

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](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A full description of the statistical parameters 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) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted <i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.

Data analysis

PRFFECT Toolbox for Spectroscopists (Smith BR, Baker MJ, Palmer DS. PRFFECT: a versatile tool for spectroscopists. Chemometrics and intelligent laboratory systems. 2018 Jan 15. <https://doi.org/10.1016/j.chemolab.2017.10.024>), developing in-house using the R Project for Statistical Analysis, and the open source caret package for classification and regression training.

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. 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 are available on request due to privacy or other restrictions

Field-specific reporting

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

x

Life sciences study design

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

| | |
|-----------------|---|
| Sample size | For retrospective study of 724 patients, no statistical analysis was performed as we aimed to simply create the largest and most varied brain cancer serum database possible. For prospective study presented here based upon over 200 patients, this data represents over half of the statistically powered measure of diagnostic accuracy study. This study was powered by calculations conducted by the Edinburgh Clinical Trials Unit director, Professor John Norrie. |
| Data exclusions | A thorough quality test on spectral data was implemented, but no data was excluded from this study. |
| Replication | All patients were analysed in triplicate to account for technical variance, with each sample also being measured in triplicate to account for instrument variation. A number of pre-analytical experiments have been conducted in-house to characterise variation that arises from biological and technique variation in our protocol, and this is largely understood in this presented study. |
| Randomization | For retrospective studies, patient samples were randomised process. For prospective studies, samples were analysed chronologically when acquired; best replicating our real-world application. |
| Blinding | For retrospective studies, patient samples were analysed in a double-blind randomised process. For prospective studies, samples were analysed chronologically when acquired, but all analysis was conducted double-blind, as we had no information regarding disease status; best replicating our real-world application. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clinical data |

Methods

| n/a | Involved in the study |
|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

| | |
|-----------------|---|
| Antibodies used | <i>Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.</i> |
| Validation | <i>Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.</i> |

Eukaryotic cell lines

Policy information about [cell lines](#)

| | |
|--|--|
| Cell line source(s) | <i>State the source of each cell line used.</i> |
| Authentication | <i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i> |
| Mycoplasma contamination | <i>Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.</i> |
| Commonly misidentified lines (See ICLAC register) | <i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i> |

Palaeontology

| | |
|---------------------|---|
| Specimen provenance | <i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).</i> |
|---------------------|---|

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

For the retrospective study presented here, we obtained data from patients confirmed with brain cancer (prior to resection/biopsy) and controls with confirmed no history of cancer. The patient demographic represents the expected prevalence of brain cancer in age and sex. In this study we investigate age-matching this population and this does not impact our overall diagnostic capability.

For prospective studies, patients were recruited based upon presentation which "query: brain tumour" symptoms, that were then referred to secondary care for imaging using the "Open Access CT" scanning program running in the Lothian NHS trust. This is the targeted population for this test. The demographic again is representative of the patient population shown in the retrospective study.

Recruitment

Retrospectively, patients were recruited at three distinct locations from a range of biobanked sources. Prospectively, patients are recruited at their outpatient appointment prior to CT scanning, or following presentation in emergency care. Patients are recruited by the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, or by the EMERGE (Emergency Department) Clinical Research Team at the Royal Infirmary, Edinburgh.

Ethics oversight

University of Edinburgh and the LREC Lothian NHS Trust.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the [ICMJE guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from [ClinicalTrials.gov](#) or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

ChIP-seq

Data deposition

Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission *Provide a list of all files available in the database submission.*

Genome browser session (e.g. [UCSC](#)) *Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument *Identify the instrument used for data collection, specifying make and model number.*

Software *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*

Gating strategy *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type *Indicate task or resting state; event-related or block design.*

Design specifications *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI Used Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: Whole brain ROI-based Both

Statistic type for inference (See [Eklund et al. 2016](#))

Correction

Models & analysis

n/a | Involved in the study

Functional and/or effective connectivity

Graph analysis

Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis