nature research

Corresponding author(s): John Le Quesne, Charles Swanton, Yinyin Yuan

Last updated by author(s): Apr 11, 2020

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	Main source of data collection is microscopic HE and IHC slide scanning. NanoZoomer Digital Pathology System version 3.1.7 (Hamamatsu, Japan) was used for digitalizing HE and IHC slides.
Data analysis	The deep learning pipeline for digital pathology image analysis is available for non-commercial research purposes: https://github.com/qalid7/compath. The pipeline runs on Python (version 3.5), TensorFlow (version 1.3) and MATLAB (version R2018b). All code used for statistical analyses of image data was developed in R (version 3.5.1) and is available: https://github.com/qalid7/tx100_compath.

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 digital pathology images from the TRACERx study generated or analysed during this study are not publicly available and restrictions apply to its use. A test subset of such digital pathology images are available through the Cancer Research UK & University College London Cancer Trials Centre (ctc.tracerx@ucl.ac.uk) for non-commercial research purposes and access will be granted upon review of a project proposal that will be evaluated by a TRACERx data access committee and entering into an appropriate data access agreement, subject to any applicable ethical approvals. Digital pathology images for LATTICe-A samples with expert pathologist's annotations used for validation is available: https://github.com/qalid7/compath. Request for data access for the remaining LATTICe-A samples can be submitted to J.L.Q.

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 statistical methods were used to predetermine sample size. For TRACERx, all patient tumor regions with histology of sufficient quality were included in this study. This represents the largest multi-region histology-omics study cohort and is sufficient for exploratory analyses. For LATTICe-A cohort all patient samples with histology of sufficient quality were included in this study. This represents the largest sample histology study cohort and is sufficient for validation analyses.
Data exclusions	Sample availability is the main reason of exclusion. 15 patients were excluded from the TRACERx regional histology cohort for "too little tissue" and 24 patients were excluded from the LATTICe-A cohort due to missing clinical data and/or not having a single tumor block. More details are provided in Extended Data Fig. 1.
Replication	Deep learning histology analysis and immune regional classification developed for TRACERx were applied directly to the external cohort (LATTICe-A) allowing the replication of the prognostic value of the number of immune cold regions in 970 patients. Other reported findings were not replicated.
Randomization	Not relevant to the study, as samples were split into immune hot and immune cold using a cohort median-based classification scheme.
Blinding	Not relevant to the study, as there was no control and treatment arms involved.

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

Antibodies

Antibodies used	For TRACERx multiplex T cell IHC slides (CD8/CD4/FOXP3) these antibodies were used: anti-CD8 (type: Rabbit Monoclonal, clone: SP239, cat. no.: ab178089, source: Abcam Plc, Cambridge, UK, used at 1:100); anti-CD4 (type: Rabbit Monoclonal, clone: SP35, cat. no.: ab213215, source: Abcam Plc, Cambridge, UK, used at 1:50); anti-FOXP3 (type: Mouse, clone: 236A/E7, source: kind gift from Dr G Roncador, CNIO, Madrid, Spain, used at: 1:100). For validating deep learning results against sequential IHC staining using LATTICe-A samples, the following antibodies were used: anti-human CD45 (type: Mouse Monoclonal, clone: 2B11 + PD7/26, cat. no.: M0701, source: Agilent DAKO, USA, used at 1:200); SMA (myofibroblast marker, type: Mouse Monoclonal antibody Smooth Muscle Actin (1A4), cat. no.: 760-2833, source: Roche, Switzerland), this is a ready to use antibody; TTF-1 (type: Novocastra Liquid Mouse Monoclonal antibody thyroid transcription factor 1, clone: SPT24, cat. no.: NCL-L-TTF-1, source: Leica biosystems, Germany, used at 1:100).
Validation	Antibodies with prior manufacturer validation were used where possible, data is available at the manufacturer's website according to the antibodies information provided above. The CD8/CD4/FOXP3 multiplex staining was validated on negative and positive controls. Immunohistochemistry and protein reactivity patterns were assessed by a pathologist with expertise in multiplex-immunostaining. CD45 was validated at several concentrations at both low and high pH to find the optimal assay parameters. Control tissue used for validation was FFPE human tonsil. This antibody was validated by the manufacturer for routine diagnostic use, but on different staining platform. SMA is a clinical antibody in routine diagnostic use, the control tissue used was FFPE human uterus, a suggested positive control. Both high and low ph. retrieval solutions were tested with a range of

incubation times to find the optimal assay parameters. TTF1 was validated at multiple concentrations using both high and low ph. retrieval conditions to find the optimal assay parameters. The control tissue used was FFPE human lung.

Human research participants

Policy information about studies involving human research participants

Population characteristics	There were 68 male and 32 female non-small cell lung cancer patients in the TRACERx study, with a median age of 68. The cohort is predominantly early-stage: Ia(26), Ib(36), IIa(13), IIb(11), IIIa(13), IIIb(1). Seventy-two had no adjuvant treatment and 28 had adjuvant therapy. Patients were recruited into TRACERx according to the following eligibility criteria (taken from the study protocol). Inclusion criteria: _Written Informed consent _Patients ≥18 years of age, with early stage I-IIIA disease who are eligible for primary surgery _Histopathologically confirmed NSCLC, or a strong suspicion of cancer on lung imaging necessitating surgery (e.g. diagnosis determined from frozen section in theatre) _Primary surgery in keeping with NICE guidelines planned (see section 9.3) _Agreement to be followed up in a specialist centre _Performance status 0 or 1 _Suspected tumour at least 15mm in diameter on pre-operative imaging Exclusion criteria: _Any other current malignancy or malignancy diagnosed or relapsed within the past 5 years (other than non-melanomatous skin cancer, stage 0 melanoma in situ, and in situ cervical cancer) _Psychological condition that would preclude informed consent _Treatment with neo-adjuvant therapy for current lung malignancy deemed necessary _Adjuvant therapy other than platinum-based chemotherapy and/or radiotherapy _Known Human Immunodeficiency Virus (HIV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) or syphilis infection. _Sufficient tissue, i.e. a minimum of two tumour regions, is unlikely to be obtained for the study based on pre-operative imaging Patient ineligibility following registration: _There is insufficient tissue _The patient is unable to comply with protocol requirements _There is a change in histology from NSCLC following surgery, or NSCLC is not confirmed during or after surgery. _Change in staging to IIIB/IV following surgery _The operative criteria are not met (e.g. incomplete resection with macroscopic residual tumours (R2)); see section 9.3 for a list of accepted surgical procedures. Patients with microscopic residual tumours (R1) are eligible and should remain in the study _Adjuvant therapy other than platinum-based chemotherapy and/or radiotherapy is administered. The external validation cohort was obtained from the Leicester Archival Thoracic Tumor Investigatory Cohort – Adenocarcinoma (LATTICe-A) study which consists of 970 University Hospitals of Leicester (UHL) Trust patients who underwent surgical treatment with curative intent for primary lung adenocarcinoma. 455 were men and 515 were women with a median age of 69. Most clinical data (age, sex, adjuvant therapy status and time to recurrence or death) were available for all patients, with complete pathological stage for 827 and smoking history for 651. LATTICe-A patients were initially identified using diagnostic histopathology reports from surgical specimens within UHL's local histopathology database. Inclusion criteria: • Surgical resection procedures included are wedge resection; lobectomy; bilobectomy; segmentectomy; and pneumonectomy. • Patients diagnosed with synchronous primary lung tumours, including non-adenocarcinoma tumours were considered eligible for inclusion • Patients >= 18 years of age. Exclusion criteria: • Patients must not have been diagnosed with any other lung tumour in the previous five years before surgical treatment for the primary lung adenocarcinoma. • Patients who have been diagnosed with metastatic lung adenocarcinoma. • Lung adenocarcinoma tumours diagnosed as a recurrence of a previous primary lung adenocarcinoma. • Patients who have only had biopsy surgery performed i.e. core biopsies/EBUS. • Patients who have been diagnosed with combined tumour types e.g. adenosquamous carcinoma or combined adenocarcinoma/high grade neuroendocrine carcinoma. • Patients who have been diagnosed with undifferentiated non-small cell lung tumours.
Recruitment	For TRACERx, patients seen with a new diagnosis of lung cancer in lung cancer units across the United Kingdom, according to the eligibility criteria above, were recruited. No selection bias has been identified to date. All patient tumor regions with sufficient histology quality were analyzed in this study. All patients were assigned a study ID that was known to the patient. These were subsequently converted to linked study Ids such that the patients could not identify themselves in study publications. All human samples, tissue and blood, were linked to the study ID and barcoded such that they were anonymised and tracked on a centralised database overseen by the study sponsor only. Informed consent for entry into the TRACERx study was mandatory and obtained from every patient.
Ethics oversight	TRACERx study was approved by the NRES Committee London with the following details: Study title: TRAcking non small cell lung Cancer Evolution through therapy (Rx) REC reference: 13/LO/1546 Protocol number: UCL/12/0279 IRAS project ID: 138871

The validation cohort (LATTICe-A study) was ethically approved by an NHS research ethics committee (ref. 14/EM/1159), more information can be found at: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/characterisation-of-thoracic-malignancies-using-archival-human-tissue/

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

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	TRACERx Lung https://clinicaltrials.gov/ct2/show/NCT01888601, approved by an independent Research Ethics Committee, 13/LO/1546
Study protocol	TRACERx: https://clinicaltrials.gov/ct2/show/NCT01888601
Data collection	For TRACERx, Clinical and pathological data is collected from patients during study follow up - this period is a minimum of five years. Data collection is overseen by the sponsor of the study (Cancer Research UK & UCL Cancer Trials Centre) and takes place in hospitals across the United Kingdom. A centralised database called MACRO is used for this purpose. For LATTICe-A, data was collected from local and national databases and compiled into a bespoke cohort database.
Outcomes	For TRACERx, disease-free outcome data were gathered for all patients. Last updated on 15 May 2018. For LATTICe-A, data on recurrence, cancer-specific survival and overall survival were gathered for all patients and revised periodically.