

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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 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)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	https://www.cancer.gov/ccg/research/genome-sequencing/tcga , https://gdc.cancer.gov/about-gdc/contributed-genomic-data-cancer/research/clinical-proteomic-tumor-analysis-consortium-cptac
Data analysis	https://github.com/Samiran-Dev/PathGen

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

For all our experiments, we use data from two anonymised publicly available open-access datasets (with ethics in place) - The Cancer Genomic Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC). The TCGA data can be accessed at <https://www.cancer.gov/tcga>, and the CPTAC data can be accessed at <https://gdc.cancer.gov/about-gdc/contributed-genomic-data-cancer-research/clinical-proteomic-tumor-analysis-consortium-cptac>. We use the corresponding

transcriptomic data downloaded from cBioPortal <https://www.cbioportal.org> for all cohorts and datasets. The gene categories used are found at https://github.com/mahmoodlab/MCAT/blob/master/datasets_csv_sig/signatures.csv. No new primary data was collected as part of this study. Source data are provided with this paper at <https://docs.google.com/spreadsheets/d/1jpXvQ5Gr7Tn09VVZ8KMLDShp6X9dV5NDmIKDV3WuYvs/edit?usp=sharing>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We have performed sex based fairness analysis of our models and presented the results in the Result section and Supplementary Tables.
Reporting on race, ethnicity, or other socially relevant groupings	We used the information from the clinical data provided in the open-access public datasets -TCGA and CPTAC.
Population characteristics	All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.
Recruitment	All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.
Ethics oversight	All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.

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

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.

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	The publicly available data was split into 70%, 10%, 10%, and 10% for train, validation, calibration and test. For TCGA-GBMLGG, 912 whole slide images (WSIs) coming from 745 cases are used, of which 183 cases belong to TCGA-GBM and 562 cases belong to TCGA-LGG. We consider 485 WSIs from 462 cases for TCGA-KIRC, 294 WSIs from 267 cases for TCGA-UCEC, 1010 WSIs from 946 cases for TCGA-BRCA, 242 WSIs from 62 cases for CPTAC-GBM, and 364 WSIs from 71 cases for CPTAC-UCEC for our experiments. All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

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

Methods

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

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

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	All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.
Study protocol	N/A
Data collection	All data used are anonymised publicly available secondary data with ethics in place. No new data has been collected.
Outcomes	N/A

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A