nature research

Corresponding author(s): Sergio A. Quezada and Charles Swanton

Last updated by author(s): Mar 12, 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	Flow cytometry data was acquired using BD DIVA V8.0.1 for Microsoft windows 10.
Data analysis	All analysis carried out in R version 3.4.3, code is available on request. Versions of all packages used were as follows; flowAI 1.6.3, flowCore 1.44.1, FlowSOM 1.10.0, ConsensusClusterPlus 1.42.0, uwot 0.0.0.9010, edger 3.20.8, NetMHCpan 2.8, NetMHC 4.0, Decombinator 4.0.1, Trimgalore – 0.4.5STAR – 2.5.4, RSEM – 1.3.1, scimpute 0.0.9, ComplexHeatmap 1.99.5, fgsea 1.4.1, CellPhoneDB 2.1.1, XCell 1.1.0, nlme 3.1-131, survival 2.41-3, survminer 0.4.2

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 tumour region RNA sequencing data, bulk RNA sequencing data from sorted T cells, single cell RNA sequencing data from sorted neoantigen-reactive T cells, TCR sequencing data from sorted T cells, and flow cytometry data (in each case from the TRACERx study) used or analysed during this study 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.

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	Samples for flow cytometry were selected from the TRACERx 100 cohort. All samples with available single cell digest material were used and those with paired exome sequencing prioritized. No sample size calculation was done. Available samples were of sufficient size to permit multiplicity corrected univariate and multivariate analyses described in the text.
Data exclusions	Precision of subset frequency in flow cytometry analysis is dominated by gated event number up to 10 events (Roederer Cytometry Part A 2008; DOI 10.1002/cyto.a.20549). To minimize capture of 10 events or less per cluster (20-26 clusters) we applied a sample QC threshold of 2000 live CD3+ events or 1000 live CD3+8+ or CD3+8- cells in lineage specific analyses . Regions with fewer than noted thresholds of live cells were excluded from subsequent analysis.
Replication	We sampled a second set of patients from the TRACERx 100 cohort for flow cytometry with an overlapping antibody panel; this cohort was not subjected to unsupervised clustering. Analyses carried out at the RNA level were replicated between TRACERx and TCGA cohorts.
Randomization	Patients were not grouped
Blinding	Blinding was not relevant for this observational study

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

Cells were stained in BD Brilliant stain buffer (BD cat# 563794) using fixable Viability Dye eFluor 780 (Thermo, cat# 65-0865-14) according to the manufacturer's instructions. The following monoclonal antibodies were used to surface stain samples in cohort 1: BUV395 conjugated antibody to human CD45RO (clone UCLH1; BD cat# 564291); BUV496 conjugated antibody to human CD8 (clone RPA-T8; BD cat# 564804); BUV563 conjugated antibody to human CD45RA (clone HI100; BD cat# 565702); BUV661 conjugated antibody to human CD4 (clone SK3; BD cat# 566003); BUV737 conjugated antibody to human CD28 (clone 28.2; BD cat# 564438); BUV805 conjugated antibody to human CD3 (clone SK7; BD cat# 565511); BV421 conjugated antibody to human PD-1 (clone EH12; BD cat# 562516); BV605 conjugated antibody to human CD57 (clone NK-1; BD cat# 563896); BV786 conjugated antibody to human CD27 (clone L128; BD cat# 563327), BV650 conjugated to 41BB (4B4-1 cat# 564092), BB515 conjugated antibody to TIM3 (7D3 cat# 565568), APC-R700 conjugated antibody to CD25 (2A3 cat# 565106), all from BD and anti-ICOS conjugated to PE-CY7 (C398.4A cat# 313520) from Biolegend. Intracellular staining was performed with FOXP3 AF647 (259D/C7 cat# 560045), TBET PE (4B10 cat# 644810), GZMB PE-CF594 (GB11 cat# 562462), Ki67 BV480 (B56 cat# 566109) all from from BD and EOMES PerCP-e710 (WD1928 cat# 46-4877-42) from ThermoFisher Scientific. All antibodies were used at 2ul per 50ul staining volume with the exception of CD8 (1ul), CD45RA (0.5ul), Ki67 (1ul), 4-1BB (2.5ul), GZMB (0.5ul), FOXP3 (10ul).

For cohort 2, cells were stained with the following antibodies to surface markers; CD8 BUV496 (clone RPA-T8 cat# 564804), CD45RA BUV563 (clone HI100 cat# 565702), HLA-DR BUV661 (clone G46-6 cat# 565073), Fas BUV737 (clone DX2 cat #612790), CD3 BUV805 (SK7 cat# 565511), PD1 BV421 (EH12.1 cat# 562516), CD57 BV605 (NK-1 cat# 563896), CD127 BB515 (HIL-7R-M21cat# 564423), CD28 APC-R700 (28.2, cat# 564438) all from BD and CD4 biotin (OKT4, cat# 317046), CD27 BV510 (O323 cat# 563092), CCR7 BV650 (G043H7 cat# 563407), CD103 BV711 (Ber-ACT8 cat#561162), ICOS BV786 (C398.4A cat#313534) from Biolegend. Streptavidin BUV395 was purchased from BD (cat#564176). Intracellular staining was performed using antibodies

specific for EOMES PerCP-e710 (WD1928 cat# 46-4877-42) and FOXP3 PE (PCH101 cat#15-4776-42) from Thermo Fisher scientific, CTLA4 PE-CY7 (L3D10 cat#349913) from Biolegend and GZMB PE-CF594 (GB11 cat# 562462) from BD. All antibodies were used at 2ul per 50ul staining volume with the exception of CD8 (1ul), CD45RA (0.5ul), GZMB (0.5ul).

For the TCF7 and CD39 staining panel, the indicated samples were stained with CD3, PD-1, CD45RA, CD8 and FOXP3 as above and CD4 BB790 (SK3, custom synthesis, BD), CD57 APC (QA17A04, cat# 393306 Biolegend), TCF7 PE (7F11A10 cat# 655208 Biolegend) CD39 BUV737 (TU66 cat# 654726 BD) using 2ul in 50ul total volume.

Phenotypic characterization of neoantigen-specific CD8 T cells in L011, L012 and L021 was performed using antibodies including CD3 BUV395 (SK7 cat#364000), CD4 BV785 (SK3 cat#563877), CD8 BV510 (RPA-T8 cat#563256), CD45RA BV711 (HI100 #cat563733), CD57 BV421 (NK-1 cat#563896), Ki67 FITC (clone #556026) from BD and PD1 BV605 (EH12.2H7 cat#329924) ICOS PE-Cy7 (C398.4A cat#313520) from Biolegend. Staining was performed as above using 2ul of each antibody in 50ul total.

Staining prior to sorting for TCRseq analysis was performed using the following antibodies across two panels. BV421 conjugated antibody to human PD-1 (clone EH12; BD cat# 562516); BV605 conjugated antibody to human CD57 (clone NK-1; BD cat# 563896); and CD127 BB515 (HIL-7R-M21cat# 564423) and CD25 APC-R700 (2A3 cat#565106) from BD and CD4 BV650 (OKT4 cat#317436), CD3 BV785 (SK7 cat#344842) or FITC (#344804), CD8 BV785 (SK1 cat#344740), ICOS PE (C398.4A cat#313508), CD45RA APC (HI100 cat#304112) from Biolegend. Volumes used were 2ul in 50ul unless otherwise indicated for the same target above. Gating strategy for CD8 and CD4 T cell subset sorts are indicated in Figure 3.

Validation

Panel optimization, set up and validation was performed using UltraComp ebeads (01-222-41, Thermo Fisher) and multiple aliquots of CRUK0079 NTA and region 4 TIL single-cell digests with unstained controls, FMO and antibody titrations where required. Data were acquired on a BD Symphony flow cytometer and cells gated for size, singlets, viability and CD3+CD8+ or CD3+CD8+ T cells in FlowJo v10 (Treestar) for further analysis.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Full patient demographics are available in Extended figure 1. TRACERX100 cohort demographics are available in Hajani et al NEJM 2017 referenced in the main text (ref 33). 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. The details of TRACERx recuritment are avialble at http://www.cruklungcentre.org/Research/TRACERx and https://clinicaltrials.gov/ct2/show/NCT01888601. Criteria
Inclusion Criteria:
• Written Informed consent
• Patients ≥18 years of age, with early stage IIA-IIIB disease (according to TNM 8th edition) who are eligible for primary surgery. Patients with a radiological staging of IB (N0) who could be upstaged to IIA-IIIB following surgery (due to the presence of possible nodal involvement on the pre-operative scan) may also be included, but will be withdrawn if post-surgical staging remains IB. 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 at a TRACERx site
• Performance status 0 or 1
• Minimum tumour diameter at least 15mm to allow for sampling of at least two tumour regions (if 15mm, a high likelihood of nodal involvement on pre-operative imaging required to meet eligibility according to stage, i.e. T1N1-3)
Exclusion Criteria:
• Any other* malignancy diagnosed or relapsed at any time, which is currently being treated (including by hormonal therapy).
• Any other* current malignancy or malignancy diagnosed or relapsed within the past 3 years**.
*Exceptions are: non-melanomatous skin cancer, stage 0 melanoma in situ, and in situ cervical cancer
**An exception will be made for malignancies diagnosed or relapsed more than 2, but less than 3, years ago only if a pre-operative biopsy of the lung lesion has confirmed a diagnosis of NSCLC.
• Psychological condition that would preclude informed consent
• Treatment with neo-adjuvant therapy for current lung malignancy deemed necessary
• Post-surgery staging is not IIA-IIIB
• 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

Recruitment

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 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

The TRACERx study (Clinicaltrials.gov no: NCT01888601) is sponsored by University College London (UCL/12/0279) and has been approved by an independent Research Ethics Committee (13/LO/1546).

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	NCT01888601
Study protocol	See https://www.ctc.ucl.ac.uk/TrialDetails.aspx?Trial=102. Alternatively, the study protocol is available at NEJM.org linked to Jamal-Hanjani et al NEJM 2017 (PMID: 28445112)
Data collection	Clinical data is primarily collected from the patient records of 17 UK NHS Trusts by site research staff and entered into a trial database. Some clinical data is obtained directly from patients through a one-off questionnaire. Data is also generated from blood and tissue samples that are collected longitudinally across 5 years of follow-up for each patient. Following patient recruitment, data is collected at baseline and during the surgery. After surgery/completion of adjuvant chemotherapy (if given), patients will be followed up according to national guidelines, which is every 3 months for the first 2 years, then every 6 months in years 3-5. Once the 5-year on-study assessments have been completed, minimal follow-up for date of death and date(s) of relapse(s) will continue until death, withdrawal or end of study.
Outcomes	The primary outcome measures are intratumour heterogeneity quantified by the ratio index ITB, disease-free survival and overall survival. These will be assessed by analysing the longitudinal samples and clinical data collected from patients and their medical records. ITB is defined as the mean number of non–ubiquitious mutations per region or lymph node divided by the number of ubiquitious mutations. DFS is the time from study registration until disease recurrence, new primary cancer or death from any cause, whichever occurred first (patients who had none of these were censored at the date they were last known to be alive from the case report forms). OS is the time from study registration until death from any cause. For secondary outcome measures: as TRACERx is an observational study in which a large amount of data will be collected, accordingly there will be a wide range of subsidiary analyses using data that has already been collected from patients. Examples of these include: developing and validating the intratumour heterogeneity ratio index ITB as a prognostic or predictive biomarker, and inferring a complete picture of NSCLC evolutionary dynamics.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Detailed in Methods
Instrument	Detailed in Methods
Software	Detailed in Methods
Cell population abundance	Detailed in Methods
Gating strategy	Detailed in Methods

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.