ARTICLE

doi:10.1038/nature22364

Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution

A list of authors and their affiliations appears in the online version of the paper.

The early detection of relapse following primary surgery for non-small-cell lung cancer and the characterization of emerging subclones, which seed metastatic sites, might offer new therapeutic approaches for limiting tumour recurrence. The ability to track the evolutionary dynamics of early-stage lung cancer non-invasively in circulating tumour DNA (ctDNA) has not yet been demonstrated. Here we use a tumour-specific phylogenetic approach to profile the ctDNA of the first 100 TRACERx (Tracking Non-Small-Cell Lung Cancer Evolution Through Therapy (Rx)) study participants, including one patient who was also recruited to the PEACE (Posthumous Evaluation of Advanced Cancer Environment) post-mortem study. We identify independent predictors of ctDNA release and analyse the tumour-volume detection limit. Through blinded profiling of postoperative plasma, we observe evidence of adjuvant chemotherapy resistance and identify patients who are very likely to experience recurrence of their lung cancer. Finally, we show that phylogenetic ctDNA profiling tracks the subclonal nature of lung cancer relapse and metastasis, providing a new approach for ctDNA-driven therapeutic studies.

Lung cancer is the leading cause of cancer-related deaths[1,2]. Metastatic non-small-cell lung cancer (NSCLC) cannot be cured with systemic chemotherapy, yet clinical studies have shown a 5% benefit of post-operative (adjuvant) chemotherapy on overall survival[3]. This modest survival benefit may reflect a vulnerability of low-volume disease in the context of reduced intra-tumour heterogeneity[4]. ctDNA detection in plasma following resection of breast[5,6] and colorectal[7] tumours has been shown to identify patients who are very likely to relapse post-operatively in advance of established clinical parameters. Identifying, monitoring and genomically characterizing residual disease following primary lung cancer surgery may improve outcomes in the adjuvant setting. This would create a therapeutic setting in which only patients who are very likely to have cancer recurrence would receive treatment and intervention could be directed to the evolving tumour subclone that seeds metastatic recurrence.

Here, we report a bespoke multiplex-PCR next-generation sequencing (NGS) approach to ctDNA profiling in the context of the prospective tumour evolutionary NSCLC TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601). We address determinants of ctDNA detection in early-stage NSCLC and investigate the ability of ctDNA to identify and genomically characterize postoperative NSCLC relapse within a tumour phylogenetic framework.

Phylogenetic ctDNA profiling

The TRACERx study monitors the clonal evolution of NSCLC from diagnosis through to death[8,9]. Using multi-region exome sequencing (M-seq)-derived tumour phylogenetic trees developed through prospective analysis of 100 patients from the TRACERx cohort, we conducted a phylogenetic approach to ctDNA profiling in early-stage NSCLC (Fig. 1). Bespoke multiplex-PCR assay panels were synthesized for each patient, targeting clonal and subclonal single-nucleotide variants (SNVs) selected to track phylogenetic tumour branches in plasma (Fig. 1). SNV detection in plasma was established through a calling algorithm using negative control samples (see Methods). Analytical validation of the multiplex-PCR NGS platform demonstrated a sensitivity of above 99% for the detection of SNVs at frequencies above 0.1% and the specificity of detecting a single SNV was 99.6% (Extended Data Fig. 1a). At least two SNVs were detected in ctDNA from early-stage NSCLCs that had been analysed in our published discovery cohort data[10], suggesting biological sensitivity of a two-SNV threshold for ctDNA detection. Therefore, we selected a threshold of two detected SNVs for calling a sample ctDNA-positive for validation within this study, to minimize type I errors when testing up to 30 tumour-specific SNVs per time point in a single patient (see Extended Data Fig. 1b for justification).

Determinants of ctDNA detection in NSCLC

We sought to identify clinicopathological determinants of ctDNA detection in early-stage NSCLC by profiling preoperative plasma samples in 100 TRACERx patients. Samples from four patients could not be analysed (see cohort design in Extended Data Fig. 2a, patient characteristics in Extended Data Table 1a–c and Supplementary Table 1). Individual assay panels were designed to target a median of 18 SNVs (range, 10–22), comprising a median of 11 clonal SNVs (range, 2–20) and a median of 6 subclonal SNVs (range, 0–16) per patient (Extended Data Fig. 2b, e).

At least two SNVs were detected in ctDNA preoperatively in 46 out of 96 (48%) early-stage NSCLCs and a single SNV was detected in 12 additional cases (Fig. 2a). Centrally reviewed pathology data revealed that ctDNA detection was associated with histological subtype: 97% (30 out of 31) of lung squamous cell carcinomas (LUSCs) and 71% (5 out of 7) of other NSCLC subtypes were ctDNA-positive, compared with 19% (11 out of 58) of lung adenocarcinomas (LUADs) (Fig. 2a). ctDNA was detected in 94% (16 out of 17) of stage I LUSCs compared with 13% (5 out of 39) of stage I LUADs (Extended Data Fig. 3a). Passive release of ctDNA into the circulation may be associated with necrosis[11]. As expected, LUSCs were significantly more necrotic than LUADs[12] and ctDNA-positive LUADs formed a sub-group of more necrotic tumours compared with ctDNA-negative LUADs (Extended Data Fig. 3b). Necrosis, lymph-node involvement, lymphovascular invasion, pathologic tumour size, Ki67-labelling indices, non-adenocarcinoma histology and total cell-free DNA input predicted ctDNA detection in univariable analyses (Extended Data Fig. 3c). Multivariable analysis revealed non-adenocarcinoma histology, the

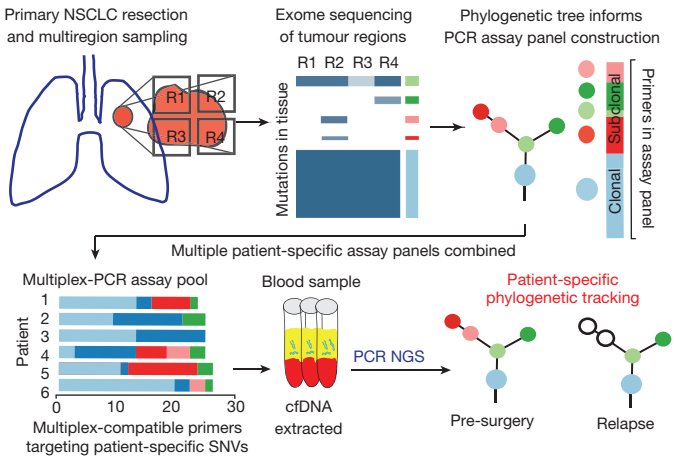


Figure 1 | Phylogenetic ctDNA tracking. Overview of the study methodology. Multi-region sequencing of NSCLC was performed as part of the TRACERx study. PCR assay panels were designed on the basis of the phylogenetic analysis, targeting clonal and subclonal SNVs to facilitate non-invasive tracking of the patient-specific tumour phylogeny. Assay panels were combined into multiplex assay pools containing primers from up to 10 patients. Cell-free DNA (cfDNA) was extracted from pre- and postoperative plasma samples and multiplex-PCR was performed, followed by sequencing of the amplicons. Findings were integrated with M-seq exome data to track tumour evolution.

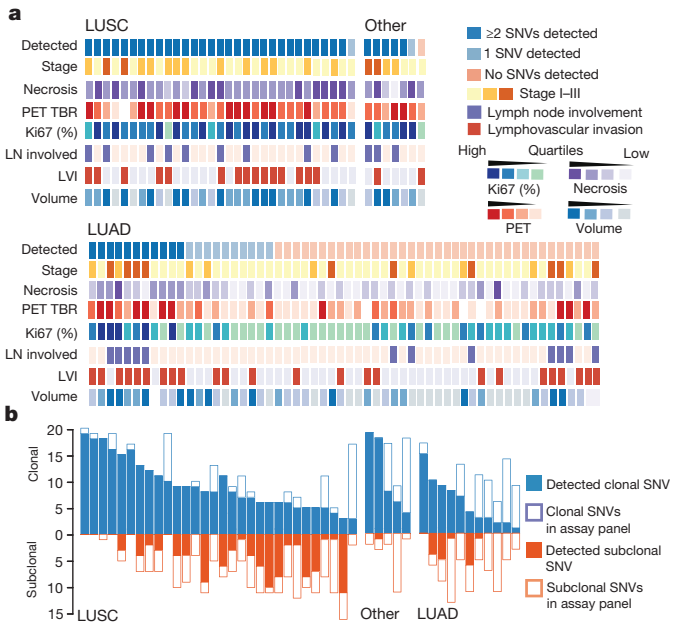


Figure 2 | Clinicopathological predictors of ctDNA detection. a, Heat map showing clinicopathological and ctDNA detection data; continuous variables are shown in quartiles. LN, lymph node; TBR, tumour-to-background ratio; LVI, lymphovascular invasion. Raw data and patient identifications are shown in the Source Data. **b**, Detection of clonal and subclonal SNVs within 46 patients with two or more SNVs detected in the plasma. Histology indicated in panels as LUSC, LUAD and other.

presence of lymphovascular invasion and a high Ki67 proliferation index as independent predictors of ctDNA detection (Extended Data Fig. 3c). Because ^{18}F-fluoro-deoxyglucose (FDG) avidity on positron emission tomography (PET) scans correlates with proliferative indices in early-stage NSCLC[13,14], we investigated tumour PET FDG avidity and ctDNA detection. PET FDG avidity predicted ctDNA detection (area under the curve = 0.84, $P < 0.001$, $n = 92$) (Extended Data Fig. 3d). Within LUADs, driver events in *KRAS*, *EGFR* or *TP53* were not associated with ctDNA detection (Extended Data Fig. 3e).

We analysed the distribution of clonal and subclonal SNVs in ctDNA-positive patients. Clonal SNVs were detected in all 46 ctDNA-positive patients and a median of 94% (range, 11–100%) of clonal SNVs targeted by assay panels were detected in the ctDNA of these patients. Of 46 ctDNA-positive patients, 40 had subclonal SNVs targeted by the assay panels, and subclonal SNVs were detected in 27 (68%) of these patients. A median of 27% (range, 0–91%) of subclonal SNVs within individual assay panels were detected in ctDNA-positive patients (Fig. 2b). The mean plasma variant allele frequency (VAF) of clonal SNVs was higher than that of subclonal SNVs (Extended Data Fig. 4a, within patient comparison, Wilcoxon signed-rank test, $P < 0.001$, $n = 27$, Supplementary Table 2), supporting the use of clonal alterations as a more sensitive method of ctDNA detection than subclonal alterations[10,15].

In ctDNA-positive patients, pathologic tumour size correlated with the mean plasma VAF of clonal SNVs (Spearman's $\rho = 0.405$, $P = 0.005$, $n = 46$, Extended Data Fig. 4b). Computed tomography (CT) scan volumetric analyses were evaluated in 37 out of 46 ctDNA-positive patients (see Extended Data Fig. 4c). Tumour volume from CT analyses correlated with mean clonal plasma VAF (Spearman's $\rho = 0.63$, $P < 0.001$, $n = 37$, Fig. 3a). A linear relationship between log-transformed volume and log-transformed mean clonal VAF values was observed (Fig. 3a). The line of best fit applied to the data was consistent with the line fitted to NSCLC volumetric data and ctDNA plasma VAFs that have been reported previously[16] (Extended Data Fig. 4d). Linear modelling based on the TRACERx data predicted that a primary tumour burden of 10 cm^3 would result in a mean clonal plasma VAF of 0.1% (95% confidence interval, 0.06–0.18%) (Fig. 3b). Tumour purity was multiplied by tumour volume to control for stromal contamination, to determine the relationship between cancer cell volume and

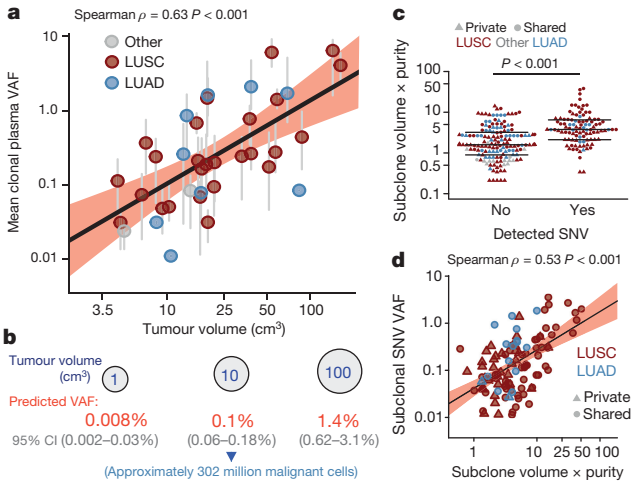


Figure 3 | Tumour volume predicts plasma VAF. a, Tumour volume (cm^3) measured by CT volumetric analysis correlates with mean clonal plasma VAF. $n = 37$, grey vertical lines represent range of clonal VAF, red shading indicates 95% confidence intervals (CIs). **b**, The predicted mean clonal VAF at hypothetical volumes ranging from 1 cm^3 to 100 cm^3 is based on the model in **a**, the predicted cancer cell number is based on the model in Extended Data Fig. 4e. **c**, Estimated effective subclone size, defined as the mean cancer cell fraction (CCF) of the subclone multiplied by tumour volume and purity, influences subclonal SNV detection. For negative calls, median effective subclone size was 1.70 cm^3 (range, 0.21–24.11 cm^3, $n = 163$); for positive calls, median effective subclone size was 4.06 cm^3 (range, 0.31–49.20 cm^3, $n = 109$). Wilcoxon rank-sum test, $P < 0.001$; data from 34 patients (passed volumetric filters with subclonal SNVs represented in assay panel). Median and 25th and 75th quartiles are indicated by thin horizontal lines. **d**, Estimated effective subclone size correlates with subclonal plasma VAF. $n = 109$ subclonal SNVs, data from 34 patients (passed volumetric filters with detected subclonal SNVs in plasma). Red shading indicates 95% confidence interval.

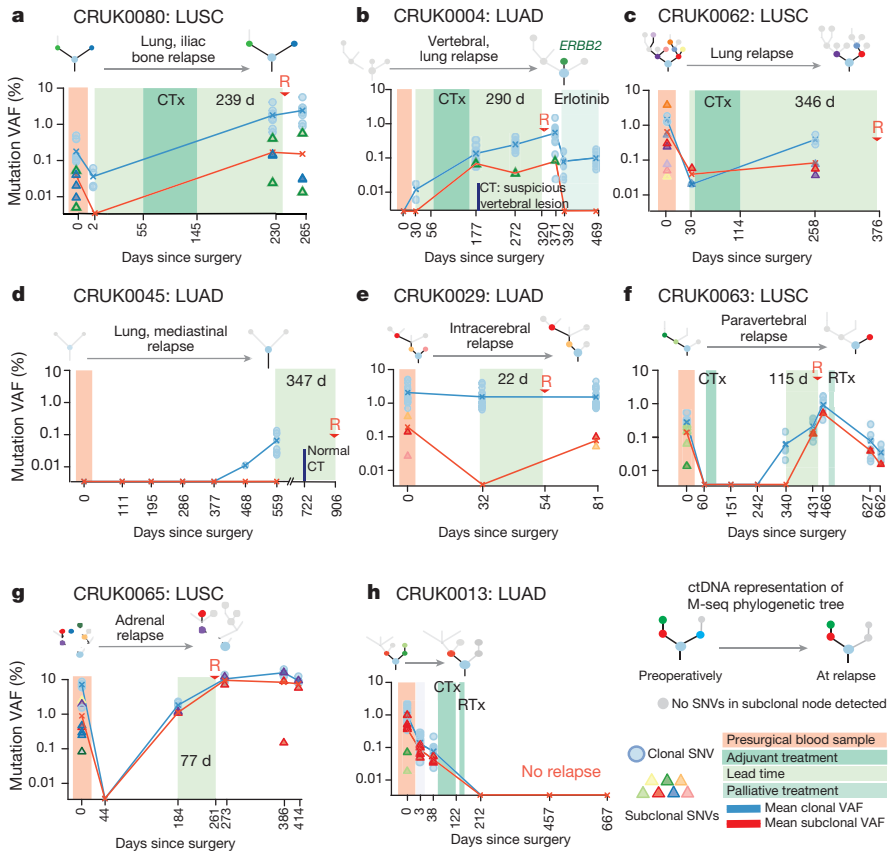


Figure 4 | Postoperative ctDNA detection predicts and characterizes NSCLC relapse. a–h, Longitudinal cell-free DNA profiling of indicated patients. ctDNA detection in plasma was defined as the detection of two tumour-specific SNVs. Detected clonal (circles, light blue) and subclonal (triangles, colours indicate different subclones) SNVs from each patient-specific assay panel are plotted on graphs coloured by M-seq-derived tumour phylogenetic nodes. Mean clonal (blue) and mean subclonal (red) plasma VAF are indicated on graphs as connected lines. Preoperative and relapse M-seq-derived phylogenetic trees represented by ctDNA are illustrated above each graph. CTx, chemotherapy; RTx, radiotherapy; R, relapse. Lead time is indicated by the number of days in the panels.

clonal plasma VAF (Extended Data Fig. 4e). On the assumption that a cancer cell volume of 1 cm^3 contains 9.4×10^7 cells[17], a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 302 million tumour cells (Fig. 3b and Extended Data Fig. 4f).

To investigate predictors of subclone detection, detected subclonal SNVs were mapped back to M-seq-derived tumour phylogenetic trees. Of 57 shared subclones (identified in more than one tumour region through M-seq analysis), 35 (61%) were identified in ctDNA, compared with 26 out of 80 (33%) private subclones (detected in only a single tumour region) (Extended Data Fig. 4g). This suggested that subclone volume influences subclonal ctDNA detection. Subclone volume was estimated on the basis of mean regional subclone cancer cell fraction and cancer cell volume. Detected subclonal SNVs mapped to subclones with significantly higher estimated volumes than did subclones containing undetected SNVs (Wilcoxon rank-sum test, $P < 0.001$, $n = 272$, Fig. 3c) and subclone volume significantly correlated with subclonal SNV plasma VAF (Spearman's $\rho = 0.53$, $P < 0.001$, $n = 109$, Fig. 3d).

Detecting and characterizing NSCLC relapse

The longitudinal phase of the study aimed to determine whether ctDNA profiling with patient-specific assay panels could detect and characterize the branched subclone(s) seeding the NSCLC relapse. Pre- and post-surgical plasma ctDNA profiling was performed blinded to relapse status in a sub-group of 24 patients (cohort characteristics, Extended Data Table 1d, e). This included relapse-free patients who had been followed up for a median of 775 days (range, 688–945 days; $n = 10$) and confirmed NSCLC-relapse cases ($n = 14$) (cohort design, Extended Data Fig. 2c). Assay panels were redesigned in this phase of the study to target additional clonal SNVs in LUADs to attempt to improve ctDNA detection. A median of 18.5 SNVs (range, 12–20) were targeted by the LUSC assay panels, comprising a median of 8.5 clonal SNVs (range, 3–18) and a median of 9.5 subclonal SNVs (range, 2–17) (Extended Data Fig. 2d, e). A median of 28 SNVs (range, 25–30) were targeted by the LUAD assay panels, comprising a median of 20.5 clonal SNVs (range, 13–26) and a median of 6 subclonal SNVs (range, 0–13) (Extended Data Fig. 2d, e).

Patients were followed up every three months for the first two years following study enrolment and every six months thereafter with clinical assessment and chest radiographs. At least two SNVs were detected in 13 out of 14 (93%) patients with confirmed NSCLC relapse before, or at, clinical relapse (Fig. 4a–g and Extended Data Fig. 5). At least two SNVs were detected in 1 out of 10 (10%) patients (CRUK0013) with no clinical evidence of NSCLC relapse (Fig. 4h and Extended Data Fig. 6). Excluding a single case where no postoperative plasma was taken before clinical relapse (CRUK0041), the median interval between ctDNA detection and NSCLC relapse that was confirmed by CT imaging indicated by clinical and chest radiograph follow-up (lead time) was 70 days (range, 10–346 days). Of the 13 relapse cases, 4 exhibited lead times of more than six months (Fig. 4a–d). In two cases ctDNA detection preceded CT imaging inconclusive for NSCLC relapse by 157 days (CRUK0004, Fig. 4b) and 163 days (CRUK0045, Fig. 4d). ctDNA profiling reflected adjuvant chemotherapy resistance; CRUK0080, CRUK0004 and CRUK0062 had detectable ctDNA in plasma within 30 days of surgery. The number of detectable SNVs increased in all three cases despite adjuvant chemotherapy, with disease recurring within

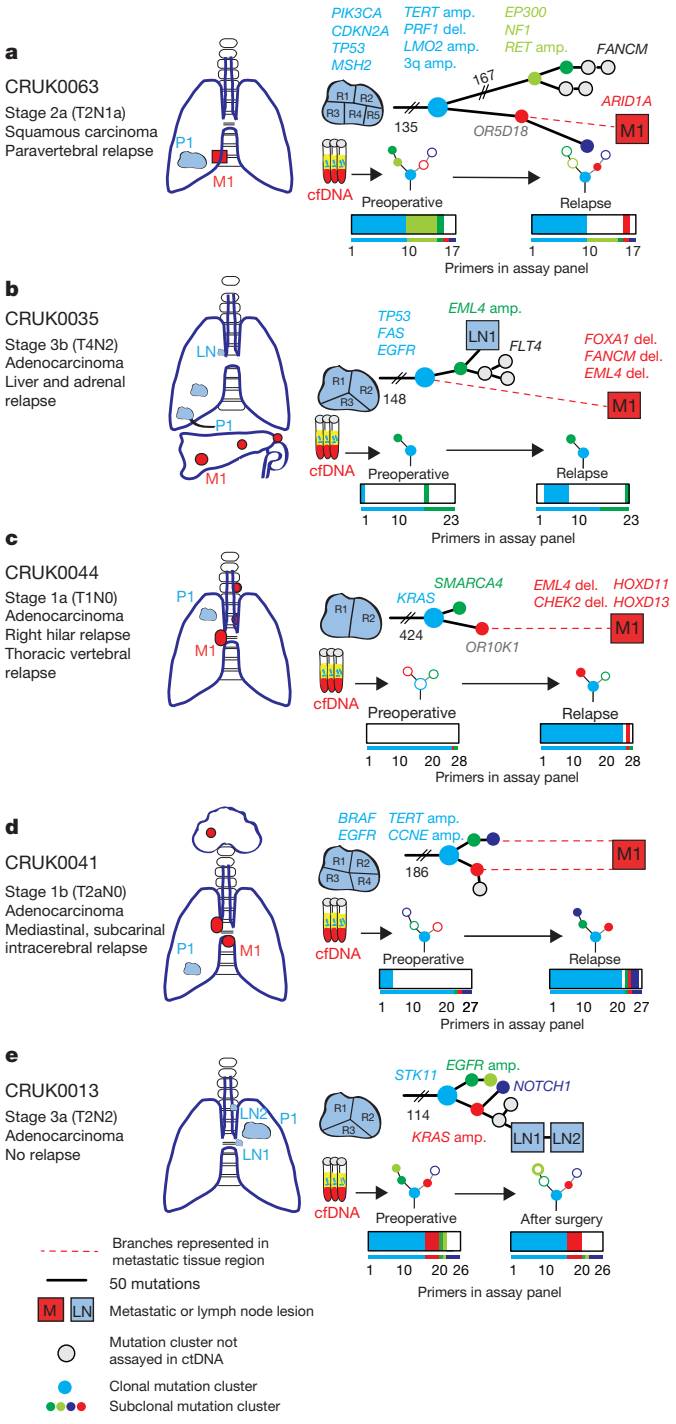


Figure 5 | Phylogenetic trees incorporating relapse tissue sequencing data. a–e, Phylogenetic trees were based on the mutations found in primary and metastatic tissue (**a**–**d**), or primary tumour and lymph node biopsies (**e**). Coloured nodes in the phylogenetic trees indicate cancer clones containing mutations assayed for in ctDNA, grey indicates a clone not assayed. Branch length is proportional to the number of mutations unless crossed. Dashed red lines show branches leading to metastatic relapse. Coloured bars show the number of assays per sample detected preoperatively and at relapse (**a**–**d**) or, in the absence of relapse, preoperatively and after surgery (**e**). The thin coloured bar shows the number of assays in total. Colours match clones on the phylogenetic trees. amp., amplification; del., deletion; M, metastasis; LN, lymph node; P1, primary site; R1–R5, region within the primary tumour. Cancer stages are indicated on the basis of the pathological tumour, node and metastasis (TNM) nomenclature.

one year after surgery (Fig. 4a–c). In contrast, CRUK0013 had 20 SNVs detectable in ctDNA 72 h after surgery and 13 SNVs detectable before adjuvant chemoradiotherapy; 51 days after completion of adjuvant treatment and at postoperative days 457 and 667, no SNVs were detectable and the patient remains relapse-free 688 days post-surgery (Fig. 4h). ctDNA profiling detected relapse from the central nervous system; CRUK0029 had a PET scan performed 50 days before surgery demonstrating normal cerebral appearances. ctDNA remained detectable following surgery; 54 days postoperatively the patient was diagnosed with intracerebral metastasis; no extracranial disease was evident on CT imaging (Fig. 4e).

In cases where subclonal SNVs were detected in ctDNA postoperatively, we predicted subclonal clusters involved in the relapse process by mapping SNVs detected in plasma back to primary M-seq data (Fig. 4 and Extended Data Fig. 5b, c). Subclonal SNVs displaying plasma VAFs similar to those of clonal SNVs from clusters confined to a single phylogenetic branch were detected postoperatively in the ctDNA of four patients who suffered NSCLC relapse (CRUK0004, CRUK0063, CRUK0065 and CRUK0044) (Fig. 4b, f, g and Extended Data Fig. 5b). This suggested a relapse process dominated by one subclone represented in our assay panel. The subclone implicated by ctDNA as driving the relapse in the case of CRUK0004 contained an *ERBB2* (also known as *HER2*) amplification event (more than 15 copies, triploid background), which may be targetable in NSCLC[18] (Fig. 4b). Relapses involving subclones from more than one phylogenetic branch were evident in patients CRUK0080, CRUK0062 (Fig. 4a, c) and CRUK0041 (Extended Data Fig. 5c).

Validation of phylogenetic characterization

To validate subclonal ctDNA analyses, data acquired from sequencing metastatic tissue at recurrence were integrated with M-seq primary tumour data (for biopsy details, see Supplementary Table 3). Patient CRUK0063 suffered para-vertebral relapse of their NSCLC. Postoperative ctDNA analysis revealed the detection of the same subclonal SNV (*OR5D18*) on four consecutive occasions over a 231-day period (Extended Data Fig. 7a). The *OR5D18* SNV traced back to a subclonal cluster private to primary tumour region three (Fig. 5a). CT-guided biopsy tissue was acquired from the para-vertebral metastasis at postoperative day 467. Exome sequencing of the relapse tissue revealed that the subclonal cluster in the primary tumour containing the *OR5D18* SNV probably gave rise to the metastatic subclone (Fig. 5a), consistent with ctDNA phylogenetic characterization of the relapse. The para-vertebral biopsy contained 88 SNVs not called as present in the primary tumour, including an *ARID1A* stop-gain driver SNV. Re-examination of primary tumour region M-seq data with a lower SNV calling threshold revealed that 16 out of 88 SNVs, including *ARID1A*, were detectable in primary tumour region three, compared with a maximum of 2 out of 88 in other tumour regions (Extended Data Fig. 7b). These data suggest that ctDNA profiling can resolve the primary tumour region from which a low-frequency metastatic subclone derives. CRUK0035 developed one adrenal and two liver metastases (Fig. 5b). Sequencing of the metastatic liver deposit revealed that only 109 out of 149 SNVs classed as clonal in the primary tumour were detectable in the metastasis. This was suggestive of an ancestral branching event that had not been resolved through the primary M-seq analysis (Fig. 5b). Postoperative ctDNA profiling not only identified clonal SNVs present in the liver metastasis biopsy, but also revealed SNVs representing a subclone from the primary tumour (Extended Data Fig. 7c). This subclone was not present in the metastatic liver deposit (Fig. 5b). These data may reflect ctDNA identified from the non-biopsied metastases, suggesting multiple metastatic events. CRUK0044 suffered a vertebral and right hilar relapse. Postoperatively, the same subclonal SNV (*OR10K1*) was detected in ctDNA on two occasions that were 85 days apart (Extended Data Fig. 7d). This SNV was represented in a single subclone detected through sequencing of the hilar lymph node metastatic tissue, supporting the ctDNA findings

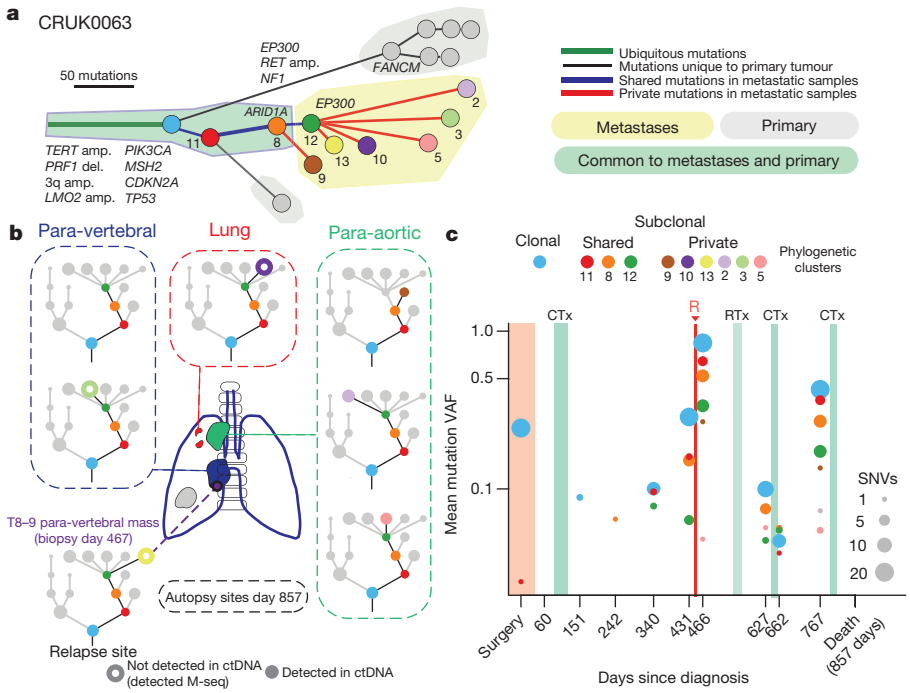


Figure 6 | ctDNA tracking of lethal cancer subclones in CRUK0063. Phylogenetic analysis of one relapse biopsy (day 467) and five metastatic biopsies (post-mortem). **a**, To-scale phylogenetic tree generated by multi-region sequencing of metastatic and primary tumour regions. Branch length is proportional to number of mutations in each subclone. **b**, Phylogenetic trees matching metastatic lesions. Coloured nodes represent mutation clusters found at each site and assayed for in ctDNA. Open circles represent mutation clusters not detected in ctDNA. **c**, Tracking plot showing mean VAF of identified mutation clusters in ctDNA. Size of dots indicates number of assays detected. Colours correspond to mutation clusters and match in **a** and **b**.

(Fig. 5c). CRUK0041 suffered an intracerebral, hilar and subcarinal lymph node relapse. Four subclonal SNVs representing both branches of the tumour phylogenetic tree were detectable in the ctDNA at relapse. Consistent with these data, sequencing of subcarinal metastatic tissue revealed the presence of subclonal SNVs mapping to both phylogenetic branches (Fig. 5d and Extended Data Fig. 7e). Patient CRUK0013 had 20 SNVs detectable in ctDNA at day 3 postoperatively and 13 SNVs detectable at day 38. Following adjuvant chemoradiotherapy for lymph node metastases identified in the pathological specimen, ctDNA levels became undetectable (Fig. 4h). Two lymph nodes involved were sampled for exome analysis together with M-seq of the primary tumour. Four subclonal SNVs detected in the ctDNA postoperatively mapped to an ancestral subclone (describing a subclone that existed during the tumour's evolution) (Extended Data Fig. 7f). This ancestral subclone contained a *KRAS* amplification (more than 15 copies, triploid background) and was identified as present in the primary tumour and sampled lymph nodes by M-seq (Fig. 5e). These data provide phylogenetic characterization of postoperative residual disease that responded to adjuvant chemoradiotherapy (Fig. 4h).

ctDNA profiling in the metastatic setting

Patient CRUK0063 was examined through the PEACE post-mortem study 24 h after death. M-seq data from the six post-mortem tumour regions (para-aortic, para-vertebral and lung metastases, day 857), the para-vertebral relapse biopsy (day 467) and five primary tumour regions (day 0) were combined to infer the phylogenetic structure of this patient's NSCLC (Fig. 6a). All seven metastatic tumour regions arose from a single ancestral subclone represented by phylogenetic cluster 8. Six metastatic regions shared a later phylogenetic origin, cluster 12 (Fig. 6b). The single tumour region not containing phylogenetic cluster 12 was sampled from the para-aortic metastasis at autopsy and contained a private subclone represented by phylogenetic cluster 9 (Fig. 6b).

We designed a bespoke ctDNA assay panel to retrospectively track metastatic subclonal burden. Twenty clonal SNVs and a median of 8 subclonal SNVs (range, 4–15) in each of 9 metastatic subclonal clusters were targeted by the assay panel (Extended Data Fig. 8). Because 103 variants per time point were profiled, SNV detection thresholds were increased to maintain platform specificity (see Methods). This resulted in a predicted false-positive rate of 0.0011, translating to a 10.7% risk of a single false-positive SNV at each time point, and a 0.5% risk of two false-positive SNVs at each time point when testing 103 SNVs.

Two clonal SNVs were detected by the 103-SNV assay panel 151 days after surgery (Fig. 6c and Extended Data Fig. 8); 189 days before the time point at which ctDNA was detected using the 19-SNV assay panel in Fig. 4f. At day 242 a single subclonal SNV was detected from phylogenetic cluster 8 (Fig. 6c and Extended Data Fig. 8); in the context of a 10.7% false-positive risk, a single SNV call could represent a type I error. At day 466, following clinical relapse at the thoracic para-vertebral site, 18 out of 20 SNVs mapping to phylogenetic clusters (8, 11 and 12) were detected in ctDNA; these subclonal clusters were shared between six out of seven metastatic sites (Fig. 6b, c and Extended Data Fig. 8). Single SNVs from two private subclones (phylogenetic clusters 5 and 9) were also detectable in ctDNA at day 466 (Fig. 6c and Extended Data Fig. 8). These subclones were not identified in the CT-guided para-vertebral biopsy taken at day 467 (Fig. 6b). The mean plasma VAF of the SNVs detected in phylogenetic clusters 11, 8, 12, 9 and 5 mirrored their proximity to the clonal cluster (light blue) in the M-seq-derived phylogenetic tree (Fig. 6a, c). This suggested a tiered burden of subclonal disease concordant with M-seq phylogenetic inferences. The mean clonal VAF decreased in response to palliative radiotherapy and chemotherapy, but increased at day 767 (Fig. 6c). Single SNVs mapping to phylogenetic clusters 5 and 9 and two SNVs mapping to phylogenetic cluster 2 were now detectable in ctDNA 90 days before death (Fig. 6a–c and Extended Data Fig. 8). These three phylogenetic clusters represented subclones private to the para-aortic metastases (Fig. 6a, b). Consistent with these data, significant para-aortic progression was observed post-mortem compared with the most recent CT imaging performed 112 days before death, which showed no evidence of para-aortic disease.

Discussion

In summary, we find predictors of ctDNA detection in early-stage NSCLC characterized by non-adenocarcinoma histology, necrosis, increased proliferative indices and lymphovascular invasion. Triple-negative breast cancers display necrosis[19] and high proliferative indices[20,21], and are associated with increased ctDNA levels compared with other breast cancer subtypes[6], suggesting the possibility of extending these observations beyond NSCLC.

Tumour volume correlated with the mean plasma VAF of clonal SNVs in ctDNA-positive NSCLCs (Fig. 3a). A primary NSCLC tumour volume of 10 cm^3 predicted a ctDNA plasma VAF of 0.1%. The sensitivity of the multiplex-PCR NGS platform was in excess of 99% at VAFs of 0.1% and above, suggesting optimum platform sensitivity with tumour burdens in excess of 10 cm^3. Low-dose CT lung screening can identify lung nodules with diameters from 4 mm (ref. 22). Assuming a spherical nodule this would translate to a tumour volume of 0.034 cm^3. On the basis of the relationship between tumour volume and ctDNA plasma VAF observed in this study, a tumour volume of 0.034 cm^3 would equate to a plasma VAF of 1.8×10^{-4}% (95% confidence interval, 9.8×10^{-6}–0.0033%), which is at the extreme of detection limits of current ctDNA platforms[23]. Therefore, using current technologies, the sensitivity of clonal-SNV ctDNA-directed early-NSCLC screening may be constrained by tumour size.

A limitation to targeted ctDNA profiling is cost, which is estimated at US$1,750 per patient for sequencing of a single tumour region, synthesis of a patient-specific assay panel and profiling of five plasma samples. Adjuvant platinum-based chemotherapy in NSCLC improves cure rates after surgery in only 5% of patients, and 20% of patients receiving chemotherapy experience acute toxicities[3]. There is a need to increase the efficacy of adjuvant therapy and better target its use. Bespoke ctDNA profiling can characterize the subclonal dynamics of relapsing NSCLC and identify adjuvant chemotherapy resistance. These findings indicate that drug development guided by ctDNA platforms to identify residual disease, define adjuvant treatment response and target emerging subclones before clinical recurrence in NSCLC is now feasible.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 January; accepted 13 April 2017.
Published online 26 April 2017.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90 (2011).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA Cancer J. Clin.* **67,** 7–30 (2017).
3. Pignon, J.-P. *et al.* Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J. Clin. Oncol.* **26,** 3552–3559 (2008).
4. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152,** 714–726 (2013).
5. Beaver, J. A. *et al.* Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin. Cancer Res.* **20,** 2643–2650 (2014).
6. Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl. Med.* **7,** 302ra133 (2015).
7. Tie, J. *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8,** 346ra92 (2016).
8. Jamal-Hanjani, M. *et al.* Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol.* **12,** e1001906 (2014).
9. Jamal-Hanjani, M. *et al.* Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* http://dx.doi.org/10.1056/NEJMoa1616288 (2017).
10. Jamal-Hanjani, M. *et al.* Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. *Ann. Oncol.* **27,** 862–867 (2016).
11. Diaz, L. A. Jr & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J. Clin. Oncol.* **32,** 579–586 (2014).
12. Caruso, R. *et al.* Histologic coagulative tumour necrosis as a prognostic indicator of aggressiveness in renal, lung, thyroid and colorectal carcinomas: a brief review. *Oncol. Lett.* **3,** 16–18 (2012).
13. Vesselle, H. *et al.* Lung cancer proliferation correlates with [F-18] fluorodeoxyglucose uptake by positron emission tomography. *Clin. Cancer Res.* **6,** 3837–3844 (2000).
14. Higashi, K. *et al.* FDG PET measurement of the proliferative potential of non-small cell lung cancer. *J. Nucl. Med.* **41,** 85–92 (2000).
15. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat. Commun.* **6,** 8760 (2015).
16. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20,** 548–554 (2014).
17. Del Monte, U. Does the cell number 10^9 still really fit one gram of tumor tissue? *Cell Cycle* **8,** 505–506 (2009).
18. Peters, S. & Zimmermann, S. Targeted therapy in NSCLC driven by HER2 insertions. *Transl. Lung Cancer Res.* **3,** 84–88 (2014).
19. Livasy, C. A. *et al.* Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod. Pathol.* **19,** 264–271 (2006).
20. Keam, B. *et al.* Prognostic impact of clinicopathologic parameters in stage II/III breast cancer treated with neoadjuvant docetaxel and doxorubicin chemotherapy: paradoxical features of the triple negative breast cancer. *BMC Cancer* **7,** 203 (2007).
21. Rhee, J. *et al.* The clinicopathologic characteristics and prognostic significance of triple-negativity in node-negative breast cancer. *BMC Cancer* **8,** 307 (2008).
22. Aberle, D. R. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365,** 395–409 (2011).
23. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34,** 547–555 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge R. Macina for developing the collaboration between Natera and TRACERx. We thank S. Navarro and A. Tin for facilitating the PEACE ctDNA analysis. We thank the members of the TRACERx and PEACE consortia for participating in this study. C.S. is Royal Society Napier Research Professor. This work is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001169, FC001202), the UK Medical Research Council (FC001169, FC001202) and the Wellcome Trust (FC001169, FC001202). C.S. is funded by Cancer Research UK (TRACERx and CRUK Cancer Immunotherapy Catalyst Network), the CRUK Lung Cancer Centre of Excellence, Stand Up 2 Cancer (SU2C), the Rosetrees Trust, NovoNordisk Foundation (ID 16584), the Prostate Cancer Foundation, the Breast Cancer Research Foundation and the European Research Council (THESEUS), and support was provided to C.S. by the National Institute for Health Research, the University College London Hospitals Biomedical Research Centre and the Cancer Research UK University College London Experimental Cancer Medicine Centre. P.V.L. is a Winton Group Leader in recognition of the Winton Charitable Foundation's support towards the establishment of the Francis Crick Institute. The TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601) is sponsored by University College London (UCL/12/0279) and has been approved by an independent Research Ethics Committee (13/LO/1546). TRACERx is funded by Cancer Research UK (grant number C11496/A17786) and coordinated through the Cancer Research UK and UCL Cancer Trials Centre. The PEACE study (https://clinicaltrials.gov/ct2/show/NCT03004755) is sponsored by University College London (UCL/13/0165) and has been approved by an independent Research Ethics Committee (13/LO/0972). PEACE is funded by Cancer Research UK (C416/A21999) and coordinated through the Cancer Research UK and UCL Cancer Trials Centre.

Author Contributions C.A., N.J.B., G.A.W., M.J.-H., T.C., R.Sa. and J.L.Q. contributed equally to this work. D.A.M. and S.V. contributed equally to this work as joint second authors. C.A. and C.S. co-wrote the manuscript. C.A., M.J.-H. and C.S. conceived the study design. C.A., N.J.B., G.A.W. and R.R. integrated clinicopathological data, exome data and ctDNA data. M.R., B.G.Z., C.J.L., T.C., R.S., E.K., N.S., D.H., A.Nai. and A.G. conducted and analysed multiplex-PCR NGS experimental work for ctDNA data. N.J.B., G.A.W., T.B.K.W., M.A.B., R.R. and N.M. conducted M-seq analyses of exome data. J.L.Q., T.M. and D.A.M. conducted the pathological review. F.F., R.E. and F.Z. conducted radiological review of PET scans. H.J.W.L.A., W.L.B., F.M.F. and N.J.B. conducted radiomic analyses. S.V., D.J., J.L., S.S., J.C.-K., A.R., T.C., D.O. and A.U.A. conducted TRACERx sample processing. G.E., S.W., N.M. and G.A.W. conducted exome sequencing. L.M., J.R. and J.A.S. conducted ctDNA cross-platform validation. M.J.-H., C.D., J.A.S. and C.S. designed the study protocols. C.H., S.M.L., M.D.F., T.A., M.F., E.B., D.L., M.H., S.Ko., N.P., S.M.J., R.T., A.A., F.B., Y.S., R.Sh., L.J., A.M.Q., P.A.C., B.N., G.M., G.L., S.T., M.N., H.R., K.K., M.C., L.G., D.A.F., A.Nak., S.R., G.A., S.Kh., P.R., V.E., B.I., M.I.-S., V.P., J.F.L., M.K., R.A., H.A., H.D. and S.L. are clinical members of TRACERx study sites. J.A.H. and H.L.L. run the UCL GCLP facility. A.H., H.B., N.I. and Y.N. were involved in study oversight. J.A.S., J.L.Q., Z.S., E.G., S.Ka., S.T., M.A.B., R.F.S., J.H., A.S., S.A.Q., P.V.L., C.D., C.J.L. and B.G.Z. gave advice and reviewed the manuscript. A.H. gave statistical advice. C.S. provided overall study oversight.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.S. (Charles.Swanton@crick.ac.uk).

Reviewer Information *Nature* thanks S. Lippman, R. Rosell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Christopher Abbosh[1]*, Nicolai J. Birkbak[1,2]*, Gareth A. Wilson[1,2]*, Mariam Jamal-Hanjani[1]*, Tudor Constantin[3]*, Raheleh Salari[3]*, John Le Quesne[4]*, David A. Moore[4], Selvaraju Veeriah[1], Rachel Rosenthal[1], Teresa Marafioti[1,5], Eser Kirkizlar[3], Thomas B. K. Watkins[1,2], Nicholas McGranahan[1,2], Sophia Ward[1,2,6], Luke Martinson[4], Joan Riley[4], Francesco Fraioli[7], Maise Al Bakir[2], Eva Grönroos[2], Francisco Zambrana[1], Raymondo Endozo[7], Wenya Linda Bi[8,9], Fiona M. Fennessy[8,9], Nicole Sponer[3], Diana Johnson[1], Joanne Laycock[1], Seema Shafi[1], Justyna Czyzewska-Khan[1], Andrew Rowan[2], Tim Chambers[2,6], Nik Matthews[6,10], Samra Turajlic[2,11], Crispin Hiley[1,2], Siow Ming Lee[1,12], Martin D. Forster[1,12], Tanya Ahmad[12], Mary Falzon[5], Elaine Borg[5], David Lawrence[13], Martin Hayward[13], Shyam Kolvekar[13], Nikolaos Panagiotopoulos[13], Sam M. Janes[1,14,15], Ricky Thakrar[14], Asia Ahmed[16], Fiona Blackhall[17,18], Yvonne Summers[18], Dina Hafez[3], Ashwini Naik[3], Apratim Ganguly[3], Stephanie Kareht[3], Rajesh Shah[19], Leena Joseph[20], Anne Marie Quinn[20], Phil A. Crosbie[21], Babu Naidu[22,58], Gary Middleton[23], Gerald Langman[24], Simon Trotter[24], Marianne Nicolson[25], Hardy Remmen[26], Keith Kerr[27], Mahendran Chetty[28], Lesley Gomersall[29], Dean A. Fennell[4], Apostolos Nakas[30], Sridhar Rathinam[30], Girija Anand[31], Sajid Khan[32,33], Peter Russell[34], Veni Ezhil[35], Babikir Ismail[36], Melanie Irvin-Sellers[37], Vineet Prakash[38], Jason F. Lester[39], Malgorzata Kornaszewska[40], Richard Attanoos[41], Haydn Adams[42], Helen Davies[43], Dahmane Oukrif[1], Ayse U. Akarca[1], John A. Hartley[44], Helen L. Lowe[44], Sara Lock[45], Natasha Iles[46], Harriet Bell[46], Yenting Ngai[46], Greg Elgar[2,6], Zoltan Szallasi[47,48,49], Roland F. Schwarz[50], Javier Herrero[51], Aengus Stewart[52], Sergio A. Quezada[53], Karl S. Peggs[53,80], Peter Van Loo[54,55], Caroline Dive[1,56], C. Jimmy Lin[3], Matthew Rabinowitz[3], Hugo J. W. L. Aerts[8,9,57], Allan Hackshaw[46], Jacqui A. Shaw[4], Bernhard G. Zimmermann[3], The TRACERx consortium†, The PEACE consortium‡ & Charles Swanton[1,2]

[1]Cancer Research UK Lung Cancer Centre of Excellence London and Manchester, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK. [2]Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [3]Natera Inc., 201 Industrial Road, San Carlos, California 94070, USA. [4]Cancer Studies, University of Leicester, Leicester LE2 7LX, UK. [5]Department of Pathology, University College London Hospitals, 21 University Street, London WC1 6JJ, UK. [6]Advanced Sequencing Facility, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [7]Department of Nuclear Medicine, University College London Hospitals, 235 Euston Road, Fitzrovia, London, NW1 2BU, UK. [8]Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [9]Harvard Medical School, Boston, Massachusetts 02115, USA. [10]Tumour Profiling Unit Genomics Facility, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. [11]Renal and Skin Units, The Royal Marsden Hospital, London SW3 6JJ, UK. [12]Department of Oncology, University College London Hospitals, 250 Euston Road, London NW1 2BU, UK. [13]Department of Cardiothoracic Surgery, University College London Hospitals, 235 Euston Road, Fitzrovia, London NW1 2BU, UK. [14]Department of Respiratory Medicine, University College London Hospitals, 235 Euston Road, Fitzrovia, London NW1 2BU, UK. [15]Lungs for Living Research Centre, UCL Respiratory, Division of Medicine, Rayne Building, University College London, 5 University Street, London WC1E 6JF, UK. [16]Department of Radiology, University College London Hospitals, 235 Euston Road, Fitzrovia, London NW1 2BU, UK. [17]Institute of Cancer Studies, University of Manchester, Oxford Road, Manchester M13 9PL, UK. [18]The Christie Hospital, Manchester M20 4BX, UK. [19]Department of Cardiothoracic Surgery, University Hospital South Manchester, Manchester M23 9LT, UK. [20]Department of Pathology, University Hospital South Manchester, Manchester M23 9LT, UK. [21]North West Lung Centre, University Hospital South Manchester, Manchester M23 9LT, UK. [22]Department of Thoracic Surgery, Birmingham Heartlands Hospital, Birmingham B9 5SS, UK. [23]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham B15 2TT, UK. [24]Department of Cellular Pathology, Birmingham Heartlands Hospital, Birmingham B9 5SS, UK. [25]Department of Medical Oncology, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen AB25 2ZN, UK. [26]Department of Cardiothoracic Surgery, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen AB25 2ZD, UK. [27]Department of Pathology, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen AB25 2ZD, UK. [28]Department of Respiratory Medicine, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen AB25 2ZN, UK. [29]Department of Radiology, Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen AB25 2ZN, UK. [30]Department of Thoracic Surgery, Glenfield Hospital, Leicester LE3 9QP, UK. [31]Department of Radiotherapy, North Middlesex University Hospital, London N18 1QX, UK. [32]Department of Respiratory Medicine, Royal Free Hospital, Pond Street, London NW3 2QG, UK. [33]Department of Respiratory Medicine, Barnet and Chase Farm Hospitals, Wellhouse Lane, Barnet EN5 3DJ, UK. [34]Department of Respiratory Medicine, The Princess Alexandra Hospital, Hamstel Road, Harlow CM20 1QX, UK. [35]Department of Clinical Oncology, St.Luke's Cancer Centre, Royal Surrey County Hospital, Guildford GU2 7XX, UK. [36]Department of Pathology, Ashford and St. Peter's Hospital, Guildford Road, Chertsey, Surrey KT16 0PZ, UK. [37]Department of Respiratory Medicine, Ashford and St. Peter's Hospital, Guildford Road, Chertsey, Surrey KT16 0PZ, UK. [38]Department of Radiology, Ashford and St. Peter's Hospital, Guildford Road, Chertsey, Surrey KT16 0PZ, UK. [39]Department of Clinical Oncology, Velindre Hospital, Cardiff CF14 2TL, UK. [40]Department of Cardiothoracic Surgery, University Hospital of Wales, Cardiff CF14 4XW, UK. [41]Department of Cellular Pathology, University Hospital of Wales and Cardiff University, Heath Park, Cardiff, UK. [42]Department of Radiology, University Hospital Llandough, Cardiff CF64 2XX, UK. [43]Department of Respiratory Medicine, University Hospital Llandough, Cardiff CF64 2XX, UK. [44]University College London Experimental Cancer Medicine Centre GCLP Facility, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK. [45]Department of Respiratory Medicine, The Whittington Hospital NHS Trust, London, N19 5NF, UK. [46]University College London, Cancer Research UK and UCL Cancer Trials Centre, London W1T 4TJ, UK. [47]Centre for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark. [48]Computational Health Informatics Program (CHIP), Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [49]MTA-SE-NAP, Brain Metastasis Research Group, 2nd Department of Pathology, Semmelweis University, 1091 Budapest, Hungary. [50]Berlin Institute for Medical Systems Biology, Max Delbrueck Center for Molecular Medicine, Berlin, Germany. [51]Bill Lyons Informatics Centre, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK. [52]Department of Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [53]Cancer Immunology Unit, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK. [54]Cancer Genomics Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [55]Department of Human Genetics, University of Leuven, B-3000 Leuven, Belgium. [56]Cancer Research UK Manchester Institute, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK. [57]Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215-5450, USA. [58]Institute of Inflammation and Ageing, University of Birmingham, Birmingham B15 2TT, UK.

*These authors contributed equally to this work.

†A list of participants and their affiliations appears in the online version of the paper.

‡A list of participants and their affiliations appears in the Supplementary Information.

The TRACERx consortium

Charles Swanton[1,2], Mariam Jamal-Hanjani[1], Christopher Abbosh[1], Selvaraju Veeriah[1], Seema Shafi[1], Justyna Czyzewska-Khan[1], Diana Johnson[1], Joanne Laycock[1], Leticia Bosshard-Carter[1], Gerald Goh[1], Rachel Rosenthal[1], Pat Gorman[1], Nirupa Murugaesu[1], Robert E. Hynds[1,15], Gareth A. Wilson[1,2], Nicolai J. Birkbak[1,2], Thomas B. K. Watkins[1,2], Nicholas McGranahan[1,2], Stuart Horswell[2], Maise Al Bakir[2], Eva Grönroos[2], Richard Mitter[2], Mickael Escudero[2], Aengus Stewart[52], Peter Van Loo[54,55], Andrew Rowan[2], Hang Xu[2], Samra Turajlic[2,11], Crispin Hiley[1,2], Jacki Goldman[2], Richard Kevin Stone[2], Tamara Denner[2], Nik Matthews[6,10], Greg Elgar[2,6], Sophia Ward[1,2,6], Jennifer Biggs[2,6], Marta Costa[2], Sharmin Begum[2], Ben Phillimore[2], Tim Chambers[2,6], Emma Nye[2], Sofia Graca[2], Kroopa Joshi[1], Andrew Furness[1], Assma Ben Aissa[1], Yien Ning Sophia Wong[1], Andy Georgiou[1], Sergio A. Quezada[53], Karl S. Peggs[53,80], John A. Hartley[44], Helen L. Lowe[44], Javier Herrero[51], David Lawrence[13], Martin Hayward[13], Nikolaos Panagiotopoulos[13], Shyam Kolvekar[13], Mary Falzon[5], Elaine Borg[5], Teresa Marafioti[5], Celia Simeon[58], Gemma Hector[58], Amy Smith[58], Marie Aranda[58], Marco Novelli[5], Dahmane Oukrif[1], Ayse U. Akarca[1], Sam M. Janes[1,14,15], Ricky Thakrar[14], Martin D. Forster[1,12], Tanya Ahmad[12], Siow Ming Lee[1,12], Dionysis Papadatos-Pastos[12], Dawn Carnell[58], Ruheena Mendes[58], Jeremy George[58], Neal Navani[58], Asia Ahmed[16], Magali Taylor[58], Junaid Choudhary[58], Yvonne Summers[18], Raffaele Califano[59], Paul Taylor[59], Rajesh Shah[19], Piotr Krysiak[59], Kendadai Rammohan[59], Eustace Fontaine[59], Richard Booton[59], Matthew Evison[59], Phil A. Crosbie[21], Stuart Moss[59], Faiza Idries[59], Leena Joseph[20], Paul Bishop[59], Anshuman Chaturvedi[59], Anne Marie Quinn[20], Helen Doran[59], Angela Leek[60], Phil Harrison[60], Katrina Moore[60], Rachael Waddington[60], Juliette Novasio[60], Fiona Blackhall[17,18], Jane Rogan[60], Elaine Smith[59], Caroline Dive[1,56], Jonathan Tugwood[56], Ged Brady[56], Dominic G. Rothwell[56], Francesca Chemi[56], Jackie Pierce[56], Sakshi Gulati[56], Babu Naidu[22], Gerald Langman[24], Simon Trotter[24], Mary Bellamy[61], Hollie Bancroft[61], Amy Kerr[61], Salma Kadiri[61], Joanne Webb[61], Gary Middleton[23], Madava Djearaman[61], Dean A. Fennell[4], Jacqui A. Shaw[4], John Le Quesne[4], David A. Moore[4], Anne Thomas[4], Harriet Walter[4], Joan Riley[4], Luke Martinson[4], Apostolos Nakas[30], Sridhar Rathinam[30], William Monteiro[62], Hilary Marshall[62], Louise Nelson[63], Jonathan Bennett[63], Lindsay Primrose[63], Girija Anand[31], Sajid Khan[32,33], Anita Amadi[64], Marianne Nicolson[25], Keith Kerr[27], Shirley Palmer[65], Hardy Remmen[26], Joy Miller[65], Keith Buchan[65], Mahendran Chetty[28], Lesley Gomersall[29], Jason F. Lester[39], Alison Edwards[66], Fiona Morgan[67], Haydn Adams[42], Helen Davies[67], Malgorzata Kornaszewska[40], Richard Attanoos[41], Sara Lock[45], Azmina Verjee[45], Mairead MacKenzie[68], Maggie Wilcox[68], Harriet Bell[46], Natasha Iles[46], Allan Hackshaw[46], Yenting Ngai[46], Sean Smith[46], Nicole Gower[46], Christian Ottensmeier[69], Serena Chee[69], Benjamin Johnson[69], Aiman Alzetani[69], Emily Shaw[69], Eric Lim[70], Paulo De Sousa[70], Monica Tavares Barbosa[70], Alex Bowman[70], Simon Jordan[70], Alexandra Rice[70], Hilgardt Raubenheimer[70], Chiara Proli[70], Maria Elena Cufari[70], John Carlo Ronquillo[70], Angela Kwayie[70], Harshil Bhayani[70], Morag Hamilton[70], Yusura Bakar[70], Natalie Mensah[70], Lyn Ambrose[70], Anand Devaraj[70], Silviu Buderi[70], Jonathan Finch[70], Leire Azcarate[70], Hema Chavan[70], Sophie Green[70], Hillaria Mashinga[70], Andrew G. Nicholson[70,71], Kelvin Lau[72], Michael Sheaff[72], Peter Schmid[72], John Conibear[72], Veni Ezhil[35], Babikir Ismail[36], Melanie Irvin-Sellers[37], Vineet Prakash[38], Peter Russell[34], Teresa Light[73], Tracey Horey[73], Sarah Danson[74], Jonathan Bury[74], John Edwards[74], Jennifer Hill[74], Sue Matthews[74], Yota Kitsanta[74], Kim Suvarna[74], Patricia Fisher[74], Allah Dino Keerio[74], Michael Shackcloth[75], John Gosney[75], Pieter Postmus[75], Sarah Feeney[75], Julius Asante-Siaw[75], Tudor Constantin[3], Raheleh Salari[3], Nicole Sponer[3], Ashwini Naik[3], Bernhard G. Zimmermann[3], Matthew Rabinowitz[3], Hugo J. W. L. Aerts[8,9,57], Stefan Dentro[76] & Christophe Dessimoz[77,78,79]

[58]University College London Hospitals NHS Foundation Trust, London, UK. [59]University Hospital South Manchester, Manchester, M23 9LT, UK. [60]Manchester Cancer Research Centre Biobank, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK. [61]Heart of

England NHS Foundation Trust, Birmingham, UK. [62]NIHR Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK. [63]University Hospitals of Leicester NHS Trust, Leicester, UK. [64]Barnet and Chase Farm Hospitals, Wellhouse Lane, Barnet EN5 3DJ, UK. [65]Aberdeen Royal Infirmary, Aberdeen, AB25 2ZN, UK. [66]Clinical Trials Unit, Velindre Cancer Centre, Whitchurch, Cardiff, CF14 2TL, UK. [67]Cardiff & Vale University Health Board, Cardiff, UK. [68]Independent Cancer Patients' Voice, 17 Woodbridge Street, London, EC1R 0LL, UK. [69]University of Southampton and Southampton University Hospitals, Southampton, UK. [70]Royal Brompton and Harefield NHS Foundation Trust, London, UK. [71]National Heart and Lung Institute, Imperial College, Sydney Street, London, UK. [72]Barts Health NHS Trust, London, UK. [73]The Princess Alexandra Hospital NHS Trust, Harlow, UK. [74]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. [75]Liverpool Heart and Chest Hospital NHS Foundation Trust, Liverpool, UK. [76]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK. [77]Bioinformatics Group, Department of Computer Science, University College London, UK. [78]University of Lausanne, 1015 Lausanne, Switzerland. [79]SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [80]Research Department of Haematology, University College Cancer Institute, London WC1E 6DD, UK.

METHODS

Patients and samples. The cohort of 100 patients evaluated within this study comprises the first 100 patients prospectively analysed by the lung TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601, approved by an independent Research Ethics Committee, 13/LO/1546) and mirrors the prospective 100 patient cohort described in ref. 9.

Multi-region tumour sampling was performed as previously described[9]. Relapse tissue samples, excess to diagnostic requirements, were acquired via clinical procedures detailed in Supplementary Table 3. For patient CRUK0063 post-mortem examination was performed through the PEACE study 24 h after death (https://clinicaltrials.gov/ct2/show/NCT03004755, approved by an independent Research Ethics Committee, 13/LO/0972). Informed consent was obtained from all subjects for procedures conducted in these studies. The experiments were not randomized.

Tissue microarray creation and Ki67 immunohistochemistry. Tissue microarrays were created for 100 NSCLC cases for $Ki67^+$ immunohistochemistry. Representative primary tumour areas were defined by examination of sections stained with haematoxylin and eosin from TRACERx cases. Two 2-mm cores were selected from different regions within each specimen and re-embedded in recipient blocks. This resulted in a tissue microarray of 200 cores with four normal lung cores as negative control. 2–5 μm sections from tissue microarrays containing tumours were cut. Immunohistochemistry with anti-Ki67 monoclonal antibody (dilution, 1:100; clone MIB-1; DAKO Agilent Technologies LDA) was performed using BenchMark Ultra (Ventana/Roche). The percentage of $Ki67^+$ cells was averaged across two tumour sections for each case. Detection was performed using the peroxidase-based detection reagent conjugate (OptiView DAB IHC Detection kit; Ventana).

Central histopathological review. Digital images of diagnostic tumour sections from all cases were reviewed in detail centrally by at least one pathologist, and in cases of uncertainty, by two. Histological subtype, percentage of necrosis and the presence of lymphovascular invasion were all evaluated on digital images from scanned diagnostic slides blinded to the ctDNA detection status of the patient in question.

Central radiology review and volume estimation. 92 out of 96 anonymized diagnostic PET–CT scans were reviewed by a nuclear medicine physician, blinded to the initial PET–CT reports. Scan images were not available in three cases (CRUK0025, CRUK0039 and CRUK0023) and in one case a preoperative PET–CT was not performed (CRUK0082). CT and PET images were matched and fused into transaxial, coronal and sagittal images and reviewed on a dedicated PET–CT software visualizer (AW 4.1/4.2 GE medical systems). The semiquantitative parameter standardized uptake value (SUV) maximum for the primary tumour mass was calculated and recorded along with the SUV_{max} of mediastinal background uptake. The tumour-to-background ratio (TBR) was calculated on the basis of the SUV_{max} of the tumour divided by the mediastinal background uptake[24,25]. Tumour volume was determined on the basis of tumour CT scans. CT slices of the primary tumour were measured with 3D Slicer applying the 'lung algorithm window' settings, tumour contours were segmented on each axial CT slice. These steps were performed by an experienced resident (W.L.B.), and all contours were confirmed and edited where necessary, by a radiologist with 14 years of experience in cancer imaging (F.M.F.). Cancer cell volume was defined as tumour volume multiplied by the mean purity of the tumour on the basis of the M-seq results, purity estimates derived from the ASCAT analysis as previously described[9]. Effective subclone size was defined as the mean cancer cell fraction (CCF) across the regions of the mutation cluster multiplied by tumour volume and mean tumour purity.

Tissue-exome sequencing and processing. Whole-exome sequencing was performed on DNA purified from tumour tissue and normal blood as previously described[9], with the exception of CRUK0063_BR_T1-R1. This capture was performed according to the manufacturer's 200 ng DNA protocol (Agilent). Annotated SNV calls from primary tumours are available in ref. 9. For this study, metastatic tissue biopsies from each of four patients (CRUK0035, CRUK0041, CRUK0044 and CRUK0063) and six metastatic samples acquired at post-mortem examination of CRUK0063 were obtained. Genomic DNA was purified from all tissue samples, and processed through the TRACERx bioinformatics pipeline as previously described[9]. Annotated SNV calls are available in Supplementary Table 4.

cfDNA extraction and quantification. Blood samples were collected in K_2-EDTA tubes. Samples were processed within 2 h of collection by double centrifugation of the blood, first for 10 min at 1,000*g*, then the plasma for 10 min at 2000*g*. Plasma was stored in 1 ml aliquots at −80 °C. Up to 5 ml of plasma per case was available for this study (range, 1–5 ml; median 5 ml). The entire volume of plasma was used for cfDNA extraction. cfDNA was extracted using the QIAamp Circulating Nucleic Acid kit (Qiagen) and eluted into 50 μl DNA Suspension Buffer (Sigma). Every cfDNA sample was analysed on the Bioanalyzer High Sensitivity (Agilent) and quantified by interpolation of the mononucleosomal peak height on a calibration curve prepared from a pure cfDNA sample that was quantified previously.

cfDNA library preparation. Subsequently, 40 μl of cfDNA from each plasma sample, which is present as fragments of mononucleosomal and polynucleosomal length, was used as input into Library Prep using the Natera Library Prep kit; in two samples with extremely high cfDNA amounts, input was restricted to approximately 50,000 genome equivalents (165 ng). cfDNA was end-repaired and A-tailed. Natera custom adapters were ligated. The libraries were amplified for 15 cycles to plateau and then purified using Ampure beads following the manufacturer's protocol. The purified libraries were run on the LabChip. Successful libraries had a single peak at around 250 bp.

SNV assay design and optimization. Natera's standard assay design pipeline was used to generate forward and reverse PCR primers for somatic SNVs detected in tumour samples. For every pair of primers, the probability of forming a primer-dimer was calculated and assays were combined into pools such that any primer combination in a pool is not predicted to form primer-dimers. For each patient, assays were prioritized such that (1) assays covering driver SNVs had highest priority and (2) there was uniform sampling of the phylogenetic tree. For the baseline cohort, 10 balanced pools were created, each containing on average 18 assays for 10 patients' SNVs. For the longitudinal cohort, up to 10 extra assays were generated for samples. For patient CRUK0063 post-mortem analysis, new assays were designed on the basis of the M-seq of the metastatic biopsy retrieved on day 467 and of metastatic lesions collected post-mortem. A total of 103 new assays were designed compared with the 19 that were based on the primary tumour alone. Primer details are available in Supplementary Table 5 (baseline, preoperative cohort), Supplementary Table 6 (longitudinal cohort) and Supplementary Table 7 (extended longitudinal assays for CRUK0063).

SNV assays were ordered from IDT. Each pool was optimized by running the multiplex-PCR and sequencing protocol using one plasma cfDNA library from a healthy subject. For optimization, PCR parameters (primer concentration and annealing temperature) that yielded the best percentage of on-target reads, depth-of-read uniformity (measured as the ratio of the 80th percentile to the 20th percentile), and number of drop-out assays (defined as assays with $<1{,}000$ reads) were determined by sequencing. The PCR conditions that yield the best percentage of on-target reads, depth-of-read uniformity, and the lowest number of drop-outs were determined. For all pools, the optimal conditions were 10 nM primers and 60 °C or 62.5 °C annealing temperatures. Primer pairs contributing to dimer formation were removed from each final pool.

Analytical validation. Synthetic spikes representing twenty SNVs that were randomly selected from primer assay pool 1 were designed and synthesized (IDT) as 160-bp oligos with the respective SNV placed in the middle (position 80). These synthetic spikes were mixed at equimolar ratios and used to prepare a library. This library was titrated into a library prepared from mononucleosomal DNA (10,000 copies) from a normal cell line (AG16778 from Coriell). The library of 20 synthetic spikes was titrated into the mononucleosomal DNA library at 2.5%, 0.5%, 0.25%, 0.1%, 0.05% and 0% (each in triplicate), and 0.01%, 0.005% and 0.001% (each in quadruplicate). Because preparing spiked samples at such low levels is either subject to sampling noise (0.01% spikes into 10,000 genomic copies background is equivalent to one mutant copy), or is not possible (at levels less than 0.01%), samples were mixed as libraries. Following library mixing and sequencing, data was analysed to detect all the targets in assay pool 1 using the same parameters as used for the patient samples.

The measured VAF of each spike for the samples with 2.5% nominal input was used to calculate an input correction factor (measured VAF/2.5%). This correction factor was applied to the other inputs of the corresponding spike titration series. The measured VAF differed from the nominal input most likely because the mononucleosomal fragmentation pattern is not entirely random. Because of this, the actual input levels differ from the nominal input levels. Therefore, analytical sensitivity and specificity were measured on the basis of the corrected input intervals (see Extended Data Fig. 1a).

Plasma SNV multiplex-PCR NGS workflow. The library material from each plasma sample was used as input into the multiplex-PCR (mPCR) using the relevant assay pool and an optimized plasma mPCR protocol. Optimal mPCR conditions were as previously described[10]. Each PCR assay pool was used to amplify the SNV targets from the 10 corresponding samples and 20 negative control samples (plasma libraries prepared from healthy subjects; BioMed IRB 601-01 and E&I West Coast Board IRB00007807, study 13090-01A and 13090-04A). The mPCR products were barcoded in a separate PCR step. Each amplicon pool was sequenced on one Illumina HiSeq 2500 Rapid Run with 50 cycles paired-end reads using the Illumina Paired End v1 kit with an average target read depth of around 40,000 per assay.

Bioinformatics pipeline. All the paired-end reads were merged using Pear[26]. Merged reads were mapped to the hg19 reference genome with Novoalign version 2.3.4 (http://www.novocraft.com/) and sorted and indexed using SAMtools[27]. Bases that did not match in forward and reverse reads or that have Phred quality score <20 were filtered out to minimize sequencing errors in subsequent steps. Merged reads with mapping quality >30 and at most one mismatch under the sequence of the primers were marked as on-target. Targets with <1,000 reads were considered to have failed and were filtered from further analyses. Quality control was performed using an in-house program checking for a wide list of statistics per sample that included total numbers of reads, mapped reads, on-target reads, number of failed targets and average error rate.

Plasma SNV calling. For each target SNV a position-specific error model was built (see Methods section 'SNV calling algorithm'). Samples with high plasma VAF (>20%) among the putative negatives were considered to have possible germline mutations and were excluded from the error model. A confidence score was calculated for each target SNV on the basis of the error model and a positive plasma SNV call was made if the confidence score passed a threshold of 95% for transitions and 98% for transversions. There was no difference in read depth between called and not called SNVs (Extended Data Fig. 1c).

Because the post-mortem analysis of CRUK0063 involved a larger number of target SNVs per time point being analysed (103 versus 19 targets in previous samples), updated calling thresholds were applied to control for false positives. The new updated thresholds were chosen such that the average number of false positives in the 30 negative samples in the run becomes around 1 per sample. All multiplex-PCR NGS ctDNA SNV assays with confidence score data are available in Supplementary Table 5 (baseline, preoperative cohort assays), Supplementary Table 6 (longitudinal assays), and Supplementary Table 7 (extended longitudinal assays for CRUK0063).

SNV calling algorithm. The PCR process was modelled as a stochastic process, estimating the error parameters using a set of 28–30 control plasma samples and making the final SNV calls on the target cancer samples. For each target SNV, we built a target-specific background-error model by estimating the following parameters from the control samples: PCR efficiency (p), probability that each molecule is replicated in a PCR cycle; error rate (p_e), error rate per cycle for mutation type e (for example, wild-type allele A to mutant allele G); initial number of molecules (X_0).

The target-specific error propagation model was used to characterize the distribution of error molecules. As a molecule is replicated over the course of the PCR process, more errors occur. If an error occurs in cycle i and there are X_i wild-type molecules in the system, that error molecule is duplicated in next cycle with probability p, and new error molecules are produced from the wild-type background molecules according to a binomial process, $B(X_i, p_e)$. Using a recursive relation, we computed the mean and variance of the number of total molecules X_n and number of error molecules E_n after n PCR cycles.

Algorithm steps are as follows. (1) Estimate the PCR efficiency and per-cycle error rate using the normal control samples. (2) Using the efficiency estimate, compute the starting number of molecules in the test set. (3) Use the starting number of molecules and the prior efficiency distribution from the training set to estimate the PCR efficiency in the test sample. (4) For a range of potential real mutant fraction values θ between 0 and 1 (we used 0.15 as upper bound), estimate the mean and variance for the total number of molecules, background error molecules and real mutation molecules using the described error propagation model and parameters estimated in steps (1)–(3). (5) Use the mean and variance estimated in step (4) to compute the likelihood $L(\theta)$ for each potential real mutant fraction, select the value of θ that maximizes this likelihood (denoted by $\hat{\theta}_{\mathrm{MLE}}$) and compute the confidence score as $\frac{L(\hat{\theta}_{\mathrm{MLE}})}{L(0)+L(\hat{\theta}_{\mathrm{MLE}})}$. (6) Call a mutation positive if the confidence score passes a predefined threshold.

Cross-platform validation using generic PCR NGS panel section. Cross-platform validation was performed in 28 patients with M-seq-confirmed SNV(s) within one or more hotspots targeted by a generic multiplex PCR NGS panel (Extended Data Table 2a, b and Supplementary Table 8). 20 ng of isolated cfDNA was used for library preparation using the Oncomine Lung cfDNA assay (ThermoFisher Scientific), according to the manufacturer's instructions. Automated template preparation and chip loading was conducted on the Ion Chef instrument using the Ion 520 & Ion 530 Kit-Chef (ThermoFisher Scientific). Ultimately, samples were sequenced on Ion 530 chips using the Ion S5 System (ThermoFisher Scientific). Sequencing data was accessed on the Torrent suite version 5.2.2. Reads were aligned against the human genome (hg19) using Alignment version 4.0-r77189, and variants were called using the coverage Analysis version 4.0-r77897 plugin. All 18 bespoke-panel ctDNA-negative patients had no tumour SNVs detectable in plasma preoperatively by the generic panel, supporting biological specificity of the bespoke targeted approach, 7 out of 10 bespoke-panel ctDNA-positive patients had tumour SNVs detected in plasma by the generic panel (Extended Data Table 2a, b). SNVs detected by the hotspot panel not identified by M-seq are displayed in Extended Data Table 2c.

Processing and phylogenetic analysis of relapse and primary tumour multi-region whole-exome data. Biopsies from multiple regions from the primary tumour ($n=327$), metastatic biopsies ($n=4$) and matching blood germline samples ($n=100$) were subjected to multi-region whole exome sequencing and analysis including estimation of copy number, purity and ploidy, and phylogenetic tree construction as previously described[9]. In brief, phylogenetic analysis was performed on the basis of the CCF determined for SNVs and clustered across tumour regions using a modified version of Pyclone[9] into clusters with similar CCF values, filtered and processed as previously described[9]. Mutation clusters are assumed to represent tumour subclones, either current or ancestral, and are used as input for construction of the phylogenetic relationship. Phylogenetic trees were primarily constructed using the published tool CITUP (0.1.0)[28]. However, in a small number of cases, including all relapse/autopsy cases, manual tree construction was required and performed as previously described[9]. Complete details of primary tumour tree construction can be found in ref. 9. Relapse tree construction was performed as follows. CRUK0063: clustering was performed twice, once across five primary tumour regions and once across five primary, one relapse, and six autopsy regions. To ensure consistency, when deriving a phylogenetic tree based on all tumour regions, for CCF clusters based on clustering, only the primary tumour regions were maintained for mutations not involved in metastatic relapse. A phylogenetic tree was constructed based on 17 mutation clusters. CRUK0035: clustering primary tumour regions with the relapse region revealed one cluster private to the relapse, and one cluster shared with the relapse and all other regions. CRUK0044: clustering primary tumour regions with the relapse region revealed a cluster private to the relapse, descended from a cluster private to region 1 in the primary tumour. CRUK0041: clustering primary tumour regions with the relapse region revealed cluster 4 as private to the relapse. This cluster must have evolved from cluster 3 only found in the relapse and in region 4. A private cluster 6 in region 4 must have evolved from cluster 4. However, this conflicts with clusters 2 and 5, found in the relapse and regions 1–3, but not region 4. This can be reconciled by assuming a polyclonal relapse, seeded primarily from regions 1–3, but with some contribution from cluster 3, private to region 4. Cluster data are available in Supplementary Table 4 under 'PyClonePhyloCluster'.

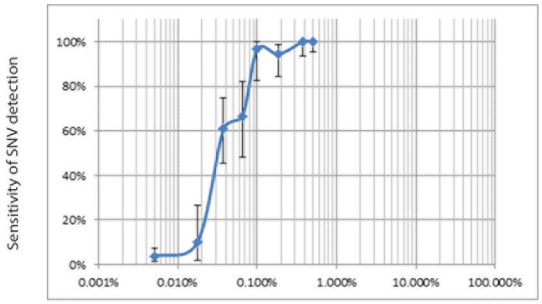
Statistical data analysis. No statistical methods were used to predetermine sample size. Analysis was performed in the R statistical environment version 3.2.3 and SPSS version 24. All statistical tests were two-sided, unless expressly stated. Multivariate logistic regression used detection of ctDNA (the dependent variable) classified as detection of two or more patient-specific variants in ctDNA and the covariates listed in Supplementary Table 1. All predictors were entered simultaneously into the regression. All continuous independent variables were found to be linearly related to the logit of the dependent variable (assessed using the Box–Tidwell procedure). The logistic regression model was statistically significant, $\chi^2_{10}=81.35$, $P<0.001$ and the Hosmer–Lemeshow P value was 0.9858, indicating that the model was not a poor fit. To determine the ability of PET TBR to predict whether or not tumour ctDNA was identified in plasma, PET TBR estimates were analysed by receiver operating characteristic (ROC) curve analysis against binary detection of ctDNA in plasma at baseline on the basis of at least two variants detected; significance was based on theWilcoxon rank-sum test. For analysis involving longitudinally detected variants (Fig. 4 and Extended Data Fig. 5), only subclonal variants from Pyclone clusters present in phylogenetic trees were displayed; this did not affect ctDNA detection status of any time points. In non-relapse cases presented in Extended Data Fig. 6 all detected subclonal SNVs were plotted. To determine the relationship between tumour volume and ctDNA VAF, ctDNA assays against clonal SNVs were selected. For each patient, the mean ctDNA VAF of the clonal SNVs was determined as baseline for 38 out of 46 patients with at least two SNVs detected in plasma. As detailed in Extended Data Fig. 4c, 9 out of 46 patients were not included in the analysis: CRUK0036 had no preoperative CT scan available; CRUK0087 and CRUK0096 had a large cavity inside the primary cancer; CRUK0099 had a collapsed lung making volume assessment inaccurate; CRUK0100, CRUK0077 and CRUK0052 had a CT slice spacing of >5 mm (CT slice spacing for all volumetric analyses detailed in Supplementary Table 1); and finally CRUK0088 and CRUK0091 had a total tumour volume of $<3.5\,\mathrm{cm}^3$. Linear regression was performed on log-transformed mean VAF and tumour volume. The log transformation was justified as it symmetrized the residuals in the model. An independent analysis was performed where tumour volume was multiplied with tumour purity to estimate the cancer cell volume. The same log transformation and analysis was applied to

data acquired from ref. 16, where ctDNA VAF was determined based on CAPP-seq analysis with matched tumour volume data available. To analyse clone size versus ctDNA VAF for subclonal SNVs, the mean CCF of the mutations within a subclonal mutation cluster was multiplied with tumour volume, and as a second independent analysis, with tumour purity.

Data availability. Sequence data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the The European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under accession numbers EGAS00001002247 (primary tumour data) and EGAS00001002415 (metastatic tumour data). Further information about EGA can be found at https://ega-archive.org (the European Genome-phenome Archive of human data consented for biomedical research).

24. Hofheinz, F. *et al.* An investigation of the relation between tumor-to-liver ratio (TLR) and tumor-to-blood standard uptake ratio (SUR) in oncological FDG PET. *EJNMMI Res.* **6,** 19 (2016).
25. Bütof, R. *et al.* Prognostic value of pretherapeutic tumor-to-blood standardized uptake ratio in patients with esophageal carcinoma. *J. Nucl. Med.* **56,** 1150–1156 (2015).
26. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620 (2014).
27. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
28. Malikic, S., McPherson, A. W., Donmez, N. & Sahinalp, C. S. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics* **31,** 1349–1356 (2015).

a Multiplex-PCR platform sensitivity and specificity



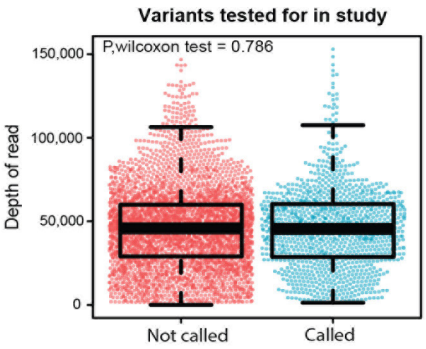
SNV spike input range	Eligible positive positions	True positive calls	Sensitivity	95% CI
>0.5	81	81	100.0%	95.6-100%
0.25-0.5%	54	54	100.0%	93.4-100%
0.1-0.25%	66	64	97.0%	89.5-99.6%
0.05-0.1%	51	43	84.3%	71.4-93.0%
0.01-0.05%	76	35	46.1%	34.6-57.9%
<=0.01%	212	9	4.2%	2-7.8%
	Eligible negative positions	False positive calls	Specificity	95% CI
	5099	19	99.6%	99.4-99.8%

b

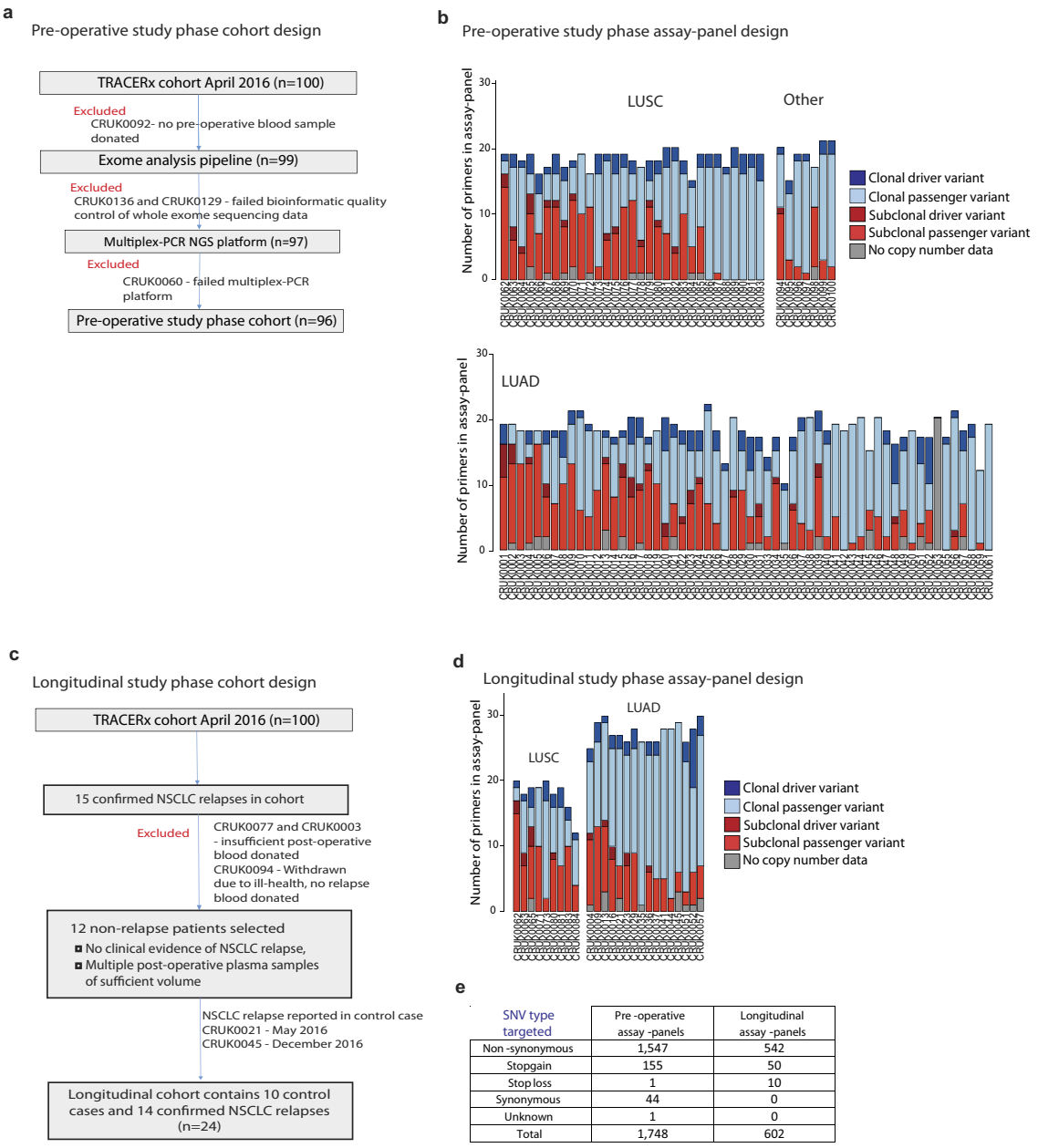
SNVs tested for in ctDNA per patient	Predicted specificity for ctDNA detection 1 SNV threshold	2 SNV threshold
1	99.63%	
5	98.16%	99.99%
10	96.36%	99.94%
20	92.85%	99.75%
30	89.48%	99.44%
50	83.08%	98.51%
100	69.03%	94.66%
200	47.65%	83.04%

2 SNV threshold caclulated based on binomial probability using false positive rate of 0.0037.

c



Extended Data Figure 1 | Multiplex-PCR NGS platform analytical validation. a, Analytical validation of the multiplex-PCR NGS platform was performed by spiking synthetic SNVs into control cfDNA. Sensitivity and specificity of the platform at different spike concentrations was ascertained, 95% binomial confidence interval displayed as error bars. **b**, Specificity of ctDNA detection based on a one-SNV and two-SNV call threshold taking into account parallel testing of multiple SNVs. **c**, The median read depth across a position did not vary depending on whether an SNV position was called or not called using the platform error model. Wilcoxon rank-sum test, $P = 0.786$. Median read depth at uncalled positions, 45,777; range, 0–146,774; $n = 3{,}745$. Median read depth at called positions, 45,478; range, 1,354–152,974; $n = 1{,}124$. Whiskers represent 1.5× the interquartile range, two-sided test.



e

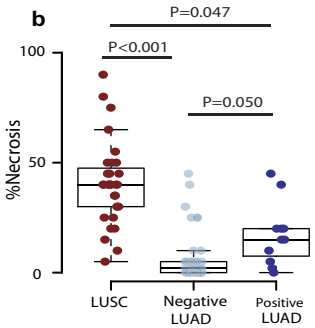
SNV type targeted	Pre -operative assay -panels	Longitudinal assay -panels
Non -synonymous	1,547	542
Stopgain	155	50
Stop loss	1	10
Synonymous	44	0
Unknown	1	0
Total	1,748	602

Extended Data Figure 2 | Study construction and assay panel design. **a**, The preoperative study phase cohort consisted of 100 TRACERx patients present in the first 100 patient TRACERx cohort in April 2016. Preoperative plasma samples were profiled for 96 patients for the reasons listed. **b**, Contents of patient-specific assay panels designed in the preoperative study phase for LUSC, LUAD and other. **c**, The longitudinal study phase cohort consisted of patients with confirmed NSCLC relapse and patients without relapse. **d**, Contents of patient-specific assay panels designed in the longitudinal phases of this study. **e**, SNV type that was targeted.

a

Histological subtype			TNM Stage I	II	III	Total
LUAD	ctDNA detected	Yes	5	2	4	11
		No	34	7	6	47
LUSC	ctDNA detected	Yes	16	12	2	30
		No	1	0	0	1
Other	ctDNA detected	Yes	1	2	2	5
		No	2	0	0	2

b



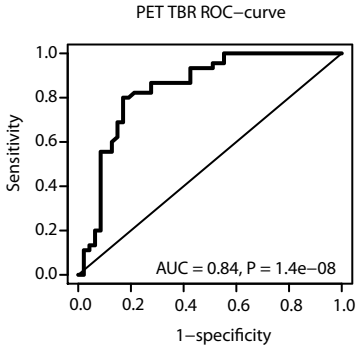
c

Predictors of ctDNA detection by multiplex-PCR NGS in early stage NSCLC

	Univariable analysis		Multivariable analysis	
	OR (95% CI)	P-value	OR (95% CI)	P-value
Clinicopathological variables				
Non-adenocarcinoma histology	49.85 (12.93 -192.19)	<0.001	40.76 (4.55 - 365.14)	0.001
%Ki67$^+$ cells (10% increase)	1.72 (1.40 - 2.12)	<0.001	1.40 (1.05 - 1.84)	0.022
Lympho-vascular invasion	2.53 (1.10 - 5.80)	0.028	5.84 (1.07 - 32.03)	0.042
Necrosis (10% increase)	2.16 (1.58 - 2.97)	<0.001	1.04 (0.64 - 1.71)	0.862
Path tumour size (10mm increase)	1.45 (1.13 - 1.86)	0.004	1.32 (0.91-1.91)	0.134
Lymph-node involvement	3.60 (1.33 - 9.77)	0.012	3.82 (0.61 - 23.99)	0.153
Male gender	1.80 (0.78 - 4.16)	0.172	1.06 (0.21-5.39)	0.941
Age (years)	0.96 (0.92 - 1.01)	0.115	0.99 (0.92-1.07)	0.820
Technical variables				
cfDNA input (ng)	1.01 (1.00 - 1.03)	0.028	1.01 (1.00 - 1.02)	0.229
Ubiquitous SNVs in assay-panel	0.96 (0.88 - 1.05)	0.341	1.00 (0.84 - 1.19)	0.975

Ubiquitous SNVs (SNVs present in all tumour regions sequenced).

d



e

LUAD

		TP53 driver Yes	No	Total
ctDNA detected	Yes	4	7	11
	No	24	23	47
Total		28	30	58

Fishers Exact Test P=0.508

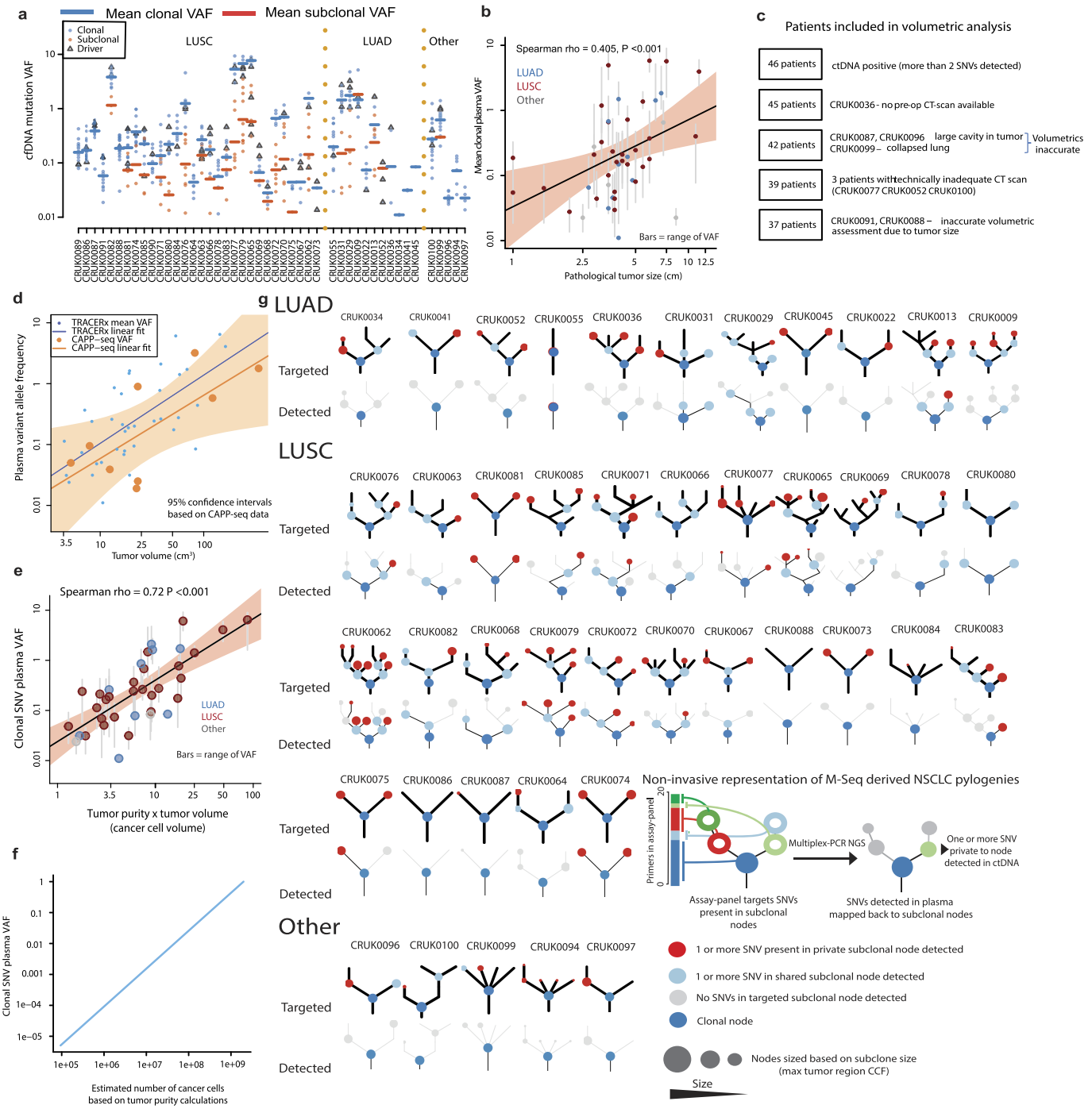
		KRAS driver Yes	No	Total
ctDNA detected	Yes	3	8	11
	No	20	27	47
Total		23	35	58

Fishers Exact Test P=0.499

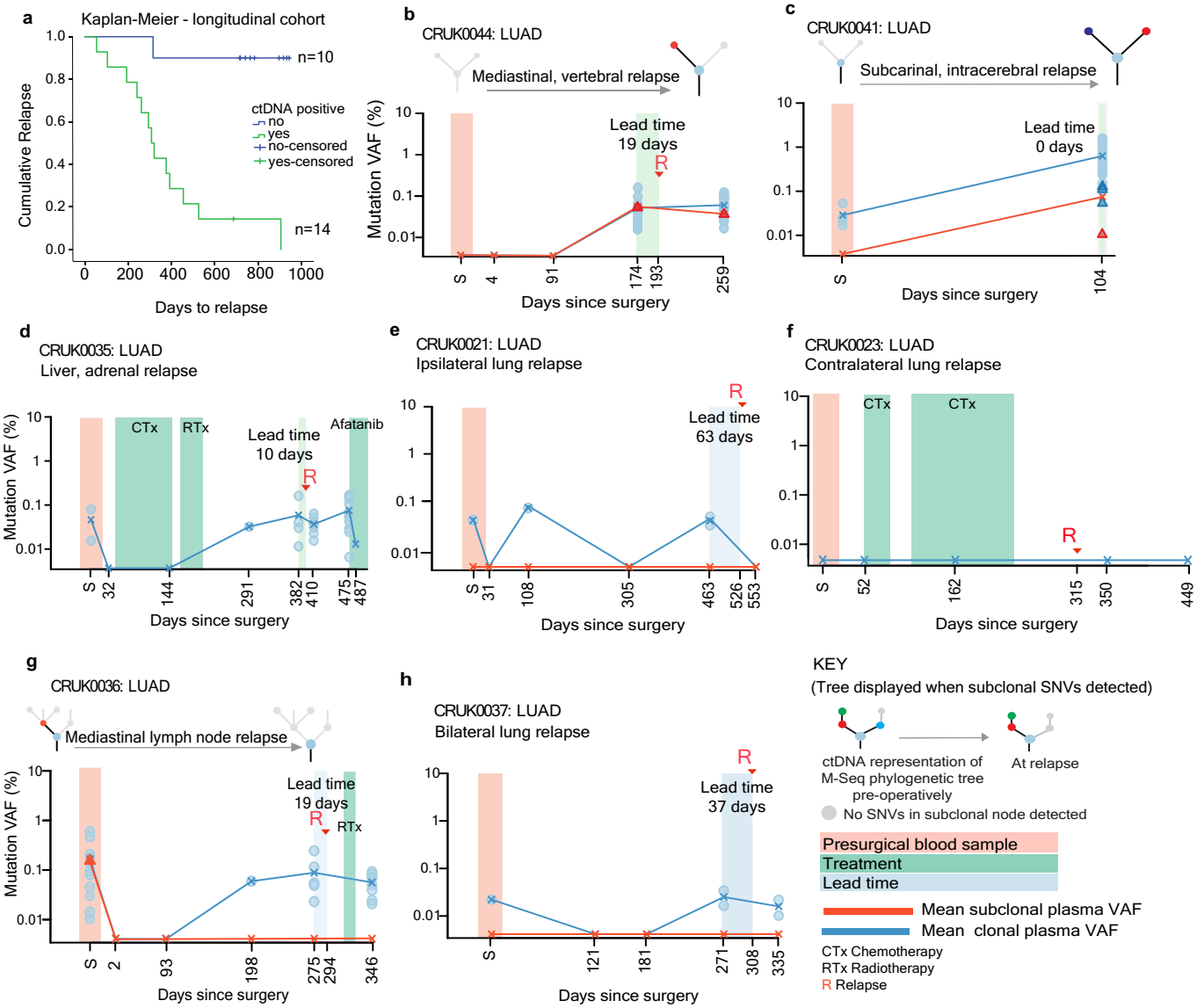
		EGFR driver Yes	No	Total
ctDNA detected	Yes	1	10	11
	No	11	36	47
Total		12	46	58

Fishers Exact Test P=0.429

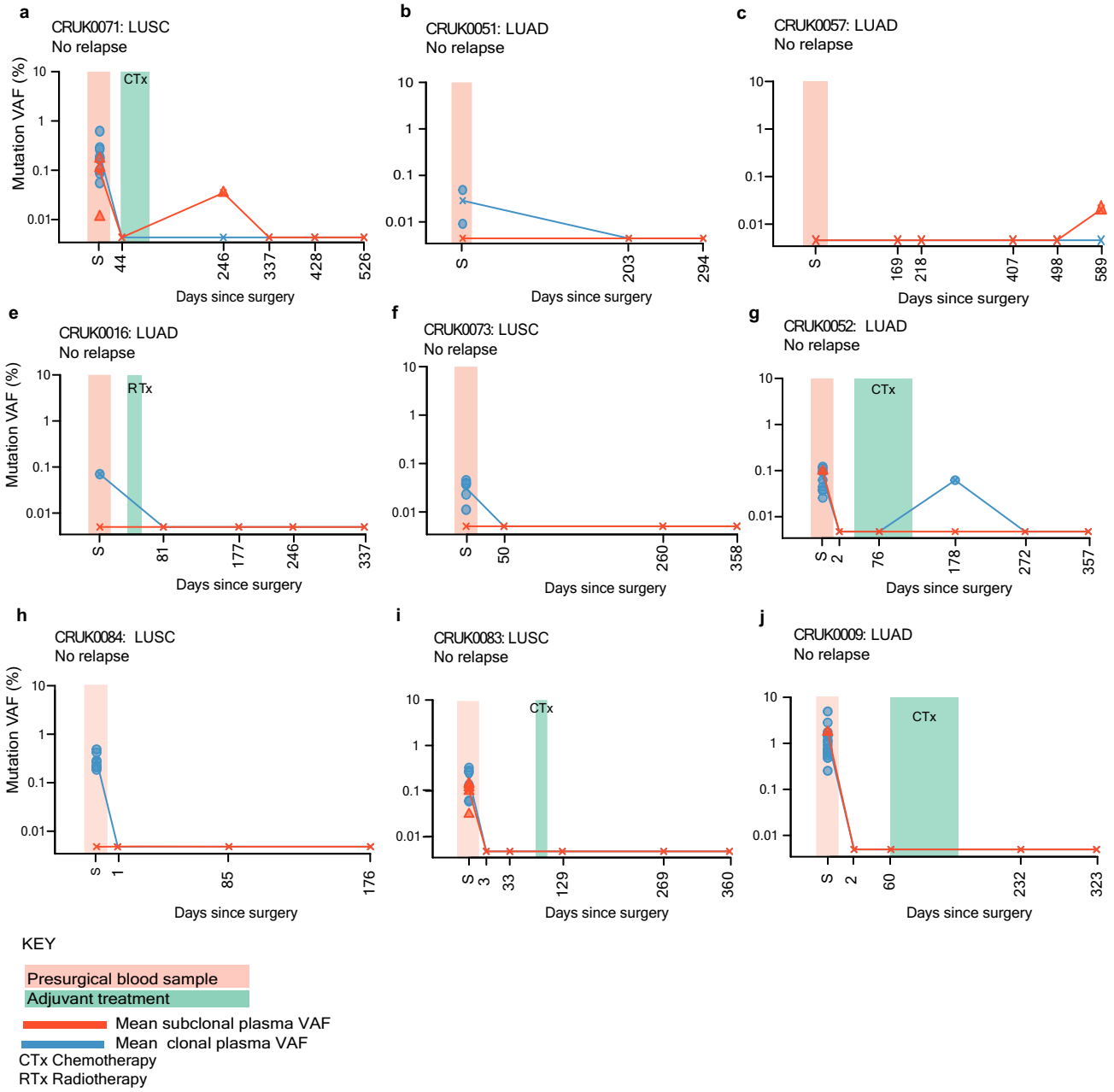
Extended Data Figure 3 | Clinicopathological predictors of ctDNA detection. a, 96 patients in the preoperative cohort stratified by pathological tumour, node and metastasis (TNM) stage. **b**, LUSCs and ctDNA-positive LUADs are significantly more necrotic that ctDNA-negative LUADs. Significant differences in necrosis between groups: LUSCs (median necrosis, 40%; $n = 31$), ctDNA-positive LUADs (median necrosis, 15%; $n = 11$) and ctDNA-negative LUADs (median necrosis, 2%; $n = 47$), Kruskal–Wallis test, $P < 0.001$, two-sided pairwise comparisons were performed using Dunn's procedure with Bonferroni correction. **c**, Univariate (left) and multivariate analyses (right) were performed, by logistic regression to determine significant predictors of ctDNA detection in early-stage NSCLC. ctDNA detection was defined as detection of two or more SNVs in preoperative plasma samples. Details regarding multivariable analysis methodology are in the Methods. **d**, Receiver operating characteristic (ROC) curve analysis of preoperative PET scan FDG avidity (normalized to tumour background ratio (TBR), see Methods), as a predictor of ctDNA detection (92 out of 96 PET scans were available for central review), determined by the area under the curve (AUC). Median PET TBR of detected tumours = 9.01, $n = 45$. Median PET TBR of undetected tumours = 3.64, $n = 47$. *P* value based on Wilcoxon rank-sum test. **e**, LUAD subtype analyses based on ctDNA detection and the presence of an *EGFR*, *KRAS* or *TP53* driver mutation.



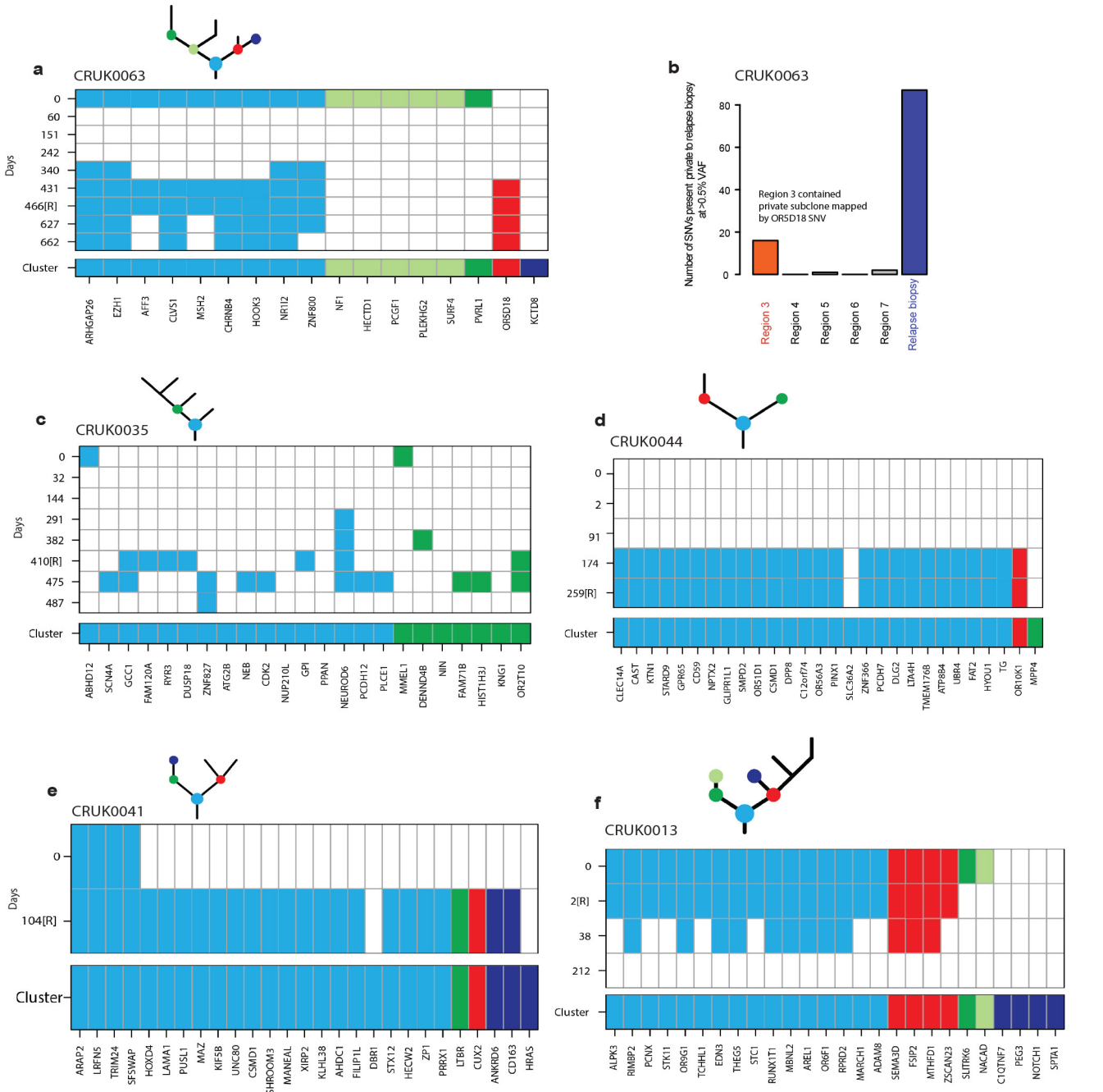
Extended Data Figure 4 | Predictors of plasma VAF. a, Plasma VAFs of SNVs detected in plasma in 46 patients who were ctDNA-positive (two or more SNVs detected). Clonal (blue) and subclonal (red) VAFs are indicated, mean is shown as a horizontal line. Driver variants are shown as triangles. **b**, Mean clonal VAF correlated with maximum tumour size measured in a post-surgical specimen (pathologic size, $n = 46$). Grey vertical bars represent range of clonal VAF; shaded red background indicates the 95% confidence interval. **c**, Filtering steps taken to define a group of ctDNA-positive patients with volumetric data considered adequate to model tumour volume and plasma VAF. **d**, Scatter plot showing mean clonal VAF relative to tumour volume for TRACERx (blue dots and fitted blue line, $n = 37$) and VAF relative to volume for previously published data[16] based on CAPP-seq analysis of ctDNA (orange dots and orange fitted line, $n = 9$). Orange shaded background indicates the 95% confidence interval based on CAPP-seq data. **e**, Mean clonal VAF correlated with tumour volume × tumour purity (cancer cell volume), $n = 37$. Shaded red background indicates the 95% confidence interval. **f**, Association between the number of cancer cells and VAF of clonal SNVs in plasma based on linear modelling of Extended Data Fig. 4e and the assumption that a cancer cell volume of 1 cm^3 contains 9.4×10^7 cells[17]. **g**, Detected subclonal SNVs were mapped back to M-seq-derived tumour phylogenetic trees (process illustrated in the key). Detected private subclones (subclones identified within only a single tumour region) are coloured red. Shared subclones (subclones detected in more than one tumour region) are light blue. Subclonal nodes were sized on the basis of the maximum recorded CCF. The top row of the phylogenetic trees represent subclonal nodes targeted by primers within that patient's assay panel, the bottom row represent subclonal nodes detected in ctDNA, within this row grey subclonal nodes represent subclones that were not detected in ctDNA.



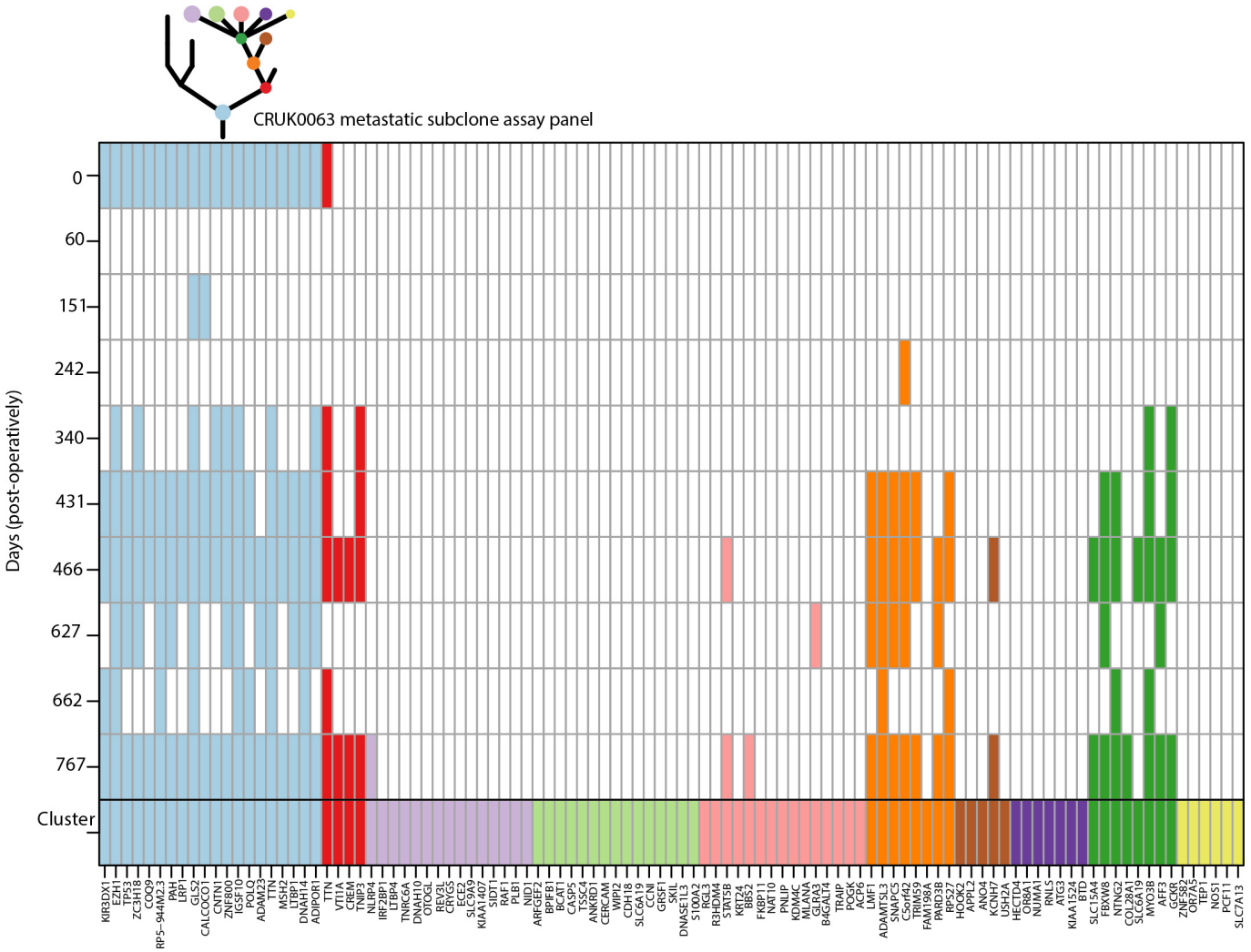
Extended Data Figure 5 | Longitudinal ctDNA profiling, remaining relapse cases. a, Kaplan–Meier curve demonstrating relapse-free survival for patients in whom ctDNA was detected versus patients in whom ctDNA was not detected. **b**–**h**, Longitudinal cfDNA profiling. ctDNA detection in plasma was defined as the detection of two tumour-specific SNVs. Relapse was based on imaging-confirmed NSCLC relapse, imaging performed as clinically indicated. Detected clonal (circles, light blue) and subclonal (triangles, colours indicate different subclones) SNVs from each patient-specific assay panel are plotted coloured by M-seq-derived tumour phylogenetic nodes. Mean clonal (blue) and mean subclonal (red) VAFs are indicated on graphs. Preoperative and relapse M-seq-derived phylogenetic trees represented by ctDNA are illustrated above each graph in cases where subclonal SNVs were detected.



Extended Data Figure 6 | Longitudinal ctDNA profiling, non-relapse cases. a–j, Detected clonal (circles, light blue) and subclonal (red triangles) SNVs from each patient-specific assay panel are plotted. Mean clonal (blue) and mean subclonal (red) VAFs are indicated.



Extended Data Figure 7 | Heat maps illustrating detection of SNVs in bespoke panels at each sampled time point. a, **c**–**f**, Bespoke assay panels for CRUK0063 (**a**), CRUK0035 (**c**), CRUK0044 (**d**), CRUK0041 (**e**) and CRUK0013 (**f**). Colours indicate originating subclonal cluster based on the phylogenetic trees above the heat map. Light blue indicates clonal mutation cluster. Full panel with cluster colour is shown below each heat map. Filled squares indicate detection of a given variant in plasma ctDNA. The *y* axis shows day of sampling; *y* axis labels appended with [R] indicate day of clinical relapse. The *x* axis indicates SNVs targeted by the assay panel. **b**, Re-examination of primary tumour regions from CRUK0063 with a lowered threshold to potentially identify SNVs private to the sequenced relapse biopsy. 16 out 88 variants were found at very low VAF in region 3, indicating that this region from the primary probably gave rise to the metastasis.



Extended Data Figure 8 | Heat map illustrating the content of the metastatic bespoke panel designed for patient CRUK0063, demonstrating detection status of SNVs across all sampled time points. Colours indicate originating subclonal cluster based on patient CRUK0063's phylogenetic tree above the heat map. Light blue indicates clonal mutation cluster. Full SNV panel with cluster colour shown below each heat map. Filled squares indicate detection of a given variant in plasma ctDNA. The *y* axis shows the postoperative day of sampling.

Extended Data Table 1 | Patient characteristics

a

Clinical characteristics
96 patient pre-operative cohort

Characteristic		Total
Age	<60	19
	≥60	77
Sex	Male	60
	Female	36
ECOG PS	0	49
	1	47
Histology	Adenocarcinoma	58
	Squamous cell carcinoma	31
	Carcinosarcoma	2
	Large cell carcinoma	1
	Adenosquamous carcinoma	3
	Large cell neuroendocrine carcinoma	1
TNM stage	Ia	24
	Ib	35
	IIa	12
	IIb	11
	IIIa	13
	IIIb	1
Lymph node metastasis	Yes	24
	No	72
Pleural involvement	Yes	27
	No	69
Vascular invasion	Yes	41
	No	55
Resection margin	R0	91
	R1	5
Smoking status	Never smoked	11
	Recent ex-smoker	30
	Ex-smoker	48
	Current smoker	7
Ethnicity	White British	85
	White-other	4
	White-Irish	4
	Caribbean	3

b

		No adjuvant therapy	Adjuvant therapy
TNM Stage	Ia	24	0
	Ib	31	4
	IIa	3	9
	IIb	4	7
	IIIa	6	7
	IIIb	0	1

c

Details regarding timing of pre-operative blood sample

Pre-surgery	Number	Details
Less than 24 hours	91	
24-72 hours	2	CRUK0051, 0003
8 days	2	CRUK0073, 0096
31 days	1	CRUK0089

d

Clinical characteristics
24 patient longitudinal sub-cohort

Characteristic		Total
Age	<60	5
	≥60	19
Sex	Male	16
	Female	8
ECOG PS	0	12
	1	12
Histology	Adenocarcinoma	16
	Squamous cell carcinoma	8
TNM stage	Ia	3
	Ib	7
	IIa	3
	IIb	7
	IIIa	3
	IIIb	1
Lymph node metastasis	Yes	9
	No	15
Pleural involvement	Yes	7
	No	17
Vascular invasion	Yes	12
	No	12
Resection margin	R0	23
	R1	1
Smoking status	Never smoked	1
	Recent ex-smoker	5
	Ex-smoker	16
	Current smoker	2
Ethnicity	White British	21
	White-other	2
	Caribbean	1

e

		No adjuvant therapy	Adjuvant therapy
TNM Stage	Ia	3	0
	Ib	6	1
	IIa	0	3
	IIb	2	5
	IIIa	1	2
	IIIb	0	1

a, Clinical characteristics of the preoperative cohort of 96 patients. **b**, Distribution of cancer stage and whether the patient received adjuvant chemotherapy. **c**, The time points at which preoperative plasma was acquired for patients within the cohort. **d**, Clinical characteristics of the longitudinal cohort of 24 patients. **e**, The distribution of cancer stage in the longitudinal cohort and whether the patient received adjuvant chemotherapy.

Extended Data Table 2 | Cross-platform validation using a generic approach to ctDNA profiling

Table 2 a - Bespoke panel detected NSCLCs - cross platform validation

			Bespoke-panel			Generic-panel
Case	Volume cm3	Plasma VAF (mean clonal)	ctDNA positive	Histology	Hotspot SNVs tumor	Hotspot SNVs detected
CRUK0029	38.51	2.10	Yes	LUAD	1	1
CRUK0009	69.01	1.71	Yes	LUAD	1	1
CRUK0062	58.48	1.41	Yes	LUSC	1	1
CRUK0081	16.41	0.21	Yes	LUSC	1	1
CRUK0089	17.39	0.16	Yes	LUSC	1	1
CRUK0022	17.20	0.08	Yes	LUAD	1	0
CRUK0067	6.64	0.07	Yes	LUSC	1	0
CRUK0052	43.69	0.06	Yes	LUAD	2	1
CRUK0064	9.24	0.05	Yes	LUSC	1	0
CRUK0034	10.59	0.01	Yes	LUAD	1	1

Table 2 b - Bespoke panel non-detected NSCLCs - cross platform validation

			Bespoke-panel			Generic-panel
Case	Volume cm3	Predicted plasma VAF	ctDNA positive	Histology	Hotspot SNVs tumor	Hotspot SNVs detected
CRUK0037	197.42	2.96 (1.01 to 8.67)	No	LUAD	1	0
CRUK0051	27.28	0.32 (0.21 to 0.49)	No	LUAD	1	0
CRUK0004	23.30	0.27 (0.18 to 0.41)	No	LUAD	1	0
CRUK0039	21.68	0.25 (0.16 to 0.38)	No	LUAD	2	0
CRUK0025	19.06	0.22 (0.14 to 0.33)	No	LUAD	2	0
CRUK0048	17.00	0.19 (0.12 to 0.29)	No	LUAD	2	0
CRUK0026	7.45	0.08 (0.04 to 0.14)	No	LUAD	1	0
CRUK0030	7.28	0.07 (0.04 to 0.14)	No	LUAD	2	0
CRUK0057	5.95	0.06 (0.03 to 0.12)	No	LUAD	1	0
CRUK0018	4.65	0.04 (0.02 to 0.10)	No	LUAD	1	0
CRUK0027	4.61	0.04 (0.02 to 0.10)	No	LUAD	1	0
CRUK0007	4.18	0.04 (0.02 to 0.09)	No	LUAD	1	0
CRUK0049	3.61	0.03 (0.01 to 0.08)	No	LUAD	1	0
CRUK0035	3.31	0.03 (0.01 to 0.08)	No	LUAD	1	0
CRUK0058	2.76	0.03 (0.01 to 0.07)	No	LUAD	1	0
CRUK0021	2.70	0.02 (0.01 to 0.07)	No	LUAD	2	0
CRUK0093	0.73	0.01 (0.001 to 0.03)	No	LUSC	2	0
CRUK0014	0.90	0.01 (0.002 to 0.03)	No	LUAD	1	0

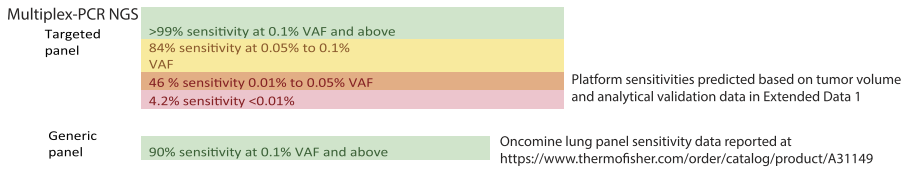


Table 2 c - Variants detected by generic PCR-NGS hotspot panel not detected in M-Seq analysis of tumor

										(unfiltered)	(unfiltered)
Case	Gene	Location	Position	Ref	Variant	AA change	Plasma VAF	DOR	ctDNA positive	Combined exome VAF	Germline VAF
CRUK0052	PIK3CA	chr3	178936091	G	A	p.E545K	0.81	60360	Yes	ND	ND
CRUK0052	PIK3CA	chr3	178952085	A	G	p.H1047R	0.12	52325	Yes	0.075	ND
CRUK0062	PIK3CA	chr3	178936091	G	A	p.E545K	0.97	89616	Yes	0.016	ND
CRUK0062	PIK3CA	chr3	178952085	A	G	p.H1047R	0.05	79205	Yes	0.005	ND
CRUK0062	TP53	chr17	7577556	C	A	p.C242F	0.05	93383	Yes	ND	ND
CRUK0089	TP53	chr17	7577121	G	A	p.R273C	0.06	59849	Yes	0.168	ND
CRUK0004	PIK3CA	chr3	178936091	G	A	p.E545K	0.59	73941	No	0.081	ND
CRUK0018	PIK3CA	chr3	178936091	G	A	p.E545K	4.44	99159	No	ND	ND
CRUK0018	PIK3CA	chr3	178952085	A	G	p.H1047R	0.81	77806	No	0.044	ND
CRUK0021	PIK3CA	chr3	178952085	A	G	p.H1047R	0.11	50107	No	ND	ND
CRUK0027	PIK3CA	chr3	178952085	A	G	p.H1047R	0.11	65449	No	ND	ND
CRUK0037	PIK3CA	chr3	178952085	A	G	p.H1047R	0.09	51071	No	ND	ND
CRUK0058	KRAS	chr12	25398284	C	A	p.G12V	3.44	63090	No	0.124	ND

ND - non detected
DOR - depth of read
Combined exome VAF (unfiltered) - Variant allele frequency across all tumor regions analysed (without call filters).

a, 7 out of 10 (70%) of bespoke-panel ctDNA-positive patients had tumour SNVs detectable in plasma preoperatively by a generic hotspot PCR NGS lung panel (Lung Oncomine, Thermofisher). The three bespoke-panel ctDNA-positive patients undetected by the generic panel had mean clonal plasma VAFs lower than the 0.1% plasma VAF limit of detection reported for the generic panel (shaded yellow). **b**, On the basis of the CT volumetric assessment of each patient's primary tumour we predicted plasma VAF corresponding to a tumour of that size (see Fig. 3 and Methods for details of the approach). This allowed us to infer platform sensitivities for each patient within the bespoke-panel non-detected cohort. Six LUADs (shaded green; CRUK0037, CRUK0051, CRUK0004, CRUK0039, CRUK0025 and CRUK0048) had tumour volumes approximating to a plasma VAF of more than 0.1%. This suggested that these tumours resided within the top platform sensitivity bracket of both the generic and bespoke-panel ctDNA platforms. No ctDNA was detected by either platform in these cases, suggesting biological specificity of the bespoke-panel. **c**, Hotspot SNVs not identified in tumour tissue through exome sequencing were identified in plasma of 9 out of 28 patients by the generic panel. This suggested a non-tumour origin of cfDNA, platform non-specificity or an evolving minor subclone or second primary. DOR, depth of read; ND, not detected. Combined exome VAF (unfiltered), VAF across all tumour regions analysed (without call filters).

CORRIGENDUM

doi:10.1038/nature25161

Corrigendum: Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution

Christopher Abbosh, Nicolai J. Birkbak, Gareth A. Wilson, Mariam Jamal-Hanjani, Tudor Constantin, Raheleh Salari, John Le Quesne, David A. Moore, Selvaraju Veeriah, Rachel Rosenthal, Teresa Marafioti, Eser Kirkizlar, Thomas B. K. Watkins, Nicholas McGranahan, Sophia Ward, Luke Martinson, Joan Riley, Francesco Fraioli, Maise Al Bakir, Eva Grönroos, Francisco Zambrana, Raymondo Endozo, Wenya Linda Bi, Fiona M. Fennessy, Nicole Sponer, Diana Johnson, Joanne Laycock, Seema Shafi, Justyna Czyzewska-Khan, Andrew Rowan, Tim Chambers, Nik Matthews, Samra Turajlic, Crispin Hiley, Siow Ming Lee, Martin D. Forster, Tanya Ahmad, Mary Falzon, Elaine Borg, David Lawrence, Martin Hayward, Shyam Kolvekar, Nikolaos Panagiotopoulos, Sam M. Janes, Ricky Thakrar, Asia Ahmed, Fiona Blackhall, Yvonne Summers, Dina Hafez, Ashwini Naik, Apratim Ganguly, Stephanie Kareht, Rajesh Shah, Leena Joseph, Anne Marie Quinn, Phil A. Crosbie, Babu Naidu, Gary Middleton, Gerald Langman, Simon Trotter, Marianne Nicolson, Hardy Remmen, Keith Kerr, Mahendran Chetty, Lesley Gomersall, Dean A. Fennell, Apostolos Nakas, Sridhar Rathinam, Girija Anand, Sajid Khan, Peter Russell, Veni Ezhil, Babikir Ismail, Melanie Irvin-Sellers, Vineet Prakash, Jason F. Lester, Malgorzata Kornaszewska, Richard Attanoos, Haydn Adams, Helen Davies, Dahmane Oukrif, Ayse U. Akarca, John A. Hartley, Helen L. Lowe, Sara Lock, Natasha Iles, Harriet Bell, Yenting Ngai, Greg Elgar, Zoltan Szallasi, Roland F. Schwarz, Javier Herrero, Aengus Stewart, Sergio A. Quezada, Karl S. Peggs, Peter Van Loo, Caroline Dive, C. Jimmy Lin, Matthew Rabinowitz, Hugo J. W. L. Aerts, Allan Hackshaw, Jacqui A. Shaw, Bernhard G. Zimmermann, The TRACERx consortium, The PEACE consortium & Charles Swanton

Nature **545,** 446–451 (2017); doi:10.1038/nature22364

For 6 of the 96 patients included in this Article (patients CRUK0014, CRUK0030, CRUK0048, CRUK0059, CRUK0096 and CRUK0097) incorrect tumour volumetric data and positron emission tomography (PET) tumour background ratio (TBR) data were analysed. This error occurred because of the incorrect assignment of patient identifiers during the anonymization mandated by the independent review board of pre-operative computed tomography (CT) scans belonging to these patients. Data relating to this error were presented in Figs 2a and 3a and b, Extended Data Figs 3d and 4c–f, Extended Data Table 2b and Supplementary Table 1. The reanalysis of correctly anonymized scans does not influence the conclusions of this Article and correlation coefficients improve following inclusion of the corrected data. These errors have been corrected online in the original Article. The authors apologize for any confusion these errors may have caused.

Quartiles for the heat map in Fig. 2a have been redefined after including the correct data to reflect changes in quartiles for 3 (of 92) PET TBR values and 7 (of 95) volume parameters. The Source Data file supplied for Fig. 2 was not uploaded on publication; the corrected Source Data file for Fig. 2 is now available in the HTML version of the original Article.

The plot and legend for Fig. 3a have been corrected to reflect updated volumetric data for the two patients affected by the correction who were analysed in this figure (CRUK0096 and CRUK0097). CRUK0096 was excluded from the updated volumetric analysis based on a criterion applied to our original analysis (large cavity within primary tumour). Consequently, the sentence in the legend to Fig. 3a "$n = 38$, grey vertical lines represent range of clonal VAF, red shading indicates 95% confidence intervals (CIs)" has been updated to read "$n = 37$" and in the Methods section 'Statistical data analysis' the line "8 out of 46 patients were not included in the analysis: CRUK0036 had no preoperative CT scan available; CRUK0087 had a large cavity inside the primary" has been updated to read "9 out of 46 patients were not included in the analysis: CRUK0036 had no preoperative CT scan available; CRUK0087 and CRUK0096 had a large cavity inside the primary cancer".

In Fig. 3b and the main text, the variant allele frequency (VAF) prediction values (based on tumour volume), confidence intervals and estimated malignant cell number contributing to a VAF of 0.1% have been updated. In the section 'Determinants of ctDNA detection in NSCLC', confidence intervals in the sentence "a primary tumour burden of 10 cm^3 would result in a mean clonal plasma VAF of 0.1% (95% confidence interval, 0.05–0.17%)" have been altered to read "a primary tumour burden of 10 cm^3 would result in a mean clonal VAF of 0.1% (95% confidence interval, 0.06–0.18%)" and the sentence "a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 326 million tumour cells" has been altered to read "a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 302 million tumour cells". In the 'Discussion' section the sentence "on the basis of the relationship between tumour volume and ctDNA plasma VAF observed in this study, a tumour volume of 0.034 cm^3 would equate to a plasma VAF of 1.4×10^{-4}% (95% confidence interval, 6.4×10^{-6}–0.0031%)", has been altered to read "on the basis of the relationship between tumour volume and ctDNA plasma VAF observed in this study, a tumour volume of 0.034 cm^3 would equate to a VAF of 1.8×10^{-4}% (95% confidence interval, 9.8×10^{-6}–0.0033%)".

Further figure corrections pertaining to the six affected patients in Extended Data Figs 3d and 4c–f, Extended Data Table 2b and Supplementary Table 1 of the original Article are described and corrected in the Supplementary Information of this Corrigendum, which also shows the original, wrong Figs 2a and 3a and b. The Supplementary Data (containing Supplementary Table 1) of the original Article has been corrected.

Supplementary Information is available in the online version of this Corrigendum.