the generation of models was performed exclusively using samples in the training cohort, and model performance was tested in samples not used for model generation (that is, we left out samples in the test sets).

To determine whether gliomas can be distinguished from extracranial cancers and healthy controls, the compilation of 447 samples across 9 pathologies was partitioned into 50 independent training and testing cohorts in an 80%:20% manner. Then 50 binomial RF classifiers (gliomas versus all other cancers) were developed using the top 300 DMRs identified using the moderated t statistic with limma-trend in each training cohort to estimate the probability of a sample being a glioma. Model optimization was performed using three rounds of tenfold cross-validation in the training cohort only. Model performance was evaluated for each held-out sample in test cohorts by computing the area under the

receiver operating characteristic (ROC) curve (AUROC). To evaluate the capacity of glioma ctDNA to drive effective classification, we generated similar models using hematopoietic cell-unmethylated windows and evaluated model performance in a similar fashion.

To determine whether cfMeDIP-seq samples could be used to distinguish common brain tumors, a total of 161 cfMeDIP-seq samples (IDH mutant and wild-type gliomas, low grade glial–neuronal tumors, brain metastases of unknown primary, meningiomas and hemangiopericytomas) was used for analysis. This cohort was split into 50 different training (80%) and test (20%) sets balanced for each class. Using only the training cohorts, the top 300 DMRs for each class compared with other classes were selected to train a series of one versus another, regularized, generalized linear model (default alpha and lambda parameters). The performance of each model was evaluated for each held-out sample in test cohorts by computing the AUROC.

Reporting Summary.

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

Extended Data

Extended Data Fig. 1 | cfMeDiP-seq signals of gliomas compared to extracranial cancers and healthy controls.

a, Bar-chart showing the distribution of samples in the 447 sample cfMeDiP cohort. **b**, Flowchart of machine learning algorithm used to train and evaluate cfMeDiP-seq in glioma detection and classification **c**, heatmap showing cfMeDiP-seq signals (log2 counts per million) of all DMRs (rows) derived from training sets for patients (columns) in the machine learning analyses detailed in (b). **d**, MDS plot of the features depicted in the heatmap in

(c) in gliomas (n = 59) as well as other cancers and healthy controls samples (n = 388). **e**, scatterplot showing difference in plasma cfDNA methylation signals of gliomas vs healthy controls after restricting to windows typically unmethylated in healthy plasma (n = 138,328 windows) against differences in methylation levels of glioma tumors vs healthy control plasma, with associated density contours. Pearson correlation coefficient ($r = 0.42$) and two-tailed p values ($p < 2.2 \times 10^{-16}$) are shown. **f**, Boxplots showing the distribution of per-sample median signal (counts per million) in gliomas (n = 59) as well as other cancers and healthy controls samples (n = 388) of windows unmethylated in healthy plasma and hypermethylated in glioma cell lines compared to cell lines from 33 other cancer types ($\Delta\text{-Beta} > 0.3$, $\text{FDR} < 0.01$). Central bars indicate medians, the box defines the upper and lower quartiles of the distribution, and whiskers define the 1.5x interquartile range. Two-tailed p-value ($p = 1.767 \times 10^{-12}$) from Wilcoxon's Rank Sum Test is shown.

Extended Data Fig. 2 |. Algorithm for machine-learning analysis of plasma-based brain tumor classifier.

a, Bar-chart showing the distribution of samples in brain cfMeDIP cohort.

Hemangiopericytoma (n = 9), Meningioma (n = 60), low-grade glioneuronal (n = 14), IDH mutant glioma (n = 41), IDH-wildtype glioma (n = 22), brain metastases (n = 15). **b**, Flowchart of machine learning algorithm used to train and evaluate cfMeDIP-seq in brain tumor detection and classification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank N. Pirouzmand for her technical assistance with experimental protocols. F.N. is supported by the Canadian Institute of Health Research (CIHR) Vanier Scholarship, AANS/CNS Section on Tumors & NREF Research Fellowship Grant, and Hold'em for Life Oncology Fellowship. A.C. is supported by a CIHR Banting Fellowship. The Northwestern Nervous System Tumor Bank (C.H.) is supported by the P50CA221747 SPORE for Translational Approaches to Brain Cancer. C.H. is funded by the National Institutes of Health (grant no. R01NS102669). G.Z. is funded by a CIHR project grant award (grant no. 159452), and the Brain Tumor Charity UK Quest for Cures grant (grant no. GN-000430). D.D.C. is funded by the CIHR New Investigator salary award (201512MSH360794-228629), Helen M. Cooke professorship and the Gattuso-Slaight Personalized Cancer Medicine Fund from Princess Margaret Cancer Foundation, Canada Research Chair, CIHR Foundation grant (grant no. FDN 148430), CIHR Project grant (grant no. PJT 165986), NSERC (grant no. 489073) and Ontario Institute for Cancer Research with funds from the province of Ontario.

Data availability

The data used to deconvolute healthy plasma cell type were previously published and available in the Gene Expression Omnibus (GEO) repository under accession code GSE122126.

All the cell line datasets analyzed during the present study were previously published and available in the GEO repository under accession code GSE68379.

Processed cfMeDIP-seq data and intermediate data objects are available in a Zenodo archive at <https://doi.org/10.5281/zenodo.3715312>. Preprocessed cfMeDIP-seq data generated in this manuscript are available on request from the corresponding authors (D.D.C. and G.Z.) to comply with the Princess Margaret Cancer Center Institute ethics regulations to protect patient privacy. All requests will be promptly reviewed by the Technology Development and Commercialization team to verify whether the request is subject to any intellectual property or confidentiality obligations. Any data and materials that can be shared will be released subject to a data transfer agreement.

References

1. Heitzer E, Haque IS, Roberts CES & Speicher MR Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* 20, 71–88 (2019). [PubMed: 30410101]
2. De Mattos-Arruda L et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat. Commun.* 6, 8839 (2015). [PubMed: 26554728]
3. Newman AM et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* 20, 548–554 (2014). [PubMed: 24705333]
4. Shen SY et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* 563, 579–583 (2018). [PubMed: 30429608]
5. Shen SY, Burgener JM, Bratman SV & De Carvalho DD Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat. Protoc.* 14, 2749–2780 (2019). [PubMed: 31471598]
6. Moss J et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* 9, 5068 (2018). [PubMed: 30498206]
7. Iorio F et al. A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754 (2016). [PubMed: 27397505]
8. Capper D et al. DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474 (2018). [PubMed: 29539639]

Fig. 1 | Tumor-specific plasma methylomes can distinguish gliomas from extracranial cancers and healthy controls.

a, Ensemble of ROC curves for 50 iterations of trained glioma versus other classifiers. **b,c**, Boxplots showing distribution of AUROC statistics for glioma-versus-other classifiers stratified by IDH mutation status ($n = 50$ iterations) (**b**) and World Health Organization grade ($n = 50$ iterations) (**c**). Central bars indicate medians, the box defines the upper and lower quartiles of the distribution, and the whiskers define the $1.5 \times$ interquartile range (IQR). **d**, Scatterplot showing difference in plasma cfDNA methylation signals of gliomas

versus healthy controls ($n = 186,437$ windows) against difference in methylation levels of glioma tumors versus healthy control plasma, with associated density contours. Pearson's correlation coefficient ($r = 0.37$) and two-tailed P values ($P < 2.2 \times 10^{-16}$) are shown. **e**, Ensemble of ROC curves for 50 iterations of trained glioma versus other classifiers that are trained with windows restricted to unmethylated regions in healthy plasma (**a**). **f,g**, Heatmap (**f**) and multidimensional scaling plot (**g**) from the subset of features in **e** that intersect with signatures derived from glioma cell lines in gliomas ($n = 59$), as well as other cancers and healthy control samples ($n = 388$). **h**, Boxplots showing the distribution of per-sample summed cfMeDIP-seq signals (c.p.m.) in gliomas ($n = 59$) as well as other cancers and healthy control samples ($n = 388$), using windows that are unmethylated in healthy plasma and hypermethylated in glioma cell lines in comparison to 33 other cancer types, which also intersect with plasma-derived DMRs. Central bars indicate medians, the box defines the upper and lower quartiles of the distribution, and whiskers define the $1.5 \times$ IQR. The two-tailed P value ($P < 2.2 \times 10^{-16}$) from Wilcoxon's rank-sum test is shown. AML, acute myeloid leukemia; BLCA, bladder cancer; BRCA, breast cancer; CRC, colorectal cancer; LUC, lung cancer; PDAC, pancreatic ductal adenocarcinoma; RCC, renal cell carcinoma.

Fig. 2 | Plasma cfDNA methylomes can discriminate common intracranial tumors with similar cells of origin.

a, Ensemble of ROC curves from 50 iterations of trained one-class-versus-other models.

b, A t -distributed stochastic neighbor embedding (t -SNE) plot, generated using the top 100 DMRs for each class-versus-other model in **a** with associated density contours. **c**,

Distribution of median class probabilities for two extra-axial tumors and three intra-axial tumors that cannot be reliably distinguished by routine MRI. The results of the different

one-class-versus-other models are displayed in a column for a single case. Dark-purple boxes indicate accurate prediction of tumor diagnosis with the highest median probability.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript