

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 (n) 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. 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> |
| ☒ | ☐ | 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 |

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 Custom LABVIEW codes (LABVIEW 2012) were used for automated high-throughput near-infrared fluorescence spectroscopy.

Data analysis Custom MATLAB codes (MATLAB R2019b) were used for background subtraction, spectral corrections, and feature vector extraction. Custom Python codes, written using PyCharm 2020.1.2 software, were used for the machine-learning processes.

The custom Python and MATLAB codes for the machine learning and the data analyses reported in this study are not yet publicly available owing to intellectual-property-filing issues, yet they are available for research purposes from the corresponding author on reasonable request.

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](#)

The main data supporting the results in this study are available within the paper and its Supplementary Information. Source data for the figures are provided with this paper. The raw datasets generated during the study are too large to be publicly shared, yet they are available for research purposes from the corresponding author on reasonable request.

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	269 patient serum samples were used to train and validate the machine-learning algorithms. No statistical methods were used to predetermine sample size. The sample size of the training set (n = 215) was determined on the basis of prediction scores of the cross-validation being larger than 0.90.
Data exclusions	No data were excluded from the analyses.
Replication	Fluorescence spectroscopy on each sample was done with triplicate, to confirm the consistency of the measurements.
Randomization	The order and sites of the blood draws, and the collection mechanism (pre-operative and post-operative blood draws) were randomized for each participant. For model development, the datasets were repeatedly (ten times) partitioned randomly into ten subsamples for cross-validation. For the test set, we used a new set of patient samples that was not used for model development.
Blinding	Blinding was not possible for method development, because of the need for knowledge of the disease status for cross validation.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Human research participants

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

Population characteristics	The study was limited to patient serum samples from adult women. Men were not included since they are not affected by ovarian cancer. There was no restriction or discrimination based on age, weight or other underlying health conditions. However, children were not included, because they are rarely diagnosed with ovarian cancer.
Recruitment	Waste blood samples were collected from female patients diagnosed with ovarian cancer, other diseases and healthy controls, under a protocol approved by the Memorial Sloan Kettering Cancer Center Institutional Review Board. No self-selection bias was present.
Ethics oversight	Memorial Sloan Kettering Cancer Center Institutional Review Board.

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