

1 ***Pulmonary venous circulating tumour cell dissemination before***
2 ***tumour resection and disease relapse***

3 Francesca Chemi^{1,2*}, Dominic G Rothwell^{1*}, Nicholas McGranahan^{2,3,4*}, Sakshi
4 Gulati¹, Chris Abbosh³, Simon P Pearce¹, Cong Zhou¹, Gareth Wilson^{3,4}, Mariam
5 Jamal-Hanjani^{2,3}, Nicolai Birkbak^{3,4}, Jackie Pierce¹, Chang Sik Kim¹, Saba Ferdous¹,
6 Deborah Burt¹, Daniel Slane-Tan¹, Fabio Gomes⁵, David Moore³, Rajesh Shah⁶,
7 Maise Al Bakir⁴, Crispin Hiley^{2,3}, Selvaraju Veeriah³, Yvonne Summers⁵, Philip
8 Crosbie^{2,5,7}, Sophia Ward⁴, Barbara Mesquita¹, Marek Dynowski¹, Dhruva Biswas⁴,
9 Jonathan Tugwood¹, Fiona Blackhall^{2,5}, Crispin Miller⁸, Allan Hackshaw³, Ged
10 Brady^{1,2#}, Charles Swanton^{2,3,4#} and Caroline Dive^{1,2#} on behalf of the TRACERx
11 consortium.

12

13 **Affiliations**

14 ¹ Clinical and Experimental Pharmacology Group, Cancer Research UK Manchester Institute, The
15 University of Manchester, Alderley Park, SK10 4TG, UK

16 ² Cancer Research UK Lung Cancer Centre of Excellence at The University of Manchester and
17 University College London, UK.

18 ³University College London Cancer Institute, London, UK

19 ⁴Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK

20 ⁵ Division of Cancer Sciences, Faculty of Biology Medicine and Health, University of Manchester,
21 Manchester, UK

22 ⁶ Department of Thoracic Surgery, Wythenshawe Hospital, Manchester University NHS Foundation
23 Trust, Manchester, UK

24 ⁷North West Lung Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust,
25 Manchester, UK

26 ⁸ RNA Biology, Cancer Research UK Manchester Institute, The University of Manchester, Alderley
27 Park, SK10 4TG, UK

28

29

30 *Equal contribution, #co-corresponding authors

31

32 (caroline.dive@cruk.manchester.ac.uk, Charles.Swanton@crick.ac.uk,
33 Ged.Brady@cruk.manchester.ac.uk)

34

35 Approximately 50% of early-stage non-small cell lung cancer (NSCLC) patients that
36 undergo surgery with curative intent will relapse within 5 years^{1,2}. Detection of
37 circulating tumour cells (CTCs) at the time of surgery may represent a tool to identify
38 patients at higher risk of recurrence where more frequent monitoring is advised.
39 Here, we asked whether CellSearch detected pulmonary venous CTCs (PV-CTCs)
40 at surgical resection of early stage NSCLC represent subclones responsible for
41 subsequent disease relapse. PV-CTCs were detected in 48% of 100 patients
42 enrolled into the TRACERx study³ and were associated with lung cancer specific
43 relapse, and remained an independent predictor of relapse in multivariate analysis
44 adjusted for tumour stage. In a case study, genomic profiling of single PV-CTCs
45 collected at surgery revealed a higher mutation overlap with a metastasis detected
46 10 months later (91%) compared to the primary tumour (79%), suggesting that early
47 disseminating PV-CTCs were responsible for disease relapse. Together, PV-CTC
48 enumeration and genomic profiling highlight the potential of PV-CTCs as early
49 predictors of NSCLC recurrence after surgery. However, limited sensitivity of PV-
50 CTCs to predict relapse suggests further studies using a larger, independent cohort
51 are warranted to confirm and better define this potential clinical utility of PV-CTCs in
52 early stage NSCLC.

53
54 Lung cancer is the leading cause of cancer related deaths worldwide with a 5 year
55 relative survival rate of 4% in the metastatic setting⁴. NSCLC is the most common
56 form of lung cancer. Patients presenting with early-stage NSCLC may undergo
57 surgery with or without adjuvant chemotherapy and/or adjuvant radiotherapy in an
58 attempt to achieve cure. However, disease recurrence following surgery is common,
59 with 5-year relapse rates ranging from ~20% in patients with stage I disease to ~50%
60 in those with stage III disease^{1,2}.

61 Strategies to understand the biology of early dissemination and to identify patients at
62 high risk of relapse may inform novel therapeutic approaches for adjuvant treatment
63 to improve cure rates. CTCs are the assumed ‘foundations of metastasis’⁵, though
64 this has not been formally proven in NSCLC. CTCs enriched from breast cancer,
65 melanoma, NSCLC and small cell lung cancer (SCLC) patients’ peripheral blood can
66 form tumours in immune compromised mice confirming their tumorigenic potential⁶⁻⁹.
67 CTC number, measured using the CellSearch® platform, is a Food and Drug
68 Administration (FDA) approved prognostic test in breast, colorectal and prostate
69 cancers and is also prognostic in NSCLC¹⁰. Although peripheral blood CTCs (using
70 CellSearch that captures only cells expressing EpCAM and Cytokeratin) are rare in
71 early stage NSCLC patients, we previously demonstrated in a pilot study that
72 CellSearch CTCs obtained from the draining pulmonary vein of the cancer-affected
73 lung (PV-CTCs) are more frequent and we observed a trend towards worse disease-
74 free survival (DFS) and overall survival (OS)¹¹. To determine whether our preliminary
75 findings that PV-CTCs at resection are associated with relapse holds in a larger
76 patient cohort, we enumerated PV-CTCs from 100 NSCLC patients enrolled onto the
77 TRACERx study¹².

78
79 In our current cohort of 100 TRACERx patients (46% stage I, 34% stage II and 20%
80 stage III; median follow-up 993 days), (Figure 1a, Table 1 and Supplementary Table
81 1), 48% (48/100) harboured at least 1 PV-CTC per/7.5mL blood (mean \pm SD, 42.2 \pm
82 127.3, median 0, range 0-896) (Figure 1b). PV-CTC count was not significantly
83 associated with clinicopathological factors such as age, gender, pathological stage,
84 smoking status and treatment received (Figure 1c and Supplementary Table 2). In
85 contrast to circulating tumour DNA (ctDNA)¹³, PV-CTC count was not significantly
86 different between adenocarcinoma (LUAD) and non-LUAD (p=0.554, t-test)

87 suggesting that factors that control release of ctDNA and the dissemination of intact
88 CTCs are distinct.

89 In our previous pilot study of 30 patients¹¹, there was an association between the
90 PV-CTC 'high' count (≥ 18 PV-CTCs/7.5ml blood) and DFS ($p=0.055$). When we
91 applied this cutpoint in the TRACERx cohort there was a significant association with
92 poorer DFS ($p=0.019$ log-rank, HR=2.28, Figure 2a) and this remained an
93 independent predictor in multivariate analysis when adjusted for tumour stage
94 ($p=0.021$, HR=2.4, 95% CI 1.14-5.2, Figure 2b). However, the performance of this
95 cutpoint in predicting DFS, defined by time-dependent receiver operating
96 characteristic (tdROC) curves, revealed limited sensitivity (sensitivity = 31.7%,
97 specificity = 84.9%). We therefore conducted further exploratory analysis to refine the
98 'PV-CTC high' cutpoint to predict lung cancer specific relapse events and investigate
99 the biological relevance of PV-CTCs in NSCLC metastasis. Briefly, of the 37
100 recorded DFS events in the TRACERx cohort, 22 were due to lung cancer specific
101 relapse. The remaining events occurred either without evidence of lung cancer
102 relapse before death ($n=9$, Supplementary Table 3) due to a second non-lung
103 primary cancer ($n=4$, confirmed by histology, imaging and clinical discussion,
104 Supplementary Table 3) or lacked sufficient clinical information to determine cause
105 ($n=2$, Supplementary Table 3). tdROC curves showed that the sensitivity and
106 specificity in predicting lung cancer specific relapse at two years was optimal when a
107 75th quantile cutpoint was applied (≥ 7 PV-CTCs/7.5ml blood, Extended Data Fig.1a).
108 A 'PV-CTC high' status of ≥ 7 PV-CTCs/7.5ml blood showed significant association
109 with lung cancer relapse in Kaplan-Meier analysis ($p=0.009$ log-rank, HR=2.78,
110 Extended Data Fig.1b) and remained an independent predictor in multivariate

111 analysis when adjusted for tumour stage ($p=0.027$, HR = 2.6, 95% CI 1.1-6.2,
112 Extended Data Fig.1c).

113 Analysis of PV-CTCs as a continuous variable showed that each doubling of PV-
114 CTC count was a significant prognostic factor for DFS when modelled as a sole
115 covariate ($p=0.035$, HR=1.113) and when modelled with other significant prognostic
116 factors ($p=0.040$, HR=1.116, 95% CI 1.005-1.239, Wald test two-sided,
117 Supplementary Table 4). Each doubling of PV-CTC count was also significantly
118 associated with lung cancer specific relapse in both uni-variate ($p=0.029$, HR=1.148)
119 and multi-variate analysis ($p=0.024$, HR=1.170, 95% CI 1.021-1.341, Wald test two-
120 sided, Supplementary Table 5). We also noted a significant association between PV-
121 CTCs as a continuous variable and intracranial disease present at clinical relapse
122 ($p=0.028$, t-test two-sided, Supplementary Table 2).

123 Collectively, these data raise the possibility that patients with a 'high' CellSearch PV-
124 CTC count at resection may benefit from increased minimal residual disease (MRD)
125 monitoring post-surgery. We have shown that increasing PV-CTC count as a
126 continuous variable is associated with poor prognosis. To use PV-CTCs in a clinical
127 setting, a pre-defined cutpoint will be required to prospectively stratify patients.
128 Although the previously defined cutpoint of ≥ 18 PV-CTCs/7.5ml blood¹¹ was verified
129 here and the further exploratory analysis of a ≥ 7 PV-CTC/7.5ml blood increased the
130 performance of PV-CTCs in predicting lung cancer specific relapse, sensitivity
131 remained modest (45.2% for ≥ 7 PV-CTCs/7.5ml blood vs 32.8% for ≥ 18 PV-
132 CTCs/7.5ml blood, Extended Data Fig.2a) and further studies are clearly required
133 before clinical utility can be evaluated.

134

135 We next sought to assess the degree of genomic similarity between early-
136 disseminated PV-CTCs and metastatic disease by comparing the primary tumour,
137 the PV-CTCs and subsequent metastatic disease from the same patient. Single PV-
138 CTCs and white blood cell (WBCs) controls were successfully isolated from 14
139 patients (Extended Data Fig. 2a and online methods). Of these 14 patients, five
140 experienced a lung cancer specific relapse event (Extended Data Fig.2b); with an
141 evaluable metastatic tissue biopsy available for one patient (CRUK0242). This 74-
142 year old male was diagnosed with stage IIIA, invasive adenocarcinoma in the right
143 lung and underwent tumour resection, at which point 28 PV-CTCs were detected.
144 The patient received adjuvant chemotherapy and radiotherapy and at 10 months
145 post-surgery, positron-emission tomography (PET) identified relapse involving the
146 right pleura. At this time a biopsy from the right pleural lesion was sequenced and
147 peripheral blood samples collected for circulating free DNA (cfDNA) analysis. After
148 receiving palliative chemotherapy and radiotherapy, the patient progressed and died
149 the following year (Figure 3a). In this case study, three spatially-separated primary
150 tumour regions, PV-CTCs, cfDNA isolated from pulmonary and peripheral veins at
151 resection and again from the periphery at disease relapse, and the pleural
152 metastasis were genetically profiled and compared.

153 From the 28 PV-CTCs detected by CellSearch, we successfully isolated and
154 amplified six single PV-CTCs (Extended Data Fig. 2c). Low-pass whole genome
155 sequencing was performed which revealed that 3/6 PV-CTCs harboured copy
156 number alterations (CNA) that matched the primary tumour. The remaining cells,
157 although phenotypically CTC candidates by CellSearch criteria, showed flat copy
158 number profiles as observed in WBC controls (Figure 3b, Extended Data Fig. 2d).
159 We have termed these cells 'circulating epithelial cells' (CECs) and propose these

160 are likely to be normal epithelial cells that enter the blood along with PV-CTCs;
161 similar cells have recently been described in non-cancer patients¹⁴. In order to
162 identify somatic mutations present in the PV-CTCs, we performed whole exome
163 sequencing (WES) followed by targeted deep sequencing of the 3 PV-CTCs, 3 CECs
164 and 2 WBC controls. This identified 198 mutations (single nucleotide variants, SNVs)
165 in the PV-CTCs and none in the CECs (Figure 3c). After accounting for technical
166 drop-out due to the single cell sequencing approach (loci drop-out = 102/441 in
167 tumour, 81/342 in metastasis)¹⁵ (Supplementary Table 6 and 7, methods online),
168 46% (157/339) of all primary tumour mutations were also detected in PV-CTCs
169 (Figure 3c and Extended Data Fig. 3a). Along with the CNA data this confirms the
170 tumour origin of the PV-CTCs, but the presence of PV-CTC mutations not detected
171 in the primary tumour suggests these cells may represent a minor subclone of the
172 tumour. Although a resolvable tumour specific CNA pattern was not observed in the
173 metastasis (Figure 3b), due to low tumour content, WES and targeted deep
174 sequencing revealed 91% of the PV-CTC mutations were seen in the metastasis
175 (181/198), which is a higher mutational overlap than between the PV-CTCs and
176 primary tumour (157/198, 79%) (Figure 3c and Extended Data Fig.3a). In addition,
177 96.8% (120/124) of the primary tumour mutations that were not detected in the
178 metastasis were also not detected in the PV-CTCs (Figure 3c). Strikingly, of the 41
179 PV-CTC private mutations that were not detected in the primary tumour, 28 (68.3%)
180 were identified in the relapse biopsy WES (Figure 3c and Extended Data Fig.3a)
181 suggesting that the PV-CTCs present in the patient's blood at surgery share a
182 common progenitor with the metastasis that was detected 10 months later. The
183 evolutionary origin of the PV-CTCs and metastasis was confirmed by phylogenetic
184 analysis that revealed both PV-CTCs and metastasis are part of the same specific

185 branch, which is distinct from all other subclones of the primary tumour (Figure 3d).
186 The identification of PV-CTC specific mutations that are undetectable by bulk tumour
187 analysis, yet are present in the relapse samples, strongly suggests that the PV-CTCs
188 belong to a minor tumour subclone which is responsible for eventual relapse.
189 Examination of the mutations shared between PV-CTCs and the metastatic biopsy
190 yet absent from the primary tumour has the potential to give insight into the
191 mechanisms of metastasis. In this patient, the 28 PV-CTC/metastatic associated
192 mutations not detected in the primary tumour included a putative inactivating driver
193 mutation in the tumour suppressor gene *LZTS1* (p.Pro104His) (Supplementary Table
194 8) which has been shown to inhibit tumour migration and whose lower expression
195 has been linked to poor overall survival in NSCLC¹⁶.
196 Finally, to address the question whether the 13 private PV-CTC mutations not initially
197 detected in the primary tumour or relapse biopsy, were in fact present at low
198 frequency, additional targeted deep-sequencing of the tumour and metastasis was
199 performed. All 13 mutations were present in either the primary tumour (5/13), the
200 metastasis (12/13) and/or relapse cfDNA (7/13) (Figure 3e, Extended Data Fig.3b
201 and Extended Data Fig.4). Interestingly, even using targeted deep-sequencing none
202 of the 520 pre-identified mutations were detected in either baseline pulmonary or
203 peripheral blood cfDNA samples (Extended Data Fig.4), highlighting the unique
204 aspect of molecular analysis of PV-CTCs at resection.
205 Previous studies have shown a genetic link between CTCs, primary tumour and
206 metastasis with clonal and subclonal mutations detected in CTCs in both colorectal
207 and prostate cancer^{17,18}. However, these studies were performed in metastatic
208 patients and to our knowledge, this case report is the first to show that CTCs at
209 surgery are phylogenetically linked to subsequent metastatic disease. This is

210 exemplified by the larger mutational overlap between the PV-CTCs and the
211 metastatic tumour that arose 10 months post PV-CTC isolation, than between the
212 PV-CTCs and primary tumour which were collected at the same time in our case
213 study. Comprehensive molecular analysis of early disseminating PV-CTCs also
214 raises the opportunity to identify putative mechanisms of metastatic spread from the
215 primary tumour prior to establishment of recurrent disease.

216

217 In early-stage NSCLC disease recurrence post-surgery occurs frequently and in this
218 scenario survival is dismal; therefore, strategies that enable the identification of
219 patients at higher risk of recurrence are an unmet medical need. We show here PV-
220 CTC count (using the CellSearch platform) is associated with DFS and lung cancer
221 specific survival in the TRACERx cohort, reinforcing the biological importance of PV-
222 CTCs as founders of NSCLC metastasis. However, the clinical strength of a PV-CTC
223 count to predict lung cancer specific relapse is modest and requires further validation
224 in an independent and prospective patient cohort. Reasons underpinning the
225 modest predictive strength of PV-CTC counts for NSCLC specific relapse could
226 include the co-existence of CECs and bonafide epithelial CTCs as seen in the blood
227 sample of the case study. This mixed population of EpCAM positive cells could
228 confound the true PV-CTC count and the inability of CellSearch to detect
229 mesenchymal CTCs further reduce the sensitivity of this approach. Additional
230 detailed investigations are warranted to differentiate between epithelial and
231 mesenchymal CTCs and CECs and to incorporate this greater understanding into
232 NSCLC relapse prediction models. This study highlights the benefit of combining PV-
233 CTC, tumour and cfDNA analysis to unearth new biological insights into the process
234 of NSCLC metastasis.

235 **Acknowledgements**

236 We sincerely thank the patients and their families for donating of blood samples for
237 research. We thank Ekram Aidaros-Talbot for administrative assistance with the
238 manuscript. We also thank Professor Jacqui Shaw for kindly providing plasma
239 (relapse time point) of patient CRUK0242. TRACERx is funded by Cancer Research
240 UK (grant number C11496/A17786). This research was supported by Cancer
241 Research UK - Core funding to CRUK Manchester Institute (A27412) Centre funding
242 to Manchester (A25254), and funding of the CRUK Lung Cancer Centre of
243 Excellence. Support was also received from the Manchester Experimental Cancer
244 Medicine Centres and Manchester NIHR Biomedical Research Centre, FC is funded
245 by the CANCER-ID Consortium (115749- Cancer-ID). BM is funded by Menarini
246 Biomarkers Singapore PTE Ltd. CSK is funded by The Manchester MRC Single Cell
247 Research Centre (MR/M008908/1). CS is Royal Society Napier Research Professor.
248 This work was supported by the Francis Crick Institute that receives its core funding
249 from Cancer Research UK (FC001169,FC001202), the UK Medical Research
250 Council (FC001169, FC001202), and the Wellcome Trust (FC001169, FC001202).
251 CS is funded by Cancer Research UK (TRACERx, PEACE and CRUK Cancer
252 Immunotherapy Catalyst Network), the CRUK Lung Cancer Centre of Excellence,
253 Stand Up 2 Cancer (SU2C), the Rosetrees Trust, NovoNordisk Foundation
254 (ID16584) and the Breast Cancer Research Foundation (BCRF).
255 The research leading to these results has received funding from the European
256 Research Council under the European Union's Seventh Framework Programme
257 (FP7/2007-2013)/ ERC grant agreement n°FP7 – 617844 (PROTEUS) and Marie
258 Curie Network PloidyNet. Support was also provided to CS by the National Institute
259 for Health Research, the University College London Hospitals Biomedical Research

260 Centre, and the Cancer Research UK University College London Experimental
261 Cancer Medicine Centre. All funders and sponsors had no role in this study.

262

263 **Authors Contributions**

264 CS, CD, PC, DGR and GB developed the clinical study, directed research, and co-
265 wrote the manuscript. FC designed and conducted experiments, analysed data and
266 drafted the manuscript with assistance of DGR and NMG. SG, SPP, GW, NB, NMG,
267 CSK, SF, CM and MD provided bioinformatic support for the study. CA provided
268 support for the clinical interpretation of the data. CZ performed statistical analysis.
269 CA and DM performed centrally pathology review. DB, DST and BM provided
270 support for single cell isolation. MJ-H, JP, FG, RS, MAB, CH, SV, YS, PC, SW, DB,
271 JT, FB and AH provided support for patients' recruitment, samples 'management and
272 clinical support for the study.

273

274 **Competing Interests statement**

275 CD receives research grants/support from Menarini and research grants are also
276 received from AstraZeneca, Astex Pharmaceuticals, Bioven, Amgen, Carrick
277 Therapeutics, Merck AG, Taiho Oncology, GSK, Bayer, Boehringer Ingelheim,
278 Roche, BMS, Novartis, Celgene, Epigene Therapeutics Inc., all outside the scope of
279 this paper. CD acts in a consultant or advisory role for Biocartis and AstraZeneca,
280 again outside the scope of this work. CS has received honoraria, consultancy, or
281 SAB Member fees for Pfizer, Novartis, GlaxoSmithKline, MSD, BMS, Celgene,
282 AstraZeneca, Illumina, Sarah Canon Research Institute, Genentech, Roche-Ventana
283 and GRAIL. Advisor for Dynamo Therapeutics. CS has also received research
284 grants/support from Pfizer, AstraZeneca, BMS, Ventana, Roche and is a stock

285 shareholder of Apogen Biotechnologies, Epic Bioscience, Achilles Therapeutics and
286 GRAIL.

287 **References**

- 288 1 Uramoto, H. & Tanaka, F. Recurrence after surgery in patients with NSCLC.
289 *Translational Lung Cancer Research* **3**, 242-249, doi:10.3978/j.issn.2218-
290 6751.2013.12.05 (2014).
- 291 2 Taylor, M. D. *et al.* Tumor recurrence after complete resection for non-small cell lung
292 cancer. *Ann Thorac Surg* **93**, 1813-1820; discussion 1820-1811,
293 doi:10.1016/j.athoracsur.2012.03.031 (2012).
- 294 3 Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer.
295 (2017).
- 296 4 Siegel, R. L., Miller, K. D. & Jemal, A. *CA Cancer J Clin* **67**, 7-30 (2017).
- 297 5 Aceto, N., Toner, M., Maheswaran, S. & Haber, D. A. En Route to Metastasis:
298 Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. *Trends in*
299 *Cancer* **1**, 44-52, doi:10.1016/j.trecan.2015.07.006 (2015).
- 300 6 Hodgkinson, C. *et al.* Circulating tumour cells from small cell lung cancer patients are
301 tumourigenic. *Lung Cancer* **87**, S1-S1 (2015).
- 302 7 Morrow, C. J. *et al.* Tumourigenic non-small-cell lung cancer mesenchymal circulating
303 tumour cells: a clinical case study. *Annals of Oncology* **27**, 1155-1160 (2016).
- 304 8 Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from
305 breast cancer patients that initiates metastasis in a xenograft assay. *Nat Biotechnol*
306 **31**, 539-544 (2013).
- 307 9 Girotti, M. R. *et al.* Application of Sequencing, Liquid Biopsies, and Patient-Derived
308 Xenografts for Personalized Medicine in Melanoma. *Cancer Discovery* **6**, 286-299,
309 doi:10.1158/2159-8290.CD-15-1336 (2016).
- 310 10 Mohan, S., Chami, F. & Brady, G. Challenges and unanswered questions for the next
311 decade of circulating tumour cell research in lung cancer. *Transl Lung Cancer Res* **6**,
312 454-472, doi:10.21037/tlcr.2017.06.04 (2017).
- 313 11 Crosbie, P. A. *et al.* Circulating Tumor Cells Detected in the Tumor-Draining
314 Pulmonary Vein Are Associated with Disease Recurrence after Surgical Resection of
315 NSCLC. *Journal of Thoracic Oncology* **11**, 1793-1797 (2016).
- 316 12 Jamal-Hanjani, M. *et al.* Tracking genomic cancer evolution for precision medicine:
317 the lung TRACERx study. *PLoS Biol* **12** (2014).
- 318 13 Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer
319 evolution. *Nature* **545**, 446, doi:10.1038/nature22364 (2017).
- 320 14 Romero-Palacios, P. J. *et al.* Liquid biopsy beyond of cancer: Circulating pulmonary
321 cells as biomarkers of COPD aggressivity. *Crit Rev Oncol Hematol* **136**, 31-36,
322 doi:10.1016/j.critrevonc.2019.02.003 (2019).
- 323 15 Deleye, L., Vander Plaetsen, A. S., Weymaere, J., Deforce, D. & Van Nieuwerburgh, F.
324 Short Tandem Repeat analysis after Whole Genome Amplification of single B-
325 lymphoblastoid cells. *Scientific reports* **8**, 1255, doi:10.1038/s41598-018-19509-5
326 (2018).
- 327 16 Lin, C. W. *et al.* MicroRNA-135b promotes lung cancer metastasis by regulating
328 multiple targets in the Hippo pathway and LZTS1. *Nature communications* **4**, 1877,
329 doi:10.1038/ncomms2876 (2013).
- 330 17 Heitzer, E. *et al.* Complex Tumor Genomes Inferred from Single Circulating Tumor
331 Cells by Array-CGH and Next-Generation Sequencing. *Cancer Research* **73**, 2965
332 (2013).

333 18 Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a
334 window into metastatic prostate cancer. *Nat Biotech* **32**, 479-484,
335 doi:10.1038/nbt.2892 (2014).
336

337 **Figure Legends**

338 **Fig. 1: PV-CTC detection in early NSCLC.** **a**, TRACERx consort diagram. 149
339 patients consented for pulmonary vein blood sampling between June 2014 and
340 March 2017. 27 samples were excluded because of failures in CellSearch®
341 enrichment and enumeration. 22 patients were defined ineligible post-surgery and
342 the remaining 100 patients constituted the final cohort for PV-CTC
343 enumeration. **b**, Distribution of the number of PV-CTCs enumerated by CellSearch®
344 from 100 patients with early NSCLC. LUAD (blue circle) and non-LUAD (red circle)
345 patients are indicated. **c**, Heat map showing clinicopathological and PV-CTC
346 detection data; Patients are stratified according to PV-CTC detection. Histological
347 disease type is indicated by coloured bar above the heatmap.

348
349 **Fig.2: PV-CTCs as independent predictors of disease-free survival.** **a**, Kaplan–
350 Meier curves showing disease-free survival (DFS) of 100 patients stratified as PV-
351 CTC high or low based on the previously published threshold from our pilot study
352 (≥ 18 PV-CTCs/7.5ml blood)¹¹. The number of patients at risk for every time point is
353 indicated below the time point and colour coded according to the high or low groups.
354 P value, HR and relative 95% confidence intervals (CI) (two-sided log-rank test) are
355 indicated. **b**, Forest plot showing the results of multivariable regression analysis for
356 PV-CTC high or low patients (≥ 18 PV-CTCs/7.5ml blood). The x-axis represents the
357 hazard ratio with the reference line (dashed) and significance is calculated using a
358 Cox proportional hazards model. The estimated hazard ratios and their 95% CI are
359 presented as error bars. The log-rank test used was two-sided.

360
361 **Fig.3: Mutations present in the relapse tumour are detected 10 months earlier**
362 **in PV-CTCs and not in the primary tumour.** **a**, Patient timeline from diagnosis to
363 death (FU=follow up; PET=positron emission tomography; MR=magnetic
364 resonance). **b**, Heat map showing the comparison between CNA detected in PV-
365 CTCs or circulating epithelial cells (CECs), in primary tumour regions (R1-3), in
366 relapse tumour (Met) and in a WBC control. Regions of loss are coloured blue,
367 regions of gain are coloured red. Chromosomes are indicated at the top of the figure.

368 **c**, Heat map showing the comparison of SNVs detected in PV-CTCs, primary tumour
369 regions and the metastasis. Mutations are ordered according to their clonality as
370 established by primary tumour analysis. Green dashed boxes indicate mutations that
371 are seen in the primary tumour, but not metastasis or PV-CTCs. Blue dashed box
372 indicates the overlap between mutations considered metastatic private by primary
373 tumour analysis and PV-CTCs. No mutations were found in the three CECs and two
374 WBCs. **d**, Evolutionary tree encompassing tumour and PV-CTCs: the relationships
375 between identified subclones is depicted, with size of circle reflecting the number of
376 mutations in each subclone relative to largest. Length of lines connecting tumor
377 subclones does not carry information. The beehive plots indicate the subclonal
378 architecture of each tumour region, with 100 representative cells shown for each
379 region and the nested colours corresponding to the ancestry of each cell. **e**, Heat
380 map showing PV-CTC private mutations that are detected in primary tumour,
381 metastasis and cfDNA following targeted deep sequencing.

382

383 **Extended Data Fig.1: a**, Time-dependent receiver operating characteristic (ROC)
384 curves showing true positive and false positive rates for the 65th, 75th, 85th PV-CTC
385 quantiles (≥ 3 , ≥ 7 and ≥ 39 PV-CTCs/7.5ml blood respectively) alongside the
386 previously published threshold from our pilot study (≥ 18 PV-CTCs/7.5ml blood)¹¹. All
387 predictions were made at 720 days. Sensitivity and specificity of each category is
388 shown along with area under ROC curve (AUROC) value. **b**, Kaplan–Meier curve
389 showing lung cancer specific relapse free survival for 98 patients stratified as PV-
390 CTC high or low according to the 75th quantile (≥ 7 PV-CTCs/7.5ml blood). The
391 number of patients at risk for every time point is indicated below the time point and
392 colour coded according to the high or low groups. P value, HR and relative 95%
393 confidence intervals (CI) (two-sided log-rank test) are indicated. **c**, Forest plot
394 showing the results of multivariable regression analysis for PV-CTC high or low
395 patients (≥ 7 PV-CTCs/7.5ml blood). The x-axis represents the hazard ratio with the
396 reference line (dashed) and significance is calculated using a Cox proportional
397 hazards model.

398

399 **Extended Data Fig.2: a**, Consort diagram describing samples used for downstream
400 analysis. Only patients with ≥ 5 PV-CTCs (29) were processed through single cell
401 isolation (DEPArray™). Single cells were not isolated from 6 out of the 29 samples
402 due to failures during sample loading into the DEPArray™ machine. From the
403 remaining 23 samples, 7 patients whose single CTCs isolated did not meet
404 morphology criteria (see methods) were excluded. 16 samples were processed for
405 whole genome amplification (WGA) and 2 patients whose CTCs did not show good
406 quality genomic integrity index in QC post-WGA were removed (see methods). **b**,
407 Table showing cases of relapse among the patients with single PV-CTCs isolated. **c**,
408 Agarose gel showing results of a QC-PCR assay used to determine the genome
409 integrity of each sample. 0–4 bands determine the overall DNA integrity of each
410 sample. DEPArray images of corresponding PV-CTC (cytokeratin (CK)+ stained
411 green, CD45+ stained blue, DAPI+ stained purple) are shown above. **d**, Examples of
412 copy number profiles detected in single PV-CTCs, CECs and WBC control. Blue and
413 red indicate regions of copy number loss and gain respectively.

414

415 **Extended Data Fig.3: a**, Venn diagram showing the overlap of somatic mutations
416 detected between single PV-CTCs, primary and metastatic tumour. **b**, Venn diagram
417 showing the overlap of somatic mutations detected between single PV-CTCs,
418 metastatic tumour and cfDNA isolated at the time of relapse.

419

420 **Extended Data Fig.4:** Heat map showing the comparison of SNVs detected in
421 primary tumour regions, metastasis, PV-CTCs, CECs, WBCs, and cfDNA samples
422 (cfDNA pre-surgery was isolated from peripheral blood, cfDNA surgery was isolated
423 from the pulmonary vein and cfDNA relapse was isolated at the time of relapse).
424 Mutations are ordered according to their clonality established by primary tumour
425 analysis.

426

427

428

Table 1- Baseline characteristics of 100 patients and presence of PV-CTCs

Characteristics	PV-CTC positive, n (%)	PV-CTC negative, n (%)
Age, in years Average age Range	68 39-85	67 48-82
Gender Male (n=61) Female (n=39)	28 (46%) 20 (51%)	33 (54%) 19 (49%)
Tumour Histology Adenocarcinoma (n=59) Non-adenocarcinoma (n=43)	28 (47%) 20 (47%)	31 (53%) 23 (53%)
Pathological Stage I (n=47) II (n=34) III (n=19)	22 (47%) 15 (44%) 11 (58%)	25 (53%) 19 (56%) 8 (42%)
Smoking Status Current smokers (n=14) ex smokers (n=78) never smokers (n=8)	7 (50%) 38 (49%) 3 (37%)	7 (50%) 40 (51%) 5 (63%)

429

430 **Online Methods**

431 **Patients and pathology review**

432 The cohort of 100 patients evaluated here for PV-CTC detection within this study
433 comprises patients analysed by the lung TRACERx study
434 (<https://clinicaltrials.gov/ct2/show/NCT01888601>). Patient eligibility and exclusion
435 criteria for TRACERx enrolment is described in Jamal-Hanjani *et al*² but briefly
436 patients had given their informed written consent to participate in the study, were at
437 least 18 years of age, had received a diagnosis of NSCLC in stages IA through IIIA
438 and not received previous systemic therapy. The study has received a favourable
439 opinion from the NRES Committee London – Camden & Islington Research Ethics
440 Committee. The clinical data used in this study was derived from the “*February 2019*
441 *TRACERx data release*”. The NSCLC cohort in this study consisted of lung
442 adenocarcinoma (LUAD) (59%) and remaining 41% of non-adenocarcinoma
443 histology (Extended Data Fig. 1b and Supplementary Tables 1 and 2). The median
444 age of patient was 68 and the population consisted of 61 males and 39 females
445 (Table 1).

446 Digital images of diagnostic tumour sections from all cases were reviewed in detail
447 centrally by at least one pathologist, and in cases of uncertainty, by two. Histological
448 subtype and mitotic rate (number of visible mitoses per high-power field) were
449 evaluated on digital images from scanned diagnostic slides blinded to the PV-CTC
450 detection status of the patient in question.

451

452 **Statistical analysis**

453 All statistical tests were 2-sided unless otherwise stated. The association of PV-CTC
454 count with individual clinical characteristics, including gender, stage, histology,

455 smoking status, chemotherapy received, sites of relapse were evaluated using
456 ANOVA, while age and mitotic rate were evaluated using Pearson's correlation. PV-
457 CTC count was log2 transformed in all analysis.

458

459 **Cox proportional hazard regression analysis**

460 The association between PV-CTC count and patient survival (DFS or lung cancer
461 specific relapse) was assessed by including it as a sole covariate in a Cox
462 proportional hazards model. Assumption of proportionality was verified based on
463 Schoenfeld residuals¹⁹. A plot of the Martingale residuals was examined for evidence
464 of nonlinearity²⁰. The same uni-variate analysis was carried out on each clinical
465 characteristic. Significant covariates in the uni-variate analysis were selected for
466 subsequent multi-variate analysis, where a backward stepwise method was applied
467 to investigate the impact of PV-CTC count on survival with other significant clinical
468 characteristics under control.

469 Time-dependent receiver operating characteristics (tdROC) curves were applied to
470 evaluate the performance of predicting lung cancer specific relapse using PV-CTC
471 counts stratified by the 65th, 75th, 85th quantiles and the previously published
472 threshold from our pilot study¹¹ (≥ 18 PV-CTCs/7.5ml blood) within 720 days post-
473 surgery. This analysis showed the upper quartile (75th quantile) had the highest
474 AUROC (0.58, Extended Data Fig.1a). The diagnostic odds ratio (DOR) was also
475 calculated for each PV-CTC cutoff. In order to avoid data overfitting, these DOR
476 values were fitted into a polynomial curve, and the optimal cutoff for PV-CTC counts
477 was selected as the one that corresponds to the maximum point of the curve.

478 All analysis were performed according to REMARK guideline ²¹, using R version
479 3.5.1²². R packages survival (v2.38)²³, and survminer (v0.3)²⁴ and survivalROC
480 (v1.0.3)²⁵ were applied.

481

482 **Lung cancer specific relapse event analysis**

483 We collected available clinical data from all 37 patients who had been reported as
484 having experienced a DFS event (defined as the time from study enrolment until
485 recurrence of tumour or death from any cause) in the *February 2019 TRACERx data*
486 *release*. Clinical data was available for 35 of 37 patients, 2 patients without available
487 data (CRUK0005 and CRUK0770) were excluded from this analysis. We defined a
488 lung cancer specific relapse event as histological or imaging confirmed NSCLC
489 relapse. Nine of 37 patients who experienced a DFS event died without evidence of
490 a lung cancer specific relapse event (details in Supplementary Table 3). These
491 patients were either censored at the point of last computed tomography (CT) scan
492 imaging prior to death showing the absence of metastatic disease (CRUK0056,
493 CRUK0431, CRUK0416, CRUK0260, CRUK0017, CRUK0301) or in the event of
494 immediate post-operative death (death within 30 days of surgery), at the point of
495 death (CRUK0196, CRUK0223, CRUK0681). Four of 37 patients experienced
496 metastatic disease unrelated to their original lung primary (CRUK0768,
497 CRUK0068, CRUK0759, CRUK0085) and were censored at the point of last CT
498 imaging prior to death showing absence of metastatic NSCLC. These cases were
499 classified as second primary malignancies based on consensus imaging, histological
500 and clinical agreement. For 1 of 37 patients there was high clinical suspicion of a
501 second malignancy based on CT imaging but due to lack of investigation this was
502 not conclusively determined, therefore this patient was excluded from the analysis

503 (CRUK0073). Initial site of clinical relapse was defined as extracranial if no brain
504 metastasis were clinically confirmed within 60 days of clinical relapse or intracranial if
505 a patient presented with brain metastases within 60 days of clinical relapse.

506

507 **Blood collection**

508 A blood sample (10mL) was taken intra-operatively from the cancer-draining
509 pulmonary vein prior to vessel ligation and tumour resection for each patient. A
510 second sample was taken from the peripheral vein of patients recruited in
511 Manchester. Blood samples were stored at room temperature for up to 96 hours in
512 CellSave vacutainers prior to analysis.

513

514 **CTC enrichment enumeration and single cell isolation**

515 Blood samples were processed using the CellSearch system (Menarini), according
516 to the manufacturer's instructions. Epithelial CTCs (via EpCAM dependent capture)
517 were classified and counted based on an intact DAPI stained nucleus and positive
518 immunofluorescent staining for pan-cytokeratins (CK) and negative staining for the
519 WBC marker CD45. Following CellSearch® enrichment, single cells were isolated
520 using the DEPArray™ system (Menarini) according to the manufacturer's
521 instructions. Images of isolated PV-CTCs were manually inspected by two
522 independent operators to confirm that the following morphological criteria were met:
523 (1) cells were unambiguous positive for cytokeratin, (2) had an intact nucleus and (3)
524 were clear of contaminating WBCs. Cells that failed to meet any of the three criteria
525 were considered “ambiguous” and excluded from all downstream analysis.

526

527 **Whole genome amplification**

528 Whole genome amplification (WGA) was performed using the Ampli1 WGA kit
529 (Menarini) according to the manufacturer's instructions. The efficacy of WGA was
530 then evaluated by a multiplex quality control PCR (Ampli1 QC kit, Menarini) as
531 previously described²⁶ followed by visualization of PCR bands on a 1.5% (w/v)
532 agarose gel. This quality control step allowed us to establish a Genome Integrity
533 Index (GII) of 0–4 for each sample and single cells with $GII \geq 2$ were considered with
534 good quality DNA and eligible for subsequent downstream analysis.

535

536 **Circulating cell-free DNA and tumour samples preparation**

537 Plasma from CellSave blood samples was separated for cfDNA extraction as
538 previously described²⁷. Genomic DNA from primary and relapse tumours was
539 isolated as described in Jamal-Hanjani *et al*¹², sheared and quantified along with
540 cfDNA and germline samples using the TaqMan RNase P Detection Kit (Life
541 Technologies) as per manufacturer's instructions.

542

543 **DNA library preparation, targeted enrichment and next-generation sequencing**

544 DNA libraries for PV-CTCs and WBCs were prepared using NEBNext Ultra DNA
545 Library Prep Kit for Illumina (New England BioLabs) with 50 ng of DNA added per
546 library preparation. DNA libraries for cfDNA, tumour DNA and germline were
547 prepared using NEBNext Ultra II End Repair/dA-Tailing Module (New England
548 BioLabs) and KAPA Hyper Library Prep Kit (KAPA Biosystems) using an input of up
549 to 25 ng DNA. Each library was quantified (KAPA library quantification kit, KAPA
550 Biosystems) and equimolar amounts were pooled and shallow-depth whole genome
551 sequencing was performed on Illumina MiSeq or NextSeq 500 desktop sequencers
552 (paired end, 300 cycles).

553 PV-CTC and WBC Libraries from patient CRUK0242 were additionally subjected to
554 targeted exome enrichment using SureSelect Human All Exon V6 (Agilent) and
555 Whole Exome Sequencing (WES) was performed on Illumina NextSeq 500 desktop
556 sequencer for the detection of somatic mutations (paired end, 300 cycles). WES of
557 corresponding excised primary tumour regions was performed as previously
558 described³. For patient CRUK0242, libraries of cfDNA, isolated at surgery and at
559 relapse, were enriched for a panel of 520 (SureSelectXT Custom, Agilent) pre-
560 identified mutations and sequenced as above.

561 **Sequence alignment and data processing**

562 After trimming of sequencing adapters, the single cell sequencing reads (fastq
563 format) were aligned to human genome assembly 19 (hg19), using the Burrows-
564 Wheeler Aligner (BWA) mem (v0.7.13) algorithm²⁸ to generate SAM files. SAMtools
565 (v0.1.19) was used to convert the SAM files to BAM files, to remove reads with low
566 mapping quality (MQ < 10) and to merge files from the same cell. Picard tools
567 (v1.96) was used to sort the BAM files by chromosome coordinates and to remove
568 PCR duplicates. The BAM files were converted to BED files using Bedtools²⁹. A
569 combination of Picard tools, Bedtools and FastQC³⁰ was used to generate quality
570 control metrics.

571

572 **WGA Capture-rate**

573 To establish the capture-rate of the WGA process, we used targeted sequencing
574 data (described above) for comparisons of the germline (GL), WGA germline (WGA-
575 GL) and individual single cells (including WBC controls) following WGA. A list of
576 heterozygous single nucleotide polymorphisms (SNPs) detected within the targeted
577 regions of the germline sample was generated using Mutect (v1.1.7)³¹. SAMtools

578 mpileup was then used to check which of these SNPs were detected in each WGA
579 sample, requiring a minimum of ten reads to call the SNPs (average read depth in
580 successfully amplified regions is ~230 reads) and a Variant Allele Frequency (VAF)
581 of 0.2-0.8 to consider it to still be heterozygous. The WGA-GL sample shows a
582 complete locus drop-out of 18% due to lack of amplification in the WGA process. Of
583 the 113 heterozygous SNPs that are present in the WGA-GL, 51 and 54 are also
584 called as heterozygous in the two WBC controls. In addition 16 loci became
585 homozygous for the SNP in each cell, and 12 and 13 loci becoming homozygous for
586 the reference allele due to allele drop out (Supplementary Table 6). This gives an
587 estimate for the allele capture-rate of 58-61% of the 113 WGA-GL SNPs due to the
588 single cell sequencing.

589

590 **Copy Number Analysis**

591 Illumina whole-genome data for PV-CTCs, WBCs and tumour samples were aligned
592 to the human genome using BWA. For CNA analysis we only analysed samples with
593 a minimum of 2 million reads (after duplicate removal). Copy number alterations
594 were identified using the R Bioconductor package HMMcopy (v1.18)³² with the
595 genome divided into 1 Mb windows. Reads in each window were normalized by GC-
596 content and mappability, and a Hidden Markov Model-based approach was used to
597 segment the data into regions of similar copy-number profile and to predict a CNA
598 event for each segment.

599

600 **Somatic Mutation analysis from whole exome and targeted sequencing data**

601 For the tumour WES, high-confidence variant calls from tumour were obtained as
602 previously described³, using a combination of VarScan2 and MuTect.

603 MuTect (v1.1.7)³¹ was used to detect SNVs utilising annotation files contained in
604 GATK bundle. All variants called by MuTect were filtered according to the filter
605 parameter 'PASS' in the judgement column. All variants were annotated using
606 ANNOVAR³³. Only variants with at least 20 reads were considered for further
607 filtering.

608 To generate a high-confidence set of variant calls from PV-CTCs, the following filters
609 were applied:

- 610 1. Using the annotations as provided by ANNOVAR, all variants that were
611 present in either 1000g or the Exac03 databases are removed.
- 612 2. A blacklist filter, relating to the genomic location of the variant, was applied.
613 The blacklisted genomic regions were obtained from UCSC Genome Table Browser
614 and include regions excluded from the Encode project (both DAC and Duke list),
615 simple repeats, segmental duplications and microsatellite regions.
- 616 3. Variants with VAF < 0.2 are removed.
- 617 4. Variants had to be either present in the Tumour tissue (Primary or Relapse) or
618 in at least one other single PV-CTC.
- 619 5. Lastly, if any variant is called in any of the WBC controls, then those were
620 filtered out.

621 Supplementary Tables 9-10 and Supplementary Table 11 contain the information
622 relative to coverage and VAF for each mutation detected in the primary tumour and
623 single cells by WES and targeted deep sequencing respectively.

624

625 For the two WBC controls from patient CRUK0242, the first three filtering steps give
626 134 and 253 variants, none of which are shared with the tumour or any other single
627 cells while the non-matching three CECs have 254, 260 and 307 private SNVs. The

628 rate of false positives due to sequencing artefacts, with a range of 134-1960 variants
629 seen in white blood samples from two other patients (see Supplementary Table 12)
630 has very little overlap between them. The requirement that a mutation must be
631 present in two or more samples (whether tumour or single cell) therefore eliminates
632 the vast majority of false positives as a very conservative procedure.

633

634 Regions containing mutations detected in the primary tumour or metastasis which
635 were not covered in at least 1 of the three PV-CTC samples were removed for the
636 calculation of overlaps, although they are shown in the Extended Data Fig.5.

637

638 All somatic variants detected in PV-CTCs were analysed by using cancer genome
639 interpreter platform³⁴ to interpret whether the variants detected had potential as
640 drivers in NSCLC as well as in other solid cancers.

641

642 **Phylogenetic Analysis**

643 Phylogenetic analysis was performed as previously described³. In brief, using the
644 pigeon-hole principle (if the average cancer cell fraction of two subclones sums to
645 more than 1, the smaller subclone must be nested within the larger) as well as the
646 crossing rule (if the cancer cell fraction of subclone A and subclone B sums to less
647 than 1 and the cancer cell fraction of subclone A exceeds that of subclone B in one
648 region but the inverse is true in another region, subclone A and B must exist on
649 separate branches of the evolutionary tree), the evolutionary relationships between
650 subclones was determined and a phylogenetic tree inferred.

651

652 **Data Availability Statement**

653 The majority of data generated or analysed during this study are included in this
654 published article. The sequencing data are available through the Cancer Research
655 UK & University College London Cancer Trials Centre for non-commercial research
656 purposes and access will be granted upon review of a project proposal that will be
657 evaluated by a TRACERx data access committee and entering into an appropriate
658 data access agreement subject to any applicable ethical approvals.

659

660 **References (Methods)**

- 661 19 Gill, R. & Schumacher, M. A Simple Test of the Proportional Hazards Assumption.
662 *Biometrika* **74**, 289-300, doi:10.2307/2336143 (1987).
- 663 20 Therneau, T. M., Grambsch, P. M. & Fleming, T. R. Martingale-Based Residuals for
664 Survival Models. *Biometrika* **77**, 147-160, doi:10.2307/2336057 (1990).
- 665 21 Altman, D. G., McShane, L. M., Sauerbrei, W. & Taube, S. E. Reporting
666 Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and
667 elaboration. *PLoS Med* **9**, e1001216, doi:10.1371/journal.pmed.1001216 (2012).
- 668 22 Team, R. C. R: A language and environment for statistical computing. (2013).
- 669 23 Therneau, T. A Package for Survival Analysis in S. version 2.38.(2015).
- 670 24 Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves
671 using 'ggplot2'. *R package version 0.3 1* (2017).
- 672 25 Heagerty, P. J. & Zheng, Y. Survival model predictive accuracy and ROC curves.
673 *Biometrics* **61**, 92-105, doi:10.1111/j.0006-341X.2005.030814.x (2005).
- 674 26 Mesquita, B. *et al.* Molecular analysis of single circulating tumour cells following
675 long-term storage of clinical samples. *Mol Oncol* **11**, 1687-1697, doi:10.1002/1878-
676 0261.12113 (2017).
- 677 27 Rothwell, D. G. *et al.* Genetic profiling of tumours using both circulating free DNA
678 and circulating tumour cells isolated from the same preserved whole blood sample.
679 *Molecular Oncology* **10**, 566-574, doi:10.1016/j.molonc.2015.11.006 (2016).
- 680 28 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
681 transform. *bioinformatics* **25**, 1754-1760 (2009).
- 682 29 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
683 genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033
684 (2010).
- 685 30 Andrews, S. FastQC: a quality control tool for high throughput sequence data
686 (2010).
- 687 31 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and
688 heterogeneous cancer samples. *Nat Biotechnol.* **31**, doi:10.1038/nbt.2514 (2013).
- 689 32 Ha, G. *et al.* Integrative analysis of genome-wide loss of heterozygosity and
690 monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-
691 negative breast cancer. *Genome research* **22**, 1995-2007, doi:10.1101/gr.137570.112
692 (2012).
- 693 33 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
694 variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164-
695 e164 (2010).
- 696 34 Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical
697 relevance of tumor alterations. *Genome Med* **10**, 25, doi:10.1186/s13073-018-0531-
698 8 (2018).
- 699