High resolution phylogenetic ctDNA tracking predicts relapse risk and metastatic dissemination patterns in early-stage lung cancer

Authors: Christopher Abbosh[1*+], Alexander M. Frankell[1,2*], Thomas Harrison[3*], Judit Kisistok[4*], Aaron Garnett[3*], Laura Johnson[3], Selvaraju Veeriah[1], Mike Moreau[3], Adrian Chesh[3], Tafadzwa L. Chaunzwa[5,6], Jakob Weiss[5,6,22], Morgan Weichert[3], Sophia Ward[2,7], Kristiana Grigoriadas[1,2,8], Aamir Shahpurwalla[3], Kevin Litchfield[1,9], Clare Puttick[1,2,8], Dhruva Biswas[1,2,10], Takahiro Karasaki[1,2], James RM Black[1,8], Carlos Martinez-Ruiz[1,8], Maise Al Bakir[1,2], Oriol Pich[2], Thomas BK Watkins[2], Emilia Lim[2], Ariana Huebner[1,2,8], David A. Moore[1,2,11], Nadia Godin-Heymann[12], Anne L'Hernault[12], Hannah Bye[12], Aaron Odell[3], Paula Roberts[3], Fabio Gomes[14], Akshay J. Patel[15], Elizabeth Manzano[1], Crispin Hiley[1,2], Nicolas Carey[16], Joan Riley[16], Daniel Cook[2], Darren Hodgeson[12], Daniel Stetson[13], Carl Barrett[13], Roderik M. Kortlever[17], Gerard Evans[17], Allan Hackshaw[18], Bob Daber[3], Jacqui A. Shaw[16], Hugo JWL Aerts[5,6,19], Abel Licon[3], Josh Stahl[3], Mariam Jamal-Hanjani[1,20,21], Nicolai J Birkbak[4,23,24]^, Nicholas McGranahan[1,8]^+, Charles Swanton[1,2,21]^+

Lung TRACERx Consortium.

Joint first authors*, joint last authors^, corresponding authors[+]

Affiliations: [1]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK; [2]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK; [3]Invitae, San Francisco, California; [4]Department of Molecular Medicine, Aarhus University Hospital, Aarhus N, Denmark; [5]Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA; [6]Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; [7]Advanced Sequencing Facility, The Francis Crick Institute, London UK;[8]Cancer Genome Evolution Research Group, Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK;[9]The tumour Immunogenomics and Immunosurveillance (TIGI) Lab, University College London Cancer Institute, London, UK;[10]Bill Lyons Informatics Centre, University College London Cancer Institute, London, UK; [11] Department of Cellular Pathology, University College London Hospitals, London, UK; [12]AstraZeneca, Cambridge, UK; [13]AstraZeneca, Waltham, MA, USA;[14]The Christie NHS Foundation Trust, Manchester, UK;[15]University Hospital Birmingham NHS Foundation Trust, Birmingham, UK;[16]Cancer Research Centre, University of Leicester, Leicester, UK; [17]Department of Biochemistry, University of Cambridge, Cambridge, UK; [18]Cancer Research UK and UCL Cancer Trials Centre, London, UK;[19]Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, The Netherlands;[20]Cancer Metastasis Laboratory, University College London Cancer Institute, London, UK;[21]Department of Medical Oncology, University College London Hospitals, London, UK.[22]Department of Radiology, Freiburg University Hospital, 79106 Freiburg, Germany. [23]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark; [24]Bioinformatics Research Center, Aarhus University, Aarhus, Denmark

Declarations of interest: C.A. has received speaking honoraria or expenses from AstraZeneca and Bristol Myers Squibb and reports employment at AstraZeneca. C.A. and C.S. hold European patent applications relating to assay technology to detect tumour recurrence (PCT/GB2017/053289) and declare a patent application (PCT/US2017/028013) for methods to detect lung cancer. A.M.F., C.A. and C.S. are named inventors on a patent application to determine methods and systems for tumour monitoring (GB2114434.0). S.V is a co-inventor to a patent of methods for detecting molecules in a sample (U.S. Patent # 10,578,620). C.A., C.S., K.L., C.P., T.H., L.J., M.W., A.G., and A.L., intend to apply for a patent related to a ctDNA detection algorithm. T.H., A.G., M.M, A.C., A.S., A.O., L.J., P.R., M.W., B.D., A.L. and J.S. are former or current employees of Invitae or ArcherDx and report stock ownership. D.B. reports personal fees from NanoString and AstraZeneca and has a patent (PCT/GB2020/050221) application on methods for cancer prognostication. D.A.M. reports

consultancy fees from AstraZeneca, Thermo Fisher, Takeda, Amgen, Janssen and Eli Lilly; speaker fees from AstraZeneca and Takeda. N.G-H., A.L-H, H.B., D.H., D.S., and C.B., report stock ownership and employment at AstraZeneca. C.H. has received speaker fees from AstraZeneca. M.J.-H. is a CRUK Career Establishment Awardee and acknowledges grant support from CRUK, IASLC International Lung Cancer Foundation, Lung Cancer Research Foundation, Rosetrees Trust and NIHR UCLH Biomedical Research Centre. M.J-H. has consulted, and is a member of the Scientific Advisory Board and Steering Committee, for Achilles Therapeutics, has received speaker honoraria from Astex Pharmaceuticals, Oslo Cancer Cluster, and has applied for patent protection on methods for lung cancer detection (PCT/US2017/028013). H.J.W.L.A. acknowledges financial support from NIH (NIH-USA U24CA194354, NIH-USA U01CA190234, NIH-USA U01CA209414, and NIH-USA R35CA22052), and the European Union - European Research Council (HA: 866504). H.J.W.L.A. has also received personal fees and stock from Onc.AI and Love Health Inc, and speaking honoraria from Bristol Myers Squibb. C.S. acknowledges grant support from Pfizer, AstraZeneca, Bristol Myers Squibb, Roche-Ventana, Boehringer-Ingelheim, Invitae (previously Archer Dx Inc - collaboration in minimal residual disease sequencing technologies), and Ono Pharmaceutical. He is an AstraZeneca Advisory Board member and Chief Investigator for the AZ MeRmaiD 1 and 2 clinical trials and is also chief investigator of the NHS Galleri trial. He has consulted for Amgen, AstraZeneca, Pfizer, Novartis, GlaxoSmithKline, MSD, Bristol Myers Squibb, Illumina, Genentech, Roche-Ventana, GRAIL, Medicxi, Metabomed, Bicycle Therapeutics, Roche Innovation Centre Shanghai, and the Sarah Cannon Research Institute, C.S. had stock options in Apogen Biotechnologies and GRAIL until June 2021, currently has stock options in Epic Bioscience, and has stock options and is co-founder of Achilles Therapeutics. C.S. holds additional patent applications related to targeting neoantigens (PCT/EP2016/059401), identifying patent response to immune checkpoint blockade (PCT/EP2016/071471), determining HLA LOH (PCT/GB2018/052004), predicting survival rates of patients with cancer (PCT/GB2020/050221), identifying patients who respond to cancer treatment (PCT/GB2018/051912) and both a European and US patent application related to identifying insertion/deletion mutation targets (PCT/GB2018/051892).

Summary| Large, prospective patient cohorts incorporating comprehensive longitudinal plasma sampling, detailed clinical genomic annotations and long follow-up are required to decipher the role of circulating tumour DNA (ctDNA) as a phylogenetic biomarker in early-stage non-small-cell lung cancer (NSCLC). Here, we developed high-resolution ctDNA methods tracking a median of 200 tumour-specific mutations, across 1069 plasma samples collected from 197 patients enrolled in the lung TRACERx study. We show that preoperative ctDNA detection specifically identified lung adenocarcinomas with poor survival outcomes displaying chromosomal instability, copy-number aberration and upregulation of proliferation related transcriptomic pathways. Postoperative plasma analyses were performed and interpreted within the context of standard of care radiological surveillance and administration of adjuvant therapy. Landmark analysis of plasma samples collected within 120 days post-surgery revealed ctDNA detection in 23% of patients, including 45% of all patients who experienced clinical relapse. A landmark positive status associated with early-NSCLC relapse within the first postoperative year and 3 to 6 monthly ctDNA surveillance of patients (for up to 5 years) identified impending disease relapse in 22% of landmark ctDNA negative patients. Establishing a postoperative ctDNA positive status could guide definitive treatment (such as radiotherapy) at equivocal radiological sites. We developed and validated an informatic tool (ECLIPSE) for non-invasive tracking of subclonal architecture at ctDNA levels less than 1%. ECLIPSE identified patients with polyclonal metastatic dissemination, which was associated with poor clinical outcomes. Subclones seeding future metastases post-operatively could be differentiated from non-metastatic subclones through ECLIPSE estimates of cancer cell fraction in preoperative plasma. Phylogenetic ctDNA tracking will support (neo)adjuvant trial advances and ECLIPSE establishes a foundation for non-invasive biomarker strategies focused on clonal deconvolution of relapsed NSCLC using liquid biopsy.

Introduction| ctDNA is a multi-faceted biomarker, pre-surgical ctDNA levels reflect relapse risk in early-stage (stage I-III) NSCLC[1–3] and postoperative ctDNA detection highlights impending NSCLC recurrence[2–6]. TRACERx (NCT01888601) is tasked with assessing clonal evolution in early-stage NSCLC and represents a centrally-monitored, prospective study with longitudinal follow-up of patients within the UK National Health Service, treated according to common health technology appraisal approved care-pathways (NICE)[7,8]. Patients are followed for 5 years post primary NSCLC with primary tumours subject to multi-region sequencing to capture intratumoral heterogeneity. All patients undergo preoperative plasma sampling, then in the postoperative setting plasma sample collection is scheduled at 3-monthly intervals in the first 2-years then at 6-monthly intervals between year 3 and 5, with additional samples requested at disease recurrence. In a prior analysis of 96 TRACERx patients we demonstrated that postoperative ctDNA detection achieved through tracking a maximum of 30 mutations identified recurrent NSCLC in 13 / 14 patients who relapsed and provided insight into the clonal structure of relapsing disease[5]. Here, we report analyses of 1069 plasma samples from 197 TRACERx patients. All 197 patients had preoperative plasma analyses performed, 141/197 patients (including 75 with NSCLC recurrence) had longitudinal postoperative plasma analyses performed (median of 6 postoperative time points). We developed new phylogenetic tracking technologies including patient specific anchored-multiplex PCR (AMP)[9] cell-free DNA (cfDNA) enrichment chemistry tracking a median of 200 tumour-specific clonal and subclonal mutations combined with a novel informatic tool (ECLIPSE) to extract clonal composition in the context of the low ctDNA levels (<1%) encountered in patients with recurrent NSCLC[10]. Median follow-up of event-free patients in this cohort was 4.6 years, encompassing the postoperative period where the majority of NSCLC disease-free survival events (DFS) are expected[11]. We address histology-specific variation in prognostic implications of pre-operative ctDNA detection in NSCLC, postoperative ctDNA detection as an indicator of impending relapse and the capability of phylogenetic tracking to define polyclonal metastatic dissemination and identify metastatic competent subclones prior to surgery.

ctDNA detection using AMP-PCR| AMP patient specific cfDNA enrichment panels (PSPs) tracked a median of 200 mutations pre-identified in multi-region primary tumour exome data (range 72 to 201). A median of 126 clonal mutations were tracked, enabling sensitive identification of ctDNA; a median of 64 subclonal mutations (representing a median of 88% of primary tumour subclones) were tracked to infer subclone evolution at relapse including clonal sweeps, identify high-risk subclones in the preoperative setting and determine which primary tumour subclones seed recurrent disease (see Figure 1a, Extended figure 1a-b, Supplementary Table 1). Median cfDNA input into the AMP-PCR assay across 1069 analysed samples was 23ng (interquartile range [IQR] 15 to 33ng, Extended figure 1c, Supplementary Table 2). AMP incorporates Unique Molecular Identifier (UMI) indices to facilitate in-silico generation of single-stranded consensus reads (see methods), only single stranded consensus reads were analysed to reduce background sequencing error. Distribution of UMI-corrected sequencing depth and association of this parameter with cfDNA input are summarised in Extended figure 1d-e. A molecular residual disease (MRD) detection algorithm evaluated background (non-variant) sequencing positions to estimate intra-library trinucleotide context dependent error rates, to enable ctDNA detection (methods, see Figure 1a and Extended figure 1f-g). An MRD algorithm P-value threshold of 0.01 was determined optimal in terms of overall accuracy through analyses of a 10 patient TRACERx pilot cohort (described in Supplementary Note, Extended figure 2a-d). Analytical validation of variant DNA detection in 707 contrived spike-in samples at assay DNA inputs of 2ng to 80ng was performed using 50-variant PSPs (Extended figure 2e-h, Supplementary Table 3). Sensitivity for variant DNA detection at 0.01% allele frequency (AF, consistent with ctDNA levels observed in MRD positive samples[4,10]) was >90% at DNA inputs of 20ng and above, below 20ng input sensitivity for 0.01% variant DNA declined (58% at 10ng input and 17% at 5ng input). Analytical specificity of the locked MRD algorithm was 100% in analyses of 48 normal samples. Orthogonal validation of preoperative ctDNA positive calls was performed using digital droplet PCR (Supplementary Note and Extended figure 2i-j, Supplementary Table 4).

Clinical predictors of preoperative ctDNA detection| Preoperative cfDNA was analysed across 197 TRACERx patients including the 10 TRACERx pilot-patients used to set MRD algorithm detection thresholds (Supplementary Table 5-6 summarise clinical demographics, Extended figure 3a details cohort design). 189 patients had a single primary NSCLC (Figure 1b) and 8 patients had synchronous primary NSCLCs at diagnosis (Extended figure 3b). In agreement with prior findings[1,5], we observed higher rates of ctDNA detection in squamous and other NSCLC histologies compared with lung adenocarcinoma (41/99 solitary lung adenocarcinomas exhibited ctDNA detection versus 83/90 non-adenocarcinomas, Figure 1b). Within 8/197 patients with synchronous primary NSCLC at diagnosis, 5/8 had preoperative ctDNA detection. In 4/5 cases variants from both primaries were tracked yet only ctDNA from one tumour was detected, in 3 of these cases the largest primary tumour (CRUK0622) or the tumour with squamous histology in cases of synchronous adenocarcinoma and squamous cell carcinoma (CRUK0223, CRUK0555) led to ctDNA detection (Extended figure 3b). Across the complete cohort, preoperative ctDNA detection was associated with an increased number of pack-years smoked (P=0.012, median pack-year history in ctDNA negative patients 34.8 versus 44.2 in ctDNA positive patients, Wilcoxon-test, Extended figure 3c, Figure 1b). Preoperative ctDNA detection associated with clinically occult mediastinal (N2) nodal disease in adenocarcinoma (87 adenocarcinomas were clinical N0/1 stage, 86/87 underwent pathological nodal staging [no N2 nodes sampled at surgery - CRUK0406] and 14/86 were upstaged to pN2 disease; 11/14 (79%) pN2 upstaged patients were ctDNA positive versus 21 of 72 (29%) not upstaged to pN2 disease (P=0.001, Chi-square test, Figure 1b). Therefore, preoperative ctDNA detection could guide extensive mediastinal surgical resection or administration of postoperative radiotherapy.

Preoperative ctDNA detection and postoperative outcomes| Previous findings demonstrate association between preoperative ctDNA detection and postoperative outcome in NSCLC[1]. Given variation in ctDNA detection across NSCLC histology we assessed preoperative ctDNA status (negative, low or high; low equates to clonal ctDNA levels less than the cohort median) as a risk biomarker separately in single (non-synchronous) adenocarcinomas (n=94) and single non-

adenocarcinomas (n=86) evaluable for survival analyses (see methods). In patients with adenocarcinoma, ctDNA detection category was associated with OS (log-rank P <0.001, Figure 1c). Absence of preoperative ctDNA detection identified adenocarcinoma patients with excellent postoperative outcome (91% 2-year OS [95% CI:84 to 99%], n=56) compared with ctDNA low (60% 2-year OS [95% CI: 44 to 83%], n=26) or high adenocarcinoma (31% 2-year OS [95% CI: 13 to 75%], n=12, Figure 1c). There was no association of preoperative ctDNA status and OS in patients with non-adenocarcinoma histologies (log-rank p=0.31, Figure 1c, even when the 7 ctDNA negative patients were excluded from analyses [n=32 ctDNA low, versus n=47 ctDNA high, log-rank P = 0.2]). Similar findings were observed in freedom from recurrence (FFR) analyses (Extended figure 3d). In multivariable survival analysis including pTNM stage, adjuvant therapy status, age and unique sequencing coverage, preoperative ctDNA status associated with FFR and OS in adenocarcinoma, but not in non-adenocarcinoma (Extended figure 3e). In patients with adenocarcinoma, 19/22 (86%) extrathoracic recurrences occurred in preoperative ctDNA positive patients compared with only 8/19 (42%) intrathoracic-only recurrences (Chi-squared test, P=0.008, Extended figure 3f). Recurrences in preoperative ctDNA positive patients with adenocarcinoma associated with poor post-relapse survival compared with those occurring in ctDNA negative patients with adenocarcinoma (log-rank P = 0.004, Extended figure 3e). These data highlight preoperative ctDNA detection in adenocarcinoma as a predictor of future extrathoracic metastatic dissemination and poor survival outcome.

Genomic predictors of preoperative ctDNA detection in adenocarcinoma| Computed Tomography (CT) volumetric data was available in 159 / 189 patients with single primary NSCLC (volumetric exclusion details in Extended figure 4a, Supplementary Table 6). Relatively high tumour volumes were present in ctDNA negative solitary adenocarcinomas (mean volume 18.8cm^3, n=51, range 1.0 to 259.8cm^3) with 21/51 patients exhibiting tumour volumes above 10cm^3 (Extended figure 4b). Since 10cm^3 tumour volume is associated with clonal ctDNA levels of ~0.1%[1,4,5] (which lies above AMP-PCR assay limit of detection, Extended figure 2e-f), the absence of ctDNA detection in high

volume adenocarcinomas suggested a low-ctDNA shedding phenotype. This prompted investigation of biology related to absence of detectable ctDNA in lung adenocarcinoma.

Using a regression model incorporating tumour volume and NSCLC histology subtype to predict clonal ctDNA levels (developed on the 101 ctDNA positive patients in this cohort, Extended figure 4c), we predicted clonal ctDNA levels in 51 ctDNA negative adenocarcinomas categorising these tumours as either low-shedders (ctDNA detection expected based on tumour volume, but not observed [36 of 51 cases]) or probable technical negatives (low tumour volume predicted for ctDNA levels possibly below assay limit of detection [15/51 cases], Extended figure 4d and methods). The latter group was excluded from analyses of ctDNA detection and tumour biology, since absence of ctDNA detection in these cases may have been secondary to low tumour volume.

We conducted multi-region transcriptomic analyses comparing 34 ctDNA positive adenocarcinoma to 33 low-shedder adenocarcinomas (Figure 2a, Supplementary table 7-8). Genes upregulated in ctDNA positive adenocarcinomas included mitotic spindle genes and others associated with M-phase, cell cycle and DNA repair (Figure 2b-c, Supplementary table 9). Gene Set Variation Analysis (GSVA[12]) using the Hallmark genesets (which summarise 50 biological states[13]) revealed upregulation of proliferation and cell cycle associated gene sets in ctDNA positive adenocarcinomas, compared to low-shedders (Figure 2d). No difference in purity of tumour tissue samples was observed between ctDNA positive and low-shedder adenocarcinomas (Extended figure 4e). We evaluated our published prognostic biomarker associated with outcomes in lung adenocarcinoma (ORACLE[14]) and noted preoperative ctDNA positive adenocarcinomas demonstrated higher ORACLE scores relative to negative adenocarcinomas ($P < 0.0001$, Figure 2e). We explored somatic driver mutations associated with ctDNA detection in lung adenocarcinoma and observed no difference between ctDNA positive adenocarcinoma and low-shedders when we analysed subclonal and clonal somatic driver mutations (Extended figure 4f, showing top 12 most mutated cancer genes), or when summarising to pathway level (Extended figure 4g). To explore if chromosomal instability (CIN) associated with increased

ctDNA shedding, we calculated the weighted genome integrity index[15] (wGII), the fraction of loss of heterozygosity (FLOH)[16], and determined genome-doubled status for each tumour. We observed that ctDNA positive adenocarcinomas showed increased levels of both wGII and FLOH relative to low-shedders (P = 0.0100 & P = 0.0006, Figure 2f-g). We observed an increased percentage of ctDNA positive adenocarcinomas had experienced genome doubling (any GD compared to none, 86% versus 61%, P = 0.016, Extended figure 4h). We investigated the link between CIN and increased cell proliferation using GISTIC2[17] to obtain a positive-selection score (G-score) for each cytoband (see methods). We determined the G-score difference (GSD) between ctDNA positive adenocarcinomas and low shedders. We observed 21 amplified cytobands and four deleted cytobands enriched in ctDNA shedders (FDR q-value < 0.05) with a GSD of at least 0.5 (Figure 2h-i), a previously defined threshold for comparing two sample sets[18]. Within these cytobands a total of 968 genes are located with 26 listed in the COSMIC gene census[19] as cancer genes (Supplementary table 10), including proliferation-associated genes *CCND1* (11q13.3), *CDK4* (12q14.1), *MDM2* (12q15) and *CCNE1* (19q12). No cytobands were enriched in low-shedder adenocarcinomas.

To ensure results were not primarily driven by higher tumor volume in ctDNA positive versus low-shedder cases, we excluded the bottom quartile of tumor volumes from low-shedding adenocarcinomas and largely reproduced the above findings, including 847/935 significantly differentially expressed genes and 17/25 cytobands significantly altered in ctDNA positive adenocarcinomas (Supplementary Note, Extended figure 4i-k).

Postoperative ctDNA detection is specific for relapsing NSCLC| Postoperative cfDNA samples from 45 recurrence-free patients (Figure 3a) and 21 patients who subsequently developed new primary cancers (based on histological or clinical findings) during follow-up were analysed to assess AMP-PCR clinical specificity (PSPs are specific to the excised primary NSCLC and are not expected to detect new primary cancers, Figure 3b, Supplementary Table 11). 471 of 481 (97.9%) postoperative samples from these patients were negative for ctDNA detection, 3 of 66 (5%) patients exhibited

postoperative ctDNA detection (Figure 3a-b); CRUK0086 was ctDNA positive prior to radiation therapy, CRUK0269 was ctDNA positive post-surgery and developed a new primary NSCLC and CRUK0498 had false positive ctDNA detection at 7 of 8 postoperative timepoints due to PSP-mistargeting of variants associated with a lymphoid expansion, carrying clonal haematopoietic (CH) variants (Extended figure 5a). CH-variant mistargeting was identified through application of CRUK0498's PSP to peripheral blood mononuclear cells (PBMCs) isolated alongside cfDNA (Extended figure 5b, Supplementary Table 12). The CH-mutations existed at high AFs in tumour tissue but below the 1% germline exome detection threshold in PBMC DNA, explaining why they were misclassified tumour derived by tissue exome analyses and included in PSP design. For example, mutation *ATP2C1:T>G* (detected in 6 postoperative cfDNA samples) was present at an AF of 16.6% in tumour region 2; yet max AFs in postoperative cfDNA and PBMC DNA were only 0.57% and 0.86% respectively (Extended figure 5b). Histological examination of CRUK0498's tumour revealed lymphoid cells, likely representing a source of these CH mutations (Extended figure 5c). To address whether PSPs frequently mistargeted CH-variants present in tumour tissue, we evaluated PBMC DNA from 21 patients using their PSPs. Application of the MRD detection algorithm to these PBMC samples revealed no variant DNA detection (pan-library MRD detection algorithm P-values above 0.01, Extended figure 5d). At the level of individual mutations, 13 of 3978 (0.3%) PSP-targeted mutations were detected in PBMC DNA, only 2 of these 13 mutations were present in preoperative cfDNA and not filtered by the MRD detection algorithm (Extended figure 5e, Supplementary Table 13).

Postoperative ctDNA detection in recurrent NSCLC| 391 postoperative plasma samples were analysed from 75 patients who suffered either recurrence of their NSCLC (n=71) or incomplete resection (macroscopic residual disease, n=4, Figure 3c-e, Supplementary Table 11). Extended clinical follow-up and protocol-specified plasma collection 3-monthly for the first 2 years followed by 6-monthly intervals between year 3 and 5, enabled evaluation of relapse events occurring up to 1363 days post-surgery. ctDNA was detected postoperatively in 64/75 (85%) relapsing patients, 3/11

patients relapsing without postoperative ctDNA detection lacked plasma sampling within 100 days of clinical relapse (CRUK0303, 0495, 0570). In patients with plasma sampled close to recurrence, 2/11 had unresected nodal disease (CRUK0230, 234), 4/11 had intracranial recurrences (CRUK0331, 0407, 0567, 0736) and 2 patients had intrathoracic recurrences (CRUK0329, 0490, Figure 3c-e, Supplementary Table 14). Intracranial recurrence has previously been associated with lack of postoperative ctDNA detection[20], in this cohort 17 patients experienced brain metastases within 180 days of relapse; 14/17 patients had extracranial imaging at relapse allowing categorisation of brain recurrence as isolated (intracranial disease only) or non-isolated (involvement of extra- and intracranial sites). Postoperative ctDNA was detected at recurrence in 3/7 patients with isolated intracranial recurrence versus 7/7 patients with non-isolated intracranial recurrence (Figure 3c-e, Extended figure 6a).

Landmark MRD analysis| We explored postoperative ctDNA detection within a landmark analysis framework[4,21]. 118/141 patients with postoperative plasma sampling performed were evaluable for MRD landmark analysis based on ≥ 1 plasma sample obtained within 120 days of surgery, prior to adjuvant therapy or relapse (if relevant); the remaining 23/141 patients were landmark unevaluable (Figure 3a-e). Within this landmark window, 27/118 patients (23%) exhibited 1 (or more) positive ctDNA calls, Figure 3d). 25/27 landmark positive patients experienced either NSCLC recurrence (n=23) or were diagnosed with incompletely resected disease (n=2; positive predictive value of landmark for relapse 93%, sensitivity of landmark for relapse 45%, Figure 3c). Landmark positive status associated with higher pathological TNM stage (5/45 stage I, 8/39 stage II and 14/34 stage III patients landmark positive, chi-squared test P=0.006) and earlier NSCLC recurrence events (15/23 [65%] recurrence events occurring within 1 year of surgery landmark positive versus 8/29 [26%] events occurring later than 1-year post-surgery, 4 patients with incompletely resected disease excluded, chi-squared test P=0.02, Figure 3c). Median clonal ctDNA level at the initial point of ctDNA detection in landmark positive patients who recurred or had incompletely resected disease was 0.08% (range 0.002% to 2.41%, n=25, Extended figure 6b). In 112/118 patients evaluable for survival

311 analyses (see methods), landmark positive patients exhibited a HR of 4.8 (95% CI 2.7 to 8.4, P<0.001
312 log-rank test) for OS and an HR of 6.1 for FFR (95% CI 3.5 to 10.7, P<0.001 log-rank test) compared to
313 landmark negative patients (Extended figure 6c-d). 31/91 landmark negative patients experienced
314 either NSCLC recurrence (n=29) or were diagnosed with incompletely resected disease (n=2) and
315 20/31 of these patients emerged to be ctDNA positive during surveillance prior to, or at, clinical
316 relapse after a median of 2.5 negative plasma samples (range 1 to 9); a median of 337 days
317 postoperatively (range 120 to 929 days, Figure 3d) at a median clonal ctDNA level of 0.02% (range
318 0.003% to 6.67%. Extended figure 6b). Within the remaining 11/31 landmark negative recurrent
319 patients, 2/11 emerged to be ctDNA positive post-clinical relapse (CRUK0146, 0296) and 9/11
320 exhibited no postoperative ctDNA detection (Figure 3d).

321 **ctDNA lead-times|** To assess ctDNA lead-times (time from ctDNA detection to recurrence) we
322 assigned patients without postoperative ctDNA detection or with initial detection following clinical
323 relapse, lead-times of 0 days and excluded incompletely resected patients (n=4) and patients with no
324 ctDNA sampling before clinical recurrence (n=3, CRUK0516, 0557 and 0640). Landmark positive
325 patients had longer lead-times (median 228 days [0 to 1137 days], n=23) than landmark negative
326 patients (median 77 days, [0 to 980 days], n=29, P=0.004, Wilcoxon-test) and landmark unevaluable
327 patients (median 56 days, [0 to 477 days], n=16, P=0.005, Wilcoxon-test, Extended figure 6e).

328 **ctDNA kinetics during adjuvant therapy|** 12 patients were MRD positive before adjuvant therapy
329 (treatment data, Supplementary Table 15). CRUK0086 was MRD positive at day 2 post-surgery,
330 following adjuvant radiotherapy the patient was ctDNA negative at four further plasma sampling
331 timepoints (Days 130, 193, 284 and 375 postoperatively) remaining disease-free until non-cancer
332 associated death 869 days after surgery (Figure 3a, Extended figure 6f). The remaining 11 pre-
333 adjuvant MRD positive patients relapsed with a median lead-time of 302 days (78 to 1022 days);
334 despite 5/11 patients clearing ctDNA following completion of adjuvant therapy (CRUK0022, 0242,
335 0245, 0385, 0702). This suggests that ctDNA clearance with adjuvant therapy does not prevent

impending disease relapse. 4/5 of these patients re-emerged as ctDNA positive before disease recurrence whereas CRUK0245 cleared MRD and had no further plasma sampling performed (Figure 3, Extended figure 6f).

Imaging and MRD detection| We assessed utility of postoperative ctDNA detection in the context of standard of care extracranial imaging surveillance in the adjuvant setting, 386 extracranial surveillance clinical imaging-reports from 131/141 patients who underwent postoperative cfDNA analyses were reviewed (343 CT scans, 7 Magnetic Resonance Imaging scans and 36 whole-body Positron Emission Tomography scans, Figure 3a-e, Supplementary Table 16); 10/141 patients had no surveillance imaging reports available. In patients who eventually experienced NSCLC relapse we identified 44 surveillance scans from 23 patients that showed no new abnormalities compared to prior imaging (Figure 3c-e). 22/23 patients had plasma sampling performed prior to the scan(s) showing no new abnormalities, 9 of these patients were ctDNA positive before the scan and 8 / 9 patients suffered eventual recurrence at sites covered by the extracranial scans (CRUK0590 experienced intracranial recurrence, Figure 3c-e, Extended figure 6g). This suggests postoperative ctDNA positive status can precede new abnormalities on imaging in some cases. New equivocal imaging abnormalities were common, observed on 152 occasions in 90 patients (Figure 3, Extended figure 6h). Postoperative ctDNA detection before these abnormalities occurred in 23 patients, 20/23 suffered subsequent NSCLC recurrence (Extended figure 6f). Prior to surveillance scan(s) showing new equivocal lymphadenopathy, 14 patients were ctDNA positive and 21 patients were ctDNA negative. 11 / 14 (79%) ctDNA positive patients eventually relapsed with lymph node involvement at the equivocal site versus 7 / 21 (33%) patients ctDNA negative before the scan (CRUK0234 was diagnosed with an incompletely resected lymph node was ctDNA negative and included in the latter category; Fisher's test $P = 0.015$, Extended figure 6i). Therefore, ctDNA could guide early definitive therapeutic intervention such as radiotherapy or ablation at anatomical sites deemed equivocal by imaging.

ctDNA-based characterisation of subclonal architecture using ECLIPSE | To estimate tumour subclonal composition from deep targeted sequencing of plasma cfDNA we developed ECLIPSE which leverages and background noise estimates and tumour tissue derived copy number information to assess the presence or absence of specific tumour subclones and calculate their respective cancer cell fractions (CCFs) from low tumour content cfDNA data (Figure 1a, Extended figure 7, Supplementary Methods, Supplementary Note). Plasma samples with clonal ctDNA levels greater than 0.1% (64% of ctDNA positive samples in the cohort) had a minimally detectable CCF of 20% for a representative subclone (Supplementary Methods, Extended figure 8a-d). To validate the detection rate for small subclones at this threshold we leveraged mutant DNA spike-in experiments constructing 76,263 subclones *in-silico* from this data with ground truth CCFs, clonal ctDNA levels, numbers of tracked mutations and DNA assay inputs. We estimated a detected rate of 94% for 20% subclones in 0.1% clonal ctDNA level plasma with 4 tracked mutations and 10ng DNA input (Extended Figure 8e). We observed a significant drop off in detection rates when DNA input decreased below 10ng. Therefore, we considered 'high subclone sensitivity' samples as those with >0.1% clonal ctDNA level and >=10 ng cfDNA input, which were used to analyse clonal composition with ECLIPSE in further analyses.

Comparison of time-matched plasma and tissue-based estimates of subclonal architecture | ECLIPSE measures of subclonal CCF from preoperative plasma samples were proportional to tumour exome multi-region sequencing measures of subclonal CCF sampled at surgery (Pearson R = 0.78, m = 1, median clonal ctDNA level analysed = 0.9%, Extended Figure 9a-b, Supplementary Note), indicating that ECLIPSE could estimate CCF values from low ctDNA level plasma samples. The detection rate in pre-operative plasma was higher for subclones with a larger CCF and there was a correlation between the number of subclones detected and the clonal ctDNA level (Spearman's Rho 0.72, $P<0.001$, Extended Figure 9c). However, the detection rate of small subclones was lower than expected given our estimates of subclone sensitivity using spike-in experiments (39% sensitivity in preoperative plasma for 20% CCF subclones at 0.1% clonal ctDNA level vs 94% using spike-in

387 experiments for equivalent subclones). In addition, we observed that subclones discovered in a
388 single tumour region had a consistently lower CCFs in plasma compared to tissue, unlike subclones
389 which were present across multiple regions (Extended Figure 9d-e) and that detection rates
390 increased for small subclones spread across multiple tumour regions compared to those confined to
391 a single tumour region (Extended Figure 9f). These data are consistent with oversampling of small,
392 spatially constrained subclones discovered in tissue samples when compared to the whole tumour
393 mass due to detection bias, as previously described[22] (Extended Figure 9g, Supplementary Note)[22]

394 **Refining estimates of tumour clonality from a single biopsy using ctDNA** | In the TRACERx 421
395 cohort[23] a median of 12% of variants were determined to be present in all cancer cells of at least one
396 tumour region but were absent from other regions of the tumour, therefore exhibiting a clonal
397 illusion (Figure 4a). ctDNA may be released from several regions of the tumour and hence resolve
398 the true subclonal nature of variants displaying a clonal illusion tracked from a single standard of
399 care biopsy. In 71 patients with high subclone sensitivity samples available preoperatively, plasma-
400 based CCFs of clonal illusion variants were lower compared to variants ubiquitous across all tumour
401 regions (Wilcoxon test, $P<0.001$, Figure 4a) and plasma CCFs could predict clonal illusion with an AUC
402 of 0.81 (95% CIs: 0.79-0.82, Extended Figure 9h). These data show that pre-operative ctDNA is
403 derived from many different primary tumour regions and suggests that collection of plasma
404 alongside a single tumour biopsy can overcome sampling bias, potentially increasing accuracy of
405 future ITH-based clinical biomarkers such as clonal tumour mutation burden (a potential biomarker
406 of immune checkpoint inhibitor efficacy)[24,25].

407 **Preoperative subclones detected at high CCF forecast metastatic potential|** Predicting the
408 subclonal nature of the subsequent metastatic recurrence at the time of surgery could inform
409 precision adjuvant therapies against subclone(s) driving subsequent disease relapse. In 26 relapsing
410 NSCLC patients with high subclone sensitivity samples available both pre- and postoperatively,
411 primary tumour subclones detected in postoperative cfDNA displayed significantly larger CCFs in

plasma samples taken prior to surgery when compared to subclones that were not detectable postoperatively (Wilcoxon test, $P<0.0001$, Figure 4b). These metastatic subclones tended to expand further at relapse (Wilcoxon test $P=0.027$). This result indicates that primary tumour subclonal expansion measured non-invasively using ctDNA is associated with metastatic potential, complementing results in our companion manuscript[26] which show metastatic subclones are larger and more often spread across primary tumour regions compared to subclones which do not metastasise**.**

Identification of complete clonal sweeps at recurrence | In 18/42 (43%) patients with a high subclone sensitivity postoperative plasma sample available, we estimated that subclones tracked from the surgically resected tumour had undergone a complete clonal sweep at recurrence, where a subpopulation of cells expands to become clonal across all tumour sites (see methods). This phenomenon led to increased clonal tumour mutation burden (median 14% increase; Extended figure 10 a-b) and increased clonal neoantigen burden at recurrence (median of 11% increase; Extended figure 10c) which we have previously linked to the efficacy of immunotherapy[24]. Only 3/18 patients contained a known subclonal driver involved in the clonal sweep including mutations in *SMARCA4* (p.D881A & p.G1194E), KMT2D (p.G4182), *CTNNB1* (p.S37C) and an amplification in *MDM4*.

Non-invasive characterisation of metastatic seeding patterns | Comprehensive tissue sampling is challenging in the early-relapse setting, particularly when more than one metastatic site is involved. 44% of relapse patients had a tissue sample obtained at relapse in our cohort however we obtained high subclone sensitivity samples at relapse in 61% of relapse patients with a mean of 2.0 such samples per patient. 38% of patients had high subclone sensitivity samples at relapse but lacked a relapse tissue sample (Extended Figure 10d). In 26 patients with both high subclone sensitivity postoperative plasma and recurrence tissue sampled, we found high concordance between subclones detected in recurrence tissue and postoperative cfDNA samples (98% sensitivity for

relapse tissue subclones (50/51) which were tracked in primary tumour informed cfDNA panels, Extended figure 10e-f). Additional subclones detected in ctDNA, but absent from relapse tissue were found in 6/26 patients (20 subclones), which may have evaded tumour biopsy detection due to under-sampling of metastatic sites at relapse (Supplementary Note). This is consistent with our companion manuscript[26] which suggests a single metastatic biopsy is not sufficient to confidently capture the metastatic seeding events. While we have previously published methods for detecting subclones present in cfDNA at relapse[5], calculation of subclone CCFs along with increased fraction of tracked subclones and increased subclone detection sensitivity allowed for estimation of seeding patterns from the primary tumour to relapse using ctDNA (Supplementary Note 1). Tumours were categorised by the number of relapse-seeding clones (monoclonal = 1, polyclonal = >1) and relapse seeding phylogenetic tree branches (monophyletic = 1, polyphyletic = >1, see Figure 1a, methods). Longitudinal plasma- and tissue-based clonal composition estimates from surgery to relapse are presented for 44 patients with high subclone sensitivity plasma at relapse (methods) alongside clonal composition measurements from matched tissue samples (Figure 5a, Extended figure 1). We found an increased frequency of polyclonal relapse seeding when using cfDNA compared to metastatic biopsy whole exome sequencing data, driven by detection of cfDNA-unique subclones (10% polyclonal using tissue versus 24% polyclonal using cfDNA, in the 22 high subclone sensitivity plasma and relapse tissue matched patients where phylogenetic trees could be built, Extended figure 10g). Overall, 31/44 recurrent tumours were defined as monoclonal and 13/44 as polyclonal (4 polyclonal monophyletic and 9 polyclonal polyphyletic). Shorter OS from study registration and from the first MRD positive timepoint was observed in patients exhibiting polyclonal dissemination versus monoclonal dissemination (Figure 5B, post-registration OS: HR=4.07, 95% CIs=1.82 to 9.07, Extended figure 10h: post-MRD OS, HR 4.09 95% CIs = 1.88 to 8.89; univariable cox regression, N=44). OS from registration remained significant after adjustment for maximum postoperative clonal ctDNA level,pathological stage at surgery, histology and pre-operative ctDNA detection in a multivariable cox regression model (HR=3.25, 95% CIs=1.21-8.74, Extended Figure 10i).

Longitudinal tracking of clonal evolution | We addressed whether high resolution tumour informed phylogenetic tracking could detect changes in subclonal composition through adjuvant and post-recurrence treatment, which may represent therapy-induced shifts in selection pressure. We observed several shifts in clonal composition in CRUK0484 concurrent with treatment (Figure 5C, Supplementary Note) including extinguishing of a subclone present in more than half of tumour cells after surgery (represented in red) during adjuvant chemotherapy, and expansion of a minor subclonal lineage (represented in light green) during post-recurrence immunotherapy treatment which eventually outcompeted a parallel lineage represented in dark green). It's also notable that, despite three relapse tissue biopsies at different timepoints and metastatic sites, the dark green subclone, dominant in ctDNA after adjuvant therapy, was not detected in post-surgical tissue samples but only in a surgically excised lymph node. In CRUK0050 we observed a rapid increase in clonal ctDNA levels at D876, following treatment of recurrent lung disease with cytotoxic chemotherapy (docetaxel and nintedanib, Figure 4d). A multi-modal distribution of clonal variant allele frequencies (VAFs) was observed in plasma, which suggested 59/130 clonal mutations had altered their copy number state compared to samples taken at surgery (methods) including evidence for amplification of an oncogenic *KRAS* G12R mutation (84% VAF). This suggests the expansion of a new subclone during treatment, not directly tracked by our panel, which harboured significant chromosomal instability.

In summary, we found that chromosomal instability, copy-number aberration and upregulation of proliferation related transcriptional pathways associate with preoperative ctDNA detection in early-stage lung adenocarcinoma, providing a molecular rationale for our observations regarding ctDNA detection and poor clinical outcomes in this histological subtype. We demonstrated that preoperative ctDNA detection in adenocarcinoma associates with clinically occult mediastinal disease, with implications for planning extent of mediastinal surgical resection and administration of postoperative radiotherapy. These findings suggest that management of early-stage

adenocarcinomas deemed high-risk based on ctDNA detection is inadequate, with innovation urgently needed. Measurements of clonal structure in low tumour content ctDNA using ECLIPSE overcame tissue sampling bias to refine estimates of clonal structure, identified subclones with metastatic potential prior to surgery and, in combination with multiregion primary tumour sequencing, characterised polyclonal relapse seeding patterns which were associated with poor clinical outcomes.

ctDNA detection in the post-curative intent therapy setting forecasted impending NSCLC relapse, agreeing with prior findings[3–6,27–29]. Landmark MRD analysis was performed (analyses of plasma samples donated within 120 days of surgery)[4,21]. Landmark samples were ctDNA positive in 45% of patients experiencing NSCLC relapse, these patients were more likely to be higher stage at diagnosis and experience early-NSCLC relapse within the first postoperative year. Through 3 to 6 monthly plasma sampling for up to 5 years, we demonstrate the importance of longitudinal ctDNA surveillance given the identification of impending disease relapse in 22% of landmark negative patients, following up to 9 ctDNA negative plasma samples. Emergence of ctDNA during surveillance may reflect low-burden metastatic disease initially shedding ctDNA quantities below the AMP-PCR assay limit of detection (~95% sensitivity at 0.008% ctDNA level in high-cfDNA input [>30ng] samples). Landmark ctDNA detection rates will likely increase with next-generation MRD assays demonstrating ctDNA limits of detection approaching 1 part per million[30,31,32,33]. The degree of clinical impact conferred by improved assay sensitivity will require assessments of these technologies in large patient cohorts. Through annotation of postoperative ctDNA data with 386 radiological extracranial surveillance scans, we frequently observe ctDNA detection prior to scans showing indeterminate changes eventually manifesting as disease recurrence. This suggests that a postoperative ctDNA positive status may increase physician confidence to intervene definitively (with radiotherapy or ablation) at equivocal radiological sites.

Prior publications have used high tumour purity ctDNA samples (>10%) to profile tumour clonality[34,35,36]. However such samples are rare[10], comprising only 9% of ctDNA positive samples from 14/145 (10%) ctDNA detected patients in this study. ECLIPSE, combined with AMP-PCR, enabled an estimated 94% detection sensitivity for 20% CCF subclones in plasma samples with only 0.1% tumour content, providing inferences of clonal structure in the majority of ctDNA positive plasma samples in this study (64% of ctDNA positive samples were ECLIPSE evaluable). Our tumour-informed approach, however, was not able to extract clonality of novel variants in plasma and was unable to accurately assess clonality in plasma samples with <0.1% tumour content, suggesting deeper sequencing and increased variant tracking may further improve the applicability of ECLIPSE. We demonstrate at scale that ctDNA can sample clonal structure from multiple different surgically excised tissue sites and capture additional heterogeneity at relapse when compared to analysis of relapse tissue samples. Despite this, two thirds of patients who suffered disease recurrence still harboured only one ctDNA-detectable metastasising primary tumour subclone (monoclonal metastatic dissemination). Low ctDNA levels and incomplete primary tumour sampling may limit detection of additional dissemination events however. We observed a more aggressive disease course in patients with multiple metastatic dissemination events (polyclonal metastatic dissemination) suggesting that heterogeneity in the seeding population may provide an increased ability for Darwinian adaptation to different metastatic niches, driving more rapid disease spread. However a requirement for multiregional primary tumour sequencing currently limits the feasibility for measurements of metastatic dissemination patterns using our approach in a real-world clinical setting.

MRD assessment through ctDNA is poised to change (neo)adjuvant trial design, and intervention with curative intent for radiological equivocal lesions. Measurements of subclonal expansion in plasma before surgery may allow prediction of future metastatic subclones, offering the possibility for early intervention and suggesting new routes for biomarker development to target and eradicate such clones months or even years prior to relapse.

538

539

540 **Main Figure legends**

541 **Figure 1| A.** ctDNA Analysis Approach: Multi-region tumour exome data is used to generate a
542 phylogenetic representation of a patient's primary tumour where clonal mutations (mutations
543 present in all cancer cells) are represented by the red cluster and subclonal mutations (mutations
544 present in only a subset of cells) are represented by the light blue, dark blue, green and yellow
545 clusters. Bi-directional primers are synthesised to amplify genomic regions containing mutations
546 identified through the multi-region tumour sequencing data with the aim to target mutations from
547 all clusters. Unique molecular identifiers (UMIs) are incorporated into amplified DNA. Only deep
548 consensus reads (reads supported by 5 or more duplicate reads based on UMI-deconvolution) are
549 used to detect ctDNA. The position of interest where a mutation is expected (shown here as a G>T
550 mutation) is then evaluated. The position is considered for ctDNA calling if there is no evidence of
551 strand bias (that is variant bases are noted on both positive and reverse reads), no evidence of
552 sequencing bias (variant bases are noted across both read 1 and read 2 during next-generation
553 sequencing) and both forward and reverse primers are functioning. Alongside evaluating the
554 position of interest the MRD caller evaluates background sequencing positions from the targeted
555 variants and background sequencing positions from primers targeting 45 germline SNP primers (e.g.,
556 the TCG region highlighted in the figure). Any base errors observed in background reads are summed
557 based on their trinucleotide context. Trinucleotide contexts with an upper 99% binomial confidence
558 interval less than 0.01% (error rate confidently below 0.01%) are used for ctDNA calling. Passed
559 variants are subsequently evaluated by an outlier filter, if there are 3 or less outlier variants these
560 are removed from the ctDNA calling algorithm. For calling the presence or absence of ctDNA (panel-
561 wide MRD caller) all positions targeted by a panel are analysed and a one-sided Poisson test is
562 performed to determine whether the observed number of error-corrected alternate reads deviates

from the background error encountered by a panel. For individual subclones a subclonal caller evaluates the presence or absence of ctDNA on a per-subclonal cluster basis, this information is integrated with primary-tumour informed cancer cell fraction (CCF) estimates made through ctDNA analysis at relapse using the ECLIPSE tool (see methods) to categorise the lung cancer relapse process as either monoclonal (a single clone seeding the metastasis), polyclonal monophyletic (multiple clones seeding metastasis from a single branch of the primary tumour phylogenetic tree) or polyclonal polyphyletic (multiple clones seeding metastasis from multiple primary tumour branches of the phylogenetic tree). ECLIPSE is also able to detect the presence of a clonal sweep whereby a subclone achieves 100% CCF at recurrence. **B.** Heatmap summarising preoperative ctDNA analyses in 189 TRACERx patients with a single (non-synchronous) primary tumour diagnosed at enrollment. Cases are split by three histologies, adenocarcinoma, squamous cell carcinoma and other histological subtypes. The top row of the heatmap reflects the smoking pack-year history assigned to that patient. The N2 upstaging row demonstrates patients who were clinically staged with N0 (no lymph node involvement) or N1 (hilar) lymph-node involvement but after surgery were upstaged to N2 disease (mediastinal lymph node involvement, coloured dark blue), patients who did not undergo pathological mediastinal lymph-node examination are coloured grey. The pTNM row describes TNM version 7 tumour stage based on pathology examination of the excised tumour. The tumour volume row represents computed tomography measurements of tumour volume (grey fill indicates tumour volume unevaluable cases). The ctDNA detected row represents whether an MRD positive call was triggered based on the MRD Algorithm P-value cut-off of 0.01 defined in Extended figure 2. The ubiquitous/clonal and heterogenous/subclonal variant bar charts demonstrate the number of mutations in each patient's panel being interrogated, if black the mutation was undetected (based on a per variant P value >0.01), if red the mutation was filtered by the MRD caller and if blue the mutation was detected. The subclone detection (CCF >0.1 in tissue) barchart represents the proportion of subclones with cancer cell fractions more than 10% tracked by ECLIPSE that are detected in preoperative plasma. The clonal ctDNA level per patient is shown in the bottom

589 row of the heatmap. Patients with 0% clonal ctDNA level are given a white color in this row. **C.**
590 Kaplan Meier curves demonstrating overall survival outcomes in ctDNA high (dark red), ctDNA low
591 (blue) and ctDNA negative (grey) single primary adenocarcinoma patients (left) and single primary
592 non-adenocarcinoma patients (right). ctDNA high and low was categorised based on median clonal
593 ctDNA levels across ctDNA positive cases and relates to above and below 0.16%. Log-rank P values
594 are displayed on each plot alongside hazard ratios (HRs) that use ctDNA high patients as the
595 reference category.

596 **Figure 2| Genomic and transcriptomic predictors of ctDNA detection in early-stage NSCLC. A.**
597 Differential gene expression analysis comparing 106 tumour regions from 35 ctDNA positive
598 adenocarcinomas to 74 tumour regions from 32 ctDNA negative low-shedder adenocarcinomas X-
599 axis shows fold change as difference in means on log2 scale (LogFC). Red indicates significantly over-
600 expressed in ctDNA positive adenocarcinomas (n = 932 genes), blue indicates significantly over-
601 expressed in ctDNA negative low-shedder adenocarcinomas with technical non-shedders excluded (n
602 = 939 genes). Top 15 differentially expressed genes are labelled. **B** and **C.** Reactome pathway
603 enrichment analysis based on the 1,871 significant genes found in A) (n = 932 for ctDNA positives
604 and n = 939 for ctDNA negatives). X-axis shows the proportion of genes involved in the pathway, Y-
605 axis notes the significantly enriched pathways. B) shows the top 15 enriched pathways in ctDNA
606 positive patients, C) displays the only significantly enriched pathway in ctDNA negative patients. Size
607 of points represents the number of genes included and the colour shows the FDR adjusted p-value.
608 **D.** Differential enrichment analysis based on the Hallmark gene-sets. Samples, axes and colours
609 follow A). All significantly enriched pathways are labelled. **E**. ORACLE gene expression scores in
610 ctDNA positive (n = 115 tumour and lymph node regions sampled from 35 patients) versus ctDNA
611 negative (n = 101 tumour and lymph node regions sampled from 46 patients) adenocarcinomas. All
612 data points are displayed. Centre lines represent the medians. **F, G.** Violin-boxplots showing the
613 wGII (F) and FLOH (G) levels of ctDNA positive adenocarcinomas (37 patients, 177 tumour and lymph
614 node regions), ctDNA negative low-shedder adenocarcinomas (33 patients, 97 tumour regions) and

squamous or other carcinomas (79 patients, 322 tumour and lymph node regions). All data points are displayed. Boxplot centre lines represent the median, p-values are FDR-adjusted. **H, I**. GISTIC score analysis. Red indicates amplifications, blue indicates deletions, grey indicates non-significant values. Y-axis shows the adjusted p-value and X-axis shows the GISTIC score difference. All significant alterations are labelled. Alterations are considered significant when the GISTIC score difference exceeds 0.5 and the adjusted p-value is below 0.05, cutoffs are denoted with dotted lines.

Figure 3| Postoperative Minimal Residual Disease detection in early-stage NSCLC. A-D. Longitudinal data from patients with (**A**) no evidence of non-small-cell lung cancer (NSCLC) recurrence, n=45 and 371 pre- and postoperative plasma samples; (**B**) development of a second-primary cancer, n=21 and 176 pre- and postoperative plasma samples; (**C**) recurrence of NSCLC in landmark positive patients, n=25 patients and 138 pre- and postoperative plasma samples (**D**) recurrence of NSCLC in landmark negative patients, n=31 patients and 227 pre- and postoperative plasma samples and (**E**) recurrence of NSCLC in landmark unevaluable patients, n=19 patients and 101 pre- and postoperative plasma samples. and In all plots each circle represents a cfDNA sampling time point. The circles to the left of day-0 are preoperative timepoints from when the patient's tumour was still in-situ. The circles to the right of day-0 are taken following surgical excision of the primary NSCLC. If the circle is colored dark it reflects positive ctDNA detection. The light blue rectangles represent whether a patient received chemotherapy, the dark blue rectangles represent whether a patient received radiotherapy and the orange rectangles if a patient received post-recurrence surgery. The triangles represent standard of care postoperative CT, PET or MRI imaging classified as no disease (grey) equivocal images (yellow) or unequivocal imaging evidence of extracranial relapse (red). Light green triangles represent no evidence of intracranial relapse, dark green triangles indicate intracranial relapse. The vertical black lines represent the event date for a patient (if events such as death, second-primary, NSCLC recurrence occurred) otherwise the vertical line represents the TRACERx follow-up censorship date for that patient. Crosses represent patient death events. To the left of the panels the annotation plots highlight histology, pTNM (pathological

tumour node metastasis) status, relapse site and details regarding whether an intracranial relapse was isolated (brain-only) or non-isolated (brain and extracranial site).

Figure 4| Clonality measurements in preoperative plasma overcome sampling bias from a single tissue sample and predict metastatic seeding potential. A. Depiction of a clonal illusion where a dark blue subclone is found in 100% of cells in a single clinical tissue sample. Such clonal illusion mutations may be detected in a clinical setting using ctDNA derived from many different tumour regions to increase accuracy of ITH measurements in the clinic. Mutations which were clonal (CCF > 90%) in a single, randomly selected tumour region are compared using plasma-based preoperative CCFs splitting by those truly clonal across all tumour regions in TRACERx (clonal) and those which, while they were clonal in the randomly selected region, were absent from other tumour regions (Clonal illusion). Only data from a single randomly selected region was used by ECLIPSE to generate these CCFs. Only preoperative samples with at least 0.1% clonal ctDNA level (high subclone sensitivity samples) were included (N=71) **B.** Preoperative plasma subclonal CCFs split by whether a given subclone was found to be present or absent in cfDNA samples at relapse and postoperative plasma CCFs for relapse subclones at the last high subclone sensitivity timepoint. Only tumours with at least one sample >0.1% clonal ctDNA level (high subclone sensitivity) both preoperatively and postoperatively were included (N=26 tumours with CCFs from 247 subclones included).

Figure 5| Longitudinal measurements of clonal evolution in plasma from surgery, through therapy and to recurrence. ctDNA purity for each clone is calculated by multiplying the clone CCF by the ctDNA purity of the plasma sample (methods) and represents the fraction of all cells from which cfDNA was derived which harbour a given tumour clone at each timepoint. Clonal nesting is based on the phylogenetic tree for each tumour. Clone maps for each tumour tissue mass are depicted above the ctDNA based clonal structure with the phylogenetic tree. **A.** Depictions of longitudinal tumour evolution for examples of monoclonal, polyclonal monophyletic and polyclonal polyphyletic metastatic dissemination patterns. **B.** A kaplan meier plot depicting differences in overall survival

between metastatic dissemination classes (N= 44 tumours with at least 1 high subclone sensitivity sample). **C.** CCFs depicted through time and therapy for CRUK0484 who had a polyclonal polyphyletic relapse. **E.** Variant allele fractions for mutations tracked in CRUK0050 at recurrence. NAG = Neoantigen.

Extended figure Legends:

Extended figure 1| TRACERx ctDNA cohort parameters – **A.** Stacked bar plot of patient specific panels designed from primary tumour sequencing data showing the number of clonal (red) and subclonal (mauve) variants per panel. Variants lacking clonality information are displayed in gray (median of 3 variants per patient [1-20], these mutations are either no longer called by TRACERx or called by ArcherDx but not TRACERx, see methods) A median of 126 clonal variants (range 21 to 195) and 64 subclonal variants (range 2 to 174) were tracked by the patient specific panels. Clonality was determined by PyClone analyses of multi-region exome data (see methods), in the absence of PyClone data, variants present in all multi-region sequenced tumour samples were called clonal. **B.** Violin plot demonstrating the % of subclonal clusters derived from multi-region tumour exome data tracked by patient specific panels on a per patient basis, a median of 88% of the subclonal mutation clusters present in each patient's multi-region exome derived phylogenetic tree were tracked [range 0-100]. 191 tumours with phylogenetic trees were included. **C.**Distribution of cfDNA input values for the cohort, median input of 23ng, n=1069 samples. Capping at 60ng input was performed for some of the cohort, for the remainder of the cohort all cfDNA extracted was input into assay (colours represent different cfDNA input categories and match panel h. **D.** Histogram demonstrating the distribution of per variant unique coverage values across the cohort, unique coverage refers to error-controlled coverage (at least 5 unique molecular identifier (UMI) matched reads required to create an error-controlled read, see methods). The median unique coverage per-variant was 2226x (range 0 to 53789x, n=201910 variants). **E.**Correlation between cfDNA input (Y axis) into the assay and the median UMI-corrected depth across 1069 plasma timepoints (Y-axis). Spearman's R value = 0.63 and P value <0.001. **F-G.** Boxplots demonstrating the error-rates (%, y axis) per each of 96 mutation trinucleotide contexts (X axis, 192 mutation trinucleotide contexts simplified to 96 reverse-complement identical mutation types), plots divided by transition event (F) and transversion event (G); background position data from n=1069 cell-free DNA libraries utilised to generate plots, variants predicted to exhibit low background error-rates from pilot data analyses were prioritised for PSP design. Hinges correspond to first and third quartiles, whiskers extend to the largest/smallest value no further than 1.5x the interquartile range. **H.** Correlation between median deduplication ratio achieved in a sample (y axis) and cfDNA input into the assay (X axis), duplication ratio refers to the median number of duplicate UMI-supported reads within a read family. Resequencing of samples where the median duplication ratio was less than 10 was performed where possible to maximise recoverable information from cfDNA samples, given that 5 UMI-supported reads are required to make a UMI family. 17 of 1069 evaluated cfDNA samples exhibited a final median deduplication ratio less than 5 (corresponds to the horizontal line on the plot). Colours correspond to different cfDNA input categories and match panel C.

Extended figure 2| MRD calling thresholds and analytical validation A. postoperative MRD caller P-values observed in n=5 patients who did not have recurrence of their NSCLC, all n=55 samples had P-values in excess of P 0.1 threshold meaning that they were deemed negative for ctDNA. **B.**

postoperative ctDNA caller P-Values observed in n=5 patients who had relapse of their NSCLC, 1 of 13 calls was made between MRD P values of 0.1 and 0.01, the remaining 12 calls were made at a P-value less than 0.01. **C.** preoperative ctDNA calls from n=10 pilot cohort, 7 patients had positive ctDNA in plasma prior to surgery, all calls were made at a P-value <0.01. **D.** In-silico simulation analysis to assess ctDNA caller specificity, 3157 mock MRD panels were generated within the evaluable pilot patient libraries and ctDNA caller P-values assessed. At a P-Value <0.1 threshold 121/3157 simulated mock panels were ctDNA positive (*in-silico* specificity of 96.2%) at a P-Value <0.01 22/3157 simulated mock panels were ctDNA positive (*in-silico* specificity of 99.3%). **E.** Analytical validation of 50-variant MRD detection panel (representative of a clinic-ready Anchored Multiplex PCR Personal Cancer Monitoring assay). Fragmented DNA with a known single nucleotide polymorphism (SNP) profile was spiked into a second background of fragmented DNA with a different SNP profile and a 50- variant candidate patient specific panel was designed targeting 50 alternate positions in the spiked in DNA. 559 data points were generated across 8 different DNA input quantities to establish the limit of detection plots. The vertical axis demonstrates sensitivity (defined as the proportion of all repeats that resulted in MRD detection using an alpha of 0.01). The confidence intervals on the plot are Clopper-Pearson confidence intervals (95% Cis), the horizontal axis shows the quantity of variant germline DNA that was spiked into each repeat expressed as a percentage of total DNA in that sample. **F.** Circulating tumour DNA samples with high variant allele fractions were spiked into a different cell free DNA background, 3 different total cell free DNA inputs were analysed (10ng, 25ng and 60ng) using 50 variant MRD detection panels. Axis are the same as (E). **G.** Data from analyses of 48 blank samples donated by healthy participants, MRD algorithm P-values are displayed. **H.** Barplot demonstrating the intended allele frequencies and the measured allele frequencies in the different spike-ins presented in part **E)** and part **F)**, only data from ctDNA positive samples presented. The colors of the barplot represent different DNA input masses as shown by the legend. The error bars on the plot represent the mean value of all positive spike-in samples plus/minus the standard deviation of the values. Where the error bar is absent this is because at this spike-in level and DNA input mass only one positive sample was observed, where the error bar led to an observed mean AF less than 0 the error bar was stopped at 0 for visualisation purposes in the 0.05% spike-in, 2ng input mass case. The horizontal dashed lines correspond to 0.1%, 0.05% and 0.01% spike-in categories. **I.** Comparison between the content of cell-free DNA input into ddPCR reactions (yellow) and AMP PCR reactions (blue) – significantly more cell free DNA was input into ddPCR reactions (paired Wilcox test P=0.014). **J.** Orthogonal comparison between ctDNA detection based on AMP-PCR panels used in TRACERx and ddPCR against a single clonal variant. ddPCR ctDNA positive call threshold was two mutant droplets (left table) and 1 mutant droplet (right table). Percentage positive agreement (PPA) and percentage negative agreement (NPA) using ddPCR as the comparator is displayed on the table.

Extended figure 3| Preoperative ctDNA detection A. Flow diagram demonstrating different cohorts analysed in this manuscript, the top part of the flow diagram shows the total number of plasma samples that were intended to be analysed (n=1095 from 197 patients) which reduced to 1069 samples due to single nucleotide polymorphism mismatches between cfDNA and tissue exome data in 26 cases suggesting sample swap. These samples were analysed in 3 main cohorts, the pilot cohort (left), the preoperative cohort (middle) and the postoperative cohort (right). The postoperative cohort was divided into different categories based on landmark evaluability (relating to samples donated within 120 days of surgery to enable a landmark ctDNA analysis). Below the preoperative and postoperative cohorts the number of ECLIPSE evaluable patients are demonstrated. **B.**Heatmap demonstrating tumour specific clonal ctDNA fractions in patients with synchronous primaries diagnosed at baseline. The annotation rows of the heatmap show the ctDNA call present in that

sample across all variants interrogated by the MRD caller, the highest pathological TNM stage, the individual histology and individual tumour volumes of the two synchronous tumours present at baseline. **C.** Boxplot demonstrating the difference in pack-year history across all 197 preoperative ctDNA positive NSCLC patients and preoperative ctDNA negative NSCLC patients. Hinges correspond to first and third quartiles, whiskers extend to the largest/smallest value no further than 1.5x the interquartile range. **D.** Kaplan Meier curves demonstrating freedom from recurrence outcomes in ctDNA high (dark red), ctDNA low (blue) and ctDNA negative (grey) single primary adenocarcinoma patients (left) and single primary non-adenocarcinoma patients (right). ctDNA high and low was categorised based on median clonal ctDNA levels across ctDNA positive cases and relates to above and below 0.16%. Log-rank P values are displayed on each plot alongside hazard ratios (HRs) that use ctDNA high patients as the reference category. Freedom from recurrence refers to lung cancer relapse events only. **E.** Multivariate cox regression analyses of Overall Survival and freedom from recurrence (defined as recurrence only) in patients with single (non-synchronous) NSCLC, evaluating ctDNA detection status, pTNM stage (tumour Node Metastasis pathological stage version 7, categories I, II or III), whether adjuvant therapy was administered, age and log10-transformed unique sequencing depth as predictors in adenocarcinomas and non-adenocarcinomas separately. **F.** Heatmap showing the site of relapse in NSCLC cases divided by whether preoperative ctDNA was detected (right) or undetected (left). Intrathoracic (mediastinum, locoregional, ipsilateral lung, distant lung – red colors) or extrathoracic (bone, brain, liver, adrenal, extrathoracic lymph-nodes or other extrathoracic site – dark blue colors) sites of relapse are shown (sites shown are metastatic sites diagnosed within 180 days of clinical relapse). Heatmap is annotated by tumour Node Metastasis pathological version 7 stage, histology and is divided by ctDNA detection outcome. This heatmap demonstrates that patients with preoperative ctDNA detection are more likely to suffer an extrathoracic relapse. **G.** Kaplan Meier curves demonstrating post-relapse survival in recurrent adenocarcinoma patients stratified by whether they were preoperative ctDNA positive (red) or preoperative ctDNA negative (grey).Log-rank P values are displayed on each plot alongside hazard ratios (HRs) that use ctDNA negative patients as the reference category.

Extended figure 4|Volume and phenotypic analysis of ctDNA positive and ctDNA negative adenocarcinomas - A. Flow chart demonstrating patients available for volumetric analyses and reasons for exclusion. **B.** Histogram showing number of NSCLC cases by volume, with ctDNA positive samples shown as red bars, and ctDNA negative samples shown as grey bars. n=159 volume evaluable cases. **C.** Correlation plot with each individual TRACERx case that was ctDNA positive as a dot and coloured by adenocarcinoma status (dark red) and squamous or other histology (dark blue). **D.** Based on a multivariable linear regression model fitted to the data in C) (see methods), we categorised ctDNA negative adenocarcinomas as biological low-shedders or technical non-shedders, if tumour volume predicted for a clonal mutation ctDNA level above the clonal ctDNA level a library could detect (95% lower confidence interval for estimated clonal ctDNA level based on tumour volume is above detectable clonal ctDNA level the preoperative cfDNA library from that patient case classed as a probable biological low-shedder (red on histogram) otherwise the case was classed as a probable technical non-shedders (blue on histogram). Y axis represents the lower 95% confidence estimate for clonal mutation ctDNA level divided by the minimally detectable clonal mutation ctDNA level for that patient's panel. The X axis is each individual patient analysed. **E.** Violin box-plots comparing tumour purity in ctDNA low-shedder adenocarcinomas (blue, n = 97 tumour regions from 33 patients) and ctDNA positive adenocarcinomas (red, n = 177 tumour and lymph node regions from 37 patients). **F.** Barplots showing gene-level driver alterations between ctDNA positive adenocarcinomas (n = 41 patients) and ctDNA negative low-shedder adenocarcinomas (n = 36 patients). Colors denote ctDNA detection status. Y axis shows the top 12 most frequently altered

genes, X-axis shows the percentage of patients carrying an alteration in the gene per detection category. **G.** Pathway-level driver mutations between ctDNA positive adenocarcinomas (n = 41 patients) and ctDNA negative low-shedder adenocarcinomas (n = 36 patients). X axis shows patient IDs, Y axis shows pathways following the Sanchez-Vega definition. Top bar denotes ctDNA detection status (dark red represents ctDNA positives, blue represents biological low-shedders). Heatmap colours display mutations, blue denote clonal mutations and red denote subclonal mutation. No pathway showed significant enrichment in either ctDNA shedder or non-shedder adenocarcinomas. **H.** Genome doubling status per tumour comparing ctDNA positive adenocarcinomas to ctDNA negative low-shedder adenocarcinomas. Yellow represents the number of tumours subjected to genome doubling in at least one region, turquoise represents tumours without any genome doublings. **I.** Volume by ctDNA shedding status. Biological non-shedders in red represent the smallest quartile samples. After removal of these from the analysis no significant difference in tumor volume was found between ctDNA positive and ctDNA low-shedder. **J.** Venn diagram showing the overlap between significantly differentially expressed genes between ctDNA positive and ctDNA low shedder adenocarcinomas obtained from the full dataset, relative to the volume-adjusted dataset. **K.** Venn diagram showing the overlap between significantly altered cytobands as called by GISTIC, comparing ctDNA positive to ctDNA low shedder adenocarcinomas obtained from the full dataset, relative to the volume-adjusted dataset.

Extended figure 5| Exploration of unexpected MRD positive results in non-relapse patients. A) Table demonstrating details of unexpected ctDNA positive results in patients who did not suffer disease recurrence. **B.** These dot-plots represent confidently detected individual variants at pre (day-0) and postoperative cfDNA sampling timepoints (left) or various normal, control and Peripheral-blood mononuclear cell (PBMC) samples (middle) and tumour tissue mutant allele frequencies (right) from patient CRUK0498. Notably MRD calls were made at 8 of 9 timepoints in this patient yet they did not recur from their NSCLC. The four variants in the legend (variants in genes *ATP2C1*, *DDIT4L*, *EYS* and *TUSC3*) represent variants confidently called at 50% or more of timepoints across the cfDNA samples (note that confidently called means an individual variant Poisson p-value of <0.01) - these variants are also detected across the normal and PBMC DNA samples from the same patient (but not control genome in a bottle DNA [GIAB], middle plot first category), suggesting that they are not tumour derived or technical artefacts and likely to be clonal haematopoietic variants, yet the variants were present at high variant allele frequencies in tumour tissues (right) **C.** A representative haematoxylin and eosin image from patient CRUK0498's tumor where exome analysis detected the variants in genes *ATP2C1*, *DDIT4L*, *EYS* at high variant allele-frequencies. This image shows dense lymphocyte infiltration in this tumour-region again suggestive that these variants were in-fact haematopoietic in origin rather than associated with CRUK0498's neoplasm. **D.** A further 21 preoperative PBMC samples were analysed from TRACERx patients - no confident panel-wide variant DNA calls were made in these patients' PBMC samples using the MRD calling algorithm. **E.** Variant-level analyses of the 21 preoperative samples analysed in panel (**D**) highlighted that 13 of 3978 variants interrogated by the panels were detected (variant level P value <0.01), 9 of 13 detected variants were removed from MRD caller algorithm in cell-free DNA analyses (cfDNA) due to triggering filters highlighted in the heatmap annotation, only 2 of the 4 remaining variants carried deep alternate reads in the respective patients' preoperative cfDNA sample (red arrows). The heatmap shows the cfDNA variant allele frequency and the WBC variant allele frequency of the detected variants. Two mistargeted germline variants are highlighted for patient CRUK0296 – these variants were targeted in error by the industry panel design pipeline but not by the TRACERx exome pipeline (see methods) and were filtered from the MRD calling algorithm due to triggering the outlier filter (dao_imbalance, dark red) or the strand bias filter (dark green).

Extended figure 6| Expanded postoperative ctDNA and imaging surveillance analysis. A. Dotplot analysis of 10 patients who experienced intracranial relapse, had extracranial imaging and the point of relapse to define whether the brain relapse was isolated and who were positive for ctDNA in a postoperative blood sample. The X axis shows the clonal ctDNA level at the point of postoperative ctDNA detection and the Y axis shows the day of postoperative ctDNA detection. Points are coloured based on whether the intracranial relapse was solitary (green) or accompanied by another extracranial site (red). Points are shaped by landmark ctDNA status. **B.** Heatmap of clonal mutation ctDNA level data at first postoperative ctDNA detection. The annotation rows show the landmark ctDNA status of the patient (landmark positive, ctDNA detected within 120 days postoperatively; landmark negative, ctDNA negative within 120 days postoperatively; unevaluable, landmark status cannot be established), the day ctDNA was detected, the histology of the primary tumour and lead-time (time from ctDNA detection to clinical relapse, heatmap ordered based on this annotation bar) where lead time was not applicable (for example unresected disease, ctDNA detected post-relapse) lead-time is coloured grey. The next two rows (bar charts) demonstrate the number of clonal or subclonal mutations tracked by an AMP patient-specific-panel (PSP), if the bar is blue it represents confident detection of an individual variant (based on an individual variant P-value of <0.01), if the bar is black it represents absence of confident calling of a variant, if the bar is red it represents that a variant was filtered by the MRD calling algorithm. The final row represents the mean clonal ctDNA level at the first ctDNA detection time point for a patient. This is on a log-10 scale as displayed in the heatmap legend. **C-D.** Kaplan Meier curves in the landmark evaluable population (patients who donated blood within 120 days post-surgery before treatment or clinical recurrence, n=112/118 evaluable for survival analysis) showing overall survival (OS,D) or freedom from recurrence (FFR,E) outcomes for landmark positive (dark red) versus landmark negative (grey) patients. Log-rank P-values and hazard ratios with landmark negative patients as the reference displayed on curves. **E.** Boxplots showing the distribution of lead-times (times from ctDNA detection to clinical recurrence) categorised by patient landmark ctDNA status. Kruskal-Wallis test P=0.036 and Wilcoxon-tests compare individual categories. **F.** Longitudinal per patient plots in 12 patients who were ctDNA positive prior to adjuvant therapy. Plots are annotated with lead-time (L-t), scans performed and treatment administered (see legend). The vertical axis represents clonal ctDNA levels and each square on the plot represents a blood sampling timepoint. If the square is red it indicates that the blood samples was positive for ctDNA using the MRD caller. The horizontal access displays days post-surgery, samples to the left of day 0 were taken preoperatively. **G.** Pie charts demonstrate the number of occurrences of specified ctDNA detection statuses (red – MRD negative, green – MRD positive, blue – no ctDNA status preceding scan), preceding a scan showing no new changes (left) or new equivocal extracranial changes (middle). The MRD positive and negative categories are then broken down further into a patient-level analysis showing the outcomes of patients who experienced the occurrence of the specified imaging and ctDNA status event(s). **H.** Bar-chart showing the count of specific equivocal changes noted on scans showing new equivocal changes, equivocal lung lesions and lymph nodes were the most common abnormal equivocal findings on NSCLC surveillance imaging. **I.** Barplot of eventual site of relapse and MRD status in 34 patients with MRD status established prior to surveillance imaging showing new equivocal lymph node enlargement. The x axis shows the patient ctDNA detection status preceding surveillance scans (patient CRUK0090 exhibited occurrences of both negative and positive ctDNA statuses prior to separate equivocal lymphadenopathy scans and is present in both ctDNA positive and negative categories). The barcharts are filled with recurrence status of patients in these categories. Recurred with LN refers to lymph node involvement at relapse (dark red color). Recurred with no LN refers to recurrence with no lymph node involvement (green color).

Extended figure 7| ECLIPSE methodology. A) A conceptual overview of the ECLIPSE method and data input types. CCF; Cancer Cell fraction and VAF; Variant allele fraction. **B)** Equation to calculate tumour purity (also termed 'cellularity' or 'aberrant cell fraction') using clonal mutations. **C)** Equation to calculate cancer cell fraction (CCF). Multiplicity = the number of mutated DNA copies in each mutated cell, CNt = total copy number in the tumour, CNn = total copy number in normal (non-tumour) cells, VAF = variant allele fraction, *P* = tumour purity (the % of cells from which the DNA was extracted which are tumour cells or cellularity). **D)** Percentage change in mean multiplicity of clonal mutations comparing measurements in surgical excised tissue samples to tissue samples taken at relapse. **E)** A comparison between mean clonal vaf of mutations and ctDNA tumour purity as calculated by ECLIPSE where data points (samples) are coloured by the average copy number of tracked mutations (measured using tissue sequencing). Multi-tumour patients and samples with evidence of copy number of instability at relapse are excluded.

Extended figure 8| Subclone detection sensitivity of ECLIPSE. A) Minimally detectable CCF for each ctDNA positive sample compared to clonal ctDNA levels for each sample. All ctDNA positive samples included (N=354). Minimally detectable CCF was calculated using the minimum number of required reads for a positive (P<0.01) clone detection call (see supplementary methods). **B)** Minimally detectable CCF over time for each patient with a horizontal line indicating the threshold for high subclone sensitivity samples (20% CCF). All ctDNA positive samples included (N=354). 61% of preoperative MRD positive samples were considered high subclone sensitivity and 66% of postoperative samples were considered of high subclone sensitivity (overall 64% of samples). **C)** A histogram of clonal ctDNA levels for all ctDNA positive samples (N=354) with vertical lines indicating thresholds for ECLIPSE evaluability and for traditional clonal deconvolution evaluability used for TRACERx tissue samples. **D)** A histogram of maximum clonal ctDNA levels observed in post-operative samples for each patient with vertical lines indicating thresholds for ECLIPSE evaluability and for traditional clonal deconvolution evaluability used for TRACERx tissue samples. **E)** Validation of ECLIPSE detection rates across varying subclonal mutation number, clonal ctDNA level, subclone cancer cell fraction and DNA input amount into the assay. Subclones were constructed using ground truth in vitro spike-in experiments in replicates with known mutant allele fractions which were then mixed in silico to construct 76,263 subclones varying across these parameters. Data from these experimentally derived subclones were then run through ECLIPSE and subclone detection rates across each of these parameters depicted.

Extended figure 9| Time-matched comparisons between subclonal structure measured in plasma and in tissue at surgery. A) Correlation between cancer cell fractions (CCFs) as measured in preoperative plasma samples with >0.1% clonal ctDNA level & >=10ng DNA input (high subclone sensitivity samples) with ECLIPSE and those measured with multi region tissue sequencing (M-seq) at surgery. All preoperative samples with phylogenetic data and at least 0.1% clonal ctDNA level (high subclone sensitivity samples) were included (N=71 patients included). **B)** Copy number unaware CCFs calculated only using VAFs (methods) compared to tissue CCF from M-seq. All preoperative samples with phylogenetic data and at least 0.1% clonal ctDNA levels were included (N=71). **C)** A scatter plot demonstrating the relationship between clonal ctDNA level and the proportion of multi-region tumour exome (M-seq) defined subclones detected by ECLIPSE based on varying cancer cell fractions of a subclone. Spearman Rank analysis reveals an association between clonal mutation ctDNA level and percentage of >10% cancer cell fraction subclones detected in preoperative plasma by ECLIPSE (Spearman's Rho=0.7, P<0.001), loess lines are fitted to the plots. n= 117 ctDNA positive tumours. **D)** A comparison of pre-operative plasma CCFs and the average CCFs across all tissue regions sampled at surgery for clones that were unique to one tumour tissue region and for clones that were distributed across several tumour tissue regions. **E)** A comparison of pre-operative plasma CCFs and

the average CCFs across all tissue regions sampled at surgery for clones that were unique to one tumour tissue region separated between small (<20cm^3), medium (>20cm^3 & <100cm^3) and large (>100cm^3) tumours as measured on pre-operative PET/CT scans. **F)** A comparison of detection rates in pre-operative plasma for 20% CCF subclones across a range of clonal ctDNA levels split by whether the subclones were spread across multiple primary tumour tissue regions or were limited to only a single primary tumour tissue region. **G)** A map of tumour clones with areas of multiregional tissue sampling indicated and clones which are over and under sampled highlighted. Most of the undersampled clones are in fact not in the sampled areas creating a bias towards oversampling in clones which we are able to detect, an effect also called the 'winner's curse'. **H)** A ROC curve describing the sensitivity and specificity of detecting clonal illusion mutations using plasma based CCFs using 95% confidence intervals generated using bootstrapping across 500-fold cross validation (N= 71 tumours).

Extended figure 10 | Clonal composition measurements in ctDNA after surgery A). The frequency of high confidence subclonal to clonal bottlenecks (methods) at the latest possible plasma sample time point with sufficient clonal ctDNA level (high sensitivity subclone samples, N = 44 tumours) and which of these subclones harbour subclonal neoantigens (NAGs) which therefore become clonal at relapse. **B)** In cases of clonal bottlenecking at relapse, the percentage increase in the number of clonal mutations is shown with the absolute number of new clonal mutations (N = 18 tumours). **C)** In cases of clonal bottlenecking at relapse the percentage increase in the number of clonal NAGs is shown with the absolute number of new clonal NAGs (N = 18 tumours). **D)** An overview of clonal structure evaluability at relapse for TRACERx patients in our cohort (N = 75 tumours) using either cell free DNA or metastatic tissue. **E)** ctDNA detection status post operatively of subclones split by detection status in metastatic tissue. Untracked subclones were those without any mutations included in the PSP panels and therefore could not be assessed (N = 26 tumours). **F)** Clonal (detected in 100% of cells) vs subclonal (detected in <100% of cells) status at relapse of primary tumour subclones by whether they were detected in cfDNA and metastatic tissue or cfDNA alone (N = 26 tumours). **G)** Metastatic dissemination class determined by tissue and by cfDNA in 22 cases with a metastatic biopsy, a postoperative high subclone sensitivity plasma sample and a phylogenetic tree constructed. **H)** Overall survival kaplan meier plot demonstrating time from the first MRD positive timepoint to death stratified by ECLIPSE relapse dissemination class (monoclonal, light blue, polyclonal polyphyletic, purple and polyclonal monophyletic, green). **I)** A multivariable cox proportional hazards model to predict overall survival including the clonality of relapse dissemination, stage, maximum ctDNA purity (clonal ctDNA level) at relapse, average DNA assay input, histology and whether the first plasma sample after surgery was ctDNA positive including only relapse patients.

Supplementary figure 1| Longitudinal subclonal analyses across all relapsing patients with available phylogenetic trees at at least one post-operative time point with high subclone sensitivity, N = 44). ctDNA purity (CCF of each clone multiplied by ctDNA purity for each sample) is depicted for each detected clone at each time point as well as depiction of the clonal structure of matched tissue samples and the timepoints of their collections. Treatment, scan outcomes and biopsies received by each patient are also depicted.

Methods

Patients and tissue samples| The TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601) is a prospective observational cohort study that aims to transform our understanding of non-small cell lung cancer (NSCLC) the design of which has been approved by an independent research ethics committee (13/LO/1546). Informed consent for entry into the TRACERx study was mandatory and obtained from every patient. All patients were assigned a study identity number that was known to the patient. These were subsequently converted to linked study identities such that the patients could not identify themselves in study publications. All human samples (tissue and blood) were linked to the study identity number and barcoded such that they were anonymized and tracked on a centralised database, which was overseen by the study sponsor only. The ctDNA cohort represents 188 TRACERx 421 cohort eligible patients and 9 additional patients (the following 9 patients were excluded from the final TRACERx T421 cohort [after ctDNA analyses were performed] and were analysed in this manuscript: CRUK0230, 0234, 0291, 0335, 0387, 0480, 0490, 0498, 0622). Reasons for exclusion from final T421 cohort are: CRUK0480,0490: C>A artefact uncovered in exome data (excluded from ECLIPSE analyses), CRUK0291, 0234, 0230, 0387, 0622: Incomplete resection of NSCLC; CRUK0335: Concurrent oesophageal primary present at diagnosis; CRUK0498:1 of 2 tumor regions contained lymphoid associated variants. Remaining preoperative plasma from 19 patients published in Abbosh et al. 2017[5] was also analysed in this manuscript. Extended figure 3a describes the structure of the patient cohort analysed. Multi-region tumour sampling was performed as previously described[7], relapse tissue samples, excess to diagnostic requirements, were also acquired. Sample extraction from tissue and whole blood followed the protocol in the TRACERx 100 cohort and exome sequencing was performed as previously described[7].

Analyses of surveillance and relapse scan reports| Relapse site data was collected from anonymised standard of care imaging scan reports that occurred within 180 days of confirmed clinical relapse (Supplementary Table 14). Each report was reviewed by two clinicians and sites of disease documented. This central review of reports was performed blinded to a patient's disease and death status and where questions regarding interpreting the report arose there was a dialogue with the cancer centre to establish an agreed assessment. 2 patients lacked available scan reports (CRUK0311 and 0452) for these two patients data was gathered from TRACERx case report forms. Where an anatomical site was not covered by a recurrence scan this was marked as not evaluable. Anonymised surveillance (pre-relapse) scan reports were collected from patients who had donated longitudinal plasma samples. These reports were categorised as showing no new abnormality compared to previous imaging, new equivocal abnormality (an equivocal abnormality was defined as any new change compared to a previous scan, equivocal changes were categorised as being related to new lung tissue abnormality including nodules, enlarging lymph-nodes, pleural abnormality or pleural effusion, lung atelectasis or collapse or other changes) or new unequivocal abnormality: scans showing a change that was viewed as definitive malignancy and resulted in a change in clinical management (Supplementary Table 16).

Plasma samples| Blood samples were collected and processed to plasma as previously described [5], up to 4ml of plasma per case was evaluated for the study (range 0.5 to 4ml, median 4ml, see Supplementary Table 2). For 1074 of 1095 samples circulating cell-free DNA was purified from plasma using the MagMAX™ Cell-Free DNA Isolation Kit in conjunction with the KingFisher™ Flex Purification System (ThermoFisher Scientific). KingFisher™ 24-deepwell processing plates were prepared according to the manufacturer's instructions (plate setup option for KingFisher™ Flex Magnetic Particle Processor 24DW, 4 mL of plasma, 75 uL elution volume). Automated cfDNA isolation was performed on the KingFisher™ Flex. For the remaining 21 samples, cfDNA was extracted as previously described[5]. Eluted cfDNA samples were quantified on the Qubit 3.0 Fluorometer using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. The single nucleotide polymorphism (SNP) profile of cfDNA from a patient was matched back to normal exome data and samples exhibiting discordant SNP profiles were excluded as sample swaps (n=26 / 1095 plasma samples analysed).

Volumetric analyses| tumour volume was determined on the basis of pretreatment (PET-) CT scans using 3D Slicer. Contours of the primary tumour were manually segmented on each axial CT slice. Window settings were adjusted if necessary to exclude vessels, lymph nodes or adjacent mediastinal tissue. If no accurate delineation of the primary tumour was possible (e.g. pleural effusion or atelectasis), the patient was excluded from volume analysis. These steps were performed by a trained resident and all contours were confirmed and edited where necessary, by an experienced radiologist. Relevant clinical demographics including gender and tumour location were cross checked with imaging appearances for each scan analysed. Volumetric data is in Supplementary Table 6).

Library preparation using Anchored-multiplex PCR| Anchored-Multiplex PCR (AMP) is a nested multiplex – PCR enrichment chemistry that incorporates strand specific priming and the incorporation of unique molecular identifiers (UMIs) into sequenced reads[9]. Cell-free DNA, fragmented peripheral blood mononuclear cell (PBMC) DNA (60ng) or fragmented normal tissue DNA (60ng) was end-repaired phosphorylated and A-tailed. An adapter containing a universal priming site, the indexes for multiplexing and a UMI is then ligated onto DNA. One round of target specific PCR was performed with a gene-specific primer 1 (GSP1) which amplifies against the P5 primer in the adapter, and a further round of PCR was then performed with a second nested gene-specific primer (GSP2) and a primer that incorporates a second primer containing a P7 index. Strand-specific priming is performed in both rounds of amplification facilitating the identification of positive and negative strand input DNA molecules during informatic analyses.

For cfDNA libraries, indexed libraries were quantified on either the ViiA 7 Real-Time PCR System or QuantStudio Dx Real-Time PCR Instrument (ThermoFisher Scientific) using the KAPA Library Quantification Kit (Roche). Libraries were individually normalized on the Fluent 1080 Automated Workstation (Tecan), then symmetrically pooled and adjusted to a final concentration of 2 nM or 1.25 nM for standard or Xp NovaSeq loading workflows, respectively. Library pools were prepared and sequenced on the NovaSeq 6000 System (Illumina) according to the manufacturer's protocol. We aimed to sequence each library to ~10 million reads. The on-target deduplication ratio of the library, which describes the ratio of raw on-target reads to unique molecular identifier [UMI] supported on-target reads (UMI supported reads contained 5 or more supporting raw reads with a matched molecular index) was then evaluated. In samples where initial sequencing depth resulted in on target de-duplication ratio less than 10:1, additional sequencing was performed to maximise recovery of unique molecular index (UMI) families; this quality control step was introduced to maximise recovery of UMI-families (which require at least 5 UMI-supported reads) from high complexity samples to ensure maximum recoverable information from these samples, thereby reducing bias (given that only UMI families are considered in our analyses). This QC step resulted in the majority of cfDNA libraries (1052/1069) having median de-duplication ratios more than 5 (Extended Figure 1). PBMC and normal tissue libraries were either sequenced on the NovaSeq 6000 system (Illumina) or the NextSeq system (Illumina).

MRD Calling Algorithm | We generated an MRD caller that investigated background sequencing noise on an intra-library basis (Figure 1a). The MRD caller utilised the Archer informatic pipeline to clean input reads and generate deduplicated UMI supported reads. The cleaned, deduplicated, and error corrected UMI-supported reads were aligned to hg19 and used to evaluate alternate observations at predefined positions where tumour-specific variants were present in the patient's tumour (tumour-informed positions). Only "deep" consensus reads supported by 5 or more PCR duplicates (UMI-corrected) were used to infer expected sequencing noise as well as calculate signal for the MRD calling algorithm.

Alternate bases at tumour-informed positions were subject to a strict set of quality filters consisting of an off target filter, a read strand bias filter, a sequencing strand bias filter, background error rate filter, and variant allele frequency outlier filter to remove artefactual signals. The variant allele frequency outlier filter functioned by performing PAM (partitioning around medoids) clustering of the variant allele frequencies (VAFs) of the tumour informed positions that passed previously described filters, K was set to 2 in the clustering algorithm, thus yielding a high VAF group and a low VAF group. If one of the the two clusters had significantly higher VAFs, as indicated by non-overlapping confidence intervals of the highest VAF of the low VAF cluster and the lowest VAF of the higher VAF cluster, and contained 2 or fewer tumour specific variants those variants were removed from consideration downstream in the algorithm.

Next, intra-library background error-rates (ERs) were calculated. ERs were used to establish the level of noise present in each library that had to be confidently exceeded to allow an MRD call to be made. To calculate background library ERs, the number of UMI-supported alternate observations (DAOs, deep alternate observations) were tallied across the assay's region of interest (ROI) for each trinucleotide context (TNC) and for each possible alternate position based on the plus strand of the reference sequence. The ER corresponding to each TNC alternate was calculated as DAO/DDP (DDP, deep UMI-corrected depth across a TNC alternate). In order to measure only PCR and sequencing error, a position in the ROI was not included in the TNC ER calculation if the VAF at that position for a particular alternate is > 1% (on the basis this could represent a clonal haematopoiesis associated mutation or single nucleotide polymorphisms).

A mapping of tumour observed variants and their accompanied TNC ERs was generated. Any tumour observed variant with a corresponding TNC ER upper confidence interval that was above 0.01% was filtered from the MRD calling algorithm. PAM clustering was used to generate 4 "D-groups" of TNC error-rates from qualified TNCs. The population weighted average TNC error-rate was calculated for each of the four D-groups based on the product of the TNC error-rates included in each D-group cluster and the total DDP for each TNC. The generation of 4 D-groups ensured that there was sufficient intra-library DDP coverage of each D-group to make precise estimations regarding ERs for variants within each group.

To determine whether ctDNA was present in the sample, the total observed DAOs summed across tumour specific positions remaining after filters were compared to the number of DAOs that were expected due to background ERs as dictated by the D-groups. A one-tailed exact Poisson test was applied where the total remaining observed DAOs served as the value being tested and the expected number of DAOs due to error served as the lambda of the Poisson distribution. If the resulting P-value of the test is below a pre-specified alpha threshold set to 0.01 then the sample is classified as MRD positive. Supplementary Methods contain details regarding how the pre-specified alpha threshold of 0.01 used in these analyses was generated.

To investigate whether a single mutation targeted by a panel was present we utilised the specific trinucleotide error-rate corresponding to the mutation of interest and a one-tail Poisson test to assess if the number of DAOs across the mutation of interest was above expected background ER. If the number of DAOs was above higher than expected background error using an alpha threshold of 0.01 then a variant was deemed confidently detected. Supplementary Tables 11 and 18 contain sample and variant level outputs of the MRD caller pipeline.

Data inputs for ECLIPSE | For each mutation ECLIPSE requires mutation identifiers (Chromosome, position, reference allele, alternative allele), a sample identifier, the number of supporting reads, sequencing depth, estimated background error rate, clone identifier, a binary call for whether the mutation is clonal or subclonal, mutation multiplicity, total copy number at the mutated locus in tumour cells, total copy number at the mutated locus in non-tumour cells (default = 2) and several optional inputs including variants to be filtered for clone and tumour present calls due to high background error, variants that should be filtered from all analysis for a specific sample and a measurement for the maximally expected normalised standard deviation of CCF in high confidence clones used to identify clones with incoherent CCF distributions which may represent mutation clusters that are not true clones. The background error rate is the probability, for any given read, to observe the specified mutation due to sequencing error. For application of ECLIPSE to our TRACERx data we estimate this using trinucleotide context specific error rates at non-mutated loci in the deep targeted sequencing data (see MRD Calling Algorithm section). The clone identifier, clonal vs subclonal status, mutation multiplicity and total copy number in tumour cells can be calculated using standard copy number extraction and clonal deconvolution methods (ASCAT, Battenberg, Pyclone, DpCLust) used for high tumour purity (>10%) samples, for example from tissue samples, which can then be used as estimates for these variables at the time of ctDNA sampling. Clonal status can be more accurately and comprehensively extracted from the sequencing of multiple high purity samples from the same patient, as is performed in TRACERx, but is not essential. See Application of ECLIPSE to the TRACERx cfDNA data section for further details.

Stepwise description of ECLIPSE

1. **VAF denoising.** Variant allele frequencies (VAFs) are denoised by subtracting the estimated background error, provided to ECLIPSE for each variant. For a description of estimating background error in this dataset see MRD calling algorithm section. Variants in each clone are grouped into clusters (k-means clustering) with similar background error profiles where the number of groups is determined by the sum of the error estimated across all variants, so that if equally dividing the total error from all variants of a clone each group would have a combined error of at least one mutant read. Therefore, if a clone has a total combined error of less than two mutant reads only one cluster will be used. A maximum number of clusters is set to four as default which was used for this dataset. The average background error of each group per variant is subtracted from the number of supporting reads for all variants in each group and divided by the sequencing depth to calculate denoised VAFs.
2. **ctDNA tumour purity calculation**. Deniosed VAFs are used with mutation multiplicities, total copy number at the mutated locus and clonal vs subclonal mutation status for each mutation provided to ECLIPSE to calculate an estimate of ctDNA tumour purity (see above) using the equation shown in Supplementary Figure 7b for each clonal mutation. The equation shown in sup figure 7c is a rearrangement of that shown in sup figure 7b for clonal mutations where CCF = 1. We summarise the mean of these values to provide a final estimate of ctDNA tumour purity per sample.
3. **CCF calculation per mutation and subclone.** For all mutations the sample's ctDNA tumour purity, denoised VAF, multiplicity and total copy number at the mutated locus used in equation show in Supplementary Figure 7c to calculate an estimate of CCF for each mutation in a given plasma sample. The clone identities for each mutation are provided to ECLIPSE and should be calculated independently using standard methods which leverage SNP coverage applicable to high purity samples[37–39]. A mean of the per-mutation CCFs is used as a CCF estimate for each clone.
4. **Poor quality clone identification.** Mutation clustering using standard methodologies are imperfect and are fitted to high purity samples (usually matched tissue samples) used for clustering, excluding lower purity samples which ECLIPSE is able to analyse with deep targeting sequencing. Clusters which are not real, i.e. for which the mutations are not in the same cells, may not continue to track at similar CCFs in data from new samples. To capture such clusters the distribution of ECLIPSE calculated CCFs in each clone in a ctDNA sample are quantified using normalised standard deviations (SDs). This can then be compared to the expected CCF distributions of high confidence clones, for example clonal clusters in higher purity plasma samples. In our data we quantified the normalised SD of all clonal clusters in samples of greater than 5% purity and took the upper 95% confidence interval for this data calculated at 0.56. Subclonal clusters with normalised SDs for CCFs > 0.56 were considered of poor quality and were not considered for analysis. This identified 2.6% of clones in the TRACERx data as of poor quality. Expected CCF distributions will be highly dependent on the input data for ECLIPSE and should therefore be benchmarked on each data set. A function in the ECLIPSE R package is provided to calculate an upper 95% CI of normalised SDs for CCFs in clonal clusters in high purity samples as was performed for this dataset**.**
5. **Clone present calling.** Determination of whether each clone was present or absent from each sample (see High specificity subclone detection section) by comparing the sum of expected background error to the sum of observed signal across all variants in the subclone with a one-sided Poisson test. Mutations with high error that should be excluded from these calculations can be specified.
6. **tumour present calling.** Determination of whether any tumour cells were present for each sample (see above) by comparing the summed expected background error to the summed observed signal across all variants tracked in the sample with a Poisson test. Mutations with high noise that should be excluded from these calculations can be specified.
7. **Minimal detectable CCF estimation for each subclone.** Determination of the CCF equivalent to the minimal number of supporting reads across all variants in a subclone that would be required for a significant clone present Poisson test (P<0.01, see High specificity subclone detection section).
8. **Minimal detectable CCF estimation for an average subclone for each sample.** Determination of the CCF equivalent to the minimal number of supporting reads across all variants in a representative subclone that would be required for a significant clone present Poisson test (P<0.01, see High specificity subclone detection section). The background is taken as an average of the background error in all subclonal mutations tracked in a given sample and is representative for a subclone tracked by four mutations as default, the average number tracked in this dataset. This value allows comparisons of minimally detectable CCF limits across samples.

9. **Minimal detectable purity estimation for each sample.** Determination of the purity equivalent to the minimal number of supporting reads across all tracked variants that would be required for a significant tumour present Poisson test (P<0.01).
10. **Testing for the absence of a complete clonal sweep for each subclone**. A subclone which is detected in high purity samples used for mutation clustering may expand through a full clonal sweep later in the disease course. We would therefore expect to observe CCFs of 100%, indistinguishable from CCFs of clonal mutations after such an event. For each subclone in each sample, a Wilcoxon test is performed to compare the CCFs of each subclone to the CCFs of clonal mutations in the same sample. The resulting P value indicates whether there is significant evidence that the subclone has significantly <100% CCF and therefore is only present in the minority of tumour cells, without a full clonal sweep.

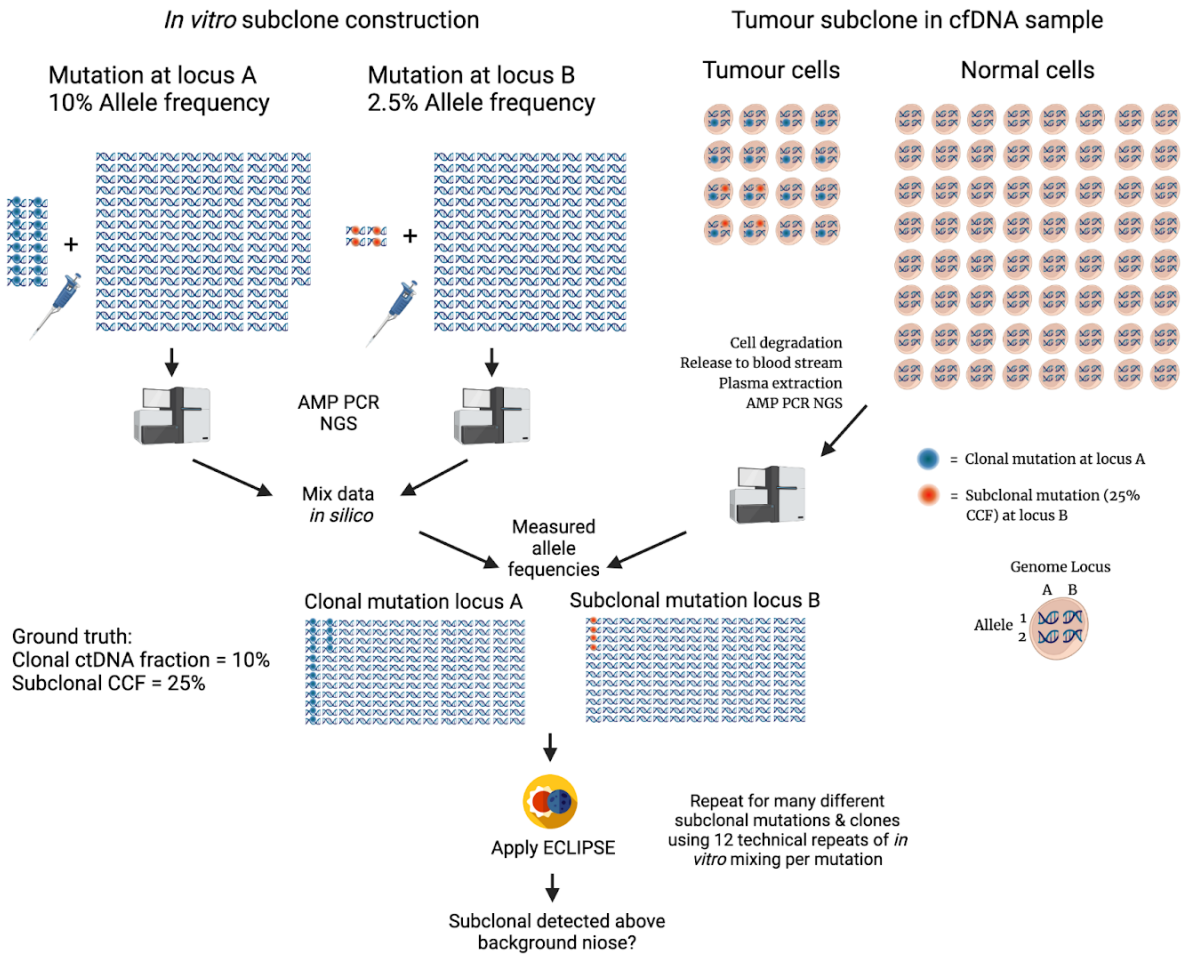
Minimal detectable CCF estimates for each subclone | To quantify our limits of detection of CCF in each sample and subclone, ECLIPSE calculates the number of supporting reads for an all mutations in each subclone that would be required for a positive clone detected call (P<0.01 threshold) based on the number of expected background error reads using the qpois function in R. This value is then divided by the mean depth of all variants in a subclone to simulate a representative minimal detectable VAF for mutations in a given subclone and these values are input into equation in Supplementary Figure 7c to calculate the equivalent CCF, using an average of the mutation multiplicity and total copy number across all mutations in the given subclone and the ctDNA purity of the sample (see Determination of 'tumour purity' in plasma section, supplementary methods). These minimally detectable CCF thresholds are highly dependant on the number of variants tracked in each subclone, hence to provide a single representative and comparable value for each plasma sample we also simulated the minimal detectable CCF for a subclone containing four mutations, which for this study is the median number of mutations tracked in each subclone in this study but can be altered as an argument in ECLIPSE. The minimal detectable number of supporting reads in these four mutations was estimated using the average background error profile of all subclonal mutations in a given sample.

High specificity subclone detection | We took a similar approach to detecting subclones with high specificity in ctDNA has has been undertaken for detection of MRD in this study by estimating the background sequencing error in a trinucleotide context specific manner leveraging non-mutated positions in the target regions of the sequencing library (see MRD calling section). These background error estimates were then provided to ECLIPSE. These background noise rates were multiplied by depth to calculate the expected number of background reads alternate at each mutated position. These expected background read counts were then summed for all variants in a clone and used as the background lambda for a Poisson test comparing the sum of the observed number of reads across the same mutations. A P-value threshold of 0.01 was chosen to call a clone present to match the threshold determined for MRD calling with *in vitro* spike in experiments and the pilot cohort of patients comparing post-surgery samples to relapse status (refer to other methods section).

Application of ECLIPSE to the TRACERx cfDNA data | Inputs to ECLIPSE were prepared from the TRACERx 421 cfDNA and exome sequencing data as follows for all analyses unless otherwise specified. For inputs extracted from matched tissue exome sequencing data all available samples were used including from relapse tissue where possible. Clonal vs subclonal status, cluster identities and multiplicity status were extracted using presence and absence informed clustering as previously described[26] which builds upon the PyClone algorithm[39]. Total copy number in each tumour sample at each mutated locus was extracted as previously described[26]. Normal CN was presumed to be diploid. For metrics calculated per sample purity-adjusted averages (sum of metric per sample multiplied by sample purity divided by sum of all sample purities) were calculated across the whole tumour for input into ECLIPSE for multiplicity and total tumour copy number. The number of variant supporting reads and depth in each cfDNA sample was calculated considering only unique reads with at least 5 supporting duplicates to minimise background error. Trinucleotide specific error estimates were used as input to the background error per variant. 'Hard filtered' variants (those excluded from all ECLIPSE analyses) were those with 'failed filters' of "primer_abundance_filter", "primer_strand_bias", 'sequence_strand_bias', 'dro_cutoff' and 'dao_imbalance'. Additionally 'mid filtered' variants were those with 'failed filters' 'tnc_error_rate' where the background error was considered to high for inclusion in estimates of MRD (see MRD determination NAME section) and were also excluded for estimates for clone presence or absence in ECLIPSE (see steps of ECLIPSE section).

Validation of ECLIPSE CCFs vs tissue exome M-seq CCFs | To compare ECLIPSE estimated CCFs to those estimated using using validated methods used for tissue sequencing at a matched time point we compared purity adjusted averages (see Application of ECLIPSE to the TRACERx cfDNA data section) of surgically excised tumour tissue based CCFs for each subclone from presence and absence informed clustering[26], a benchmarked variant of PyClone[39] to subclone CCFs estimated in ECLIPSE (supplementary figure 9a) for high subclone sensitivity preoperative samples which are defined as those with at least 0.1% clonal ctDNA level. These were samples with an estimated minimally detectable CCF of at least 20% (see power analysis in supplementary figure 8a) compromising 61% of MRD positive preoperative samples from 67 patients. While a formal method for CCF estimation in deep targeted sequencing data has not been previously published for comparison, we compared ECLIPSE to a VAF only method for CCF estimation, naive to copy number statuses where the mean VAF of each subclonal cluster was divided by the mean VAF of the clonal cluster in each sample (supplementary figure 9b) which caused a consistent underestimation of CCF relative to estimates from tissue exome sequencing driven by the higher average multiplicity of clonal mutations compared to subclonal mutations which more commonly occur before large scale copy number amplifications, for example whole genome duplication, which therefore increase mutation multiplicities.

Validation of subclone detection rates using our data and ECLIPSE | To further investigate the sensitivity of subclone detection at different frequencies using ECLIPSE, we have analysed data generated using in vitro spike-in experiments described in Supplementary Figure 2. To generate these data different mutation allele fractions were spiked into wildtype DNA and different total DNA amounts inputted into our AMP PCR NGS assay, including 12 replicates for each spike-in mutation fraction and input amount combination. In total this comprised 398 spike-in samples, each with 50 spiked in mutations, which were then subject to our AMP PCR NGS pipeline, identical to that applied to our plasma-derived cell free DNA samples. We subsampled mutations from each of these spike-in experiments in silico to represent subclones with 1, 2, 4, 10 and 20 mutations (a median of 4 mutations were tracked per subclone in our TRACERx ctDNA panels). Each of these 'subclones' was combined in silico with data from spike-in mutations at higher mutant allele fractions to represent clonal mutations. This allowed us to construct in silico subclones with various cancer cell fractions (determined by the ratio of spiked in mutant allele fraction of the subclonal mutations to the spiked in mutant allele fraction of the clonal mutations), across various clonal ctDNA levels (the spiked in mutant allele fraction of the clonal mutations) across a range of total DNA inputs to the assay (see below). Although these data derive from mixing mutations together from different experiments in silico, the concentrations of DNA are known from ground truth, hence these mixtures provide a deeper level of validation, controlling for various sources of noise in the assay and providing technical replicates. In total we constructed 76,263 subclones from these data which varied in the cancer cell fraction (CCF), clonal ctDNA level, number of mutations per subclone and assay DNA input amount. We ran these data through ECLIPSE using background noise estimates from the same libraries to determine how the rate of subclone detection varies with these four parameters. We focused on the lower DNA inputs (<= 10ng) as the greatest variety of allele fractions were spiked in for these inputs, enabling construction of a wider range of CCFs, and these samples represented the most challenging scenarios for subclone detection. We calculated the fraction of subclones detected for each experimental replicate at each specified clonal ctDNA level and at each CCF. We then used the resulting distribution of detection rates across experimental replicates, for each clonal ctDNA level and CCF, to calculate 95% confidence intervals.



Clonal illusion analyses | For analysis of clonal illusion we reran ECLIPSE for each TRACERx patient considering only data from a single randomly selected tumour sample to simulate a clinic biopsy including multiplicity and total copy number estimates. Clonal status was recalculated using a CCF 90% CCF threshold in the selected region and only mutation specific, rather than clone specific, estimates of CCF were analysed removing the requirement for subclonal cluster status. To analyse clonal illusions all mutations which would be considered clonal in the randomly selected region were selected and split by their clonal status in the presence and absence informed clustering across all TRACERx regions into those truly clonal mutations in all regions (labelled Clonal) and those mutations which were in fact subclonal when other tumour regions were considered and therefore harboured Clonal illusion in the randomly selected region. ECLIPSE estimates (using only data from the randomly selected region as described) were then plotted for these two mutation groups in Figure 4A. To calculate AUC and ROC curve for the sensitivity and specificity of clonal illusion detection all apparently clonal mutations (>90% CCF) in the randomly selected region were used with the ROCIT R package with scores inputted as the mutation specific single region ECLIPSE CCF estimates and final classes considered as the Clonal or Clonal Illusion status leveraging all tumour regions in TRACERx.

Longitudinal depictions of clonal evolution in cfDNA and tissue | Representations of clonal evolution over time were depicted using the ECLIPSE plasma CCFs per subclone, the subclonal CCFs in matched tissue samples extracted either at surgery and the phylogenetic subclone relationships calculated from tissue multi-regional exome sequencing as described[26]. ECLIPSE plasma subclone dynamics were plotted using modified code from the fishPlot R package[40] and clonal structure of tissue samples were plotted using an R package developed in house called cloneMap[41] distributed on GitHub (https://github.com/amf71/cloneMap). Only clones with at least one cfDNA tracked mutation which was not hard filtered (see Application of ECLIPSE to the TRACERx

cfDNA data section) in all samples were shown in the ctDNA and tissue clonality representations and the phylogenetic trees. Clonal dynamics in cfDNA were represented by ctDNA clone purity which was calculated by multiplying the CCF of each clones by the ctDNA tumour purity of each cfDNA sampling therefore presenting the proportion of cfDNA derived cells (including normal hematopoietic cells) which belong to a specific subclone. 44 patients which relapsed from their primary disease and where phylogenetic trees were available from tissue exome sequencing were depicted in figure 5 and supplementary figure 1. In figure 5C CCFs rather than ctDNA clone purities are plotted as purities in this patient vary over several orders of magnitude hence making depiction of change in clonal composition impossible to distinguish on a linear scale required for intuitive interpretation of such area plots. Sample purities are depicted in this case as grey circles below the CCFs.

Definition & detection of clonal sweeps at relapse | Subclones undergoing a clonal sweep were those which expanded after surgery, when they were first detected in tissue WES, to 100% CCF, ie such previously subclonal mutations were now estimated to be present in every tumour cell and parallel subclonal lineages were estimated to have been extinguished. To call instances of a clonal sweep ECLIPSE performs a Wilcoxon test comparing the CCF all mutations in a given subclone to clonal mutation in each sample. The resulting p-value indicates the probability that the subclone has undergone a clonal sweep with a null hypothesis of a clonal sweep being present. We considered a clonal sweep present when this p-value was greater than 0.05 and absolute mean subclone CCF was at least 90%. For each patient the latest possible time point with high subclone sensitivity (ie a clonal ctDNA level of at least 0.1%) was used to determine clonal sweeps at relapse. To estimate how these clonal sweeps at relapse modified the tumour trunk we added all mutations and neoantigen in relapse clonal sweep subclones (including those clustered together in exome sequencing but not tracked in cfDNA) to the clonal mutations for re-estimation of clonal tumour mutational burden and clonal neoantigen burden at relapse.

Determination of phylogenetic metastatic dissemination class | Phylogenetic metastatic dissemination classes were determined in tissue and cfDNA for each relapse patient in this study where a high subclone sensitivity postoperative samples (>0.1% clonal ctDNA level) considering relapse disseminations detected at timepoints after surgery. Work current under review[26] has focused on metastatic disseminations estimated from tissue including disseminations detected at surgery to local lymph nodes which were excised with curative intent but cannot be estimated in cfDNA alone as preoperative ctDNA may derive from either metastatic lymph nodes or the primary tumour. For tissue based metastatic dissemination classes seeding clones specifically for postoperative relapse we extracted using subclones which were shared between the primary tumour and relapse samples from tissue sequencing of metastatic tissue[26]. These clones were used to determine whether single clone or multiple clones seeded the tissue metastasis (monoclonal and polyclonal dissemination respectively) and using the phylogenetic tree in polyclonal cases we determined whether clones were direct descended from one-another in the same clade (polyclonal monophyletic) or if there is branching between the disseminating clones into different clades (polyclonal polyphyletic). In cfDNA patients if all primary tumour subclones detected in postoperative ctDNA were direct descants in the phylogenetic tree and were present at 100% CCF the relapse was considered Monoclonal, If any subclone was present at < 100% (using Wilcoxon test comparing clonal cluster CCFs to each CCFs in each subclonal cluster, $P < 0.05$ and also requiring a mean subclone CCF < 90%) then the metastatic dissemination was considered polyclonal. In polyclonal cases, if the subclones present at relapse were direct descendants of one-another the relapse dissemination was considered as polyclonal monophyletic and if they were branched into separate clades they were considered as polyclonal polyphyletic metastatic disseminations. We did not find a significant difference between the number of tracked mutations in post-operative plasma detected compared to postoperative plasma undetected primary tumour subclones (Wilcoxon text P=0.13, median number of variants tracked = 4 in both cases) suggesting power of detection did not strongly influence which clones were detected in relapse cfDNA.

Designing AMP-MRD enrichment panels| tumour-informed personalised AMP-MRD enrichment panels were designed for 197 TRACERx patients. A median of 50 variants per panel (range 0 to 50) were chosen using the ArcherDx panel design algorithm and a median of 150 variants (range 34 to 153) were chosen using variants selected from TRACERx multi-region exome sequencing data. Due to alterations in our TRACERx exome sequencing pipeline between panel design (2019 to 2020) and final analysis, a small fraction of mutations (3%) targeted by patient specific panels were no longer called with high confidence in tissue exome sequencing data. These mutations were included in MRD analyses (to align with the originally intended analysis approach plus prevent any possible bias conferred by manually removing these variants from consideration by the MRD caller) but were excluded from analyses of clonal structure. For Archer variant selection WES sequencing data from the highest purity tumour region and from the paired germline DNA were used. The algorithm then identified those variants for which there was high confidence that the variants were not artefacts and were tumour specific using the following criteria:the quality of the primers targeting the variant (to ensure high sequencing coverage of the target variant), predicted error rate for the variant in error corrected bins and known involvement in cancer. The predicted error rates for each variant was based on an analysis of AMP cfDNA libraries sequenced on a NovaSeq instrument.

The algorithm then determines which variants can be targeted using an ArcherDX AMP panel and from this set of variants the 50 most informative mutations are targeted based on these criteria: the quality of the primers targeting the variant (to ensure high sequencing coverage of the target variant), predicted error rate for the variant in error corrected bins and mappability. The predicted error rate for each variant is based on an analysis of AMP cfDNA libraries sequenced on a NovaSeq instrument. This error rate analysis was performed by performing targeted variant calling on every possible SNV in a set of Archer LiquidPlex cfDNA libraries. The TRACERx variants were selected using the TRACERx primary tumour WES pipeline[7] for ranking. SNVs were ranked based on their (1) driver designation, (2) trinucleotide context as described above, (3) mean mutation allele count in primary tumour. All SNVs were categorised as either “neoantigen”, “clonal” or “subclonal”. Up to 50 variants were picked from each category. Neoantigens were additionally ranked by binding affinity[42]. Subclonal mutations were picked to represent all phylogenetic mutation clusters, picking up to an equal number of mutations from each when possible, up to a total of 50 maximum. Finally, 50 clonal variants were picked. If the sum of subclonal and neoantigen mutations was less than 100, the difference was picked from the list of clonal mutations.

Each personalised enrichment panel also contained 90 primers targeting 45 common single nucleotide polymorphisms (SNPs). During analyses the zygosity of these SNPs in a cfDNA library is compared to their zygosity in the whole exome sequencing data for that patient to confirm that a sample swap did not occur. In addition the coverage provided by these primers helps in establishing the background PCR and sequencing error rate for a library. These 45 SNPs were chosen based on being present in each Gnomad subpopulation at a frequency of 25%-75% to maximise utility in detecting sample swaps.

ArcherDX variant choosing and panel design deviated from the standard workflow in two cases. In the case of the pilot sample CRUK0297 the tumour and non-tumour samples used in design were not properly matched and rare germline variants appeared to be tumour-specific as a result. The ArcherDX variant choices in this panel included many germline variants. For this reason the MRD libraries for CRUK0297 underwent manual blanking of the germline targeted variants to facilitate use of these samples in the pilot patient analyses . All subsequent ArcherDX panel designs included a quality control step to confirm that the common population polymorphisms in the tumour and non-tumour samples matched. The second case in which panel design deviated from the standard ArcherDX workflow occurred in the design of CRUK0296. The variant call data for a tumour-normal tissue pair could not be obtained in the standard format for this patient. In this case, the standard variant caller could not produce a result so the variant caller VarDict[43] was used and data was not available for the non-tumour sample in the standard format. As a result two germline variants (chr6:31118898:A:T and chr16:70928307:C:A) were targeted. These two variants were removed from consideration in making the MRD call automatically by the Outlier Filter in every library prepared with this panel (refer to the Library-specific MRD Calling section above) but were kept in all analyses and not manually blanked. Two patients lacked Archer picked variants (CRUK0157, 0227), CRUK0157 as the exome data could not be processed by the Archer variant picking pipeline and CRUK0227 due to an error during PSP primer ordering.

Neoantigen pipeline | HLAHD was used to determine the patient-specific HLA composition. 9-11mer peptides harbouring nonsynonymous mutations coupled with patient-specific HLA were used as input to NetMHCPan4.1. A Rnk_EL <0.5 was used to determine strong binder peptides.

Analytical validation experiments| For experiment LOD1, 634 samples of fragmented DNA with a known SNP profile (Genome in a Bottle DNA, NA24385) were added to a background of four other fragmented Genome in a Bottle inputs (NA24149, NA24631, NA24694 and NA24695). Six AMP enrichment panels were generated targeting 50 SNPS heterozygous in NA24385 and absent from the other four cell lines. To generate contrived samples NA24385 DNA was spiked into a background of the other four samples at ratios of 0.006% to 0.2% by mass to target variant allele frequencies ranging from 0.003% to 0.1% allele frequency (since heterozygous variants are present at 50% in the neat NA24538). As part of the same dilution series, admixtures with target allele frequencies of 1%, 5% and 10% were made. These mixtures were used as input for AMP library preparation to confirm that mixing based on mass achieved the desired target allele levels. The spike-in variant level was measured in these higher AF libraries by adding the number of deep alternate reads across the targeted SNPs and dividing by the total coverage of all deep reads across targeted SNPs. This analysis confirmed that the spike-ins achieved the targeted AFs. Fragmented DNA inputs from 2ng to 80ng were used in the experiment to reflect the range of DNA inputs encountered in a clinical setting. Overall 559 of 634 samples were deemed evaluable for LOD1 analysis (62 samples failed because of incorrect DNA input used, determined by on-target read per primer per ng input of <30 or >400, 8 samples failed because they had less than 10 million reads and 5 samples due to potential duplicate libraries). Clinical samples were used in validation of AMP-MRD (LOD2) and were prepared using a similar method to the Genome in a Bottle mixtures. Whole exome sequencing data from four patients was used to design patient-specific panels with the ArcherDx panel design algorithm containing 50 SNVs. The panels were used to prepare libraries using cfDNA from each patient and the overall tumour variant AF for each sample was calculated by adding the total number of deep unique reads containing a targeted tumour-specific variant and dividing by the sum of the deep unique coverage across all targeted tumour variants. All four patient cfDNA libraries had a total AF of >1%. A single mixture was made using cfDNA from healthy donors and used to dilute the patient cfDNA. These dilutions were performed as a serial dilution. First a dilution was made targeting a 1% total AF and libraries were prepared using this mixture. The total AF was measured for this sample and a dilution correction factor was calculated to account for differences in conversion efficiency between the background cfDNA. E.g. if a 1% AF was targeted and an AF of 1.3% was observed then this would indicate that the patient cfDNA is more efficiently converted to library than the background and more background DNA would need to be used. Mixtures were then made to achieve AFs of 0.1%, 0.05%, 0.01%, 0.008% and 0.005%. A total of 100 libraries were prepared at 5 AFs and 3 input masses. 48 blank samples (DNA donated from 22 healthy donors) were analysed to assess assay specificity. Panel observed allele frequencies were calculated by taking the number of deep alternate reads noted across the AMP panel, removing estimated background error and dividing by deep depth across the panel. For assay sensitivities at specific spike-in categories, Clopper-Pearson binomial two-sided 95% confidence intervals were calculated in Extended figure 2e-f using the R package DescTools and the function BinomCI.

Simulation analysis to assess specificity| Trinucleotide context of tumour-specific SNVs within each TRACERx AMP-MRD pilot cohort panel was assessed. Based on these data mock tumour signatures (genomic positions covered by the enrichment primers with positions of similar expected error rates of the targeted SNVs) were generated. A mock variant was added to a mock signature if the following criteria was met: It is bi-directionally covered by primers intended for MRD detection, It contained the same TNC-group error rate as the true MRD variant it's replacing, it was not a known population SNP variant as dictated by Ensemble's Variant Effect Predictor version 94.5, had a error-corrected coverage delta no more than 2,000 compared with the true MRD variant, and was not used within any other mock tumour signature, including itself. Thus, the resulting mock signatures targeted bases that are not mutated in the primary tumour and any positive MRD call from these mock signatures was by default a false positive. 3157 mock signatures across 91 pilot cfDNA libraries were interrogated for MRD positive calls A simulated ctDNA level was estimated for each sample by taking the number of deep alternate reads noted across the mock signature, removing estimated background error and dividing by deep depth across the mock signature. Data from this simulation is present in supplementary table 17.

Digital droplet PCR orthogonal validation| Digital droplet polymerase chain reaction (ddPCR) orthogonal analyses were performed in 30 preoperative plasma samples from TRACERx patients who also had

preoperative plasma analysed by the AMP- personalised tumour informed approach and 8 negative controls (preoperative plasma from patients diagnosed postoperatively with non-malignant disease). TRACERx patients were selected as having clonal driver mutations that could be targeted by a single ddPCR assay. Clonal driver mutations targeted included *KRAS G12R, G12D, G12V, G12S, G12A, G12C* and *EGFR L858R*. The ddPCR assays used were SAGAsafe® assays (SAGA diagnostics) and had been designed and developed on the BioRad QX200 Droplet Digital PCR system. ddPCR analyses were performed at SAGA, SAGA received plasma (median 4.8 mls, range 2.5 to 5.2mls). cfDNA was extracted using the QiaAMP MinElute ccfDNA Midi Kit (Qiagen). cfDNA was eluted in 40 ul of Buffer EB. The entirety of cfDNA material was input in each case and ddPCR analyses were run in 4 replicate reaction wells per sample. All 8 negative controls (each assay tested once, *KRAS G12A* tested twice exhibited no mutant droplets detectable in control cfDNA, Supplementary Table 4).

Transcriptional data analyses | Statistical data analysis was carried out using R version 4.1.2. Gene level transcription analysis uses limma version 3.50.1. The analysis includes 106 tumour regions sampled from 35 ctDNA positive patients and 74 tumour regions sampled from 32 biological ctDNA low-shedder patients. The analysis takes into account 18876 protein coding genes based on the HGNC database, retrieved on 03/04/2022. Genes with insufficient expression levels (count < 30) are filtered out and effective library sizes are calculated using the TMM (trimmed mean of M values) method. Count data is then transformed to logCPM values (log2-counts per million). Prior to linear modelling, a weight per observation is calculated based on the association between mean and variance. In order to take into account the association between tumour regions within patients, a per-patient consensus correlation is computed. Based on the logCPM table, the within-patient correlations and the ctDNA detection status, a linear model is fitted. A contrast matrix comparing ctDNA positives and biological ctDNA low-shedders is constructed alongside with the associated coefficients and standard errors, and the empirical Bayes method (eBayes) is used to calculate the differential expression. The resulting gene-level p-values are adjusted for multiple testing using the Benjamini-Hochberg (FDR) method. Genes are noted as significantly differentially enriched if their adjusted p-value is below 0.05.

The set of significantly overexpressed genes per detection category (n = 932 for ctDNA positives, n = 939 for biological ctDNA low-shedders) is used to calculate Reactome pathway enrichment (ReactomePA version 1.38.0). The resulting p-values are FDR-corrected and an adjusted p-value cutoff of 0.05 is employed.

Additionally, pathway enrichment with respect to the Hallmark Gene Sets from the msigDB database was investigated. Pathway enrichment analysis was carried out on logCPM data including 17815 protein-coding genes using Gene Set Variation Analysis (GSVA version 1.42.0). Fold change of GSVA enrichment scores comparing 106 tumour regions from 35 ctDNA positives and 74 tumour regions from 32 biological ctDNA low-shedders is calculated using the estimated marginal means (emmeans version 1.7.2) method, using a linear mixed-effects (lmerTest version 3.1-3) model to take into account the patient-tumour region associations, treating detection status as fixed effect and patient ID as random effect. The resulting pathway-level p-values were FDR-corrected for multiple testing.

Mutation analyses| Driver mutations in 180 genes from 77 ctDNA positive and biological ctDNA low-shedder patients (41 ctDNA positives, 36 biological ctDNA low-shedders) were included in the analysis. Clonality was determined based as part of the TRACERx WES pipeline. If a patient carries multiple mutations in the same gene with differing clonality, the clonal state was kept. In the gene-level analysis, the top 12 frequently mutated genes were considered. Genes were assigned to pathways as described by Sanchez-Vega et. al[44]. Fisher's exact test was conducted in a two-tailed manner to compare the number of ctDNA positives and ctDNA low-shedders carrying alterations in the frequently mutated genes. The resulting p-values were corrected using the FDR method.

Chromosomal instability analyses | Copy number data including allele-specific copy numbers and purity estimates were derived from the TRACERx WES pipeline and were available for 177 tumour and lymph node regions collected from 64 ctDNA positive and biological ctDNA low-shedder patients. Cytoband analysis was conducted using GISTIC 2.0, taking one sample per patient as input. For amplifications and deletions, the maximum and minimum copy number per genomic segment were used, respectively. GISTIC score difference of 0.5 was used as threshold for significance cutoff. Cytobands are overlapped with output from GISTIC2.0 to get a mean GISTIC score for each cytoband. FLOH and wGII were analyzed on the region level (591 tumour and lymph node regions from 148 patients - 177 regions from 37 ctDNA positive adenocarcinomas, 97 regions from 33 biological low-shedder adenocarcinomas and 322 regions from 79 non-adenocarcinomas). Comparing the

chromosomal instability metrics between ctDNA positive and biological ctDNA low-shedder adenocarcinomas and non-adenocarcinomas is performed using a linear mixed model, taking into account the within-sample associations. Pairwise comparisons are made using the estimated marginal means method and p-values are FDR-adjusted. Tumour regions were considered genome doubled (GD) if the fraction of the genome with major allele >= copy number 2 was >50% as per previous publications[45]. Tumours were considered to have GD if any single region harboured a GD event. GD data was available for 70 lung adenocarcinoma patients (33 biological ctDNA low-shedders, 37 ctDNA positives).

Purity analysis | Using 274 regions from 70 patients (177 regions from 37 ctDNA positives and 97 regions from 33 ctDNA negatives), we performed an estimated marginal means analysis incorporating a linear mixed model approach to account for the within-sample associations. The analysis compares ctDNA positives with biological ctDNA low-shedders.

ORACLE analysis|ORACLE scores were calculated by using the method developed by Biswas et al[14], including 216 tumour and lymph node regions from 81 ctDNA positive and ctDNA low-shedder patients (115 regions from 35 ctDNA positives, 101 regions from 46 ctDNA low-shedders). Pairwise comparison between the ctDNA shedders and ctDNA low-shedders are made using the estimated marginal means method with a linear mixed-effects model to account for the within-patient associations between tumour regions.

Volume adjustment | Biological low-shedder samples are excluded from the volume-adjusted analysis if their size falls in the lowest quartile size range (< 6415.389 mm3). Transcriptomic and GISTIC analyses are repeated using the volume-adjusted dataset as described above. Taking into account the significantly overexpressed genes and significant cytobands in both datasets, Venn-diagrams are constructed for comparison and permutation test is carried out to assess the statistical significance of the overlap.

Clonal mutation ctDNA levels| Mutations that were defined as clonal by PyCloneSC or in the absence of PyClone data that were present in every primary tumour tissue region analysed (ITH state =1) and that were unfiltered by the MRD caller were used in clonal mutation ctDNA level estimations. For each mutation the MRD caller estimated trinucleotide error rate associated with that mutation and the coverage of that mutation was used to estimate the number of expected error controlled reads we would observe due to error. Clonal mutation ctDNA level was then summarised as the total number of error-corrected reads across selected mutations minus the expected error across these positions (rounded down to the nearest whole integer) divided by total clonal deep coverage.

Identifying probable technical negative and low-shedding adenocarcinomas| We generated a linear model where log-10 transformed tumour volume and histology was used to predict log-10 transformed clonal ctDNA level in 101 ctDNA positive NSCLCs analysed in this cohort, we used this model to predict clonal ctDNA levels in 51 evaluable adenocarcinomas negative for ctDNA. We tested the capability of this model to predict clonal mutation levels in 8 independent ctDNA positive adenocarcinomas with volume data available analysed in our prior work using a separate assay[5]. In this test set 6 of 8 (75%) adenocarcinomas evaluated had mean clonal mutation levels above the lower 95% confidence interval of the model prediction. We calculated minimal detectable clonal ctDNA level (MDCF) in the 51 ctDNA negative adenocarcinomas by taking the minimum deep alternate observations needed to make a call in patient cfDNA samples and subtracting the estimated deep alternate reads that would occur due to noise in the panel (rounded down to the nearest whole integer). The resulting number was the number of clonal deep alternate reads needed to make a ctDNA positive call (we conservatively assumed all real deep alternate reads will be clonal). We divided this number by the clonal deep depth across the panel to calculate the minimum clonal ctDNA level that must be exceeded to make a call and called this value MDCF. Using the above linear model, we classified cases as probable technical negatives if the lower 95% confidence interval for predicted clonal ctDNA level was below MDCF and as probable low-shedders if the lower 95% confidence interval for predicted clonal ctDNA level resulting was above MDCF.

Survival analyses| Freedom from recurrence (FFR, events were lung cancer recurrence, patients experiencing second-primary or death were right censored at last follow-up) and Overall survival (OS, death from any cause) survival analyses were performed to assess the prognostic impact of preoperative ctDNA detection in patients with single primary NSCLCs. 186 of 197 patients were evaluable for survival analyses (6 patients were excluded as they died within 30 days of surgery - CRUK0115, 0196, 0223, 0312, 0487, 0681, 5 patients were excluded as

they had confirmed unresected disease after surgery - CRUK0230, 0234, 0291, 0387, 0622). In survival analyses related to preoperative ctDNA detection, patients with synchrnous primaries were also excluded given the emphasis on associations with tumor histology. R packages survival (v3.2-13), survivalAnalysis (v0.3.0) and survminer (v0.4.9) were used to generate hazard ratios, forest plots, 1- and 2-year survival data and cox regression models in the manuscript. Differences in overall survival between relapse dissemination classes were analysed using cox proportional hazard models from either the date of study registration and from the date of MRD detection. A multivariable cox proportion hazard model including maximum relapse ctDNA level, which is known to co-correlate with tumour burden and power for subclone detection, was used to account for this confounder relative to overall survival from the date of study registration.

Code availability statement

ECLIPSE is available as an R package in the zenodo repository here. Upon publication the ECLIPSE github repository, from which this zenodo repository is derived, will be made public.

Code used to produce the figures in this manuscript is available on request.

Data access statement

The de-duplicated and aligned cfDNA sequencing files, RNA-sequencing data and multi-region tumour exome sequencing data (in each case from the TRACERx study), used or analysed during this study are available through the Cancer Research UK & University College London Cancer Trials Centre (ctc.tracerx@ucl.ac.uk) for academic non-commercial research purposes only, and subject to review of a project proposal that will be evaluated by a TRACERx data access committee, entering into an appropriate data access agreement and subject to any applicable ethical approvals.

Contributions

C.A., A.M.F., T.H., J.K. and A.G. contributed equally to this work. C.A., A.M.F., N.J.B, N.M. and C.S. co-wrote the manuscript. C.A., A.M.F., N.J.B, J.S., and C.S. conceived the study design. C.A., J.K., K.G., C.P., D.B., T.C., J.W., C.M.R., M.A.B., O.P., T.B.K.W., E.L., A.H., D.A.M., R.S., F.G., A.P., E.M., D.C., C.H., M.J.H. and N.J.B. integrated clinicopathological data, transcriptomic data, exome data and ctDNA data. C.A., T.H., A.G., A.L., J.S., M.W., K.L., L.J., C.P., and C.S., worked to develop and validate the MRD calling algorithm used in this manuscript. A.M.F., developed ECLIPSE and performed analyses of clonal composition used in this manuscript. A.G., M.M., A.C., L.H., P.R. and B.D conducted AMP-PCR NGS experimental work for ctDNA data. S.V., S.W., N.C., J.R., B.D, M.M, A.C., and J.A.S provided oversight of TRACERx patient sample storage and/or DNA extraction and/or sequencing of patient samples. T.C., J.W., and H.J.W.L.A performed radiomic analyses of baseline computed tomography scans. T.H., M.W., A.G., A.S., A.O., and A.L., conducted ArcherDx variant selection, PSP design and informatic processing of AMP-PCR data. A.M.F., K.G., M.A.B., O.P., T.B.K.W., E.L., A.H., D.C. and N.M., conducted multi-region sequencing and phylogenetic tree analyses and identified TRACERx variants for PSP design. D.A.M. conducted the pathological review. A.L.H, A.G., L.H., P.R., H.B. and N.G.H., designed conducted analytical validation experiments of the AMP-PCR MRD assay. C.A., and T.H., designed and conducted in-silico specificity experiments for the AMP-PCR assay. D.B., and N.J.B., conducted ORACLE analyses. C.A., and T.K., conducted reviews of radiological imaging reports. R.K., D.H., D.S., G.E., and C.B., gave advice on analyses performed in this manuscript. M.J.-H., J.A.S. and C.S. designed the study protocols. A.H. gave statistical advice. C.A., N.M., M.J.H., N.J.B, and C.S. provided overall study oversight. All authors approved the final version of the manuscript.

REFERENCES

1. Chabon, J. J. *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature* **580**, 245–251 (2020).
2. Peng, M. *et al.* Circulating Tumor DNA as a Prognostic Biomarker in Localized Non-small Cell Lung Cancer. *Front. Oncol.* **10**, 561598 (2020).
3. Xia, L. *et al.* Perioperative ctDNA-based Molecular Residual Disease Detection for Non-Small Cell Lung Cancer: A Prospective Multicenter Cohort Study (LUNGCA-1). *Clin. Cancer Res.* (2021) doi:10.1158/1078-0432.CCR-21-3044.
4. Chaudhuri, A. A. *et al.* Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer Discov.* **7**, 1394–1403 (2017).
5. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
6. Gale, D. *et al.* Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer. *Annals of Oncology* (2022) doi:10.1016/j.annonc.2022.02.007.
7. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
8. NICE. Lung cancer: diagnosis and management NICE guideline [NG122]. (2019).
9. Zheng, Z. *et al.* Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).
10. Abbosh, C., Birkbak, N. J. & Swanton, C. Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection. *Nat. Rev. Clin. Oncol.* **15**, 577–586 (2018).
11. Wakelee, H. A. *et al.* Adjuvant chemotherapy with or without bevacizumab in patients with resected non-small-cell lung cancer (E1505): an open-label, multicentre, randomised, phase 3 trial. *Lancet Oncol.* **18**, 1610–1623 (2017).
12. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

13. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
14. Biswas, D. *et al.* A clonal expression biomarker associates with lung cancer mortality. *Nat. Med.* **25**, 1540–1548 (2019).
15. Burrell, R. A. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492–496 (2013).
16. Wang, Z. C. *et al.* Profiles of genomic instability in high-grade serous ovarian cancer predict treatment outcome. *Clin. Cancer Res.* **18**, 5806–5815 (2012).
17. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
18. Shih, D. J. H. *et al.* Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma. *Nat. Genet.* **52**, 371–377 (2020).
19. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
20. Garcia-Murillas, I. *et al.* Assessment of Molecular Relapse Detection in Early-Stage Breast Cancer. *JAMA Oncol* **5**, 1473–1478 (2019).
21. Moding, E. J., Nabet, B. Y., Alizadeh, A. A. & Diehn, M. Detecting Liquid Remnants of Solid Tumors: Circulating Tumor DNA Minimal Residual Disease. *Cancer Discov.* **11**, 2968–2986 (2021).
22. Dentro, S. C. *et al.* Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184**, 2239–2254.e39 (2021).
23. TRACERx consortium. TRACERx 421 primary paper. (2022).
24. Litchfield, K. *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).
25. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).

26. Maise Al Bakir, Ariana Huebner, Carlos Martinez Ruiz, Kristiana Grigoriadis, Thomas B. K. Watkins, Oriol Pich, David A. Moore, Selvaraju Veeriah, Sophia Ward, Joanne Laycock, Diana Johnson, Andrew Rowan, Maryam Razaq, Mita Akther, Cristina Naceur-Lombardelli, Sonya Hessey, Michelle Dietzen, Emma Colliver, Alexander M. Frankell, Emilia Lim, Takahiro Karasaki, Christopher Abbosh, Crispin T. Hiley, Mark S. Hill, Daniel Cook, Gareth Wilson, TRACERx consortium, Allan Hackshaw, Nicolai J. Birkbak, Simone Zaccaria, Mariam Jamal-Hanjani, Nicholas McGranahan,Charles Swanton. TRACERx: The evolution of metastases in non-small cell lung cancer. *Nature in revisions*.

27. Moding, E. J. *et al.* Circulating Tumor DNA Dynamics Predict Benefit from Consolidation Immunotherapy in Locally Advanced Non-Small Cell Lung Cancer. *Nat Cancer* **1**, 176–183 (2020).

28. Chen, K. *et al.* Perioperative Dynamic Changes in Circulating Tumor DNA in Patients with Lung Cancer (DYNAMIC). *Clin. Cancer Res.* **25**, 7058–7067 (2019).

29. Li, N. *et al.* Perioperative circulating tumor DNA as a potential prognostic marker for operable stage I to IIIA non–small cell lung cancer. *Cancer* (2021) doi:10.1002/cncr.33985.

30. Kurtz, D. M. *et al.* Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat. Biotechnol.* **39**, 1537–1547 (2021).

31. Cohen, J. D. *et al.* Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat. Biotechnol.* **39**, 1220–1227 (2021).

32. Zviran, A. *et al.* Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat. Med.* **26**, 1114–1124 (2020).

33. Gydush, G. *et al.* Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. *Nat Biomed Eng* **6**, 257–266 (2022).

34. Parikh, A. R. *et al.* Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat. Med.* **25**, 1415–1421 (2019).

35. Herberts, C. *et al.* Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature* (2022) doi:10.1038/s41586-022-04975-9.

36. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat. Commun.* **6**, 8760 (2015).
37. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16910–16915 (2010).
38. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
39. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
40. Miller, C. A. *et al.* Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* **17**, 880 (2016).
41. Frankell, A. M., Colliver, E., Mcgranahan, N. & Swanton, C. cloneMap: a R package to visualise clonal heterogeneity. *bioRxiv* 2022.07.26.501523 (2022) doi:10.1101/2022.07.26.501523.
42. Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
43. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **44**, e108 (2016).
44. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
45. Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
46. Danaher, P. *et al.* Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer* **5**, 18 (2017).
47. Bentham, R. *et al.* Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* **597**, 555–560 (2021).

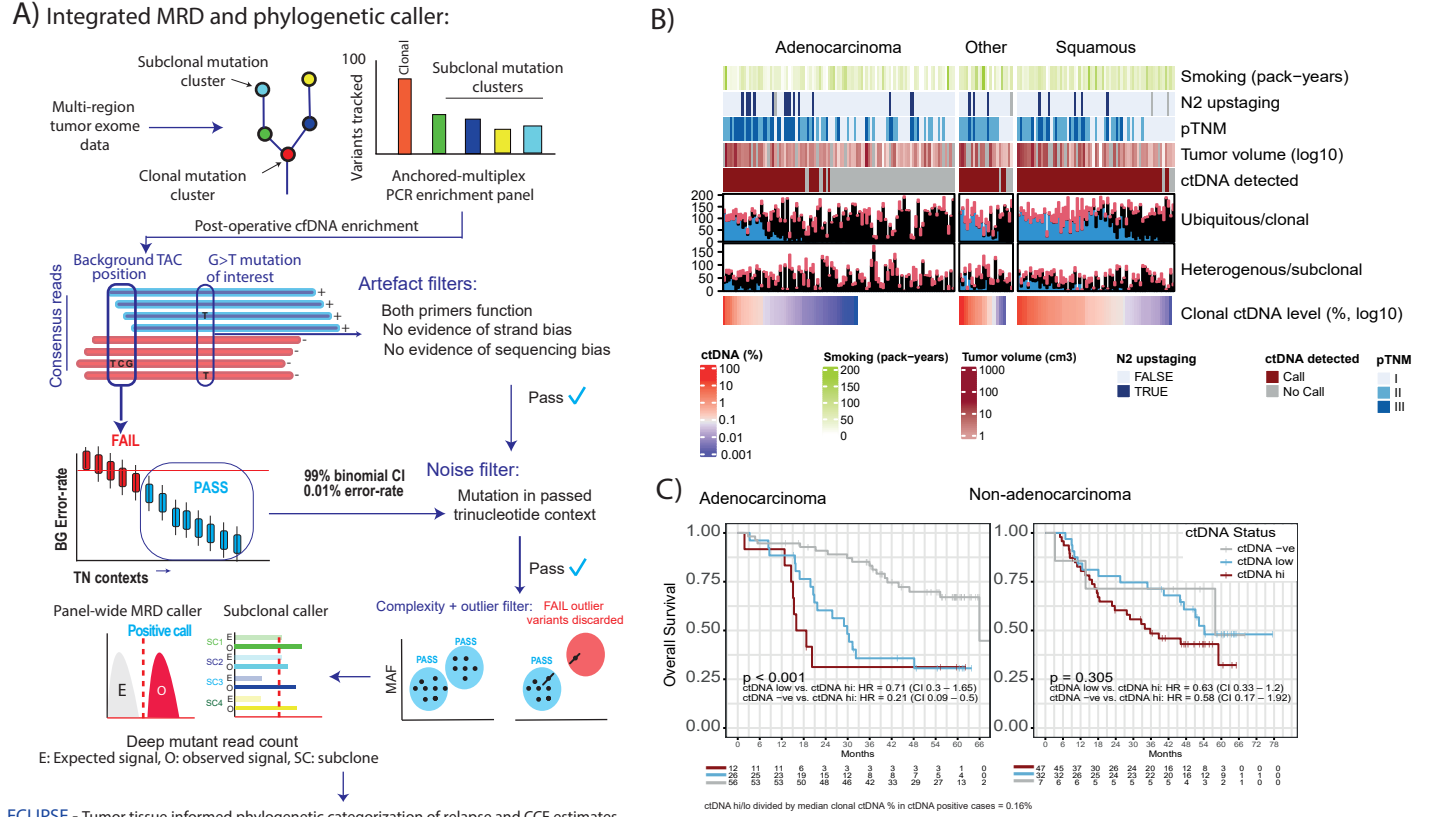

A) Integrated MRD and phylogenetic caller:
Subclonal mutation cluster
Multi-region tumor exome data
Clonal mutation cluster
100
Variants tracked
Clonal
Subclonal mutation clusters
Anchored-multiplex PCR enrichment panel
Post-operative cfDNA enrichment
Consensus reads
Background TAC position
G>T mutation of interest
Artefact filters:
Both primers function
No evidence of strand bias
No evidence of sequencing bias
Pass
FAIL
PASS
BG Error-rate
TN contexts
99% binomial CI
0.01% error-rate
Noise filter:
Mutation in passed trinucleotide context
Pass
Panel-wide MRD caller
Subclonal caller
Positive call
Complexity + outlier filter:
FAIL outlier variants discarded
MAF
Deep mutant read count
E: Expected signal, O: observed signal, SC: subclone
ECLIPSE - Tumor tissue informed phylogenetic categorization of relapse and CCF estimates
Clonal sweep
CCF
Time
Relapse class:
Monoclonal
Polyclonal monophyletic
Polyclonal polyphyletic
CCF: Cancer cell fraction (% of tumor cells corresponding to clone)
B)
Adenocarcinoma
Other
Squamous
Smoking (pack−years)
N2 upstaging
pTNM
Tumor volume (log10)
ctDNA detected
Ubiquitous/clonal
Heterogenous/subclonal
Clonal ctDNA level (%, log10)
ctDNA (%)
Smoking (pack−years)
Tumor volume (cm3)
N2 upstaging
FALSE
TRUE
ctDNA detected
Call
No Call
pTNM
C)
Adenocarcinoma
Non-adenocarcinoma
Overall Survival
ctDNA Status
ctDNA −ve
ctDNA low
ctDNA hi
p < 0.001
ctDNA low vs. ctDNA hi: HR = 0.71 (CI 0.3 – 1.65)
ctDNA −ve vs. ctDNA hi: HR = 0.21 (CI 0.09 – 0.5)
p = 0.305
ctDNA low vs. ctDNA hi: HR = 0.63 (CI 0.33 – 1.2)
ctDNA −ve vs. ctDNA hi: HR = 0.58 (CI 0.17 – 1.92)
Months
ctDNA hi/lo divided by median clonal ctDNA % in ctDNA positive cases = 0.16%

Figure 1| A. ctDNA Analysis Approach: Multi-region tumour exome data is used to generate a phylogenetic representation of a patient's primary tumour where clonal mutations (mutations present in all cancer cells) are represented by the red cluster and subclonal mutations (mutations present in only a subset of cells) are represented by the light blue, dark blue, green and yellow clusters. Bi-directional primers are synthesised to amplify genomic regions containing mutations identified through the multi-region tumour sequencing data with the aim to target mutations from all clusters. Unique molecular identifiers (UMIs) are incorporated into amplified DNA. Only deep consensus reads (reads supported by 5 or more duplicate reads based on UMI-deconvolution) are used to detect ctDNA. The position of interest where a mutation is expected (shown here as a G>T mutation) is then evaluated. The position is considered for ctDNA calling if there is no evidence of strand bias (that is variant bases are noted on both positive and reverse reads), no evidence of sequencing bias (variant bases are noted across both read 1 and read 2 during next-generation sequencing) and both forward and reverse primers are functioning. Alongside evaluating the position of interest the MRD caller evaluates background sequencing positions from the targeted variants and background sequencing positions from primers targeting 45 germline SNP primers (e.g., the TCG region highlighted in the figure). Any base errors observed in background reads are summed based on their trinucleotide context. Trinucleotide contexts with an upper 99% binomial confidence interval less than 0.01% (error rate confidently below 0.01%) are used for ctDNA calling. Passed variants are subsequently evaluated by an outlier filter if there are 3 or less outlier variants these are removed from the ctDNA calling algorithm. For calling the presence or absence of ctDNA (panel-wide MRD caller) all positions targeted by a panel are analysed and a one-sided Poisson test is performed to determine whether the observed number of error-corrected alternate reads deviates from the background error encountered by a panel. For individual subclones a subclonal caller evaluates the presence or absence of ctDNA on a per-subclonal cluster basis, this information is integrated with primary-tumour informed cancer cell fraction (CCF) estimates made through ctDNA analysis at relapse using the ECLIPSE tool (see methods) to categorise the lung cancer relapse process as either monoclonal (a single clone seeding the metastasis), polyclonal monophyletic (multiple clones seeding metastasis from a single branch of the primary tumour phylogenetic tree) or polyclonal polyphyletic (multiple clones seeding metastasis from multiple primary tumour branches of the phylogenetic tree). ECLIPSE is also able to detect the presence of a clonal sweep whereby a subclone achieves 100% CCF at recurrence. **B.** Heatmap summarising preoperative ctDNA analyses in 189 TRACERx patients with a single (non-synchronous) primary tumour diagnosed at enrollment. Cases are split by three histologies, adenocarcinoma, squamous cell carcinoma and other histological subtypes. The top row of the heatmap reflects the smoking pack-year history assigned to that patient. The N2 upstaging row demonstrates patients who were clinically staged with N0 (no lymph node involvement) or N1 (hilar) lymph-node involvement but after surgery were upstaged to N2 disease (mediastinal lymph node involvement, coloured dark blue), patients who did not undergo pathological mediastinal lymph-node examination are coloured grey. The pTNM row describes TNM version 7 tumour stage based on pathology examination of the excised tumour. The tumour volume row represents computed tomography measurements of tumour volume (grey fill indicates tumour volume unevaluable cases). The ctDNA detected row represents whether an MRD positive call was triggered based on the MRD Algorithm P-value cut-off of 0.01 defined in Extended figure 2. The ubiquitous/clonal and heterogenous/subclonal variant bar charts demonstrate the number of mutations in each patient's panel being interrogated, if black the mutation was undetected (based on a per variant P value >0.01), if red the mutation was filtered by the MRD caller and if blue the mutation was detected. The subclone detection (CCF >0.1 in tissue) barchart represents the proportion of subclones with cancer cell fractions more than 10% tracked by ECLIPSE that are detected in preoperative plasma. The clonal ctDNA level per patient is shown in the bottom row of the heatmap. Patients with 0% clonal ctDNA level are given a white color in this row. **C.** Kaplan Meier curves demonstrating overall survival outcomes in ctDNA high (dark red), ctDNA low (blue) and ctDNA negative (grey) single primary adenocarcinoma patients (left) and single primary non-adenocarcinoma patients (right). ctDNA high and low was categorised based on median clonal ctDNA levels across ctDNA positive cases and relates to above and below 0.16%. Log-rank P values are displayed on each plot alongside univariable hazard ratios (HRs) that use ctDNA high patients as the reference category.

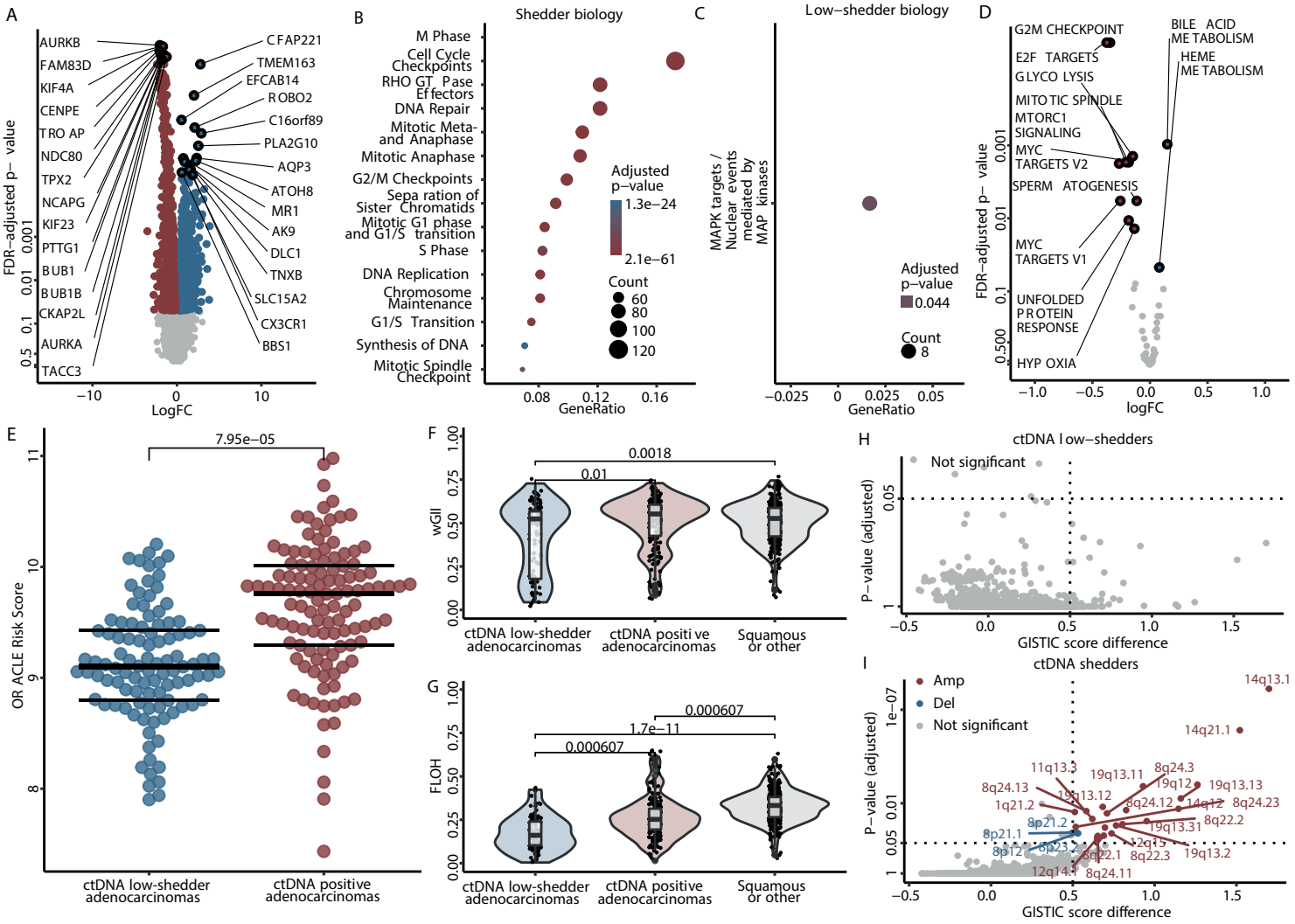

A
AURKB
FAM83D
KIF4A
CENPE
TRO AP
NDC80
TPX2
NCAPG
KIF23
PTTG1
BUB1
BUB1B
CKAP2L
AURKA
TACC3
CFAP221
TMEM163
EFCAB14
ROBO2
C16orf89
PLA2G10
AQP3
ATOH8
MR1
AK9
DLC1
TNXB
SLC15A2
CX3CR1
BBS1
FDR−adjusted p− value
LogFC
B
Shedder biology
M Phase
Cell Cycle Checkpoints
RHO GT Pase Effectors
DNA Repair
Mitotic Meta- and Anaphase
Mitotic Anaphase
G2/M Checkpoints
Sepa ration of Sister Chromatids
Mitotic G1 phase and G1/S transition
S Phase
DNA Replication
Chromosome Maintenance
G1/S Transition
Synthesis of DNA
Mitotic Spindle Checkpoint
Adjusted p−value
1.3e−24
2.1e−61
Count
60
80
100
120
GeneRatio
C
Low−shedder biology
MAPK targets / Nuclear events mediated by MAP kinases
Adjusted p−value
0.044
Count
8
GeneRatio
D
G2M CHECKPOINT
E2F TARGETS
GLYCO LYSIS
MITO TIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V2
SPERM ATOGENESIS
MYC TARGETS V1
UNFOLDED PR OTEIN RESPONSE
HYP OXIA
BILE ACID ME TABOLISM
HEME ME TABOLISM
FDR−adjusted p− value
logFC
E
7.95e−05
OR ACLE Risk Score
ctDNA low-shedder adenocarcinomas
ctDNA positive adenocarcinomas
F
0.0018
0.01
wGII
ctDNA low-shedder adenocarcinomas
ctDNA positi ve adenocarcinomas
Squamous or other
G
0.000607
1.7e−11
0.000607
FLOH
ctDNA low-shedder adenocarcinomas
ctDNA positive adenocarcinomas
Squamous or other
H
ctDNA l ow−shedders
Not significant
P−value (adjusted)
GISTIC score difference
I
ctDNA shedders
Amp
Del
Not significant
14q13.1
14q21.1
11q13.3
19q13.11
8q24.3
8q24.13
19q12
19q13.13
19q13.12
1q21.2
8q24.12
14q12
8q24.23
8p21.2
8q22.2
19q13.31
8p21.1
8p12
8p23.2
12q15
8q22.1
8q22.3
19q13.2
12q14.1
8q24.11
P−value (adjusted)
GISTIC score difference

Figure 2| Genomic and transcriptomic predictors of ctDNA detection in early-stage NSCLC. A. Differential gene expression analysis comparing 106 tumour regions from 35 ctDNA positive adenocarcinomas to 74 tumour regions from 32 ctDNA negative low-shedder adenocarcinomas X-axis shows fold change as difference in means on log2 scale (LogFC). Red indicates significantly over-expressed in ctDNA positive adenocarcinomas (n = 932 genes), blue indicates significantly over-expressed in ctDNA negative low-shedder adenocarcinomas with technical non-shedders excluded (n = 939 genes). Top 15 differentially expressed genes are labelled. B and C. Reactome pathway enrichment analysis based on the 1,871 significant genes found in A) (n = 932 for ctDNA positives and n = 939 for ctDNA negatives). X-axis shows the proportion of genes involved in the pathway, Y-axis notes the significantly enriched pathways. **B)** shows the top 15 enriched pathways in ctDNA positive patients, **C)** displays the only significantly enriched pathway in ctDNA negative patients. Size of points represents the number of genes included and the colour shows the FDR adjusted p-value. **D.** Differential enrichment analysis based on the Hallmark gene-sets. Samples, axes and colours follow A). All significantly enriched pathways are labelled. **E.** ORACLE gene expression scores in ctDNA positive (n = 115 tumour and lymph node regions sampled from 35 patients) versus ctDNA negative (n = 101 tumour and lymph node regions sampled from 46 patients) adenocarcinomas. All data points are displayed. Centre lines represent the medians. **F, G.** Violin-boxplots showing the wGII (F) and FLOH (G) levels of ctDNA positive adenocarcinomas (37 patients, 177 tumour and lymph node regions), ctDNA negative low-shedder adenocarcinomas (33 patients, 97 tumour regions) and squamous or other carcinomas (79 patients, 322 tumour and lymph node regions). All data points are displayed. Boxplot centre lines represent the median, p-values are FDR-adjusted. **H, I.** GISTIC score analysis. Red indicates amplifications, blue indicates deletions, grey indicates non-significant values. Y-axis shows the adjusted p-value and X-axis shows the GISTIC score difference. All significant alterations are labelled. Alterations are considered significant when the GISTIC score difference exceeds 0.5 and the adjusted p-value is below 0.05, cutoffs are denoted with dotted lines.

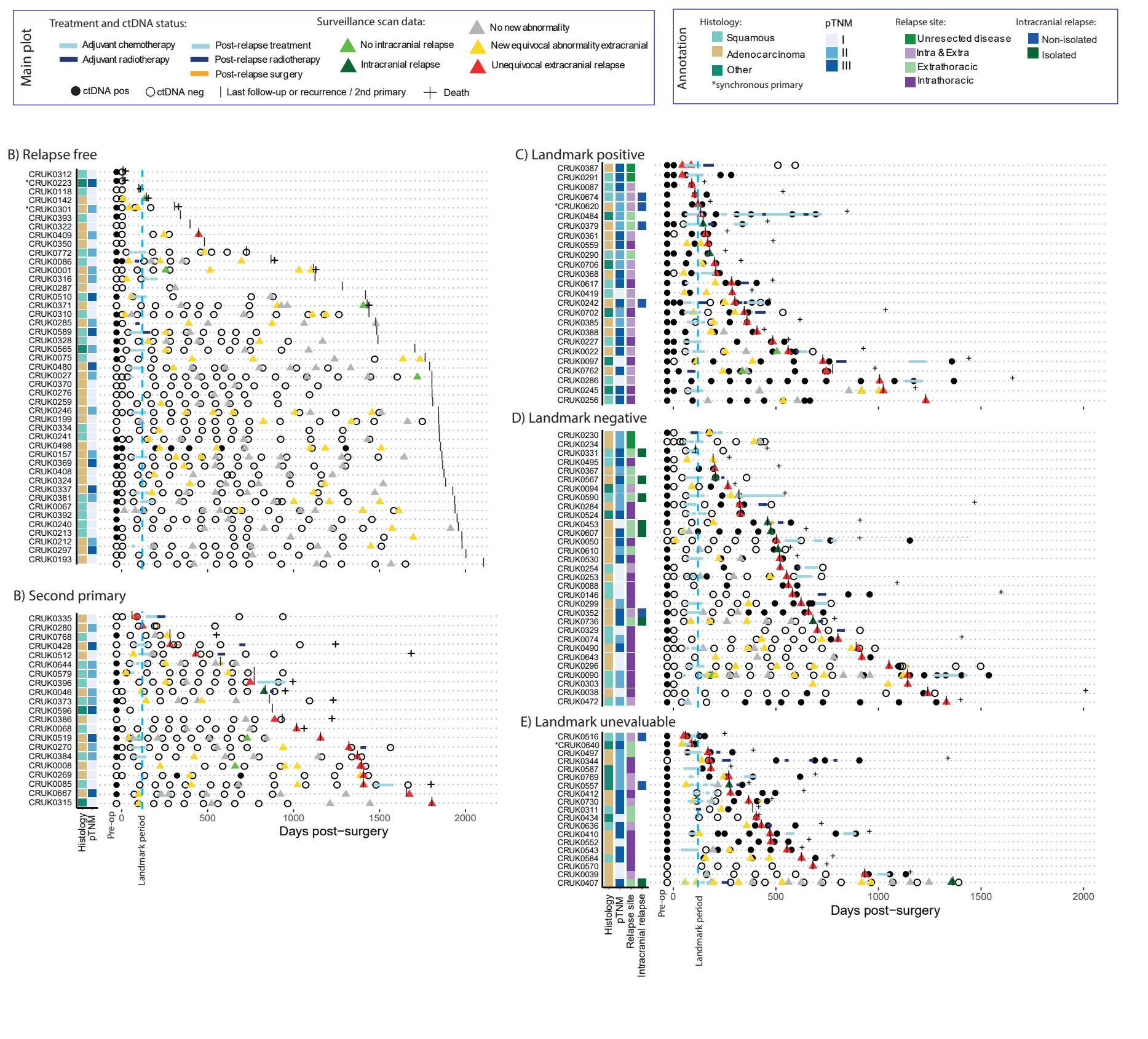

Main plot
Treatment and ctDNA status:
Adjuvant chemotherapy
Adjuvant radiotherapy
Post-relapse treatment
Post-relapse radiotherapy
Post-relapse surgery
ctDNA pos
ctDNA neg
Last follow-up or recurrence / 2nd primary
Death
Surveillance scan data:
No intracranial relapse
Intracranial relapse
No new abnormality
New equivocal abnormality extracranial
Unequivocal extracranial relapse
Annotation
Histology:
Squamous
Adenocarcinoma
Other
*synchronous primary
pTNM
I
II
III
Relapse site:
Unresected disease
Intra & Extra
Extrathoracic
Intrathoracic
Intracranial relapse:
Non-isolated
Isolated
B) Relapse free
CRUK0312
*CRUK0223
CRUK0118
CRUK0142
*CRUK0301
CRUK0393
CRUK0322
CRUK0409
CRUK0350
CRUK0772
CRUK0086
CRUK0001
CRUK0316
CRUK0287
CRUK0510
CRUK0371
CRUK0310
CRUK0285
CRUK0589
CRUK0328
CRUK0565
CRUK0075
CRUK0480
CRUK0027
CRUK0370
CRUK0276
CRUK0259
CRUK0246
CRUK0199
CRUK0334
CRUK0241
CRUK0498
CRUK0157
CRUK0369
CRUK0408
CRUK0324
CRUK0337
CRUK0381
CRUK0067
CRUK0392
CRUK0240
CRUK0213
CRUK0212
CRUK0297
CRUK0193
B) Second primary
CRUK0335
CRUK0280
CRUK0768
CRUK0428
CRUK0512
CRUK0644
CRUK0579
CRUK0396
CRUK0046
CRUK0373
CRUK0596
CRUK0386
CRUK0068
CRUK0519
CRUK0270
CRUK0384
CRUK0008
CRUK0269
CRUK0085
CRUK0667
CRUK0315
Histology
pTNM
Pre-op
Landmark period
0
500
1000
1500
2000
Days post−surgery
C) Landmark positive
CRUK0387
CRUK0291
CRUK0087
CRUK0674
*CRUK0620
CRUK0484
CRUK0379
CRUK0361
CRUK0559
CRUK0290
CRUK0706
CRUK0368
CRUK0617
CRUK0419
CRUK0242
CRUK0702
CRUK0385
CRUK0388
CRUK0227
CRUK0022
CRUK0097
CRUK0762
CRUK0286
CRUK0245
CRUK0256
D) Landmark negative
CRUK0230
CRUK0234
CRUK0331
CRUK0495
CRUK0367
CRUK0567
CRUK0094
CRUK0590
CRUK0284
CRUK0524
CRUK0453
CRUK0607
CRUK0050
CRUK0610
CRUK0530
CRUK0254
CRUK0253
CRUK0088
CRUK0146
CRUK0299
CRUK0352
CRUK0736
CRUK0329
CRUK0074
CRUK0490
CRUK0643
CRUK0296
CRUK0090
CRUK0303
CRUK0038
CRUK0472
E) Landmark unevaluable
CRUK0516
*CRUK0640
CRUK0497
CRUK0344
CRUK0587
CRUK0769
CRUK0557
CRUK0412
CRUK0730
CRUK0311
CRUK0434
CRUK0636
CRUK0410
CRUK0552
CRUK0543
CRUK0584
CRUK0570
CRUK0039
CRUK0407
Histology
pTNM
Relapse site
Intracranial relapse
Pre-op
Landmark period
0
500
1000
1500
2000
Days post−surgery

Figure 3| Postoperative Minimal Residual Disease detection in early-stage NSCLC. A-E. Longitudinal data from patients with (A) no evidence of non-small-cell lung cancer (NSCLC) recurrence, n=45 and 371 pre- and postoperative plasma samples; (B) development of a second-primary cancer, n=21 and 176 pre- and postoperative plasma samples; (C) recurrence of NSCLC in landmark positive patients, n=25 patients and 138 pre- and postoperative plasma samples (D) recurrence of NSCLC in landmark negative patients, n=31 patients and 227 pre- and postoperative plasma samples and (E) recurrence of NSCLC in landmark unevaluable patients, n=19 patients and 101 pre- and postoperative plasma samples. and In all plots each circle represents a cfDNA sampling time point. The circles to the left of day-0 are preoperative timepoints from when the patient's tumour was still in-situ. The circles to the right of day-0 are taken following surgical excision of the primary NSCLC. If the circle is colored dark it reflects positive ctDNA detection. The light blue rectangles represent whether a patient received chemotherapy, the dark blue rectangles represent whether a patient received radiotherapy and the orange rectangles if a patient received post-recurrence surgery. The triangles represent standard of care postoperative CT, PET or MRI imaging classified as no disease (grey) equivocal images (yellow) or unequivocal imaging evidence of extracranial relapse (red). Light green triangles represent no evidence of intracranial relapse, dark green triangles indicate intracranial relapse. The vertical black lines represent the event date for a patient (if events such as death, second-primary, NSCLC recurrence occurred) otherwise the vertical line represents the TRACERx follow-up censorship date for that patient. Crosses represent patient death events. To the left of the panels the annotation plots highlight histology, pTNM (pathological tumour node metastasis) status, relapse site and details regarding whether an intracranial relapse was isolated (brain-only) or non-isolated (brain and extracranial site).

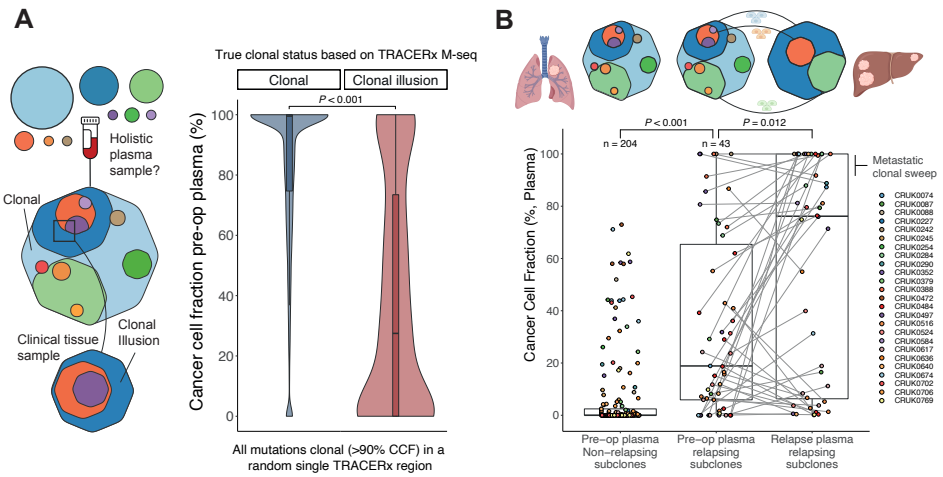
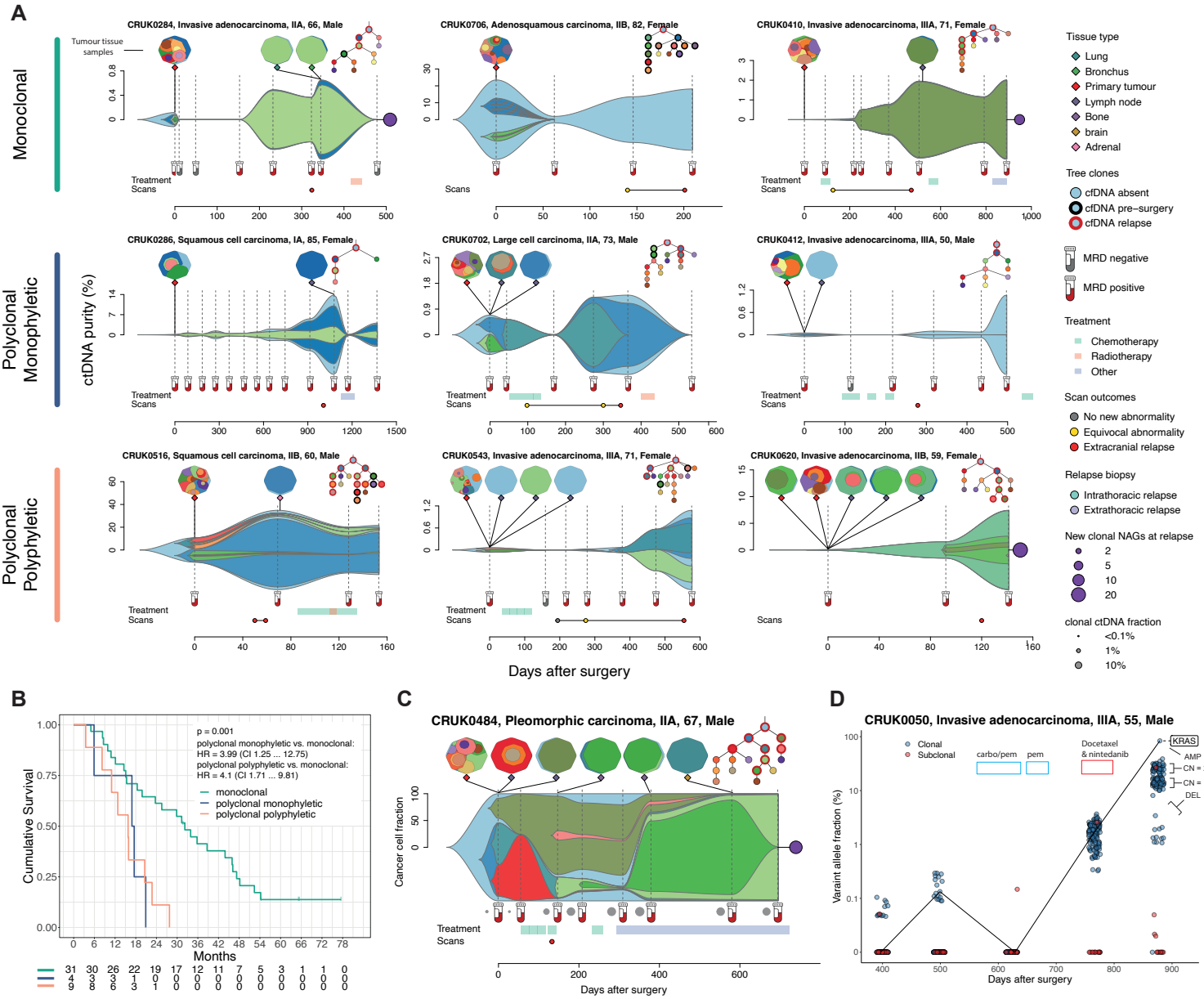

A
Holistic plasma sample?
Clonal
Clinical tissue sample
Clonal Illusion
True clonal status based on TRACERx M-seq
Clonal
Clonal illusion
P < 0.001
Cancer cell fraction pre-op plasma (%)
All mutations clonal (>90% CCF) in a random single TRACERx region
B
P < 0.001
P = 0.012
n = 204
n = 43
Cancer Cell Fraction (%, Plasma)
Metastatic clonal sweep
Pre−op plasma Non−relapsing subclones
Pre−op plasma relapsing subclones
Relapse plasma relapsing subclones

Figure 4| Clonality measurements in preoperative plasma overcome sampling bias from a single tissue sample and predict metastatic seeding potential. A. Depiction of a clonal illusion where a dark blue subclone is found in 100% of cells in a single clinical tissue sample. Such clonal illusion mutations may be detected in a clinical setting using ctDNA derived from many different tumour regions to increase accuracy of ITH measurements in the clinic. Mutations which were clonal (CCF > 90%) in a single, randomly selected tumour region are compared using plasma-based preoperative CCFs splitting by those truly clonal across all tumour regions in TRACERx (clonal) and those which, while they were clonal in the randomly selected region, were absent from other tumour regions (Clonal illusion). Only data from a single randomly selected region was used by ECLIPSE to generate these CCFs. Only preoperative samples with at least 0.1% clonal ctDNA level (high subclone sensitivity samples) were included (N=71**) B.** Preoperative plasma subclonal CCFs split by whether a given subclone was found to be present or absent in cfDNA samples at relapse and postoperative plasma CCFs for relapse subclones at the last high subclone sensitivity timepoint. Only tumours with at least one sample >0.1% clonal ctDNA level (high subclone sensitivity) both preoperatively and postoperatively were included (N=26 tumours with CCFs from 247 subclones included).


A
Monoclonal
Polyclonal Monophyletic
Polyclonal Polyphyletic
ctDNA purity (%)
Tumour tissue samples
CRUK0284, Invasive adenocarcinoma, IIA, 66, Male
CRUK0706, Adenosquamous carcinoma, IIB, 82, Female
CRUK0410, Invasive adenocarcinoma, IIIA, 71, Female
CRUK0286, Squamous cell carcinoma, IA, 85, Female
CRUK0702, Large cell carcinoma, IIA, 73, Male
CRUK0412, Invasive adenocarcinoma, IIIA, 50, Male
CRUK0516, Squamous cell carcinoma, IIB, 60, Male
CRUK0543, Invasive adenocarcinoma, IIIA, 71, Female
CRUK0620, Invasive adenocarcinoma, IIB, 59, Female
Treatment
Scans
Days after surgery
Tissue type
Lung
Bronchus
Primary tumour
Lymph node
Bone
brain
Adrenal
Tree clones
cfDNA absent
cfDNA pre−surgery
cfDNA relapse
MRD negative
MRD positive
Treatment
Chemotherapy
Radiotherapy
Other
Scan outcomes
No new abnormality
Equivocal abnormality
Extracranial relapse
Relapse biopsy
Intrathoracic relapse
Extrathoracic relapse
New clonal NAGs at relapse
2
5
10
20
clonal ctDNA fraction
<0.1%
1%
10%

B
Cumulative Survival
Months
p = 0.001
polyclonal monophyletic vs. monoclonal:
HR = 3.99 (CI 1.25 ... 12.75)
polyclonal polyphyletic vs. monoclonal:
HR = 4.1 (CI 1.71 ... 9.81)
monoclonal
polyclonal monophyletic
polyclonal polyphyletic

C
CRUK0484, Pleomorphic carcinoma, IIA, 67, Male
Cancer cell fraction
Treatment
Scans
Days after surgery

D
CRUK0050, Invasive adenocarcinoma, IIIA, 55, Male
Clonal
Subclonal
carbo/pem
pem
Docetaxel & nintedanib
KRAS
AMP
CN = 2
CN = 1
DEL
Varaint allele fraction (%)
Days after surgery

Figure 5| Longitudinal measurements of clonal evolution in plasma from surgery, through therapy and to recurrence. ctDNA purity for each clone is calculated by multiplying the clone CCF by the ctDNA purity of the plasma sample (methods) and represents the fraction of all cells from which cfDNA was derived which harbour a given tumour clone at each timepoint. Clonal nesting is based on the phylogenetic tree for each tumour. Clone maps for each tumour tissue mass are depicted above the ctDNA based clonal structure with the phylogenetic tree**. A.** Depictions of longitudinal tumour evolution for examples of monoclonal, polyclonal monophyletic and polyclonal polyphyletic metastatic dissemination patterns. **B.** A kaplan meier plot depicting differences in overall survival between metastatic dissemination classes (N= 44 tumours with at least 1 high subclone sensitivity sample). **C.** CCFs depicted through time and therapy for CRUK0484 who had a polyclonal polyphyletic relapse**. D.** Variant allele fractions for mutations tracked in CRUK0050 at recurrence. NAG = Neoantigen.