nature research

Corresponding author(s): Thomas J. Fuchs

Last updated by author(s): May 22, 2019

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	
☐	☒	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	Glass slides were digitized with Leica Aperio AT2 scanners and Philips Ultra Fast Scanner at a resolution of 0.5 microns per pixel.
Data analysis	The algorithms were written in python. We used openslide (version 3.4.1) to access the whole slide images, and pytorch (version 1.0) to train deep learning models. R (version 3.3.3) was used for the statistical analysis of the results.

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

The publicly shared MSK breast cancer metastases dataset is available at http://thomasfuchslab.org/data/ . The dataset consists of 130 de-identified whole slide images of axillary lymph node specimens from 78 patients (see Supplemental Figure 6). The tissue was stained with H&E and scanned on Leica Biosystems AT2 digital slide scanners at Memorial Sloan Kettering Cancer Center. Metastatic carcinoma is present in 36 whole slides from 27 patients and the corresponding label is included in the dataset.
The remaining data that supports the findings of this study were offered to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript upon request. The remaining data is not publicly available in accordance to institutional requirements governing human subject privacy protections.

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	No sample-size calculations were performed. Within the enrollment years listed in Figure 1a all cases with digitizes whole slides were included in the study without data curation.
Data exclusions	Less than ten whole slide images were excluded because of excessive pen ink marks present on the image. The exclusion criteria was pre-established.
Replication	Models were trained five times with each condition to ensure the stability of the training procedure. Replication was successful for all conditions for which test results were reported.
Randomization	Patients were randomly divided in three groups: training, validation, and test sets. No other covariates were controlled for.
Blinding	Since our experiments are based on digitized pathology slides, blinding is not necessary.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

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	Digital images of microscope slides from patients that were diagnosed at MSKCC over a period of at least 1 year and up to 5 years depending on the tissue type.
Recruitment	No patient recruitment was performed. All digital images that were available for the pre-established collecting period were analyzed.
Ethics oversight	Memorial Sloan Kettering Cancer Center

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