1 **Phylogenetic ctDNA analysis depicts early stage lung cancer evolution**

2

3 Christopher Abbosh[1*], Nicolai J. Birkbak[1,2*], Gareth A. Wilson[1,2*], Mariam Jamal-Hanjani[1*],

4 Tudor Constantin[3*], Raheleh Salari[3*], John Le Quesne[4*], David A Moore[4+], Selvaraju

5 Veeriah[1+], Rachel Rosenthal[1+], Teresa Marafioti[1,5], Eser Kirkizlar[3], Thomas B K Watkins[1,2],

6 Nicholas McGranahan[1,2], Sophia Ward[1,2,6], Luke Martinson[4], Joan Riley[4], Francesco Fraioli[7],

7 Maise Al Bakir[2], Eva Gronroos[2], Francisco Zambrana[1], Raymondo Endozo[7], Wenya Linda

8 Bi[8,9], Fiona M. Fennessy[8,9], Nicole Sponer[3], Diana Johnson[1], Joanne Laycock[1], Seema Shafi[1],

9 Justyna Czyzewska-Khan[1], Andrew Rowan[2], Tim Chambers[2,6], Nik Matthews[6,10], Samra

10 Turajlic[2,11], Crispin Hiley1[2], Siow Ming Lee[12,1], Martin Forster[1,12], Tanya Ahmad[12], Mary

11 Falzon[5], Elaine Borg[5], David Lawrence[13], Martin Hayward[13], Shyam Kolvekar[13], Nikolaos

12 Panagiotopoulos[13], Sam M Janes[1,14,15], Ricky Thakrar[14], Asia Ahmed[16], Fiona Blackhall[17,18],

13 Yvonne Summers[18], Dina Hafez[3], Ashwini Naik[3], Apratim Ganguly[3], Stephanie Kareht[3],

14 Rajesh Shah[19], Leena Joseph[20], Anne Marie Quinn[20], Phil Crosbie[21], Babu Naidu[22], Gary

15 Middleton[23], Gerald Langman[24], Simon Trotter[24], Marianne Nicolson[25], Hardy Remmen[26],

16 Keith Kerr[27], Mahendran Chetty[28], Lesley Gomersall[29], Dean Fennell[4], Apostolos Nakas[30],

17 Sridhar Rathinam[30], Girija Anand[31], Sajid Khan[32,33], Peter Russell[34], Veni Ezhil[35], Babikir

18 Ismail[36], Melanie Irvin-sellers[37], Vineet Prakash[38], Jason Lester[39], Malgorzata

19 Kornaszewska[40], Richard Attanoos[41], Haydn Adams[42], Helen Davies[43], Dahmane Oukrif[1],

20 Ayse U Akarca[1], John A Hartley[44], Helen L Lowe[44], Sara Lock[45], Natasha Iles[46], Harriet Bell[46],

21 Yenting Ngai[46], Greg Elgar[2,6], Zoltan Szallasi[47,48,49], Roland F Schwarz[50], Javier Herrero[51],

22 Aengus Stewart[52], Sergio A Quezada[53], Peter Van Loo[54,55], Caroline Dive[56], Jimmy Lin[3],

23 Matthew Rabinowitz[3], Hugo JWL Aerts[8,9,57], Allan Hackshaw[45], Jacqui A Shaw[4], Bernhard G.

24 Zimmermann[3], and Charles Swanton[1,2] on behalf of the TRACERx and PEACE consortia.

25 *These authors contributed equally to this work +These authors contributed equally to this work

1. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT

2. Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT

3. Natera Inc., 201 Industrial Rd., San Carlos, United States, CA 94070

4. Cancer Studies, University of Leicester, Leicester, United Kingdom, LE2 7LX

5. Department of Pathology, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

6. Advanced Sequencing Facility, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT

7. Department of Nuclear Medicine, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

8. Brigham and Women's Hospital, Boston, MA 02115, USA

9. Harvard Medical School, Boston, MA 02115, USA

10. Tumour Profiling Unit Genomics Facility, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB

11. Renal and Skin Units, The Royal Marsden Hospital, London, SW3 6JJ

12. Department of Oncology, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

13. Department of Cardiothoracic Surgery, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

14. Department of Respiratory Medicine, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

15. Lungs for Living Research Centre. Division of Medicine, Rayne Building. University College London, 5 University Street. London. WC1E 6JF

16. Department of Radiology, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

17. Institute of Cancer Studies, University of Manchester, Oxford Road, Manchester, M13 9PL

18. The Christie Hospital, Manchester, United Kingdom, M20 4BX

19. Department of Cardiothoracic Surgery, University Hospitals of South Manchester, Manchester, M23 9LT

20. Department of Pathology, University Hospitals of South Manchester, Manchester, M23 9LT

21. North West Lung Centre, University Hospital of South Manchester, Manchester, United Kingdom, M23 9LT

22. Department of Thoracic Surgery, Birmingham Heartlands Hospital, Birmingham, United Kingdom, B9 5SS

23. Department of Medical Oncology, Birmingham Heartlands Hospital, Birmingham, United Kingdom, B9 5SS

24. Department of Cellular Pathology, Birmingham Heartlands Hospital, Birmingham, United Kingdom, B9 5SS

25. Department of Medical Oncology, Aberdeen University Medical School & Aberdeen Royal Infirmary, Aberdeen, Scotland, United Kingdom, AB25 2ZN

26. Department of Cardiothoracic Surgery, Aberdeen University Medical School & Aberdeen Royal Infirmary, Aberdeen, United Kingdom, AB25 2ZD

27. Department of Pathology, Aberdeen University Medical School & Aberdeen Royal Infirmary, Aberdeen, Scotland, United Kingdom, AB25 2ZD

28. Department of Respiratory Medicine, Aberdeen University Medical School & Aberdeen Royal Infirmary, Aberdeen, United Kingdom, AB25 2ZN

29. Department of Radiology, Aberdeen University Medical School & Aberdeen Royal Infirmary, Aberdeen, Scotland, United Kingdom, AB25 2ZN

30. Department of Thoracic Surgery, Glenfield Hospital, Leicester, LE3 9QP

31. Department of Radiotherapy, North Middlesex University Hospital, London N18 1QX

32. Department of Respiratory Medicine, Royal Free Hospital, Pond Street, London, NW3 2QG

33. Department of Respiratory Medicine, Barnet and Chase Farm Hospitals, Wellhouse Lane, Barnet, United Kingdom, EN5 3DJ

34. Department of Respiratory Medicine, The Princess Alexandra Hospital, Hamstel Rd, Harlow CM20 1QX

35. Department of Clinical Oncology, St.Luke's Cancer Centre, Royal Surrey County Hospital, Guildford, GU2 7XX

36. Department of Pathology, Ashford and St. Peters' Hospital, Guildford Road, Chertsey, Surrey, KT16 0PZ

37. Department of Respiratory Medicine, Ashford and St. Peters' Hospital, Guildford Road, Chertsey, Surrey, KT16 0PZ

38. Department of Radiology, Ashford and St. Peters' Hospital, Guildford Road, Chertsey, Surrey, KT16 0PZ

39. Department of Clinical Oncology, Velindre Hospital, Cardiff, Wales, United Kingdom, CF14 2TL

40. Department of Cardiothoracic Surgery, University Hospital Llandough, Cardiff, Wales, United Kingdom, CF64 2XX

41. Department of Pathology, University Hospital Llandough, Cardiff, Wales, United Kingdom, CF64 2XX

42. Department of Radiology, University Hospital Llandough, Cardiff, Wales, United Kingdom, CF64 2XX

43. Department of Respiratory Medicine, University Hospital Llandough, Cardiff, Wales, United Kingdom, CF64 2XX

44. UCL ECMC GCLP Facility, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT

45. Department of Respiratory Medicine, The Whittington Hospital NHS Trust, United Kingdom, N19 5NF

46. University College London, Cancer Research UK & UCL Cancer Trials Centre, London, United Kingdom, W1T 4TJ

47. Centre for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark.

48. Computational Health Informatics Program (CHIP), Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

49. MTA-SE-NAP, Brain Metastasis Research Group, 2nd Department of Pathology, Semmelweis University, 1091 Budapest, Hungary.

50. Berlin Institute for Medical Systems Biology, Max Delbrueck Center for Molecular Medicine, Berlin, Germany

51. Bill Lyons Informatics Centre, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT

52. Department of Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT

53. Cancer Immunology Unit, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT

124   54. Cancer Genomics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1

125   1AT

126   55. Department of Human Genetics, University of Leuven, B-3000 Leuven, Belgium

127   56. Cancer Research UK Manchester Institute, Manchester, United Kingdom, M20 4BX

128   57. Dana-Farber Cancer Institute, 450 Brookline Ave. Boston, United States, MA 02215-

129   5450

130

131   Corresponding author:

132   Charles Swanton

133   Translational Cancer Therapeutics Laboratory

134   The Francis Crick Institute

135   3rd Floor South West

136   1 Midland Road

137   London

138   NW1 1A

139   Email: Charles.Swanton@crick.ac.uk

140   Office +44 203 796 2047

141 **Summary (156 words)**

142 Earlier detection of relapse following primary surgery for non-small cell lung cancer and the

143 characterization of emerging subclones seeding metastatic sites might offer new therapeutic

144 approaches to limit tumor recurrence. The potential to non-invasively track tumor evolutionary

145 dynamics in ctDNA of early-stage lung cancer is not established. Here we conduct a patient-

146 specific approach to ctDNA profiling in the first 100 lung TRACERx (**TRA**cking **C**ancer

147 **E**volution through therapy (**Rx**)) study participants, including one patient co-recruited to the

148 PEACE (**P**osthumous **E**valuation of **A**dvanced **C**ancer **E**nvironment) post-mortem study. We

149 identify independent predictors of ctDNA release in early-stage non-small cell lung cancer and

150 perform tumor volume limit of detection analyses. Through blinded profiling of post-operative

151 plasma, we observe evidence of adjuvant chemotherapy resistance and identify patients

152 destined to experience recurrence of their lung cancer. Finally, we show that phylogenetic

153 ctDNA profiling tracks the subclonal nature of lung cancer relapse and metastases, providing

154 a new approach for ctDNA driven therapeutic studies

155

156

Main text

Lung cancer is the leading cause of cancer death worldwide[1-2]. Established metastatic non-small cell lung cancer (NSCLC) cannot be cured with systemic chemotherapy. Yet clinical studies have shown a 5% benefit of post-operative (adjuvant) chemotherapy on overall survival[3]. This modest survival benefit may reflect a vulnerability of treating low volume disease within the context of reduced intra-tumor heterogeneity[4]. Improving adjuvant treatment of lung cancer could improve cure rates. However, achieving this objective will require the development of a diagnostic platform capable of identifying, monitoring and genomically characterizing recurring or residual disease early. This would create a therapeutic setting where only patients destined to recur would receive treatment, where intervention could be directed to the evolving tumor subclone seeding metastatic recurrence guided by clinical trials powered to determine treatment effect within smaller patient cohorts.

Circulating tumor DNA (ctDNA) detection in plasma has been shown in breast[5,6] and colorectal cancer[7] to detect minimal residual disease in the adjuvant setting and identify patients destined to relapse post-operatively in advance of established clinical parameters. Here, we report a bespoke multiplex-PCR NGS approach to ctDNA profiling within the context of the

173 prospective tumor evolutionary NSCLC study TRACERx. We address determinants of ctDNA

174 detection in early-stage NSCLC and investigate the ability of ctDNA to identify and

175 genomically characterize, at subclone resolution, post-operative NSCLC relapse using a tumor

176 phylogenetic framework.

Phylogenetic ctDNA profiling

178 The TRACERx study monitors the clonal evolution of NSCLC from diagnosis through to

179 relapse and death[8]. Using multi-region exome sequencing (M-Seq) derived tumor phylogenetic

180 trees developed through prospective analysis of the 100 patient TRACERx cohort, we

181 conducted a phylogenetic approach to ctDNA profiling in early stage NSCLC **(Fig. 1).** Bespoke

182 multiplex-PCR assay-panels were synthesised for each patient, targeting clonal and subclonal

183 single nucleotide variants (SNVs) selected to track phylogenetic tumor branches in plasma **(Fig**

184 **1.)**. Analytical validation of the multiplex-PCR NGS platform demonstrated a sensitivity of

185 above 99% for the detection of SNVs at frequencies above 0.1% and the specificity of detecting

186 a single SNV was 99.6% (**Extended Data Fig 1a**). At least two SNVs were detected in ctDNA

187 from each NSCLC analyzed in our published discovery cohort data[9], demonstrating biological

188 sensitivity of a two SNV threshold for ctDNA detection in early-stage NSCLC. Therefore, we

189   prospectively selected a threshold of two detected SNVs for calling a sample ctDNA positive

190   for validation within this study - to minimize type I error in a platform testing up to 30 tumour-

191   specific SNVs per time-point in a single patient (see **Extended Data Fig 1b** for justificity**)**.

192   Cross-platform validation was performed in 28 patients with M-Seq confirmed SNV(s) within

193   one or more hotspots targeted by a generic multiplex PCR-NGS panel **(Extended Table 1a-b,**

194   **Supplementary Table 1).** All 18 bespoke-panel ctDNA negative patients had no tumor SNVs

195   detectable in plasma pre-operatively by the generic panel supporting biological specificity of

196   the targeted approach, 7 of 10 bespoke-panel ctDNA positive patients had tumor SNVs

197   detected in plasma by the generic panel **(Extended Table 1a-b).**

198   **Determinants of ctDNA detection in NSCLC**

199   We sought to identify clinicopathological determinants of ctDNA detection in early-stage

200   NSCLC by profiling pre-operative plasma samples in 96 of 100 TRACERx patients (cohort,

201   sample characteristics **Extended Table 2a-c, Supplementary Table 2**). It was not possible to

202   analyze samples from four patients (see **Extended Data Fig 2a** for details**)**. Individual patient

203   assay-panels were designed to target a median of 18 SNVs (range 10 to 22) comprising a

204   median of 11 clonal SNVs (range 2 to 20) and a median of 6 subclonal SNVs (range 0 to 16)

205   **(Extended Data Fig 2b)**.

206   At least two SNVs were detected in ctDNA pre-operatively in 46 of 96 (48%) early-stage

207   NSCLCs, a single SNV was detected in 12 additional cases **(Fig 2a)**. Centrally reviewed

208   pathological data revealed that ctDNA detection was associated with histological subtype: 97%

209   (30/31) of lung squamous cell carcinomas (LUSCs) and 71% (5 of 7) of other NSCLC subtypes

210   were ctDNA positive, compared with 19% (11/58) of lung adenocarcinomas (LUADs) **(Fig**

211   **2a).** ctDNA detection stratified by TNM stage revealed that 94% (16 of 17) of stage I LUSCs

212   were detected compared with 13% (5 of 39) of stage I LUADs **(Extended Data Fig 3a)**.

213   Passive release of ctDNA into the circulation may be associated with apoptosis and necrosis[10].

214   As expected[11], LUSCs were significantly more necrotic than LUADs and ctDNA positive

215   LUADs formed a sub-group of more necrotic tumors compared with ctDNA negative LUADs

216   **(Extended Data Fig 3b)**. Necrosis, lymph node involvement, lymphovascular invasion,

217   pathological tumor size, Ki67 labelling indices, non-adenocarcinoma histology and total cell-

218   free DNA input predicted ctDNA detection in univariable analyses (**Extended Data Fig 3c**).

219   Multivariable analysis revealed non-adenocarcinoma histology, the presence of lympho-

220   vascular invasion and high Ki67 proliferation index as independent predictors of ctDNA

221 detection (**Extended Data Fig 3c**). Since FDG-avidity on positron emission tomography (PET)

222 scans correlates with proliferative indices in early-stage NSCLC[12,13], we investigated tumor

223 PET FDG-avidity and ctDNA detection. PET FDG-avidity predicted ctDNA detection (area

224 under curve = 0.84, P<0.001, n=92) (**Extended Data Fig 3d**). Within LUADs, common driver

225 events in *KRAS*, *EGFR* or *TP53* were not associated with ctDNA detection (**Extended Data**

226 **Fig 3e).**

227 We analyzed the distribution of clonal and subclonal SNVs in ctDNA positive patients. Clonal

228 SNVs were detected in all 46 ctDNA positive patients; a median of 94% (range 11% to 100%)

229 of clonal SNVs targeted by assay-panels were detected in ctDNA. 40 of 46 ctDNA positive

230 patients had subclonal SNVs targeted by assay-panels and subclonal SNVs were detected in 27

231 (68%) of these patients. A median of 27% (range 0% to 91%) of targeted subclonal SNVs were

232 detected in ctDNA positive patients (**Figure 2b**). The mean plasma variant allele frequency

233 (VAF) of clonal SNVs was significantly higher than that of subclonal SNVs (**Extended Data**

234 **Fig 4a)** (within patient comparison, Wilcoxon signed-rank test, P<0.001, n=27,

235 **Supplementary Table 3**) supporting the use of clonal alterations as a more sensitive method

236 of ctDNA detection than subclonal alterations[9,14].

237   In ctDNA positive patients, macroscopic tumor size correlated with mean clonal plasma VAF

238   (Spearman's Rho = 0.405, P=0.005, n=46) (**Extended Data Fig 4b**). CT scan volumetric

239   analyses were available in 38 of 46 ctDNA positive patients (see **Extended Data Fig 4c**).

240   Tumor volume correlated with mean clonal plasma VAF (**Fig 3a,** Spearman's Rho = 0.61,

241   P<0.001, n=38). A linear relationship between log- transformed volume and mean clonal VAF

242   values was observed (**Fig 3a**). The line of best fit applied to our data was consistent with the

243   line fitted to NSCLC volumetric data and ctDNA plasma VAFs reported in previously

244   published work[15] (**Extended Data Fig 4d**). Linear modelling based on the TRACERx data

245   approximated that a primary tumor volume of $11cm^3$ would result in a mean clonal plasma

246   VAF of 0.1% **(Figure 3b).** We multiplied tumor purity by tumor volume to control for stromal

247   contamination and determine cancer cell volume (**Extended Data Fig 4e**). On the assumption

248   that $1cm^3$ of effective tumor contains 9.4×10^7 cells we approximated that a plasma VAF of

249   0.1% corresponds to a tumor burden of 326 million malignant cells **(Extended Data Fig 4f).**

250   To investigate predictors of subclone detection we mapped detected subclonal SNVs back to

251   M-seq derived tumor phylogenetic trees. 35 of 57 (61%) shared subclones (identified in more

252   than one tumor region through M-Seq analysis) were identified in ctDNA, compared with 26

253   of 80 (33%) private subclones (detected in a single tumor region only) **(Extended Data Fig**

254   **4g)**. This suggested subclone volume influences subclonal ctDNA detection. We estimated

255   subclone volume based on mean regional subclone cancer cell fraction (CCF) and cancer cell

256   volume. Detected subclonal SNVs mapped to subclones with significantly higher estimated

257   volumes than subclones containing undetected SNVs (**Figure 3c)** and subclone volume

258   correlated with subclonal SNV plasma VAF (**Figure 3d**)**.**

259   **Detecting and characterizing NSCLC relapse**

260   The longitudinal phase of the study aimed to determine if ctDNA profiling with patient-specific

261   assay panels could detect and characterize the branched subclone(s) seeding NSCLC relapse.

262   Pre- and post-surgical plasma ctDNA profiling was performed blinded to relapse status in a

263   sub-group of 24 patients (cohort characteristics, **Extended Table 2d-e**). This included relapse

264   free patients who had been followed-up for a median of 775 days (range 688 to 945 days, n=10)

265   and confirmed NSCLC relapse cases (n=14) (cohort design, **Extended Data Fig 2c**). PCR

266   assays were added to panels in this phase of the study to optimize sensitivity in LUADs. A

267   median of 18.5 SNVs (range 12 to 20) were targeted by LUSC assay-panels and a median of

268   28 SNVs (range 25 to 30) were targeted by LUAD assay-panels (**Extended Data Fig 2d-e).**

269   Patients were followed up with three to six monthly clinical assessment and chest radiographs.

270   At least 2 SNVs were detected in 13 of 14 (93%) patients with confirmed NSCLC relapse prior

271   to, or at, clinical diagnosis of relapse and detected in 1 of 10 (10%) patients (CRUK0013) with

272   no clinical evidence of NSCLC relapse (**Fig 4a-k, Extended Data Fig 5a-n**). Excluding a

273   single case where no post-operative plasma was taken prior to clinical relapse (CRUK0041)

274   the median interval between ctDNA detection and NSCLC relapse confirmed on clinically

275   indicated CT imaging (lead-time) was 70 days (range 10 to 346 days). Four of 13 relapse cases

276   exhibited lead-times of more than six months **(Fig 4a-d)**. In two cases ctDNA detection

277   preceded CT imaging inconclusive for NSCLC relapse by 347 days (**Fig 4a**) and 260 days (**Fig**

278   **4d**). Post-operative ctDNA profiling reflected adjuvant chemotherapy resistance; CRUK0004,

279   CRUK0080 and CRUK0062 had detectable ctDNA in plasma within 30 days of surgery. The

280   number of detectable SNVs increased in all cases despite adjuvant chemotherapy with disease

281   recurring within 1 year of surgery (**Fig. 4a-c**). In contrast, CRUK0013 had 20 SNVs detectable

282   in ctDNA 72 hours after surgery and 13 SNVs detectable prior to adjuvant chemotherapy (**Fig**

283   **4e).** 51 days following completion of adjuvant treatment, no SNVs were detectable. Two

284   further plasma samples were profiled for this patient at day 457 and 667; ctDNA remained

285   undetectable and the patient remains relapse free 688 days post-surgery **(Fig 4e).** ctDNA

286   profiling detected intracerebral relapse; CRUK0029 had a pre-operative PET scan performed

287   50 days prior to surgery demonstrating normal cerebral appearances. Mean clonal plasma VAF

288   of detected SNVs remained above 1% 30 days post-surgery, 54 days post-operatively the

289   patient was confirmed to have intracerebral metastasis **(Fig 4f).**

290   We sought to resolve subclonal evolutionary-dynamics associated with NSCLC relapse.

291   Subclonal SNVs displaying plasma VAFs similar to clonal SNVs and mapping to phylogenetic

292   clusters confined to a single phylogenetic branch, were detected post-operatively in the ctDNA

293   of four patients who suffered NSCLC relapse (CRUK0004, CRUK0063, CRUK0065 and

294   CRUK0044) **(Fig. 4a,g-i).** These findings suggested a relapse process dominated by a subclone

295   represented in our assay-panel. Notably the subclone implicated by ctDNA as driving the

296   relapse in the case of CRUK0004 contained an ERRB2 (HER2) amplification event that may

297   be targetable in NSCLC[16]**.** This suggests ctDNA defined subclonal evolution may inform

298   precision strategies against emerging subclones **(Fig. 4a).** Relapses involving subclones from

299   more than one phylogenetic branch were evident in patients CRUK0080, CRUK0062 and

300   CRUK0041 **(Fig 4b-c,j).**

301   **Validation of phylogenetic characterization**

302 To validate subclonal ctDNA analyses, data acquired from sequencing metastatic tissue was

303 interpreted with M-seq primary tumor data **(Supplementary Table 4)**. Patient CRUK0063

304 suffered para-vertebral relapse of their NSCLC. Post-operative ctDNA analysis revealed the

305 detection of the same subclonal SNV (*OR5D18*) on four consecutive occasions over a 231-day

306 period **(Extended Data Fig 6a)**. The *OR5D18* SNV traced back to a subclonal cluster private

307 to primary tumor region three (**Fig 5a**). Exome sequencing of CT-guided biopsy tissue acquired

308 from the para-vertebral metastasis revealed the subclone implicated in the metastatic event by

309 detection of the *OR5D18* SNV in ctDNA gave rise to the metastatic clone. This supported

310 ctDNA phylogenetic characterization of relapse (**Fig 5a**). The para-vertebral biopsy contained

311 88 SNVs not present in the primary tumor including an *ARID1A* stop-gain driver SNV. Re-

312 examination of primary tumor region M-Seq data with a lower SNV calling threshold revealed

313 that 16 of 88 SNVs including *ARID1A* were detectable in primary tumor region three, compared

314 to a maximum of 2 of 88 in other tumor regions **(Extended Data Fig 6b).** Since ctDNA

315 implicated the subclone private to primary tumor region three in the relapse process, these data

316 suggest that ctDNA profiling can resolve the primary tumor region from which a low frequency

317 metastatic subclone derives. CRUK0035 developed two liver and one adrenal metastases **(Fig**

318 **5b)**. Sequencing of the metastatic liver deposit revealed that only 109 of 149 SNVs classed as

319   clonal in the primary tumor were detectable in the metastasis. This was suggestive of an

320   ancestral branching event not resolved through primary M-seq analysis (**Figure 5b**). Post-

321   operative ctDNA profiling identified clonal SNVs present in the liver metastasis biopsy but

322   also revealed SNVs representing a subclone from the primary tumor (**Extended Data Fig 6c**).

323   This subclone was not present in the metastatic liver deposit (**Fig 5b**). These data may reflect

324   ctDNA identified from the non-biopsied metastases suggesting multiple metastatic events.

325   CRUK0044 suffered a vertebral and right hilar relapse. Post-operatively the same subclonal

326   SNV (*OR10K1*), was detected in ctDNA on two occasions 85 days apart (**Extended Data Fig**

327   **6d**). This SNV represented a single subclone detected through sequencing hilar lymph-node

328   metastatic tissue, supporting ctDNA findings (**Fig. 5c**). CRUK0041 suffered an intracerebral,

329   hilar and subcarinal lymph node relapse. Four subclonal SNVs representing both branches of

330   the tumor phylogenetic tree were detectable in ctDNA at relapse. Concordant with these data,

331   sequencing of subcarinal metastatic tissue revealed the presence of subclonal SNVs mapping

332   to both phylogenetic branches (**Fig 5d, Extended Data Fig. 6e**). Patient CRUK0013 was found

333   to have lymph-node metastases following primary surgery. Two lymph node metastases were

334   sampled for exome analysis together with M-seq of the primary tumor. Subclonal SNVs

335   detected in ctDNA post-operatively mapped to an ancestral subclone (describing a subclone

336   that existed during the tumor's evolution) containing a KRAS amplification (**Extended Data**

337   **Fig 6f, Fig. 5e**). This ancestral subclone was present in the primary tumor and sampled lymph-

338   nodes **(Fig. 5e).** Given the lymph node involvement in this case these findings suggest residual

339   metastatic lymphadenopathy following surgery that responded to adjuvant chemoradiotherapy

340   (**Fig. 4e**)

341   **ctDNA profiling in the metastatic setting**

342   Patient CRUK0063 underwent examination through the PEACE post-mortem study 24 hours

343   following death. Six tumor regions were sampled from three metastatic sites (thoracic

344   vertebral, para-aortic and lung metastases). M-Seq data from the six post-mortem tumor

345   regions (day 857), the para-vertebral relapse biopsy (day 467) and five primary tumor regions

346   (day 0) were combined to infer the phylogenetic structure of this patient's NSCLC (**Fig 6a**).

347   All seven metastatic tumor regions arose from a single ancestral subclone represented by

348   phylogenetic cluster 8 **(Fig 6b).** Six metastatic regions shared a later phylogenetic origin,

349   phylogenetic cluster 12 (**Fig. 6b**). The single tumor region that had not arisen from

350   phylogenetic cluster 12 was sampled from the para-aortic metastasis at autopsy and contained

351   a private subclone represented by phylogenetic cluster 9 (**Fig 6b).** The findings could represent

352   two or more independent metastatic events arising from a single branch of the primary tumor

353   phylogenetic tree, with ongoing tumor evolution at each metastatic site **(Fig 6b)**. Or a single

354   metastatic event to the para-aortic site involving the ancestral subclone (phylogenetic cluster

355   8) prior to evolution of the subclone represented by phylogenetic cluster 9. Followed by

356   metastatic cross-seeding to para-vertebral and lung sites and ongoing clonal evolution **(Fig 6b)**.

357   We designed a bespoke ctDNA assay-panel to retrospectively track metastatic subclonal

358   burden. 20 clonal SNVs and a median of 8 subclonal SNVs (range 4 to 15) in 9 metastatic

359   subclonal clusters were targeted by the assay-panel **(Extended Data Fig 7).** Since 103 variants

360   per time-point were profiled, SNV call thresholds were increased to maintain platform

361   specificity of more than 99.2% at the 2 SNV ctDNA detection threshold (**see Extended**

362   **Methods**). ctDNA detection occurred at day 340 post-surgery (**Fig 6c, Extended Data Fig 7**).

363   At day 466 following clinical-relapse at the thoracic para-vertebral site, 18 of 20 SNVs

364   mapping to phylogenetic clusters (8,11 and 12) were detected in ctDNA (**Fig 6c**, **Extended**

365   **Data Fig 7**). These subclonal clusters were shared between six of seven metastatic sites **(Fig**

366   **6b)**. A single SNV from a private subclone (phylogenetic cluster 9) was also detectable in

367   ctDNA at day 466 **(Fig 6c, Extended Data Fig 7)**. This subclone was not identified in the CT

368   guided para-vertebral biopsy taken at day 467 (**Fig 6b)**. The mean plasma VAF of the 18 SNVs

369   detected in phylogenetic clusters 11, 8 and 12 reflected their proximity to the clonal cluster

370   (light blue) in the M-Seq derived phylogenetic tree **(Fig 6c)**. This suggested a tiered burden of

371   subclonal disease concordant with M-seq phylogenetic inferences **(Fig 6a).** Mean clonal VAF

372   fell in response to palliative radiotherapy and chemotherapy but at day 767 increased **(Fig 6c)**.

373   Single SNVs mapping to phylogenetic clusters 5 and 9 and two SNVs mapping to phylogenetic

374   cluster 2 were now detectable in ctDNA 90 days before death **(Fig 6a-c, Extended Data Fig**

375   **7).** These phylogenetic clusters represented subclones private to the para-aortic metastases **(Fig**

376   **6a-b**). Consistent with these data significant para-aortic progression was observed at post-

377   mortem compared with most recent CT imaging performed 112 days before death - which

378   showed no evidence of para-aortic disease.

379   **Discussion**

380   In summary, we find predictors of ctDNA detection in early-stage NSCLC characterized by

381   non-adenocarcinoma histology, necrosis, increased proliferative indices and lymphovascular

382   invasion **(Fig 2a)**. Triple negative breast cancers display necrosis[17], high proliferative

383   indices[18,19] and are associated with increased ctDNA levels compared with other breast cancer

384   subtypes[6] suggesting extension of observations to other tumor types.

385   We find a relationship between tumor volume and ctDNA detection. We estimate that a primary

386   NSCLC tumor with a volume of $11cm^3$ is required to achieve a ctDNA plasma VAF of 0.1%

387   (**Fig. 3b**), a VAF reflecting the optimum sensitivity of most current ctDNA platforms. Low-

388   dose CT lung screening can identify lung nodules with diameters as low as 4mm[20]. Assuming

389   a spherical nodule, this would translate to a volume of $0.034cm^3$ and an approximate plasma

390   VAF of 1.4×10^{-4} % - at the extreme of detection limits of ctDNA platforms[21]. Sensitivity of

391   ctDNA NSCLC screening may therefore be constrained by tumor size using current

392   technologies. ctDNA release dynamics may alter at disease relapse - in three LUAD cases we

393   detect no ctDNA pre-operatively, yet detect ctDNA at relapse in the absence of clinically

394   detectable disease (**Fig 4a,d,i**).

395   Targeted ctDNA profiling characterized the subclonal dynamics of relapsing NSCLC.

396   Limitations to this approach include cost, estimated at $1750 per patient for single region tumor

397   sequencing with evaluation of five plasma samples and synthesis of bespoke assay-panels

398   within a clinically relevant timeframe. Adjuvant platinum-based chemotherapy in NSCLC

399   improves cure rates following surgery in only 5% of patients and 20% patients receiving

400   chemotherapy experience acute toxicities[22]. There is an urgent need to increase adjuvant

401   therapy efficacy and better target its use. Our findings indicate that drug development guided

402 by ctDNA platforms to identify residual disease, define treatment response and target emerging

403 subclones in the adjuvant NSCLC setting, with appropriate CLIA validation, are now feasible.

404 **References**
405
406
407 1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: A
408   Cancer Journal for Clinicians 2011;61(2):69-90.
409 2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA: A Cancer Journal for Clinicians
410   2017;67(1):7-30.
411 3. Pignon J-P, Tribodet H, Scagliotti GV, Douillard J-Y, Shepherd FA, Stephens RJ, et al. Lung
412   Adjuvant Cisplatin Evaluation: A Pooled Analysis by the LACE Collaborative Group. Journal of
413   Clinical Oncology 2008;26(21):3552-9.
414 4. Landau Dan A, Carter Scott L, Stojanov P, McKenna A, Stevenson K, Lawrence Michael S, et
415   al. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia.
416   Cell;152(4):714-26.
417 5. Beaver JA, Jelovac D, Balukrishna S, Cochran RL, Croessmann S, Zabransky DJ, et al. Detection
418   of Cancer DNA in Plasma of Patients with Early-Stage Breast Cancer. Clinical Cancer Research
419   2014;20(10):2643-50.
420 6. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in
421   circulating tumor DNA predicts relapse in early breast cancer. Science Translational Medicine
422   2015;7(302):302ra133-302ra133.
423 7. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis
424   detects minimal residual disease and predicts recurrence in patients with stage II colon
425   cancer. Science Translational Medicine 2016;8(346):346ra92-ra92.
426 8. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic
427   cancer evolution for precision medicine: the lung TRACERx study. PLoS Biol
428   2014;12(7):e1001906.
429 9. Jamal-Hanjani M, Wilson GA, Horswell S, Mitter R, Sakarya O, Constantin T, et al. Detection
430   of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage
431   non-small-cell lung cancer. Annals of Oncology 2016;27(5):862-7.
432 10. Jr LAD, Bardelli A. Liquid Biopsies: Genotyping Circulating Tumor DNA. Journal of Clinical
433   Oncology 2014;32(6):579-86.
434 11. Caruso R, Parisi A, Bonanno A, Paparo D, Quattrocchi E, Branca G, et al. Histologic
435   coagulative tumour necrosis as a prognostic indicator of aggressiveness in renal, lung,
436   thyroid and colorectal carcinomas: A brief review. Oncology Letters 2012;3(1):16-8.
437 12. Vesselle H, Schmidt RA, Pugsley JM, Li M, Kohlmyer SG, Vallières E, et al. Lung Cancer
438   Proliferation Correlates with [F-18]Fluorodeoxyglucose Uptake by Positron Emission
439   Tomography. Clinical Cancer Research 2000;6(10):3837-44.
440 13. Higashi K, Ueda Y, Yagishita M, Arisaka Y. FDG PET measurement of the proliferative
441   potential of non-small cell lung cancer. The Journal of Nuclear Medicine 2000;41(1):85.
442 14. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal
443   clonal evolution characterized using circulating tumour DNA in a case of metastatic breast
444   cancer. Nature Communications 2015;6:8760.
445 15. Newman AM, Bratman SV, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive
446   method for quantitating circulating tumor DNA with broad patient coverage. Nat Med
447   2014;20(5):548-54.

448 16. Peters S, Zimmermann S. Targeted therapy in NSCLC driven by HER2 insertions. Translational
449 Lung Cancer Research 2014;3(2):84-8.
450 17. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic
451 evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol
452 2005;19(2):264-71.
453 18. Keam B, Im S-A, Kim H-J, Oh D-Y, Kim JH, Lee S-H, et al. Prognostic impact of
454 clinicopathologic parameters in stage II/III breast cancer treated with neoadjuvant docetaxel
455 and doxorubicin chemotherapy: paradoxical features of the triple negative breast cancer.
456 BMC Cancer 2007;7:203-.
457 19. Rhee J, Han SW, Oh DY, Kim JH, Im SA, Han W, et al. The clinicopathologic characteristics and
458 prognostic significance of triple-negativity in node-negative breast cancer. BMC Cancer
459 2008;8:307.
460 20. Team TNLSTR. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic
461 Screening. New England Journal of Medicine 2011;365(5):395-409.
462 21. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital
463 error suppression for improved detection of circulating tumor DNA. Nat Biotech
464 2016;34(5):547-55.
465 22. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung
466 adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. J Clin Oncol
467 2008;26(21):3552-9.

468

469 **Supplementary information** is available in the online version of the paper

485 **Authorship contribution statement**

486 C.A., N.J.B., G.A.W., M.J-H., T.C., R.S., and J.L-Q. contributed equally to this work. C.A.

487 and C.S. co-wrote the manuscript. C.A., M.J.H., and C.S. conceived study design. C.A., N.J.B.,

488 G.A.W. and R.R. integrated clinicopathological data, exome data and multiplex-PCR NGS

489 data. B.G.Z, J.L., T.C., R.S., E.K., N.S., D.H., A.N. and A.P., conducted and analysed

490 multiplex-PCR NGS experimental work. N.J.B, G.A.W, T.B.K.W, R.R., and N.M. conducted

491 bioinformatic analyses of TRACERx data. J.L-Q and D.A.M. conducted a central pathological

492 review. T.M. analyzed tissue micro-array data. F.F., R.E. and F.Z. conducted central

493 radiological review of pre-operative PET scans. H.J.W.L.A., W.L.B., F.M.F. and N.J.B.

494 conducted volumetric analyses of pre-operative CT scans. S.V., D.J., J.L., S.S., J.C-K., A.R.,

495 T.C., D.O. and A.A. conducted technical work for TRACERx sample processing. G.E., S.W.,

496 N.M. and G.A.W. conducted TRACERx sample exome sequencing. L.M. and J.R. conducted

497 cross-platform validation on TRACERx cell-free DNA samples. M.J.H., C.S. and M.Fa.

498 designed the PEACE protocol. M.J.H., C.D., J.S. and C.S. designed the TRACERx protocol.

499 C.H., S.L.M., M.F., T.A., M.Fa., E.B., D.L., M.H., S.K., N.P., S.M.J., R.T., A.A., F.B., Y.S.,

500 R.S., L.J., A.M.Q, P.C., B.N., G.M., G.L., S.T., M.N., H.R., K.K., M.C., L.G., D.F., A.N., S.R.,

501 G.A., S.K., P.R., V.E., B.I., M.I-S., V.P., J.L., M.K., R.A., H.A., H.D., S.L. are integral clinical

502 members of TRACERx study sites. J.H. and H.L. run the UCL GCLP facility. A.H., H.B., N.I.

503 and Y.N. were involved in study oversight. J.A.S., J.L-Q., Z.S., E.G., S.K., S.T., M.A.B, R.F.S.,

504 J.H., A.S., S.Q., P.V.L., C.D. and J.L. gave advice and reviewed the manuscript. A.H. gave

505 statistical advice. C.S. provided overall study oversight.

506   **Author information**

507   The authors declare competing financial interests. Reprints and permissions information is

508   available at www.nature.com/reprints. Correspondence should be addressed to C.S.

509   (Charles.Swanton@crick.ac.uk).

510

Primary NSCLC resection and multiregion sampling
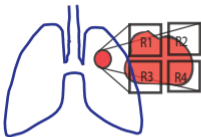
Exome sequencing of tumor regions

R1 R2 R3 R4

Mutations in tissue
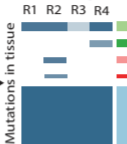
Phylogenetic tree informs PCR-assay panel construction

Primers in assay-panel

Sub-clonal

Clonal
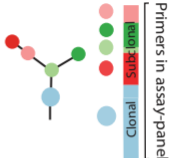
Multiple patient-specific assay panels combined

Multiplex-PCR assay-pool

Patient

1
2
3
4
5
6

0   10   20   30

Multiplex-compatible primers targeting patient-specific SNVs

Blood sample

cfDNA extracted
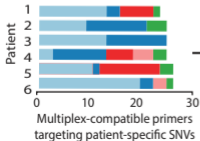
PCR-NGS

Patient-specific phylogenetic tracking

Pre-surgery

Relapse
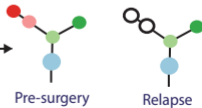
Figure 1 Phylogenetic ctDNA tracking

Overview of the study methodology. Multi-region sequencing of Stage I-IIIB non-small cell lung cancers was performed through the TRACERx study. Phylogenetic trees were constructed. PCR assay-panels were designed targeting clonal and subclonal single nucleotide variants to facilitate non-invasive tracking of the patient-specific tumor phylogeny. Based on predicted and validated primer compatibility assay-panels were combined into multiplex assay-pools containing primers from up to 10 patients. Cell-free DNA was extracted from pre and post-operative plasma samples and multiplex-PCR performed. This was followed by next generation sequencing of amplicons. Findings were integrated with M-Seq exome data to track tumor evolution.

a

LUSC | Other

Detected
Stage
Necrosis
PET TBR
%Ki67
LN involved
LVI
Volume

- ≥2 SNVs detected
- 1 SNV detected
- No SNVs detected
- Stage I-III
- Lymph node involvement
- Lymphovascular invasion

High | quartiles | Low

Ki67 % | Necrosis
PET | Volume

LUAD

Detected
Stage
Necrosis
PET TBR
%Ki67
LN involved
LVI
Volume

b

Clonal / Sub-clonal

- Detected clonal SNV
- Clonal SNVs in assay-panel
- Detected subclonal SNV
- Sub-clonal SNVs in assay-panel

LUSC | Other | LUAD

Figure 2 – Clinicopathological predictors of ctDNA detection

a) Heatmap showing clinicopathological and ctDNA detection data, continuous variables quartiled. Raw data and patient IDs in Supplementary Table 1. b) Detection of clonal and subclonal single nucleotide variants within 46 patients with two or more single nucleotide variants detected in plasma. Histology indicated in panels as LUSC, LUAD and Other. Other histology refers to large cell carcinoma (1/1 ctDNA positive), adenosquamous carcinoma (2/3 ctDNA positive), large cell neuroendocrine carcinoma (1/1 ctDNA positive) and carcinosarcoma (1/2 ctDNA positive).

a Spearman rho = 0.61, P <0.001

- Other
- LUSC
- LUAD

Mean clonal plasma VAF (y-axis)
Tumor Volume cm³ (x-axis)

b

VAF: 0.01% VAF: 0.1% VAF: 1% VAF: 10%

Predicted tumor burden (cm³)

1.4   11   78   584

c

LUSC
LUAD
Other

▲ Private
● Shared

P <0.001

Subclone volume (cm³) x purity (y-axis)
Detected SNV (x-axis): No / Yes

d Spearman rho = 0.53, P <0.001

Subclonal SNV VAF (y-axis)
Subclone volume (cm³) X purity (x-axis)

LUSC
LUAD

▲ Private
● Shared

Figure 3. Tumor volume predicts plasma variant allele frequency

a) Tumor volume (cm^3) measured by CT volumetric analysis correlates with mean clonal plasma VAF, n=38, grey vertical lines represent range of clonal VAF, line of best fit estimated in log-space, 95% confidence intervals indicated by red shading. b) Predicted tumor burden at hypothetical clonal VAF intervals ranging from 0.01% to 10% based on linear model shown in panel a. c) Estimated effective subclone size, defined as mean CCF of subclone across sampled tumor regions multiplied by effective tumor volume (volume × purity), influences subclonal SNV detection. For negative calls, median effective subclone size was 1.60 cm^3, range= 0.21-14.11, n=163 for positive calls, median effective subclone size = 3.97 cm^3, range = 0.33 – 45.09, n=109. Wilcoxon rank sum test, P<0.001, data from 34 patients (who passed volumetric filters and had subclonal SNVs represented in assay-panel). d) Estimated effective subclone size correlates with subclonal plasma VAF, n=109 subclonal SNVs, data from 24 patients (who passed volumetric filters with detected subclonal SNVs in plasma).

a CRUK0004: LUAD

ERBB2 amp.

Vertebral, lung relapse

Lead time 290 days

Relapse

Chemo

Erlotinib

CT - suspicious vertebral lesion

Mutation VAF

Days since surgery

S 30 56 177 272 320 371 392 469

b CRUK0080: LUSC

Lung, iliac bone relapse

Relapse

Chemo

Lead time 239 days

Mutation VAF

Days since surgery

S 2 55 145 230 265

c CRUK0062: LUSC

Lung relapse

Chemo

Lead time 346 days

Relapse

Mutation VAF

Days since surgery

S 30 114 258 376

d CRUK0045: LUAD

Lung, mediastinal relapse

Lead time 347 days

Relapse

Normal CT

Mutation VAF

Days since surgery

S 111 195 286 377 468 559 722 906

e CRUK0013: LUAD

Chemo RTx

No relapse

Mutation VAF

Days since surgery

S 3 38 122 212 457 667

f CRUK0029: LUAD

Intracerebral relapse

Lead time 22 days Relapse

Mutation VAF

Days since surgery

S 32 54 81

g CRUK0063: LUSC

Paravertebral relapse

Relapse

Chemo

Lead time 115 days

RTx

Mutation VAF

Days since surgery

S 60 151 242 340 431 466 627 662

h CRUK0065: LUSC

Adrenal relapse

Relapse

Lead time 77 days

Mutation VAF

Days since surgery

S 44 184 261 273 386 414

i CRUK0044: LUAD

Mediastinal, vertebral relapse

Lead time 19 days

Relapse

Mutation VAF

Days since surgery

S 4 91 174 193 259

j CRUK0041: LUAD

Subcarinal, intracerebral relapse

Lead time 0 days

Mutation VAF

Days since surgery

S 104

k CRUK0035: LUAD

Liver, adrenal relapse

Chemo RTx

Lead time 10 days

Afatinib

Relapse

Mutation VAF

Days since surgery

S 32 144 291 382 410 475 487

KEY

ctDNA representation of M-Seq phylogenetic tree pre-operatively

ctDNA representation of M-Seq phylogenetic tree at relapse

No SNVs in subclonal node detected

Clonal SNV

Subclonal SNVs

Presurgical blood sample

Adjuvant treatment

Palliative treatment

Lead time

Mean clonal VAF

Mean subclonal VAF

RTx - Radiotherapy

Figure 4 – Post-operative ctDNA detection predicts and characterizes NSCLC relapse –
2a-k) Longitudinal cell-free DNA profiling. Circulating tumor DNA (ctDNA) detection in
plasma was defined as the detection of two tumor-specific SNVs. Relapse was based on
imaging-confirmed NSCLC relapse, imaging was performed as clinically indicated. Detected
clonal (circles, light blue) and subclonal (triangles, colors indicates different subclones)
SNVs from each patient-specific assay-panel are plotted on graphs colored by M-Seq derived
tumor phylogenetic nodes. Mean clonal (blue) and mean subclonal (red) plasma VAF are
indicated on graphs as connected lines. Pre-operative and relapse M-Seq derived
phylogenetic trees represented by ctDNA are illustrated above each graph.

Relapse

a CRUK0063

PIK3CA
CDKN2A
TP53
SOX2 amp
TERT amp

NF1
FANCM

R1 R2
R3 R4 R5

ARID1A

M1

CD44 amp
LMO2 amp

Stage 2a (T2N1a) LUSC
Paravertebral relapse

cfDNA

Pre-operative

Relapse

Primers in assay-panel
1   10   17

Primers in assay-panel
1   10   17

b CRUK0035

LN

R1 R2
R3

TP53
FAS
EGFR
HOXA9

EML4 amp

FLT4

LN1

RNF217 del
FOXA1 del
FANCM del
EML4 del

M1

Stage 3b (T4N2) LUAD
Liver and adrenal relapse

cfDNA

Pre-operative

Relapse

Primers in assay-panel
1   10   23

Primers in assay-panel
1   10   23

c CRUK0044

R1 R2

KRAS

SMARCA4

EML4 del
CHEK2 del
TSC1 del
HOXD11
HOXD13

M1

Stage 1a (T1N0) LUAD
Right hilar relapse, thoracic
vertebral metastasis

cfDNA

Pre-operative

Relapse

Primers in assay-panel
1   10   20   28

Primers in assay-panel
1   10   20   28

d CRUK0041

R1 R2
R3 R4

BRAF
EGFR
TERT amp
CCNE amp

M1

Stage 1b (T2aN0) LUAD
Mediastinal, subcarinal,
intracerebral relapse

cfDNA

Pre-operative

Relapse

Primers in assay-panel
1   10   20   **27**

Primers in assay-panel
1   10   20   27

e CRUK0013

R1 R2
R3

EGFR amp

NOTCH1

STK11
KRAS amp

LN2
LN1

LN1 LN2

Stage 3a (T2N2) LUAD
No relapse

cfDNA

Pre-operative

Post surgery

Primers in assay-panel
1   10   20   26

Primers in assay-panel
1   10   20   26

Signature 4 (smoking)

Signature 2/13 (APOBEC)

Signature 5 (unknown)

Signature 1A (age)

No predominant signature

Clonal mutation cluster

Subclonal mutation cluster

Mutation cluster not assayed in ctDNA

M LN Metastatic or lymph node lesion

Figure 5. **Re-design of phylogenetic trees to incorporate relapse tissue sequencing data to benchmark post-operative ctDNA analyses**

Phylogenetic trees based on mutations found in primary and metastatic tissue (a-d), or primary tumor and lymph node biopsies (e). Colored nodes in phylogenetic trees indicate cancer clones harboring mutations assayed for in ctDNA, grey indicates a clone not assayed. Thick colored bar shows number of assays per sample detected preoperatively and at relapse (a-d) or in the absence of relapse, post surgery (e). Thin colored bar shows number of assays in total. Colors matches clones on the phylogenetic trees.

a CRUK0063

50 mutations

EP300
RET amp
NF1
FANCM

Ubiquitous mutations
Mutations unique to primary tumor
Shared mutations in metastatic samples
Private mutations in metastatic samples

ARID1A
EP300

Metastases   Primary

Common to metastases and primary

TERT amp   PIK3CA
PRF1 del   MSH2
3q amp   CDKN2A
LMO2 amp   TP53

b Para-vertebral   Lung   Para-aortic

Relapse site
T8-9 para-vertebral mass
Biopsy day 467

Autopsy sites

Not detected in ctDNA
(detected M-Seq)   Detected in ctDNA

c   Subclonal

Clonal   Shared   Private   Phylogenetic
clusters

11 8 12   9 10 13 2 3 5

Chemo   Relapse   RTx Chemo   Chemo

Mean mutation VAF

SNVs
1
5
10
20

Surgery 60 151 242 340 431 466   627 662   767 Death
857 days

Days since diagnosis

Figure 6. ctDNA tracking of lethal cancer subclones in CRUK0063

Sampling and sequencing was performed of one relapse biopsy at day 467 and five metastatic tissue samples from three lesions at time of death through the PEACE (Posthumous Evaluation of Advanced Cancer Environment) post mortem study. Phylogenetic analysis revealed cancer evolution and identified private subclones at each site. **a).** To-scale phylogenetic tree of CRUK0063 including M-seq based on metastatic and primary tumor regions. Branch length is proportional to number of mutations in each subclone. Colors represents mutation clusters, light blue node representing the clonal cluster. **b)** Tissue-specific phylogenetic trees for metastatic lesions, highlighted nodes in color represents mutation clusters found at each site and assayed for in ctDNA. Open circles represents mutation clusters not detected at any time in ctDNA. **c)** Tracking plot of identified subclones in ctDNA, showing mean VAF of identified subclones. Size of dots indicates number of assays detected. Colors corresponds to mutation clusters and matches panels a and b. Tiered burden of subclonal disease can be observed, with clusters representing earlier cancer subclones present at higher VAF, likely reflecting a larger cancer burden carrying shared relative to private mutations.