nature research

Corresponding author(s): Lyng FM - NP-P200545B

Last updated by author(s): Mar 5, 2021

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 our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	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 *Give P values as exact values whenever suitable.*
☒	☐	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	LabSpec 6 (HORIBA Scientific)
Data analysis	PLS Toolbox (Eigenvector) in MATLAB (Mathworks)

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 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

The datasets for Figures 3 – 8 presented in the Anticipated Results section are available as source data files.

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	Sample sizes were not determined as limited patient samples (n= 24 - 64) were available for each of the studies shown in Anticipated Results.
Data exclusions	Spectral data was excluded if it had poor signal to noise or was otherwise of poor quality. Outliers were removed using standard methods such as Grubb's test, Mahalanobis distance or Principal Components Analysis.
Replication	Data was reproduced successfully for n=50-64 (cervical), n=40 (oral) and n=24 (lung) patient samples
Randomization	Cervical samples were collected from patients attending a colposcopy clinic, lung samples were collected from patients attending for bronchoscopy and oral samples were collected from patients attending a dysplasia clinic. Age and gender was not found to significantly affect the spectral data.
Blinding	No blinding as the studies included were pilot studies aimed at developing classification models for pre-cancer.

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	Cervical participants were of screening age, 25-65 years and were female. Oral participants were 25 - 60 years (healthy volunteers) and 40 - 90 years (patients) and were approx. 55/45 male/female. Lung participants were approx 60/40 male/female and no ages were included in the data provided.
Recruitment	Patients were recruited by collaborating clinicians at colposcopy clinic (cervical), dysplasia clinic (oral) and bronchoscopy clinic (lung) and gave informed consent to be included.
Ethics oversight	Coombe Women and Infants University Hospital Research Ethics Committee (cervical), St James' Hospital/Tallaght University Hospital Joint Research Ethics Committee (oral and lung) and Technological University Dublin Research Ethics Committee

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