nature research

Corresponding author(s): Caroline Dive

Last updated by author(s): Aug 4, 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	NGS data was aligned using Burrows-Wheeler Aligner (BWA) (v0.7.13) algorithm.
Data analysis	Data analysis was performed using R version 3.5.1. , R packages survival (v2.38), survminer (v0.3) survivalROC (v1.0.3),Picard(v 1.96), Bedtools(v 2.20.1), Samtools(0.1.19), GATK(v 3.5), HMMcopy (version 1.8.0), MuTect(v1.1.7), Annovar (hg19 reference) (hg19_20170228), https://www.cancergenomeinterpreter.org/home.

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 majority of data generated or analysed during this study are included in this published article. The sequencing data are available through the Cancer Research UK & University College London Cancer Trials Centre 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.

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. All the patients recruited from the start of the study until April 2017 who had given consent for pulmonary vein blood sampling, in the clinical sites having this procedure in place, were analysed for PV-CTC count. The final number of patients (n=100) was sufficient for the exploratory analysis performed in this study.
Data exclusions	Reasons for patient exclusion accordingly to the TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601) are indicated in the consort diagram in the manuscript. For the mutation analysis, several filters were applied in order to generate high-confidence set of variant calls from single PV-CTCs. The filters applied are indicated in the Methods. Exclusion criteria were pre-established on the basis of previous publications reporting these criteria as reliable to call confident mutations in single cells.
Replication	To verify the accuracy of the WES findings, a further sequencing, targeting the mutations identified by WES, was performed in all single cells.
Randomization	No randomization was required as this study was not comparing different treatments.
Blinding	No blinding was required for this study as no comparative arms were included.

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	The NSCLC cohort in this study consisted of lung adenocarcinoma (LUAD) (59%) and remaining 41% of non-adenocarcinoma histology. The median age of patient was 68 and the population consisted of 61 males and 39 females. 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 -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.
Recruitment	Patients with a diagnosis of NSCLC (stages IA-IIIA) were recruited according to the eligibility criteria (indicated above). No selection bias has been identified to date. Informed consent for entry into the TRACERx study was mandatory and obtained from every patient.
Ethics oversight	The study is conducted in accordance with the World Medical Association Declaration of Helsinki entitled 'Ethical Principles for Medical Research Involving Human Subjects' (1996 version) and in accordance with the terms and conditions of the ethical approval given to the study. The study has received a favourable opinion from the NRES Committee London – Camden & Islington Research Ethics Committee.

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