nature research

Corresponding author(s): Stefan M. Pfister

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

The present work focuses on predictive modelling and no inferential statistic was performed. The minimal sample size to include a tumor class into the model was determined empirically by training and testing models with different sample sizes. The minimal class size of 8 allowed us to include rare tumor classes without losing prediction performance.

2. Data exclusions

Describe any data exclusions.

Tumor content was required to be above 70% (as described in the Methods), otherwise data was not generated. This criteria was pre-established.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The separation of samples into the defined DNA methylation classes was reliably reproduced by iterative random downsampling of the reference cohort.
The rate of establishment of a new diagnosis by methylation profiling was confirmed by the data of the external centres. The interlaboratory comparison demonstrated a reliable reproduction of the results of the original laboratory.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

The construction of the methylation classifier reference cohort was done in a supervised fashion to recapitulate the entities established in the WHO classification of tumours of the central nervous system, no randomization was performed. For the clinical implementation in the prospective samples also no randomization was performed as all cases with sufficient material were subjected to the analysis. For the technical validation samples were also not randomized, instead 51 samples of a wide selection of histological classes were chosen to increase the validity for a broader range of tumours.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The initial pathological diagnosis of the prospective series was done fully blinded as the methylation data was not generated before the finalisation of pathological diagnosis of a given case (usually two weeks after the pathological diagnosis).

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	R: A language and environment for statistical computing.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Distribution of material of human research participants is restricted

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	does not apply
c. Report whether the cell lines were tested for mycoplasma contamination.	does not apply
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	does not apply

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Reference cohort: tumor samples of 2801 individual research participants: 1278 female, 1466 male, 57 sex not available; age range 0-93 years, median 24 years; Prospective cohort: tumor samples of 1104 individual research participants: 481 female, 591 male, 32 sex not available; age range 0-85 years, median 38 years. External centre cohort: tumor samples of 401 individual research participants: 202 female, 199 male; Age range 0-86 years, median 53 years.