ORCA – Online Research @ Cardiff

Abbosh, Christopher, Birkbak, Nicolai J., Wilson, Gareth A., Jamal-Hanjani, Mariam, Constantin, Tudor, Salari, Raheleh, Quesne, John Le, Moore, David A, Veeriah, Selvaraju, Rosenthal, Rachel, Marafioti, Teresa, Kirkizlar, Eser, Watkins, Thomas B K, McGranahan, Nicholas, Ward, Sophia, Martinson, Luke, Riley, Joan, Fraioli, Francesco, Bakir, Maise Al, GrÖnroos, Eva, Zambrana, Francisco, Endozo, Raymondo, Bi, Wenya Linda, Fennessy, Fiona M., Sponer, Nicole, Johnson, Diana, Laycock, Joanne, Shafi, Seema, Czyzewska-Khan, Justyna, Rowan, Andrew, Chambers, Tim, Matthews, Nik, Turajlic, Samra, Hiley, Crispin, Lee, Siow Ming, Forster, Martin D., Ahmad, Tanya, Falzon, Mary, Borg, Elaine, Lawrence, David, Hayward, Martin, Kolvekar, Shyam, Panagiotopoulos, Nikolaos, Janes, Sam M, Thakrar, Ricky, Ahmed, Asia, Blackhall, Fiona, Summers, Yvonne, Hafez, Dina, Naik, Ashwini, Ganguly, Apratim, Kareht, Stephanie, Shah, Rajesh, Joseph, Leena, Quinn, Anne Marie, Crosbie, Phil, Naidu, Babu, Middleton, Gary, Langman, Gerald, Trotter, Simon, Nicolson, Marianne, Remmen, Hardy, Kerr, Keith, Chetty, Mahendran, Gomersall, Lesley, Fennell, Dean A., Nakas, Apostolos, Rathinam, Sridhar, Anand, Girija, Khan, Sajid, Russell, Peter, Ezhil, Veni, Ismail, Babikir, Irvin-sellers, Melanie, Prakash, Vineet, Lester, Jason, Kornaszewska, Malgorzata, Attanoos, Richard, Adams, Haydn, Davies, Helen, Oukrif, Dahmane, Akarca, Ayse U, Hartley, John A, Lowe, Helen L, Lock, Sara, Iles, Natasha, Bell, Harriet, Ngai, Yenting, Elgar, Greg, Szallasi, Zoltan, Schwarz, Roland F, Herrero, Javier, Stewart, Aengus, Quezada, Sergio A, Van Loo, Peter, Dive, Caroline, Lin, C. Jimmy, Rabinowitz, Matthew, Aerts, Hugo JWL, Hackshaw, Allan, Shaw, Jacqui A, Zimmermann, Bernhard G., Swanton, Charles, Jamal-Hanjani, Mariam, Abbosh, Christopher, Veeriah, Selvaraju, Shafi, Seema, Czyzewska-Khan, Justyna, Johnson, Diana, Laycock, Joanne, Bosshard-Carter, Leticia, Goh, Gerald, Rosenthal, Rachel, Gorman, Pat, Murugaesu, Nirupa, Hynds, Robert E, Wilson, Gareth, Birkbak, Nicolai J, Watkins, Thomas B K, McGranahan, Nicholas, Horswell, Stuart, Al Bakir, Maise, GrÖnroos, Eva, Mitter, Richard, Escudero, Mickael, Stewart, Aengus, Van Loo, Peter, Rowan, Andrew, Xu, Hang, Turajlic, Samra, Hiley, Crispin, Goldman, Jacki, Stone, Richard Kevin, Denner, Tamara, Matthews, Nik, Elgar, Greg, Ward, Sophia, Biggs, Jennifer, Costa, Marta, Begum, Sharmin, Phillimore, Ben, Chambers, Tim, Nye, Emma, Graca, Sofia, Al Bakir, Maise, Joshi, Kroopa, Furness, Andrew, Aissa, Assma Ben, Wong, Yien Ning Sophia, Georgiou, Andy, Quezada, Sergio, Hartley, John A, Lowe, Helen L, Herrero, Javier, Lawrence, David, Hayward, Martin, Panagiotopoulos, Nikolaos, Kolvekar, Shyam, Falzon, Mary, Borg, Elaine, Marafioti, Teresa, Simeon, Celia, Hector, Gemma, Smith, Amy, Aranda, Marie, Novelli, Marco, Oukrif, Dahmane, Akarca, Ayse U, Janes, Sam M, Thakrar, Ricky, Forster, Martin, Ahmad, Tanya, Lee, Siow Ming, Papadatos-Pastos, Dionysis, Carnell, Dawn, Mendes, Ruheena, George, Jeremy, Navani, Neal, Ahmed, Asia, Taylor, Magali, Choudhary, Junaid, Summers, Yvonne, Califano, Raffaele, Taylor, Paul, Shah, Rajesh, Krysiak, Piotr, Rammohan, Kendadai, Fontaine, Eustace, Booton, Richard, Evison, Matthew, Crosbie, Phil, Moss, Stuart, Idries, Faiza, Joseph, Leena, Bishop, Paul, Chaturved, Anshuman, Marie Quinn, Anne, Doran, Helen, leek, Angela, Harrison, Phil, Moore, Katrina, Waddington, Rachael, Novasio, Juliette, Blackhall, Fiona, Rogan, Jane, Smith, Elaine, Dive, Caroline, Tugwood, Jonathan, Brady, Ged, Rothwell, Dominic G, Chemi, Francesca, Pierce, Jackie, Gulati, Sakshi, Naidu, Babu, Langman, Gerald, Trotter, Simon, Bellamy, Mary, Bancroft, Hollie, Kerr, Amy, Kadiri, Salma, Webb, Joanne, Middleton, Gary, Djearaman, Madava, Fennell, Dean, Shaw, Jacqui A, Le Quesne, John, Moore, David, Thomas, Anne, Walter, Harriet, Riley, Joan, Martinson, Luke, Nakas, Apostolos, Rathinam, Sridhar, Monteiro, William, Marshall, Hilary, Nelson, Louise, Bennett, Jonathan, Riley, Joan, Primrose, Lindsay, Martinson, Luke, Anand, Girija, Khan, Sajid,

Amadi, Anita, Nicolson, Marianne, Kerr, Keith, Palmer, Shirley, Remmen, Hardy, Miller, Joy, Buchan, Keith, Chetty, Mahendran, Gomersall, Lesley, Lester, Jason, Edwards, Alison, Morgan, Fiona, Adams, Haydn, Davies, Helen, Kornaszewska, Malgorzata, Attanoos, Richard, Lock, Sara, Verjee, Azmina, MacKenzie, Mairead, Wilcox, Maggie, Bell, Harriet, Iles, Natasha, Hackshaw, Allan, Ngai, Yenting, Smith, Sean, Gower, Nicole, Ottensmeier, Christian, Chee, Serena, Johnson, Benjamin, Alzetani, Aiman, Shaw, Emily, Lim, Eric, De Sousa, Paulo, Tavares Barbosa, Monica, Bowman, Alex, Jordan, Simon, Rice, Alexandra, Raubenheimer, Hilgardt, Proli, Chiara, Elena Cufari, Maria, Ronquillo, John Carlo, Kwayie, Angela, Bhayani, Harshil, Hamilton, Morag, Bakar, Yusura, Mensah, Natalie, Ambrose, Lyn, Devaraj, Anand, Buderi, Silviu, Finch, Jonathan, Azcarate, Leire, Chavan, Hema, Green, Sophie, Mashinga, Hillaria, Nicholson, Andrew G, Lau, Kelvin, Sheaff, Michael, Schmid, Peter, Conibear, John, Ezhil, Veni, Ismail, Babikir, Irvin-sellers, Melanie, Prakash, Vineet, Russell, Peter, Light, Teresa, Horey, Tracey, Danson, Sarah, Bury, Jonathan, Edwards, John, Hill, Jennifer, Matthews, Sue, Kitsanta, Yota, Suvarna, Kim, Fisher, Patricia, Keerio, Allah Dino, Shackcloth, Michael, Gosney, John, Postmus, Pieter, Feeney, Sarah, Asante-Siaw, Julius, Constatin, Tudor, Salari, Raheleh, Sponer, Nicole, Naik, Ashwini, Zimmermann, Bernhard, Rabinowitz, Matthew, Aerts, Hugo J. W. L., Dentro, Stefan, Dessimoz, Christophe, Swanton, Charles, Jamal-Hanjani, Mariam, Abbosh, Christopher, Shiu, Kai-Keen, Bridgewater, John, Hochauser, Daniel, Van Loo, Peter, Quezada, Sergio, Beck, Stephan, Parker, Peter, Walczak, Henning, Enver, Tariq, Falzon, Mary, Proctor, Ian, Sinclair, Ron, Lok, Chi-wah, Novelli, Marco, Marafioti, Teresa, Borg, Elaine, Mitchison, Miriam, Trevisan, Giorgia, Lynch, Mark, Brandner, Sebastian, Gishen, Faye, Tookman, Adrian, Stone, Paddy, Sterling, Caroline, Larkin, James, Turajlic, Samra, Attard, Gert, Eeles, Ros, Foster, Chris, Bova, Steve, Sottoriva, Andrea, Chowdhury, Simon, Ashish, Chandra, Spicer, James, Stares, Mark, Lynch, Joanna, Caldas, Carlos, Brenton, James, Fitzgerald, Rebecca, Jimenez-Linan, Merche, Provenzano, Elena, Cluroe, Alison, Stewart, Grant, Watts, Colin, Gilbertson, Richard, McDermott, Ultan, Tavare, Simon, Maughan, Tim, Tomlinson, Ian, Campbell, Peter, McNeish, Iain, Biankin, Andrew, Chambers, Antony, Fraser, Sioban, Oien, Karin, Krebs, Matt, Blackhall, Fiona, Summers, Yvonne, Dive, Caroline, Marais, Richard, Carter, Louise, Nonaka, Daisuke, Quinn, Anne Marie, Dhomen, Nathalie, Fennell, Dean, Le Quesne, John, Moore, David, Shaw, Jacqui, Naidu, Babu, Baijal, Shobhit, Tanchel, Bruce, Langman, Gerald, Collard, Martin, Cockcroft, Peter, Taylor, Joanne, Bancroft, Hollie, Kerr, Amy, Middleton, Gary, Webb, Joanne, Kadiri, Salma, Colloby, Peter, Olisemeke, Bernard, Wilson, Rodelaine, Ottensmeier, Christian, Harrison, David, Loda, Massimo, Flanagan, Adreinne, Wilcox, Maggie, McKenzie, Mairead, Hackshaw, Allan, Lederman, Jonathan, Sharp, Abby, Farrelly, Laura and Swanton, Charles 2017. Phylogenetic ctDNA analysis depicts early stage lung cancer evolution. Nature 545 (7655) , pp. 446-451. 10.1038/nature22364

ARTICLE

Phylogenetic ctDNA analysis depicts early stage lung cancer evolution

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

ARTICLE

Phylogenetic ctDNA analysis depicts early stage lung cancer evolution

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

The early detection of relapse following primary surgery for non-small cell lung cancer and the characterization of emerging subclones seeding metastatic sites might offer new therapeutic approaches to limit tumor recurrence. The potential to non-invasively track tumor evolutionary dynamics in ctDNA of early-stage lung cancer is not established. Here we conduct a tumour-specific phylogenetic approach to ctDNA profiling in the first 100 TRACERx (TRAcking

[1]Cancer Research UK Lung Cancer Centre of Excellence London and Manchester, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 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, 21 University street, London, United Kingdom, WC16JJ. [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, 250 Euston Rd, 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, UCL Respiratory, 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 Hospital South Manchester, Manchester, M23 9LT. [20]Department of Pathology, University Hospital South Manchester, Manchester, M23 9LT. [21]North West Lung Centre, University Hospital South Manchester, Manchester, United Kingdom, M23 9LT. [22]Department of Thoracic Surgery, Birmingham Heartlands Hospital, Birmingham, United Kingdom, B9 5SS. [23]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, United Kingdom, B15 2TT. [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 Cellular Pathology, University Hospital of Wales and Cardiff University, Heath Park Cardiff, Wales U.K. [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]University College London Experimental Cancer Medicine Centre 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. [54]Cancer Genomics Laboratory, The Francis Crick Institute, 1 Midland Rd London NW1 1AT. [55]Department of Human Genetics, University of Leuven, B-3000 Leuven, Belgium. [56]Cancer Research UK Manchester Institute, University of Manchester, Wilmslow Road, Manchester, United Kingdom, M20 4BX. [57]Dana-Farber Cancer Institute, 450 Brookline Ave. Boston, United States, MA 02215-5450.

* These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Lists of participants and their affiliations appear at the end of the paper.

non–small cell lung Cancer Evolution through therapy (Rx)) study participants, including one patient co-recruited to the PEACE (Posthumous Evaluation of Advanced Cancer Environment) post-mortem study. We identify independent predictors of ctDNA release and perform tumor volume limit of detection analyses. Through blinded profiling of post-operative plasma, we observe evidence of adjuvant chemotherapy resistance and identify patients destined to experience recurrence of their lung cancer. Finally, we show that phylogenetic ctDNA profiling tracks the subclonal nature of lung cancer relapse and metastases, providing a new approach for ctDNA driven therapeutic studies

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

Here, we report a bespoke multiplex-PCR NGS approach to ctDNA profiling within the context of the prospective tumor evolutionary NSCLC TRACERx study. We address determinants of ctDNA detection in early-stage NSCLC and investigate the ability of ctDNA to identify and genomically characterize post-operative NSCLC relapse within a tumor phylogenetic framework.

Phylogenetic ctDNA profiling

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

Determinants of ctDNA detection in NSCLC

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

At least two SNVs were detected in ctDNA pre-operatively in 46 of 96 (48%) early-stage NSCLCs and a single SNV was detected

in 12 additional cases (Fig. 2a). Centrally reviewed pathological data revealed that ctDNA detection was associated with histological subtype: 97% (30/31) of lung squamous cell carcinomas (LUSCs) and 71% (5 of 7) of other NSCLC subtypes were ctDNA positive, compared with 19% (11/58) of lung adenocarcinomas (LUADs) (Fig. 2a). 94% (16 of 17) of stage I LUSCs were detected compared with 13% (5 of 39) of stage I LUADs (Extended Data Fig. 3a). Passive release of ctDNA into the circulation may be associated with necrosis[11]. As expected LUSCs were significantly more necrotic than LUADs[12] and ctDNA positive LUADs formed a sub-group of more necrotic tumors compared with ctDNA negative LUADs (Extended Data Fig. 3b). Necrosis, lymph node involvement, lymphovascular invasion, pathological tumor size, Ki67 labelling indices, non-adenocarcinoma histology and total cell-free DNA input predicted ctDNA detection in univariable analyses (Extended Data Fig. 3c). Multivariable analysis revealed non-adenocarcinoma histology, the presence of lympho-vascular invasion and high Ki67 proliferation index as independent predictors of ctDNA detection (Extended Data Fig. 3c). Since FDG-avidity on positron emission tomography (PET) scans correlates with proliferative indices in early-stage NSCLC[13,14], we investigated tumor PET FDG-avidity and ctDNA detection. PET FDG-avidity predicted ctDNA detection (area under curve = 0.84, P < 0.001, n = 92) (Extended Data Fig. 3d). Within LUADs, driver events in KRAS, EGFR or TP53 were not associated with ctDNA detection (Extended Data Fig. 3e).

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

In ctDNA positive patients, pathologic tumor size correlated with mean clonal plasma VAF (Spearman's Rho = 0.405, P = 0.005, n = 46, Extended Data Fig. 4b). CT scan volumetric analyses were available in 38 of 46 ctDNA positive patients (see Extended Data Fig. 4c). Tumor volume correlated with mean clonal plasma VAF (Fig. 3a, Spearman's Rho = 0.61, P < 0.001, n = 38). A linear relationship between log-transformed volume and log-transformed mean clonal VAF values was observed (Fig. 3a). The line of best fit applied to the data was consistent with the line fitted to NSCLC volumetric data and ctDNA plasma VAFs reported in previously published work[16] (Extended Data Fig. 4d). Linear modelling based on the TRACERx data predicted that a primary tumor burden of 10 cm^3 would result in a mean clonal plasma VAF of 0.1% (95% C.I. 0.05 to 0.17%) (Fig. 3b). Tumor purity was multiplied by tumor volume to control for stromal contamination to determine cancer cell volume corresponding to clonal plasma VAF (Extended Data Fig. 4e). On the assumption that 1 cm^3 of pure tumor contains 9.4 x 10^7 cells[17], a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 326 million cells (Fig. 3b, Extended Data Fig. 4f).

To investigate predictors of subclone detection, detected subclonal SNVs were mapped back to M-seq derived tumor phylogenetic trees. 35 of 57 (61%) shared subclones (identified in more than one tumor region through M-Seq analysis) were identified in ctDNA, compared with 26 of 80 (33%) private subclones (detected in a single tumor region only) (Extended Data Fig 4g). This suggested subclone volume influences subclonal ctDNA detection. Subclone volume was estimated based on mean regional subclone cancer cell fraction and cancer cell volume. Detected subclonal SNVs mapped to subclones with higher estimated volumes than subclones containing undetected SNVs (Fig. 3c) and subclone volume correlated with subclonal SNV plasma VAF (Fig. 3d).

Detecting and characterizing NSCLC relapse

The longitudinal phase of the study aimed to determine if ctDNA profiling with patient-specific assay panels could detect and characterize the branched subclone(s) seeding NSCLC relapse. Pre- and post-surgical plasma ctDNA profiling was performed blinded to relapse status in a sub-group of 24 patients (cohort characteristics, Extended Table 1d-e). This included relapse free patients who had been followed-up for a median of 775 days (range 688 to 945 days, n = 10) and confirmed NSCLC relapse cases (n = 14) (cohort design, Extended Data Fig. 2c). Additional PCR assays were added to panels in this phase of the study to attempt to improve ctDNA detection in LUADs, a median of 18.5 SNVs (range 12 to 20) were targeted by LUSC assay-panels and a median of 28 SNVs (range 25 to 30) were targeted by LUAD assay-panels (Extended Data Fig. 2d-e).

Patients were followed up with three to six monthly clinical assessment and chest radiographs. At least 2 SNVs were detected in 13 of 14 (93%) patients with confirmed NSCLC relapse prior to, or at, clinical relapse (Fig. 4a-g, Extended Data Fig. 5). At least two SNVs were detected in 1 of 10 (10%) patients (CRUK0013) with no clinical evidence of NSCLC relapse (Fig. 4h, Extended Data Fig. 6). Excluding a single case where no post-operative plasma was taken prior to clinical relapse (CRUK0041), the median interval between ctDNA detection and NSCLC relapse confirmed on clinically indicated CT imaging (lead-time) was 70 days (range 10 to 346 days). Four of 13 relapse cases exhibited lead-times of more than six months (Fig. 4a-d). In two cases ctDNA detection preceded CT imaging inconclusive for NSCLC relapse by 157 days (CRUK0004, Fig. 4b) and 163 days (CRUK0045, Fig. 4d). ctDNA profiling reflected adjuvant chemotherapy resistance - CRUK0080, CRUK0004 and CRUK0062 had detectable ctDNA in plasma within 30 days of surgery. The number of detectable SNVs increased in all three cases despite adjuvant chemotherapy, with disease recurring within 1 year of surgery (Fig. 4a-c). In contrast, CRUK0013 had 20 SNVs detectable in ctDNA 72 hours after surgery and 13 SNVs detectable prior to adjuvant chemotherapy; 51 days following completion of adjuvant treatment and at post-operative days 457 and 667 no SNVs were detectable and the patient remains relapse free 688 days post-surgery (Fig. 4h). ctDNA profiling detected intracerebral relapse; CRUK0029 had a PET scan performed 50 days prior to surgery demonstrating normal cerebral appearances. ctDNA remained detectable following surgery, 54 days post-operatively the patient was diagnosed with intracerebral metastasis, no extracranial disease was evident on CT imaging (Fig. 4e).

In cases where subclonal SNVs were detected in ctDNA post-operatively we predicted subclonal clusters involved in the relapse process by mapping SNVs detected in plasma back to primary M-Seq data (Fig. 4, Extended Data Fig. 5b-c). Subclonal SNVs displaying plasma VAFs similar to clonal SNVs from clusters confined to a single phylogenetic branch, were detected post-operatively in the ctDNA of four patients who suffered NSCLC relapse (CRUK0004, CRUK0063, CRUK0065 and CRUK0044) (Fig. 4b,f-g, Extended Data Fig 5b). This suggested a relapse process dominated by one subclone represented in our assay-panel. The subclone implicated by ctDNA as driving the relapse in the case of CRUK0004 contained an ERRB2 (HER2)

amplification event (>15 copies, triploid background), that may be targetable in NSCLC[18] (Fig. 4b). Relapses involving subclones from more than one phylogenetic branch were evident in patients CRUK0080, CRUK0062 (Fig. 4a,c) and CRUK0041 (Extended Data Fig 5c).

Validation of phylogenetic characterization

To validate subclonal ctDNA analyses, data acquired from sequencing metastatic tissue at recurrence was integrated with M-seq primary tumor data (for biopsy details, Supplementary Table 3). Patient CRUK0063 suffered para-vertebral relapse of their NSCLC. Post-operative ctDNA analysis revealed the detection of the same subclonal SNV (OR5D18) on four consecutive occasions over a 231-day period (Extended Data Fig. 7a). The OR5D18 SNV traced back to a subclonal cluster private to primary tumor region three (Fig. 5a). CT-guided biopsy tissue was acquired from the para-vertebral metastasis at post-operative day 467. Exome sequencing of relapse tissue revealed the subclonal cluster containing the OR5D18 SNV gave rise to the metastatic subclone (Fig. 5a), this supported ctDNA phylogenetic characterization of relapse. The para-vertebral biopsy contained 88 SNVs not called as present in the primary tumor including an ARID1A stop-gain driver SNV. Re-examination of primary tumor region M-Seq data with a lower SNV calling threshold revealed that 16 of 88 SNVs, including ARID1A, were detectable in primary tumor region three, compared to a maximum of 2 of 88 in other tumor regions (Extended Data Fig. 7b). These data suggest that ctDNA profiling can resolve the primary tumor region from which a low frequency metastatic subclone derives. CRUK0035 developed two liver and one adrenal metastases (Fig. 5b). Sequencing of the metastatic liver deposit revealed that only 109 of 149 SNVs classed as clonal in the primary tumor were detectable in the metastasis. This was suggestive of an ancestral branching event not resolved through primary M-seq analysis (Fig. 5b). Post-operative ctDNA profiling identified clonal SNVs present in the liver metastasis biopsy but also revealed SNVs representing a subclone from the primary tumor (Extended Data Fig 7c). This subclone was not present in the metastatic liver deposit (Fig. 5b). These data may reflect ctDNA identified from the non-biopsied metastases suggesting multiple metastatic events. CRUK0044 suffered a vertebral and right hilar relapse. Post-operatively the same subclonal SNV (OR10K1), was detected in ctDNA on two occasions 85 days apart (Extended Data Fig. 7d). This SNV was represented in a single subclone detected through sequencing hilar lymph-node metastatic tissue, supporting ctDNA findings (Fig. 5c). CRUK0041 suffered an intracerebral, hilar and subcarinal lymph node relapse. Four subclonal SNVs representing both branches of the tumor phylogenetic tree were detectable in ctDNA at relapse. Consistent with these data, sequencing of subcarinal metastatic tissue revealed the presence of subclonal SNVs mapping to both phylogenetic branches (Fig. 5d, Extended Data Fig. 7e). Patient CRUK0013 had detectable ctDNA 3 and 38 days post-operatively. Following adjuvant chemoradiotherapy for lymph-node metastases identified in the pathological specimen, ctDNA levels became undetectable (Fig. 4h). Two involved lymph-nodes were sampled for exome analysis together with M-seq of the primary tumor. Four subclonal SNVs detected in ctDNA post-operatively mapped to an ancestral subclone (describing a subclone that existed during the tumor's evolution) (Extended Data Fig. 7f). This ancestral subclone contained a KRAS amplification (>15 copies, triploid background) and was identified as present in primary tumor and sampled lymph-nodes by M-Seq (Fig. 5e). These data provide phylogenetic characterization of post-operative residual disease that responded to adjuvant chemoradiotherapy (Fig. 4h).

ctDNA profiling in the metastatic setting

Patient CRUK0063 underwent examination through the PEACE post-mortem study 24 hours following death. M-Seq data from the six post-mortem tumor regions (para-aortic, para-vertebral and lung metastases, day 857), the para-vertebral relapse biopsy (day 467) and five primary tumor regions (day 0) were combined to infer the phylogenetic

structure of this patient's NSCLC (Fig. 6a). All seven metastatic tumor regions arose from a single ancestral subclone represented by phylogenetic cluster 8. Six metastatic regions shared a later phylogenetic origin, cluster 12 (Fig. 6b). The single tumor region not containing phylogenetic cluster 12 was sampled from the para-aortic metastasis at autopsy and contained a private subclone represented by phylogenetic cluster 9 (Fig. 6b).

We designed a bespoke ctDNA assay-panel to retrospectively track metastatic subclonal burden. 20 clonal SNVs and a median of 8 subclonal SNVs (range 4 to 15) in each of 9 metastatic subclonal clusters were targeted by the assay-panel (Extended Data Fig. 8). Since 103 variants per time-point were profiled, SNV detection thresholds were increased to maintain platform specificity (see Methods). This resulted in a predicted false positive rate (FPR) of 0.0011 translating to a 10.7% risk of a single false-positive SNV at each time point and a 0.5% risk of 2 false-positive SNVs at each time point when testing 103 SNVs.

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

Discussion

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

Tumor volume correlated with plasma ctDNA VAF (Fig. 3a.). A primary NSCLC tumor volume of $10\,cm^3$ predicted a ctDNA plasma VAF of 0.1%; the VAF conferring optimum sensitivity for most ctDNA platforms. Low-dose CT lung screening can identify lung nodules with diameters from $4\,mm$[22]. Assuming a spherical nodule this would translate to a tumor volume of $0.034\,cm^3$. Based on the relationship between tumor volume and ctDNA plasma VAF observed in this study a tumor volume of $0.034\,cm3$ would equate to a plasma VAF of 1.4×10^{-4}% (95% CI, 6.4×10^{-6} to 0.0031%), at the extreme of detection limits of current ctDNA platforms[23]. Sensitivity of clonal SNV ctDNA directed early NSCLC screening may therefore be constrained by tumor size using current technologies.

A limitation to targeted ctDNA profiling is cost, estimated at $1750 per patient for sequencing a single tumor region, synthesis of a patient-specific assay-panel and profiling of five plasma samples. Adjuvant platinum-based chemotherapy in NSCLC improves cure rates following surgery in only 5% of patients and 20% patients receiving chemotherapy experience acute toxicities[3]. There is a need to increase adjuvant therapy efficacy and better target its use. Bespoke ctDNA profiling can characterize the subclonal dynamics of relapsing NSCLC and identify adjuvant chemotherapy resistance. These findings indicate that drug development guided by ctDNA platforms to identify residual disease, define adjuvant treatment response and target emerging subclones prior to clinical recurrence in NSCLC, with appropriate CLIA validation, are now feasible.

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

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E & Forman D. Global cancer statistics. *CA: A Cancer Journal for Clinicians* 2011;**61**(2):69–90.
2. Siegel RL, Miller KD & Jemal A. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* 2017;**67**(1):7–30.
3. Pignon J-P, Tribodet H, Scagliotti GV, Douillard J-Y, Shepherd FA, Stephens RJ, et al. Lung Adjuvant Cisplatin Evaluation: A Pooled Analysis by the LACE Collaborative Group. *Journal of Clinical Oncology* 2008;**26**(21):3552–9.
4. Landau Dan A, Carter Scott L, Stojanov P, McKenna A, Stevenson K, Lawrence Michael S, et al. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell*;**152**(4):714–26.
5. Beaver JA, Jelovac D, Balukrishna S, Cochran RL, Croessmann S, Zabransky DJ, et al. Detection of Cancer DNA in Plasma of Patients with Early-Stage Breast Cancer. *Clinical Cancer Research* 2014;**20**(10):2643–50.
6. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Science Translational Medicine* 2015;**7**(302):302ra133–302ra133.
7. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Science Translational Medicine* 2016;**8**(346):346ra92–ra92.
8. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* 2014;**12**(7):e1001906.
9. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *New England Journal of Medicine* (accepted, in press) 2017; DOI:(0.1056/NEJMoa1616288.)
10. Jamal-Hanjani M, Wilson GA, Horswell S, Mitter R, Sakarya O, Constantin T, et al. Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. *Annals of Oncology* 2016;**27**(5):862–7.
11. Jr LAD & Bardelli A. Liquid Biopsies: Genotyping Circulating Tumor DNA. *Journal of Clinical Oncology* 2014;**32**(6):579–86.
12. Caruso R, Parisi A, Bonanno A, Paparo D, Quattrocchi E, Branca G, et al. Histologic coagulative tumour necrosis as a prognostic indicator of aggressiveness in renal, lung, thyroid and colorectal carcinomas: A brief review. *Oncology Letters* 2012;**3**(1):16–8.
13. Vesselle H, Schmidt RA, Pugsley JM, Li M, Kohlmyer SG, Vallières E, et al. Lung Cancer Proliferation Correlates with [F-18]Fluorodeoxyglucose Uptake by Positron Emission Tomography. *Clinical Cancer Research* 2000;**6**(10):3837–44.
14. Higashi K, Ueda Y, Yagishita M & Arisaka Y. FDG PET measurement of the proliferative potential of non-small cell lung cancer. *The Journal of Nuclear Medicine* 2000;**41**(1):85.
15. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nature Communications* 2015;6:8760.
16. Newman AM, Bratman SV, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* 2014;**20**(5):548–54.
17. Del Monte U. Does the cell number 109 still really fit one gram of tumor tissue? *Cell Cycle* 2009;**8**(3):505–6.
18. Peters S & Zimmermann S. Targeted therapy in NSCLC driven by HER2 insertions. *Translational Lung Cancer Research* 2014;**3**(2):84–8.
19. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2005;**19**(2):264–71.
20. Keam B, Im S-A, Kim H-J, Oh D-Y, Kim JH, Lee S-H, et al. Prognostic impact of clinicopathologic parameters in stage II/III breast cancer treated with neoadjuvant docetaxel and doxorubicin chemotherapy: paradoxical features of the triple negative breast cancer. *BMC Cancer* 2007;7:203-.
21. Rhee J, Han SW, Oh DY, Kim JH, Im SA, Han W, et al. The clinicopathologic characteristics and prognostic significance of triple-negativity in node-negative breast cancer. *BMC Cancer* 2008;8:307.

22. Team TNLSTR. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011;**365**(5): 395–409.
23. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotech* 2016;**34**(5):547–55.

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

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

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

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

The TRACERx consortium members and affiliations.

Charles Swanton[1,2,5], Mariam Jamal-Hanjani[1], Christopher Abbosh[1], Selvaraju Veeriah[1], Seema Shafi[1], Justyna Czyzewska-Khan[1], Diana Johnson[1], Joanne Laycock[1], Leticia Bosshard-Carter[1], Gerald Goh[1], Rachel Rosenthal[1], Pat Gorman[1], Nirupa Murugaesu[1], Robert E Hynds[1,3], Gareth Wilson[1,2], Nicolai J Birkbak[1,2], Thomas B K Watkins[2], Nicholas McGranahan[1,2], Stuart Horswell[2], Maise Al Bakir[2], Eva Grönroos[2], Richard Mitter[2], Mickael Escudero[2], Aengus Stewart[2], Peter Van Loo[2], Andrew Rowan[2], Hang Xu[2], Samra Turajlic[2,4], Crispin Hiley[2], Jacki Goldman[2], Richard Kevin Stone[2], Tamara Denner[2], Nik Matthews[2], Greg Elgar[2], Sophia Ward[2], Jennifer Biggs[2], Marta Costa[2], Sharmin Begum[2], Ben Phillimore[2], Tim Chambers[2], Emma Nye[2], Sofia Graca[2], Maise Al Bakir[2], Kroopa Joshi[1], Andrew Furness[1], Assma Ben Aissa[1], Yien Ning Sophia Wong[1], Andy Georgiou[1], Sergio Quezada[1], John A Hartley[1], Helen L Lowe[1], Javier Herrero[1], David Lawrence[5], Martin Hayward[5], Nikolaos Panagiotopoulos[5], Shyam Kolvekar[5], Mary Falzon[5], Elaine Borg[5], Teresa Marafioti[5], Celia Simeon[5], Gemma Hector[5], Amy Smith[5], Marie Aranda[5], Marco Novelli[5], Dahmane Oukrif[1], Ayse U Akarca[1], Sam M Janes[5], Ricky Thakrar[5], Martin Forster[5], Tanya Ahmad[5], Siow Ming Lee[5], Dionysis Papadatos-Pastos[5], Dawn Carnell[5], Ruheena Mendes[5], Jeremy George[5], Neal Navani[5], Asia Ahmed[5], Magali Taylor[5], Junaid Choudhary[5], Yvonne Summers[6], Raffaele Califano[6], Paul Taylor[6], Rajesh Shah[6], Piotr Krysiak[6], Kendadai Rammohan[6], Eustace Fontaine[6], Richard Booton[6], Matthew Evison[6], Phil Crosbie[6], Stuart Moss[6], Faiza Idries[6], Leena Joseph[6], Paul Bishop[6], Anshuman Chaturved[6], Anne Marie Quinn[6], Helen Doran[6], Angela leek[7], Phil Harrison[7], Katrina Moore[7], Rachael Waddington[7], Juliette Novasio[7], Fiona Blackhall[8], Jane Rogan[9], Elaine Smith[6], Caroline Dive[9], Jonathan Tugwood[9], Ged Brady[9], Dominic G Rothwell[9], Francesca Chemi[9], Jackie Pierce[9], Sakshi Gulati[9], Babu Naidu[10], Gerald Langman[10], Simon Trotter[10], Mary Bellamy[10], Hollie Bancroft[10], Amy Kerr[10], Salma Kadiri, [10] Joanne Webb[10], Gary Middleton[10], Madava Djearaman[10], Dean Fennell[11], Jacqui A Shaw[11], John Le Quesne[11], David Moore[11], Anne Thomas[11], Harriet Walter[11], Joan Riley[11], Luke Martinson[11], Apostolos Nakas[12], Sridhar Rathinam[12], William Monteiro[13], Hilary Marshall[13], Louise Nelson[12], Jonathan Bennett[12], Joan Riley[12], Lindsay Primrose[12], Luke Martinson[12], Girija Anand[14], Sajid Khan[15], Anita Amadi[16], Marianne Nicolson[17], Keith Kerr[17], Shirley Palmer[17], Hardy Remmen[17], Joy Miller[17], Keith Buchan[17], Mahendran Chetty[17], Lesley Gomersall[17], Jason Lester[18], Alison Edwards[18], Fiona Morgan[19], Haydn Adams[19], Helen Davies[19], Malgorzata Kornaszewska[20], Richard Attanoos[21], Sara Lock[22], Azmina Verjee[22], Mairead MacKenzie[23], Maggie Wilcox[23], Harriet Bell[24], Natasha Iles[24], Allan Hackshaw[24], Yenting Ngai[24], Sean Smith[24], Nicole Gower[24], Christian Ottensmeier[25], Serena Chee[25], Benjamin Johnson[25], Aiman Alzetani[25], Emily Shaw[25], Eric Lim[26], Paulo De Sousa[26], Monica Tavares Barbosa[26], Alex Bowman[26], Simon Jordan[26], Alexandra Rice[26], Hilgardt Raubenheimer[26], Chiara Proli[26], Maria Elena Cufari[26], John Carlo Ronquillo[26], Angela Kwayie[26], Harshil Bhayani[26], Morag Hamilton[26], Yusura Bakar[26], Natalie Mensah[26], Lyn Ambrose[26], Anand Devaraj[26], Silviu Buderi[26], Jonathan Finch[26], Leire Azcarate[26], Hema Chavan[26], Sophie Green[26], Hillaria Mashinga[26], Andrew G Nicholson[26, 27], Kelvin Lau[28], Michael Sheaff[28], Peter Schmid[28], John Conibear[28], Veni Ezhil[29], Babikir Ismail[29], Melanie Irvin-sellers[29], Vineet Prakash[29], Peter Russell[30], Teresa Light[30], Tracey Horey[30], Sarah Danson[31], Jonathan Bury[31], John Edwards[31], Jennifer Hill[31], Sue Matthews[31], Yota Kitsanta[31], Kim Suvarna[31], Patricia Fisher[31], Allah Dino Keerio[31], Michael Shackcloth[32], John Gosney[32], Pieter Postmus[32], Sarah Feeney[32], Julius Asante-Siaw[32], Tudor Constatin[33], Raheleh Salari[33], Nicole Sponer[33], Ashwini Naik[33], Bernhard Zimmermann[33], Matthew Rabinowitz[33], Hugo J.W.L. Aerts[34], Stefan Dentro[35], Christophe Dessimoz[36,37,38].

[1]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, United Kingdom. [2]The Francis Crick Institute, United Kingdom. [3]Lungs for Living, UCL Respiratory, University College London, United Kingdom. [4]The Royal Marsden Hospital, United Kingdom. [5]University College London Hospitals NHS Foundation Trust, United Kingdom. [6]University Hospital of South Manchester, United Kingdom. [7]Manchester Cancer Research Centre Biobank, United Kingdom. [8]Christie NHS Foundation Trust, Manchester, United Kingdom. [9]Cancer Research UK Manchester Institute, United Kingdom. [10]Heart of England NHS Foundation Trust, Birmingham, United Kingdom. [11]Cancer Studies and Molecular Medicine, University of Leicester, United kingdom. [12]Leicester University Hospitals, United Kingdom. [13]National Institute for Health Research Leicester Respiratory Biomedical Research Unit, United Kingdom. [14]North Middlesex Hospital, United Kingdom. [15]Royal Free Hospital, United Kingdom. [16]Barnet Hospital, United Kingdom. [17]Aberdeen Royal Infirmary, United Kingdom. [18]Velindre Cancer Centre, Cardiff, Wales, United Kingdom. [19]Cardiff & Vale University Health Board, Cardiff, Wales, United Kingdom. [20]University Hospital Of Wales Heath Park, Cardiff, Wales, United Kingdom. [21]Department of Pathology, University Hospital of Wales and Cardiff University, Heath Park, Cardiff, Wales, United Kingdom. [22]The Whittington Hospital NHS Trust, United Kingdom. [23]Independent Cancer Patients Voice, United Kingdom. [24]Cancer Research UK & UCL Cancer Trials Centre, United Kingdom. [25]University Hospital Southampton NHS Foundation Trust, United Kingdom. [26]Royal Brompton and Harefield NHS Foundation Trust, United Kingdom. [27]National Heart and Lung Institute, Imperial College, United Kingdom. [28]Barts Health NHS Trust, United Kingdom. [29]Ashford and St. Peter's Hospitals NHS Foundation Trust, United Kingdom. [30]The Princess Alexandra Hospital NHS Trust, United Kingdom. [31]Sheffield Teaching Hospitals NHS Foundation Trust, United Kingdom. [32]Liverpool Heart and Chest Hospital NHS Foundation Trust, United Kingdom. [33]Natera Inc., 201 Industrial Road, Suite 410, San Carlos, CA 94070. [34]Dana-Farber Cancer Institute, Brigham & Women's Hospital, Harvard Medical School, 450 Brookline Ave, JF518, Boston, MA 02115-5450, USA. [35]Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom. [36]Bioinformatics Group, Department of Computer Science, University College London. [37]University of Lausanne. [38]Swiss Institute of Bioinformatics.

The PEACE consortium members and affiliations.

Charles Swanton[1,2,3], Mariam Jamal-Hanjani[1,2,3], Christopher Abbosh[1,2,3], Kai-Keen Shiu[3], John Bridgewater[3], Daniel Hochauser[1,3], Peter Van Loo[2], Sergio Quezada[1], Stephan Beck[1], Peter Parker[2], Henning Walczak[1], Tariq Enver[1], Mary Falzon[3], Ian Proctor[3], Ron Sinclair[3], Chi-wah Lok[3], Marco Novelli[3], Teresa Marafioti[3], Elaine Borg[3], Miriam Mitchison[3], Giorgia Trevisan[3], Mark Lynch[3], Sebastian Brandner[3], Faye Gishen[23], Adrian Tookman[23,24], Paddy Stone[3], Caroline Stirling[3], James Larkin[4,5], Samra Turajlic[2,4,5], Gert Attard[4,5], Ros Eeles[4,5], Chris Foster[4,5], Steve Bova[25], Andrea Sottoriva[4,5], Simon Chowdhury[6], Chandra Ashish[6], James Spicer[6], Mark Stares[1,2,3], Joanna Lynch[4,6], Carlos Caldas[18,19], James Brenton[18], Rebecca Fitzgerald[20], Merche Jimenez-Linan[7], Elena Provenzano[7], Alison Cluroe[7], Grant Stewart[26], Colin Watts[7], Richard Gilbertson[19], Ultan McDermott[7,8], Simon Tavare[27], Tim Maughan[28], Ian Tomlinson[21], Peter Campbell[8], Iain McNeish[17], Andrew Biankin[17], Antony Chambers[17], Sioban Fraser[17], Karin Oien[17], Matt Krebs[9], Fiona Blackhall[10], Yvonne Summers[10], Caroline Dive[9], Richard Marais[9], Louise Carter[10], Daisuke Nonaka[10], Anne Marie Quinn[10], Nathalie Dhomen[9], Dean Fennell[11], John Le Quesne[11], David Moore[11], Jacqui Shaw[11], Babu Naidu[12], Shobhit Baijal[12], Bruce Tanchel[12], Gerald Langman[12], Martin Collard[12], Peter Cockcroft[12], Joanne Taylor[12], Hollie Bancroft[12], Amy Kerr[12], Gary Middleton[12], Joanne Webb[12], Salma Kadiri[12], Dr Peter Colloby[12], Bernard Olisemeke[12], Rodelaine Wilson[12], Christian Ottensmeier[22], David Harrison[29], Massimo Loda[13], Adreinne Flanagan[14], Maggie Wilcox[15], Mairead McKenzie[15], Allan Hackshaw[16], Jonathan Lederman[16], Abby Sharp[16], Laura Farrelly[16]

[1]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, United Kingdom. [2]The Francis Crick Institute, United Kingdom. [3]University College London Hospitals NHS Foundation Trust, United Kingdom. [4]The Royal Marsden Hospital, United Kingdom. [5]The Institute of Cancer Research, London, United Kingdom. [6]Guy's and St Thomas', NHS Foundation Trust, London, United Kingdom. [7]Addenbrooke's Hospital, NHS Foundation Trust, Cambridge, United Kingdom. [8]Wellcome Trust Sanger Institute, Hinxton, United Kingdom. [9]Cancer Research UK Manchester Institute, United Kingdom. [10]Christie NHS Foundation Trust, Manchester, United Kingdom. [11]Leicester University Hospitals, United Kingdom. [12]Heart of England NHS Foundation Trust, Birmingham, United Kingdom. [13]Division of Cancer Studies, King's College London, United Kingdom. [14]Royal National Orthopaedic Hospital, NHS Foundation Trust, London, United Kingdom. [15]Independent Cancer Patients' Voice, London, United Kingdom. [16]Cancer Research UK & UCL Cancer Trials Centre, United Kingdom. [17]Wolfson Wohl Cancer Research Centre, University of Glasgow. [18]Cancer Research UK Cambridge Centre, Cambridge, United Kingdom. [19]Department of Oncology, University of Cambridge, Cambridge, United Kingdom. [20]Medical Research Council (MRC) Cancer Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, United Kingdom. [21]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. [22]University of Southampton and Southampton University Hospitals, Southampton [23]Royal Free Hospital, London, United Kingdom. [24]Marie Curie Hospice, Hampstead, United Kingdom. [25]Prostate Cancer Research Centre, University of Tampere, Finland. [26]Department of Surgery, University of Cambridge, Cambridge, United Kingdom. [27]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom. [28]Cancer Research UK/ Medical Research Council Oxford Institute for Radiation Oncology, Department of Oncology, University of Oxford, Oxford, United Kingdom. [29]University of St Andrews, Fife, United Kingdom.
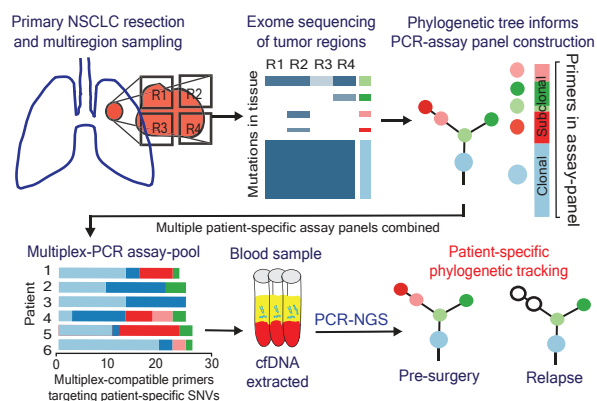
Figure 1 | Phylogenetic ctDNA tracking. Overview of the study methodology. Multi-region sequencing of NSCLC was performed as part of the TRACERx study. PCR assay-panels were designed based on phylogenetic analysis, targeting clonal and subclonal single nucleotide variants to facilitate non-invasive tracking of the patient-specific tumor phylogeny. 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, followed by sequencing of amplicons. Findings were integrated with M-Seq exome data to track tumor evolution.
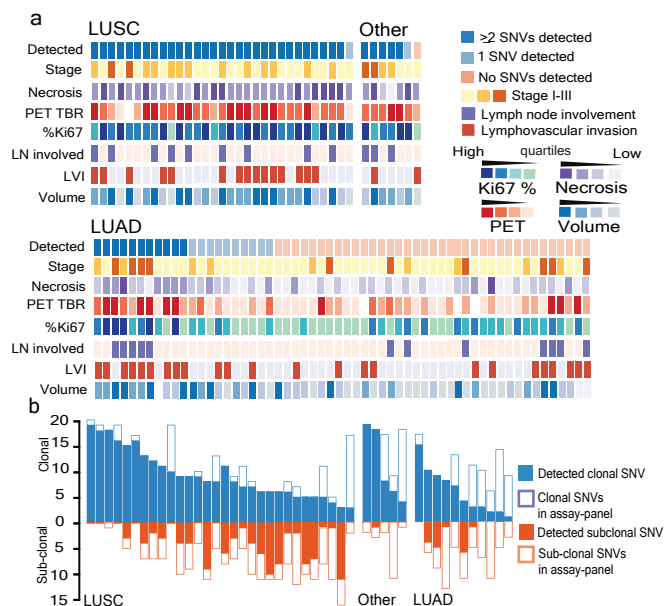
Figure 2 | Clinicopathological predictors of ctDNA detection.
a) Heatmap showing clinicopathological and ctDNA detection data, continuous variables quartiled. Raw data and patient IDs in attached worksheet. b) Detection of clonal and subclonal single nucleotide variants within 46 patients with two or more single nucleotide variants detected in plasma. Histology indicated in panels as LUSC, LUAD and Other.
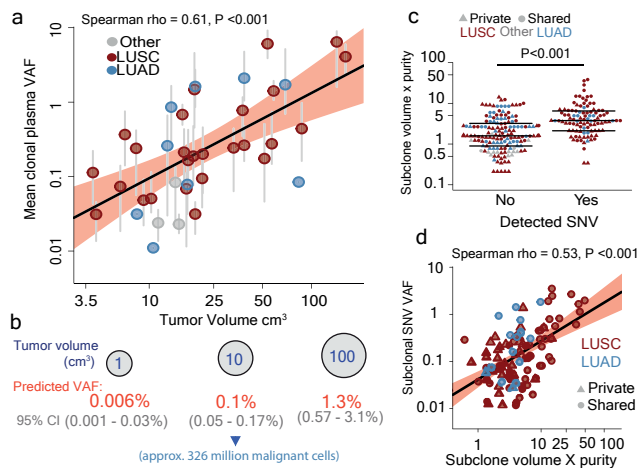
Figure 3 | Tumor volume predicts plasma variant allele frequency. a) Tumor volume (cm³) measured by CT volumetric analysis correlates with mean clonal plasma VAF, n = 38, grey vertical lines represent range of clonal VAF, red shading indicates 95% confidence intervals. b) Predicted mean clonal VAF at hypothetical volumes ranging from 1 to 100 cm³ based on model in panel a, predicted cancer cell number based on model in extended data 4e. c) Estimated effective subclone size, defined as mean CCF of subclone multiplied by tumor volume and purity, influences subclonal SNV detection. For negative calls, median effective subclone size was 1.70 cm³, range = 0.21-24.11, n = 163, for positive calls, median effective subclone size = 4.06 cm³, range = 0.31 – 49.20, n = 109. Wilcoxon rank sum test, P < 0.001, data from 34 patients (passed volumetric filters with subclonal SNVs represented in assay-panel). d) Estimated effective subclone size correlates with subclonal plasma VAF, n = 109 subclonal SNVs, data from 34 patients (passed volumetric filters with detected subclonal SNVs in plasma).
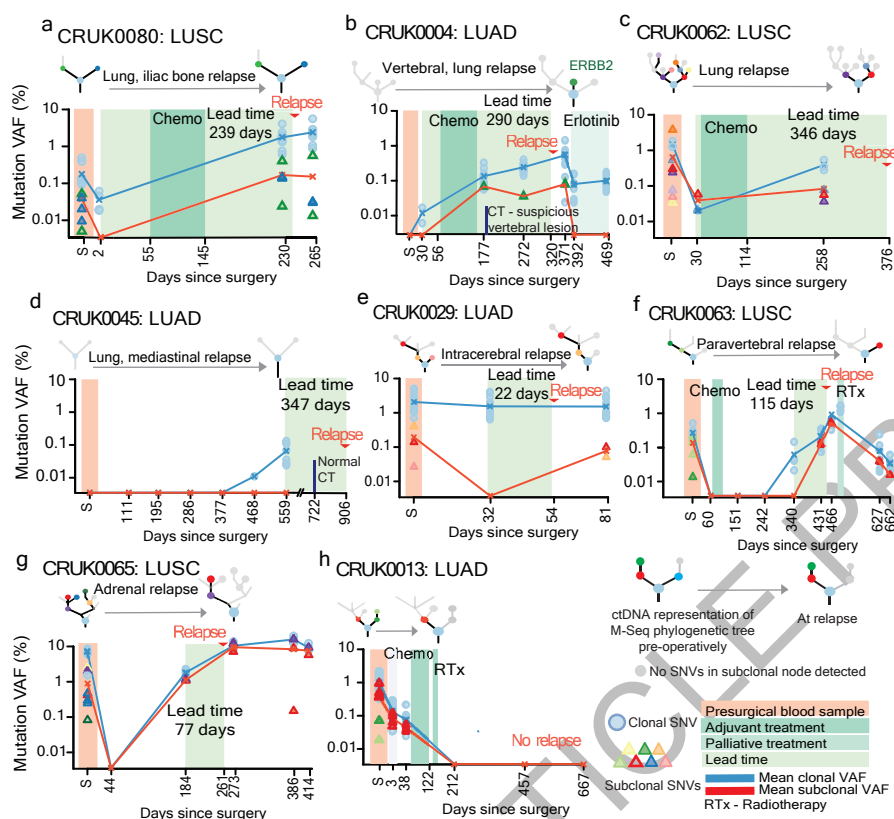
Figure 4 | Post-operative ctDNA detection predicts and characterizes NSCLC relapse. a-h) Longitudinal cell-free DNA profiling. Circulating tumor DNA (ctDNA) detection in plasma was defined as the detection of two tumor-specific SNVs. Detected clonal (circles, light blue) and subclonal (triangles, colors 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.
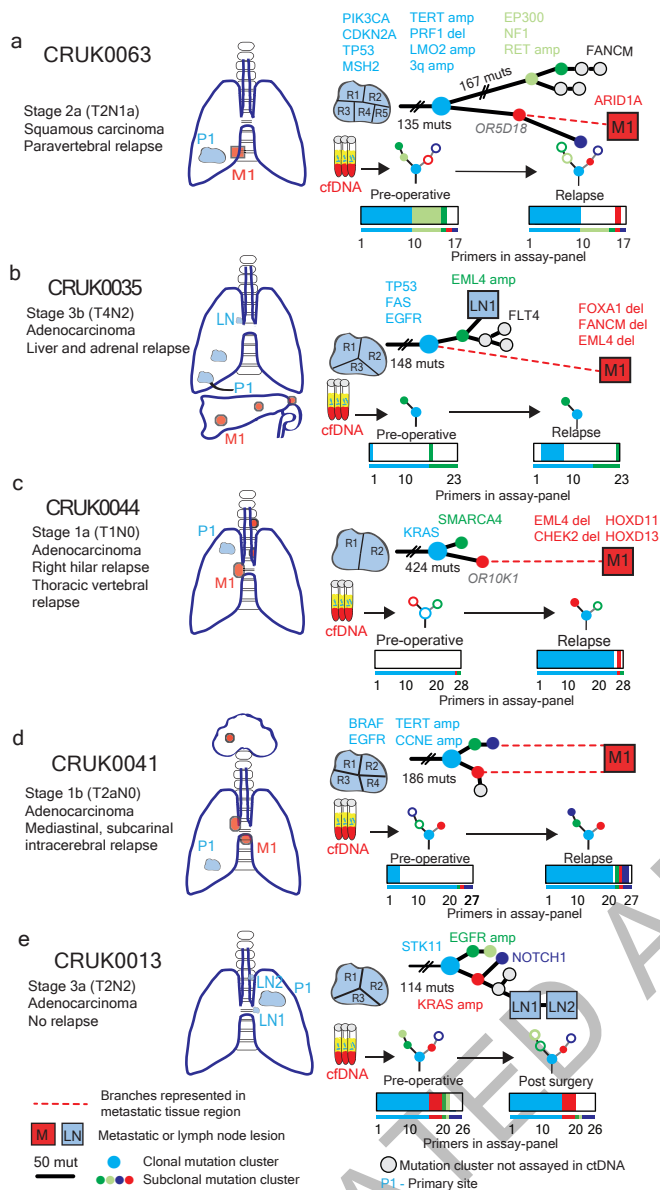
Figure 5 | Phylogenetic trees incorporating relapse tissue sequencing data. Phylogenetic trees based on mutations found in primary and metastatic tissue (a-d), or primary tumor and lymph node biopsies (e). Colored nodes in phylogenetic trees indicate cancer clones harboring mutations assayed for in ctDNA, grey indicates a clone not assayed. Branch length is proportional to number of mutations unless crossed. Dashed red lines show branches leading to metastatic relapse. Colored bars below show the 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 match clones on the phylogenetic trees.
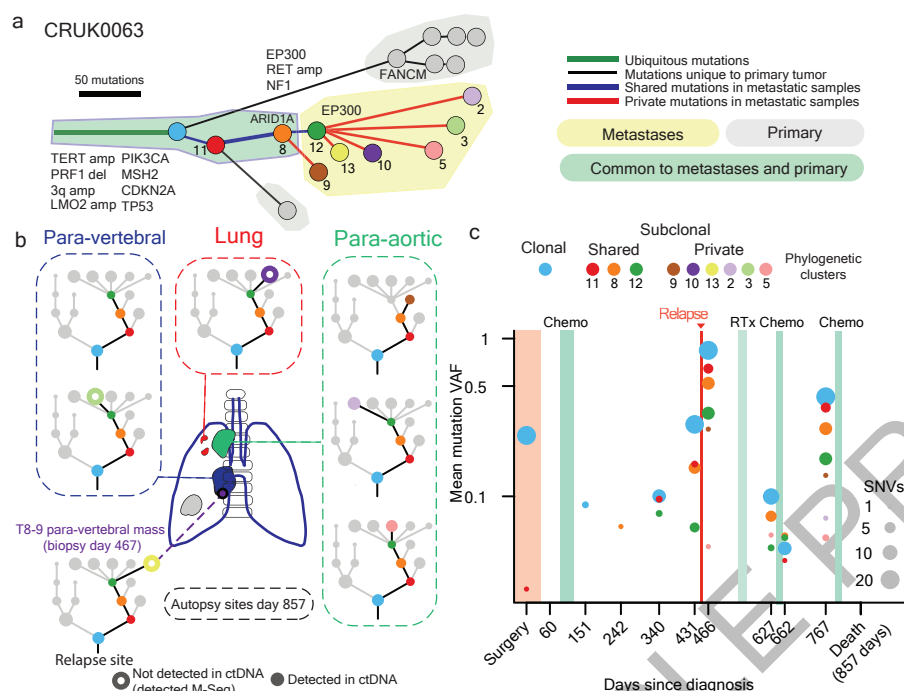
Figure 6 | ctDNA tracking of lethal cancer subclones in CRUK0063. Phylogenetic analysis of one relapse biopsy (day 467) and five metastatic biopsies (post mortem) **a)** To-scale phylogenetic tree of CRUK0063 including M-seq based on metastatic and primary tumor regions. Branch length is proportional to number of mutations in each subclone. **b)** Phylogenetic trees matching metastatic lesions, colored nodes represent mutation clusters found at each site and assayed for in ctDNA. Open circles represent mutation clusters not detected in ctDNA. **c)** Tracking plot showing mean VAF of identified mutation clusters in ctDNA. Size of dots indicates number of assays detected. Colors correspond to mutation clusters and match panels **a)** and **b)**.

METHODS

Patients and samples. The cohort of 100 patients evaluated within this study comprises the first 100 patients prospectively analyzed by the lung TRACERx study (Clinicaltrials.gov no: NCT01888601, approved by an independent Research Ethics Committee, 13/LO/1546) and mirrors the prospective 100 patient cohort by Jamal-Hanjani M et al[9].

Multi-region tumor sampling was performed as described[9]. Relapse tissue samples, excess to diagnostic requirements, were acquired via clinical procedures detailed in Supplementary Table 3. For patient CRUK0063 post mortem examination was performed through the PEACE study 24 hours following death (Clinicaltrials.gov no: NCT03004755, approved by an independent Research Ethics Committee, 13/LO/0972). Informed consent was obtained from all subjects for procedures conducted in these studies.

Tissue microarray creation and Ki67 immunohistochemistry. Tissue microarrays (TMAs) were created of 100 NSCLC cases for Ki67$^+$ immunohistochemistry. Representative primary tumor areas were defined by examination of H&E stained sections from TRACERx cases. Two 2 mm cores were selected from different regions within each specimen and re-embedded in recipient blocks. This resulted in a TMA of 200 cores with four normal lung cores as negative control. 2-5 μm sections from tissue-microarrays containing tumor were cut. Immunohistochemistry with anti-Ki67 monoclonal antibody (Dilution 1:100; clone MIB-1; DAKO Agilent Technologies LDA, UK Limited, Stockport, Cheshire SK8 3GR, UK) was performed using BenchMark Ultra (Ventana/Roche). The percentage of Ki67 positive cells were averaged across two tumor sections for each case. Detection was performed using the peroxidase-based detection reagent conjugate (OptiView DAB IHC Detection Kit; Ventana Medical Systems, Inc).

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

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

Tissue exome sequencing and processing. Whole exome sequencing was performed on DNA purified from tumor tissue and normal blood as described[9], with the exception of CRUK0063_BR_T1-R1. This capture was performed according to the manufacturer's 200 ng DNA protocol (Agilent). Annotated SNV calls from primary tumors are available in Jamal-Hanjani et al. 2017[9]. For this study, metastatic tissue biopsies from each of four patients (CRUK0035, CRUK0041, CRUK0044, CRUK0063) and six metastatic samples acquired at post mortem examination of CRUK0063 were obtained. Genomic DNA was purified from all tissue samples, and processed through the TRACERx bioinformatics pipeline as described[9]. Annotated SNV calls are available in Supplementary Table 4.

cfDNA extraction & quantification. Blood samples were collected in K$_2$-EDTA tubes. Samples were processed within 2 hours of collection by double spinning of blood first at 10 minutes at 1000 g then plasma 10 minutes at 2000 g. Plasma was stored in 1 ml aliquots at −80°C. Up to 5 ml of plasma per case was available for this study (range 1-5 ml, median 5 ml). The entire volume of plasma was used for cfDNA extraction. cfDNA was extracted using the QIAamp Circulating Nucleic Acid kit (Qiagen) and eluted into 50 μl DNA Suspension Buffer (Sigma). Every cfDNA sample was analyzed on the Bioanalyzer High Sensitivity (Agilent) and quantified by interpolation of the mononucleosomal peak height on a calibration curve prepared from a pure cfDNA sample that was quantified previously.

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

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

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

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

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

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

Bioinformatics Pipeline. All the paired-end reads were merged using Pear[26]. Merged reads were mapped to the hg19 reference genome with Novoalign v2.3.4 (http://www.novocraft.com/) and sorted and indexed using SAMtools[27]. Bases that did not match in forward and reverse reads or that have Phred quality score <20 were filtered out to minimize sequencing errors in subsequent steps. Merged reads with mapping quality >30 and at most one mismatch under the sequence of primers were marked as on-target. Targets with <1000 reads were considered failed and were filtered from further analyses. Quality control was performed using

an in-house program checking for a wide list of statistics per sample that included total numbers of reads, mapped reads, on-target reads, number of failed targets, and average error rate.

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

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

Cross-platform validation using generic PCR-NGS panel section. Cross-platform validation was performed in 28 patients with M-Seq confirmed SNV(s) within one or more hotspots targeted by a generic multiplex PCR-NGS panel (Extended Table 2a-b, Supplementary Table 8). 20 ng of isolated cfDNA was used for library preparation using the Oncomine™ Lung cfDNA Assay (ThermoFisher Scientific), according to the manufacturer's instructions. Automated template preparation and chip loading was conducted on the Ion Chef™ instrument using the Ion 520™ & Ion 530™ Kit-Chef (ThermoFisher Scientific). Ultimately, samples were sequenced on Ion 530™ chips using the Ion S5™ System (ThermoFisher Scientific). Sequencing data was accessed on the Torrent suite v5.2.2. Reads were aligned against the human genome (hg19) using Alignment v4.0-r77189, and variants were called using the coverage Analysis v4.0-r77897 plugin. All 18 bespoke-panel ctDNA negative patients had no tumor SNVs detectable in plasma pre-operatively by the generic panel supporting biological specificity of the bespoke targeted approach, 7 of 10 bespoke-panel ctDNA positive patients had tumor SNVs detected in plasma by the generic panel (Extended Table 2a-b). SNVs detected by hotspot panel not identified by M-Seq are displayed in Extended Table 2c.

Processing and phylogenetic analysis of relapse and primary tumor multiregion whole exome data. Biopsies from multiple regions from the primary tumor (n = 327), metastatic biopsies (n = 4) and matching blood germline samples (n = 100) were subjected to multi-region whole exome sequencing and analysis including estimation of copy number, purity and ploidy, and phylogenetic tree construction as described[9]. Briefly, phylogenetic analysis was performed based on CCF determined for SNVs and clustered across tumor regions using a modified version of Pyclone[9] into clusters with similar CCF values, filtered and processed as described[9]. Mutation clusters are assumed to represent tumor subclones, either current or ancestral, and are used as input for construction of phylogenetic relationship. Phylogenetic trees were primarily constructed using the published tool CITUP (0.1.0)[28]. However, in a small number of cases, including all relapse/autopsy cases, manual tree construction was required and performed as described[9]. Complete detail of primary tumor construction can be found in Jamal-Hanjani et al. 2017[9]. Relapse tree construction was performed as follow: CRUK0063: clustering was performed twice, once across 5 primary tumor regions and once across 5 primary, 1 relapse, and 6 autopsy regions. To ensure consistency, when deriving a phylogenetic tree based on all tumor regions, CCF clusters based on clustering only the primary tumor regions were maintained for mutations not involved in metastatic relapse. A phylogenetic tree was constructed based on 17 mutation clusters. CRUK0035: Clustering primary tumor regions with the relapse region revealed one cluster private to the relapse, and one cluster shared with the relapse and all other regions. CRUK0044: Clustering primary tumor regions with the relapse region revealed a cluster private to the relapse, descended from a cluster private to region 1 in the primary tumor. CRUK0041: Clustering primary tumor regions with the
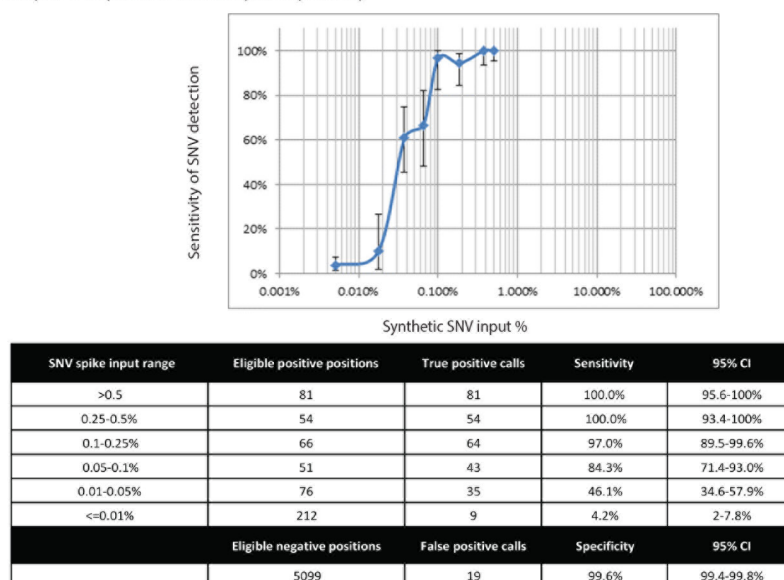
relapse region revealed cluster 4 as private to the relapse. This cluster must have evolved from cluster 3 only found in the relapse and in region 4. A private cluster 6 in region 4 must have evolved from cluster 4. However, this conflicts with clusters 2 and 5, found in the relapse and regions 1-3, but not region 4. This can be reconciled by assuming a polyclonal relapse, seeded primarily from regions 1-3, but with some contribution from cluster 3, private to region 4. Cluster data is available in Supplementary Table 4 under "PyClonePhyloCluster".

Statistical data analysis. Analysis was performed in the R statistical environment version 3.2.3 and SPSS version 24. All statistical tests were two-sided unless expressly stated. Multivariate logistic regression used detection of ctDNA (the dependent variable) classified as detection of 2 or more patient-specific variants in ctDNA and the covariates listed in Supplementary Table 1. All predictors were entered simultaneously into the regression. All continuous independent variables were found to be linearly related to the logit of the dependent variable (assessed via the Box-Tidwell procedure). The logistic regression model was statistically significant, $X^2(10) = 81.35$, $P < 0.001$ and the Hosmer-Lemeshow P value was 0.9858 indicating that the model was not a poor fit. To determine the ability of PET TBR to predict whether or not tumor ctDNA was identified in plasma, PET TBR estimates were analyzed by ROC curve analysis against binary detection of ctDNA in plasma at baseline based on at least two variants detected, significance test based on Wilcoxon rank sum test. For analysis involving longitudinally detected variants (Fig. 4, Extended Figure 5), only subclonal variants from pyclone clusters present in phylogenetic trees were displayed, this did not affect ctDNA detection status of any time-points. In non-relapse cases presented in Extended Data Fig 6 all detected subclonal SNVs were plotted. To determine the relationship between tumor volume and ctDNA VAF, ctDNA assays against clonal SNVs were selected. For each patient, the mean ctDNA VAF of the clonal SNVs was determined as baseline for 38/46 patients with at least 2 SNVs detected in plasma. As detailed in Extended Fig. 4c, 8/46 patients were not included in the analysis: CRUK0036 had no pre-op CT scan available, CRUK0087 had a large cavity inside the primary cancer, CRUK0099 had a collapsed lung making volume assessment inaccurate, CRUK0100, CRUK0077, CRUK0052 had a CT slice spacing of > 5 mm, and finally CRUK0088 and CRUK0091 had a total tumor volume < 3.5 cm³. Linear regression was performed on log-transformed mean VAF and tumor volume. The log transformation was justified as it symmetrized the residuals in the model. An independent analysis was performed where tumor volume was multiplied with tumor purity to estimate the cancer cell volume. The same log transformation and analysis was applied to data acquired from Newman et al.[16], where ctDNA VAF was determined based on CAPP-seq analysis with matched tumor volume data available. To analyze clone size versus ctDNA VAF for subclonal SNVs, the mean CCF of the mutations within a subclonal mutation cluster was multiplied with tumor volume, and as a second independent analysis, with tumor purity.

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

24. Hofheinz F, Butof R, Apostolova I, Zophel K, Steffen IG, Amthauer H, *et al.* An investigation of the relation between tumor-to-liver ratio (TLR) and tumor-to-blood standard uptake ratio (SUR) in oncological FDG PET. *EJNMMI Res* 2016;**6**(1):19.
25. Butof R, Hofheinz F, Zophel K, Stadelmann T, Schmollack J, Jentsch C, *et al.* Prognostic Value of Pretherapeutic Tumor-to-Blood Standardized Uptake Ratio in Patients with Esophageal Carcinoma. *J Nucl Med* 2015;**56**(8):1150–6.
26. Zhang J, Kobert K, Flouri T & Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics 2014;***30**(5):614–20.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**(16):2078–9.
28. Malikic S, McPherson AW, Donmez N & Sahinalp CS. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics* 2015;**31**(9): 1349–56.

a   Multiplex-PCR platform sensitivity and specificity

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

b

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

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

c
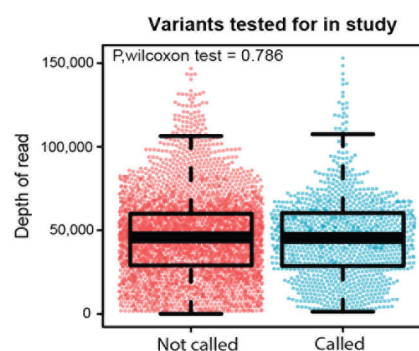
Extended Data Figure 1 | Multiplex-PCR next-generation sequencing platform analytical validation. a) Analytical validation of the multiplex-PCR NGS platform was performed by spiking synthetic single nucleotide variants into control cell-free DNA. Sensitivity and specificity of the platform at different spike concentrations was ascertained, 95% binomial confidence interval displayed as error bars. b) Specificity of ctDNA detection based on a 1 SNV and 2 SNV call threshold taking into account parallel testing of multiple SNVs. c) The median depth of read across a position did not vary depending on whether an SNV position was called or not called using the platform error-model. Wilcoxon Test, P = 0.786, median depth of read at uncalled positions = 45,777 (n = 3,745), range: 0 to 146774, median depth of read at called positions = 45,478, range = 1,354 to 152,974 (n = 1,124). Whiskers represent 1.5 times the interquartile range, 2-sided test.

a

Pre-operative study phase cohort design

bi

Pre-operative study phase assay-panel design

bii

Pre-operative study phase assay-panel design

LUAD

c

Longitudinal study phase cohort design

d

Longitudinal study phase assay-panel design

- Clonal driver variant
- Clonal passenger variant
- Subclonal driver variant
- Subclonal passenger variant
- No copy number data

e

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

Extended Data Figure 2 | Study construction and assay-panel design.
a) The pre-operative study phase cohort consisted of 100 TRACERx patients present in the first 100 patient TRACERx cohort in April 2016. Pre-operative plasma samples were profiled in 96 patients for reasons listed. bi and ii) Contents of patient-specific assay-panels designed in the pre-operative study phase. c) The longitudinal study phase cohort consisted of patients with confirmed NSCLC relapse and patients without relapse. d) Contents of patient-specific assay-panels designed in the longitudinal phases of this study. e) Single nucleotide variant type targeted.

a

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

b

c

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

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

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

d

PET TBR ROC−curve

AUC = 0.84, P = 1.4e−08

e

LUAD

TP53 driver

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

Fishers Exact Test P=0.508

KRAS driver

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

Fishers Exact Test P=0.499

EGFR driver

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

Fishers Exact Test P=0.429

Extended Data Figure 3 | Clinicopathological predictors of ctDNA detection. a) 96 patients in pre-operative cohort stratified by pathological TNM stage. b) LUSCs and ctDNA positive LUADs are significantly more necrotic that ctDNA negative LUADs. Significant differences in necrosis between groups: LUSCs (median necrosis 40%) (n = 31), ctDNA positive LUADs (median necrosis 15%) (n = 11) and ctDNA negative LUADs (median necrosis 2%) (n = 47), Kruskal-Wallis test, P < 0.001, 2-sided pairwise comparisons were performed using Dunn's procedure with Bonferroni correction. c) Univariate (left) and multivariate analyses (right) were performed, by logistic regression to determine significant predictors of ctDNA detection in early-stage NSCLC. ctDNA detection was defined as detection of two or more SNVs in pre-operative plasma samples. Details regarding multivariable analysis methodology are in methods. d) Receiver operating characteristic curve (ROC) analysis of pre-operative PET scan FDG-avidity (normalized as tumor background ratio (TBR), see methods), as a predictor of ctDNA detection (92/96 PET scans were available for central review). Median PET TBR of detected tumors = 9.01, n = 45. Median PET TBR of undetected tumors = 3.64, n = 47. P-value based on Wilcoxon Rank Sum Test. e) LUAD subtype analyses based on ctDNA detection and the presence of an EGFR, KRAS or TP53 driver mutation.
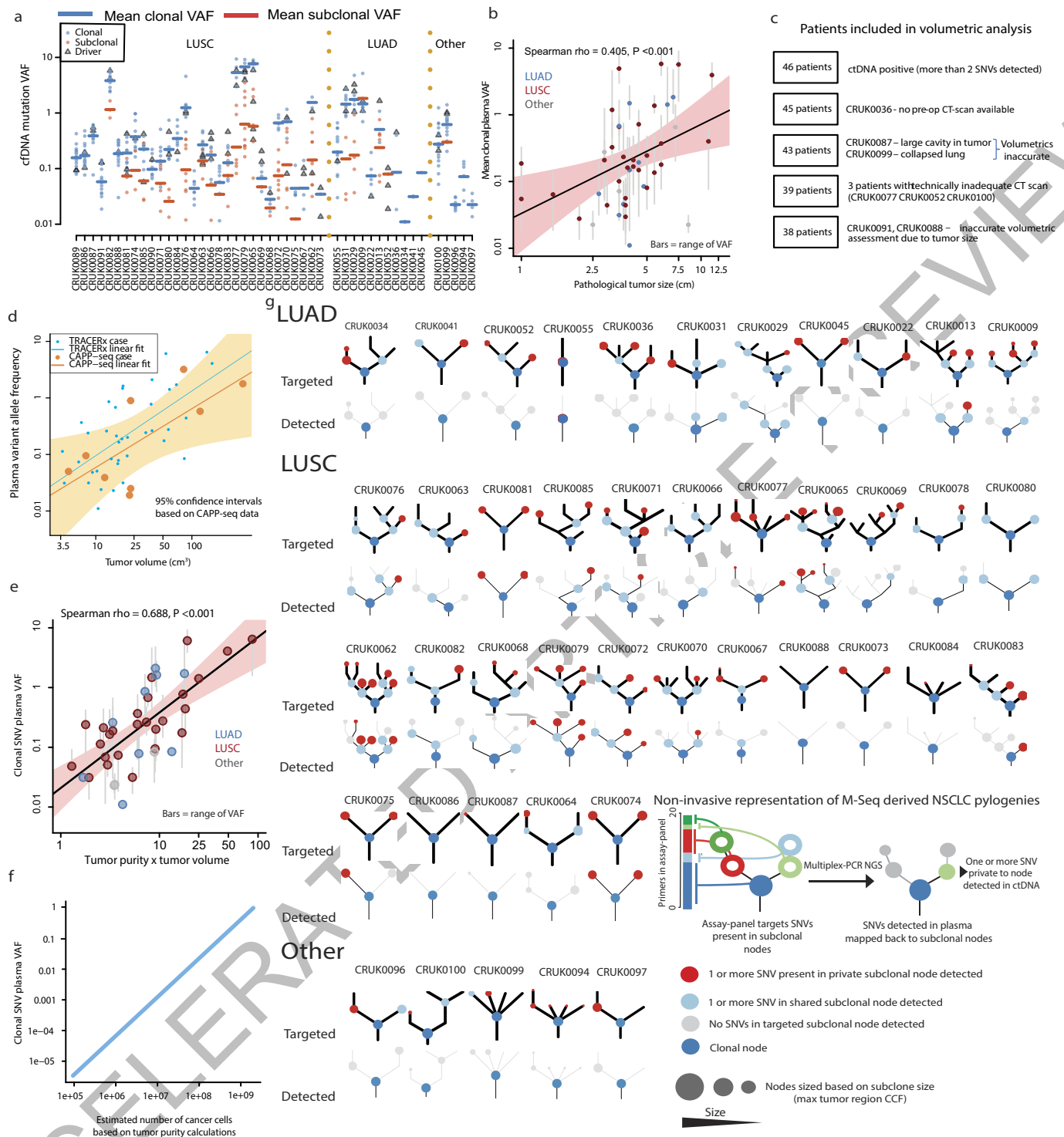
Extended Data Figure 4 | Predictors of plasma variant allele frequency.
a) Plasma variant allele frequencies of SNVs detected in plasma in 46 patients who were ctDNA positive (two or more SNVs detected). Clonal (blue) and subclonal (red) variant allele frequencies indicated, mean shown as horizontal line. Driver variants shown as triangles. b) Mean clonal VAF correlated with maximum tumor size measured in post-surgical specimen (pathological size, n = 46) grey vertical bars represent range of clonal variant allele frequency. Shaded red background indicates 95% confidence interval. c) Filtering steps taken to define a group of ctDNA positive patients with volumetric data considered adequate to model tumor volume and plasma variant allele frequency. d) Scatter plot showing mean clonal VAF relative to tumor volume for TRACERx (blue dots and fitted blue line, n = 38) and VAF relative to volume for previously published data based on CAPP-seq analysis of ctDNA (orange dots and orange fitted line, n = 9). Orange shaded background indicates

95% confidence interval based on CAPP-seq data. e) Mean clonal VAF correlated with tumor volume × tumor purity (cancer cell volume), n = 38. Shaded red background indicates 95% confidence interval. f) Association between number of cancer cells and VAF of clonal SNVs in plasma based on linear modelling of Extended Data Fig 4f. g) Detected subclonal SNVs were mapped back to M-Seq derived tumor phylogenetic trees (process illustrated in graphic). Detected private subclones (subclones identified within only a single tumor region) are coloured red. Shared subclones (subclones detected in more than one tumor regions) are light blue. Subclonal nodes were sized based on the maximum recorded cancer cell fraction (CCF). The top row of phylogenetic trees represent subclonal nodes targeted by primers within that patient's assay panel, the bottom row represent subclonal nodes detected in ctDNA, within this row grey subclonal nodes represent subclones not detected in ctDNA.
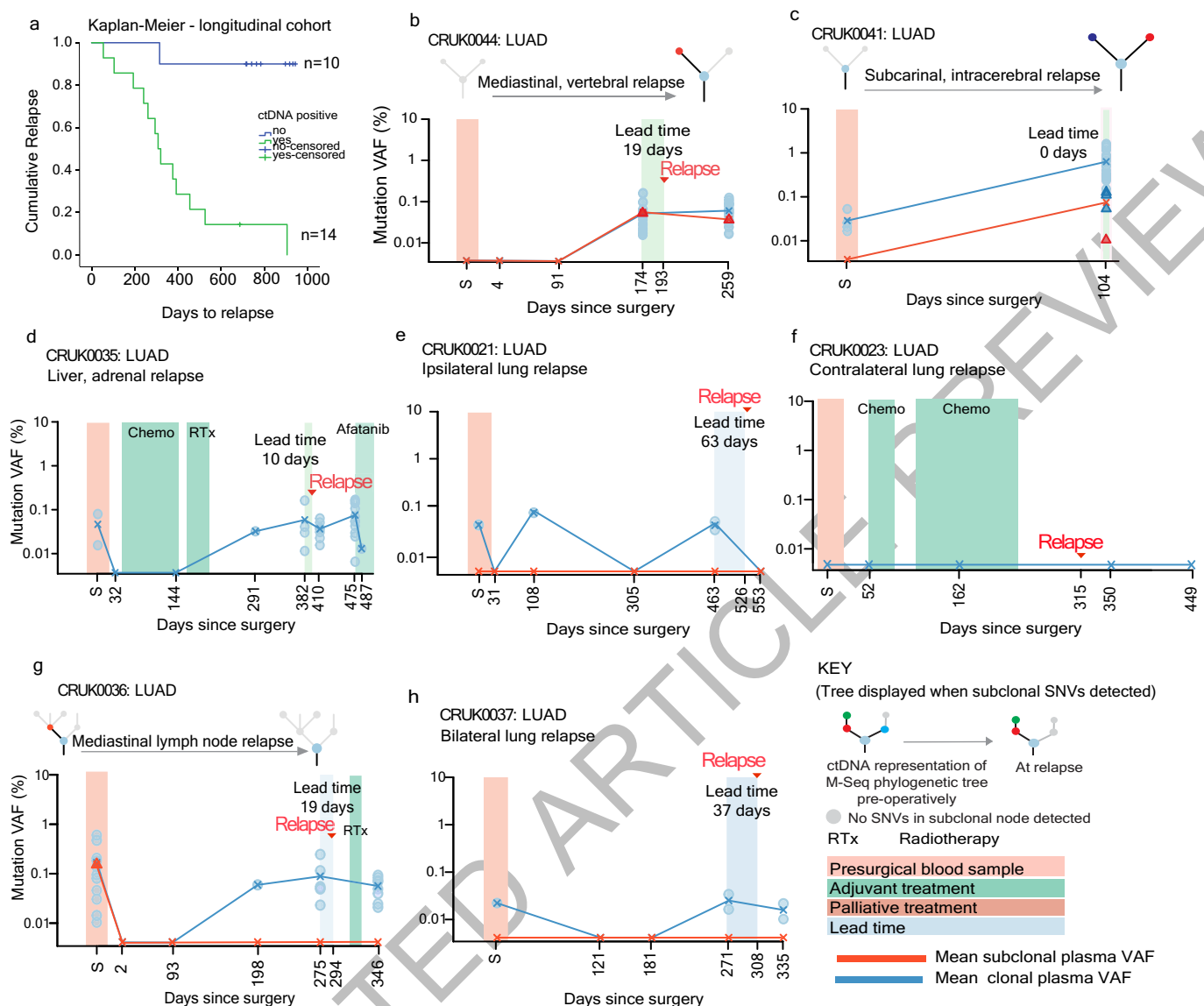
Extended Data Figure 5 | Longitudinal ctDNA profiling, remaining relapse cases. a) Kaplan-Meier curve demonstrate relapse free survival for patients in whom ctDNA was detected versus patients in whom ctDNA was not detected. b-h) Longitudinal cell-free DNA profiling. Circulating tumor DNA (ctDNA) detection in plasma was defined as the detection of two tumor-specific SNVs. Relapse was based on imaging-confirmed NSCLC relapse, imaging 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) VAF are indicated on graphs. Pre-operative and relapse M-Seq derived phylogenetic trees represented by ctDNA are illustrated above each graph in cases where subclonal SNVs were detected.
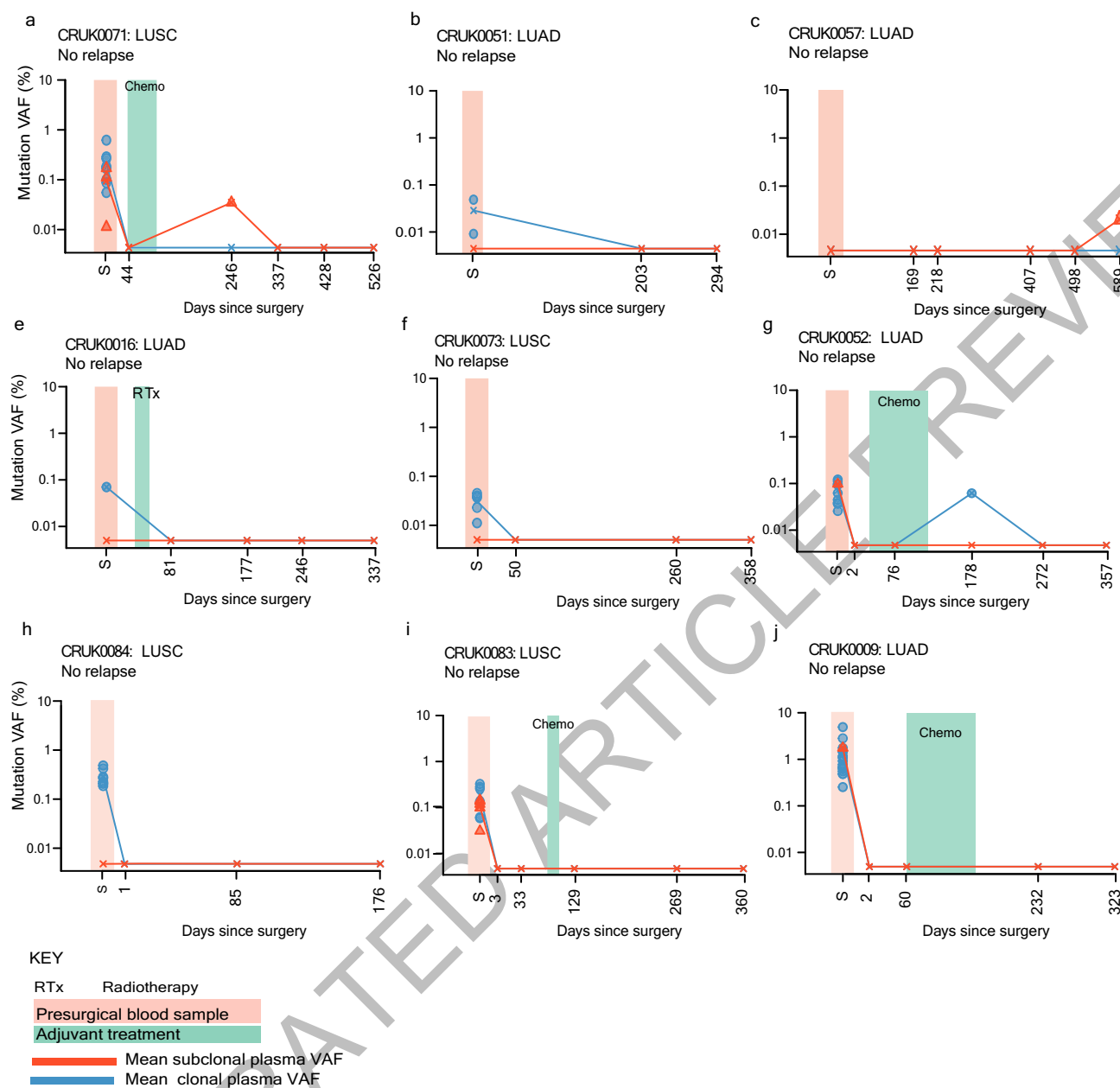
Extended Data Figure 6 | Longitudinal ctDNA profiling, non-relapse cases. a-j) Detected clonal (circles, light blue) and subclonal (red triangles) SNVs from each patient-specific assay-panel are plotted on graphs. Mean clonal (blue) and mean subclonal (red) VAF are indicated on graphs.
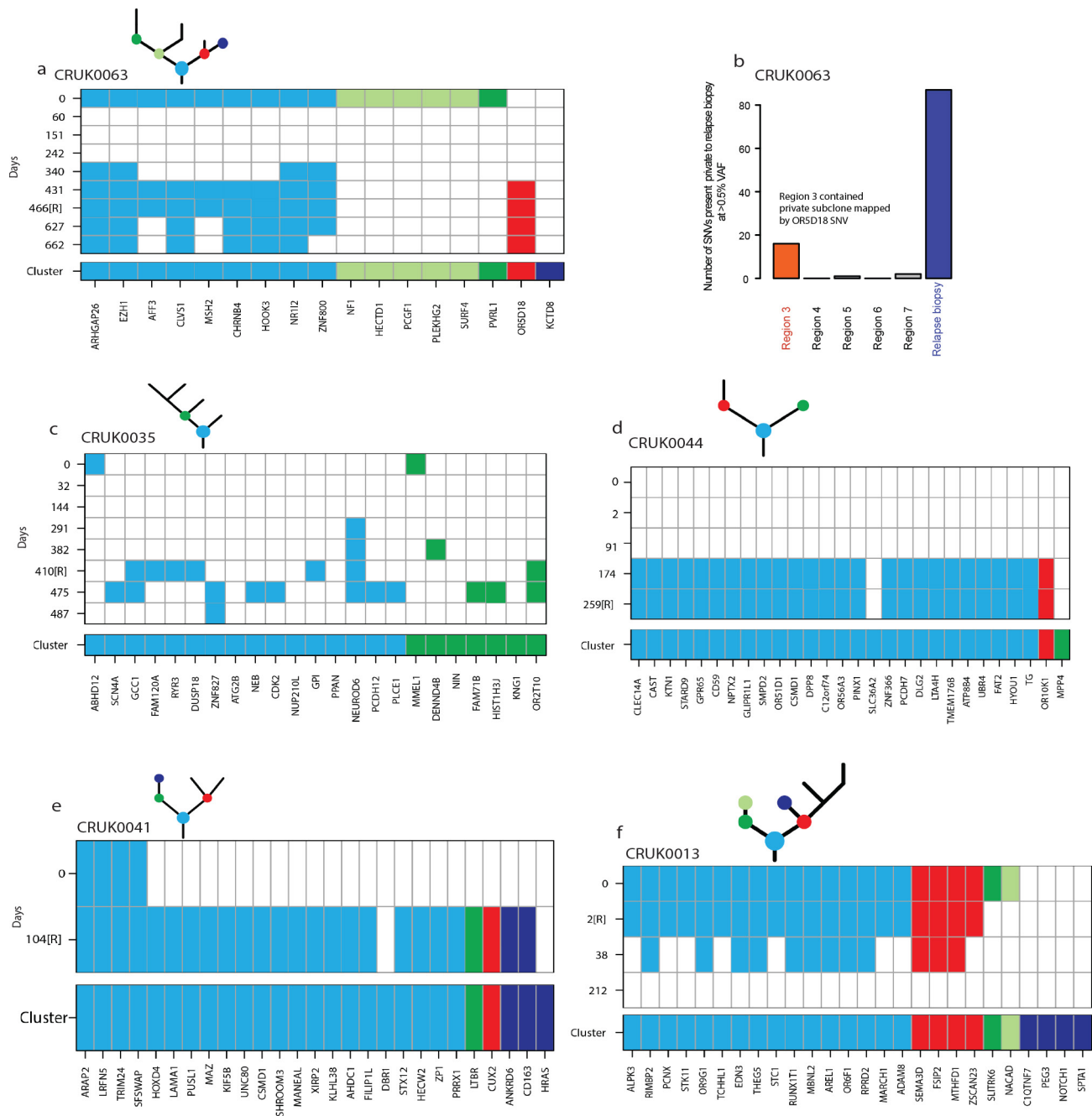
Extended Data Figure 7 | Heatmaps illustrating detection of SNVs in bespoke panel at each sampled time point. a, c-f) Bespoke assay panels for CRUK0063, CRUK0035, CRUK0044, CRUK0041 and CRUK0013. Colors indicate originating subclonal cluster based on the phylogenetic trees above the heatmap. Light blue indicates clonal mutation cluster. Full panel with cluster color shown below each heatmap. Filled squares indicates detection of a given variant in plasma ctDNA. Y-axis shows day of sampling, y-axis labels appended with [R] indicates day of clinical relapse. b) Re-examination of primary tumor regions from CRUK0063 with lowered threshold to potentially identify SNVs private to the sequenced relapse biopsy. 16/88 variants were found at very low VAF in region 3, indicating this region from the primary likely gave rise to the metastasis.
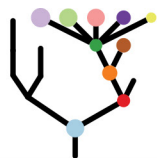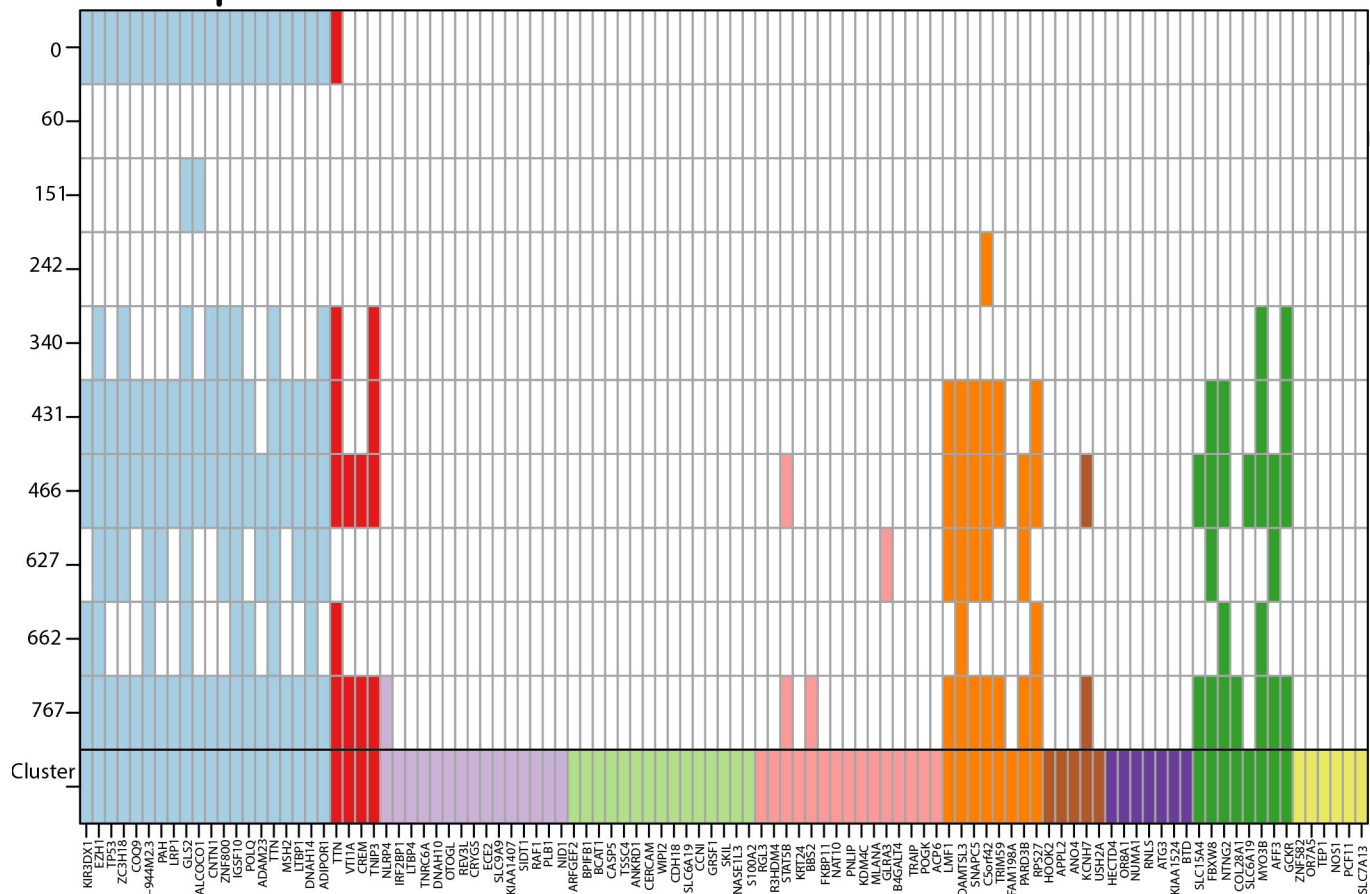
Extended Data Figure 8 | Heatmap illustrating detection of SNVs in bespoke panel based on M-seq of metastatic tumor regions for patient CRUK0063 for all sampled time points. Colors indicate originating subclonal cluster based on the phylogenetic trees above the heatmap. Light blue indicates clonal mutation cluster. Full panel with cluster color shown below each heatmap. Filled squares indicates detection of a given variant in plasma ctDNA. Y-axis shows day of sampling.

Extended Data Table 1 | Patient characteristics

a
Clinical characteristics
96 patient pre-operative cohort

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

b

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

c
Details regarding timing of pre-operative blood sample

Days pre-surgery	Number	Details
Within 24 hours	91	
24-72 hours	2	CRUK0051, 0003
8 days	2	CRUK0073, 0096
31 days	1	CRUK0089

d
Clinical characteristics
24 patient longitudinal sub-cohort

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

e

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

a) table of clinical characteristics describing the 96 patient pre-operative cohort and b) demonstrating distribution of stage and whether the patient received adjuvant chemotherapy, c) demonstrating the time-points at which pre-operative plasma was acquired for patients within the cohort, d) table of clinical characteristics describing 24 patient longitudinal cohort and e) demonstrating distribution of stage in the longitudinal cohort and whether the patient received adjuvant chemotherapy.

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

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

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

Case	Volume cm3	Predicted plasma VAF	Bespoke-panel ctDNA positive	Histology	Hotspot SNVs tumor	Generic-panel Hotspot SNVs detected
CRUK0037	197.42	2.89 (0.93 - 8.97)	No	LUAD	1	0
CRUK0051	27.28	0.30 (0.19 to 0.46)	No	LUAD	1	0
CRUK0004	23.30	0.25 (0.16 to 0.38)	No	LUAD	1	0
CRUK0039	21.68	0.23 (0.15 to 0.35)	No	LUAD	2	0
CRUK0025	19.06	0.20 (0.13 to 0.30)	No	LUAD	2	0
CRUK0014	8.58	0.08 (0.04 to 0.15)	No	LUAD	1	0
CRUK0026	7.45	0.07 (0.04 to 0.13)	No	LUAD	1	0
CRUK0057	5.95	0.05 (0.02 to 0.11)	No	LUAD	1	0
CRUK0018	4.65	0.04 (0.02 to 0.09)	No	LUAD	1	0
CRUK0027	4.61	0.04 (0.02 to 0.09)	No	LUAD	1	0
CRUK0007	4.18	0.04 (0.01 to 0.08)	No	LUAD	1	0
CRUK0049	3.61	0.03 (0.01 to 0.08)	No	LUAD	1	0
CRUK0035	3.31	0.03 (0.01 to 0.07)	No	LUAD	1	0
CRUK0058	2.76	0.02 (0.01 to 0.06)	No	LUAD	1	0
CRUK0021	2.70	0.02 (0.01 to 0.06)	No	LUAD	2	0
CRUK0048	2.16	0.02 (0.01 to 0.05)	No	LUAD	2	0
CRUK0093	0.73	0.004 (0 to 0.03)	No	LUSC	2	0
CRUK0030	0.21	0.001 (0 to 0.01)	No	LUAD	2	0

Multiplex-PCR NGS

Targeted panel:
- >99% sensitivity at 0.1% VAF and above
- 84% sensitivity at 0.05% to 0.1% VAF
- 46 % sensitivity 0.01% to 0.05% VAF
- 4.2% sensitivity <0.01%

Platform sensitivities predicted based on tumor volume and analytical validation data in Extended Data 1

Generic panel:
- 90% sensitivity at 0.1% VAF and above

Oncomine lung panel sensitivity data reported at https://www.thermofisher.com/order/catalog/product/A31149

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

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

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

a) 7/10 (70%) of bespoke-panel ctDNA positive patients had tumor SNVs detectable in plasma preoperatively by a generic hotspot PCR-NGS lung panel (Lung Oncomine, Thermofisher). The three bespoke-panel ctDNA positive patients undetected by the generic panel had mean clonal plasma variant allele frequencies lower than the 0.1% plasma variant allele frequency (VAF) limit of detection reported for the generic panel (shaded yellow). b) Based on CT volumetric assessment of each patient's primary tumor we predicted plasma VAF corresponding to a tumor of that size (see Fig. 3 and Methods for details of approach). This allowed us to infer platform sensitivities for each patient within the bespoke-panel non-detected cohort. Five LUADs (shaded green, CRUK0037, CRUK0051, CRUK0004, CRUK0039 and CRUK0025) had tumor volumes approximating to a plasma VAF of more than 0.1%. This suggested that these tumors resided within the top platform sensitivity bracket of both the generic and bespoke-panel ctDNA platforms. No ctDNA was detected by either platform in these cases, suggesting biological specificity of the bespoke-panel. c) Hotspot SNVs not identified in tumor tissue through exome sequencing were identified in plasma of 9 of 28 patients by the generic panel. This suggested non-tumor origin of cell-free DNA, platform non-specificity or an evolving minor subclone or second primary.