LETTER

doi:10.1038/nature12065

Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA

Muhammed Murtaza[1]*, Sarah-Jane Dawson[1,2]*, Dana W. Y. Tsui[1]*, Davina Gale[1], Tim Forshew[1], Anna M. Piskorz[1], Christine Parkinson[1,2], Suet-Feung Chin[1], Zoya Kingsbury[3], Alvin S. C. Wong[4], Francesco Marass[1], Sean Humphray[3], James Hadfield[1], David Bentley[3], Tan Min Chin[4,5], James D. Brenton[1,2,6], Carlos Caldas[1,2,6] & Nitzan Rosenfeld[1]

Cancers acquire resistance to systemic treatment as a result of clonal evolution and selection[1,2]. Repeat biopsies to study genomic evolution as a result of therapy are difficult, invasive and may be confounded by intra-tumour heterogeneity[3,4]. Recent studies have shown that genomic alterations in solid cancers can be characterized by massively parallel sequencing of circulating cell-free tumour DNA released from cancer cells into plasma, representing a non-invasive liquid biopsy[5–7]. Here we report sequencing of cancer exomes in serial plasma samples to track genomic evolution of metastatic cancers in response to therapy. Six patients with advanced breast, ovarian and lung cancers were followed over 1–2 years. For each case, exome sequencing was performed on 2–5 plasma samples (19 in total) spanning multiple courses of treatment, at selected time points when the allele fraction of tumour mutations in plasma was high, allowing improved sensitivity. For two cases, synchronous biopsies were also analysed, confirming genome-wide representation of the tumour genome in plasma. Quantification of allele fractions in plasma identified increased representation of mutant alleles in association with emergence of therapy resistance. These included an activating mutation in *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha) following treatment with paclitaxel[8]; a truncating mutation in *RB1* (retinoblastoma 1) following treatment with cisplatin[9]; a truncating mutation in *MED1* (mediator complex subunit 1) following treatment with tamoxifen and trastuzumab[10,11], and following subsequent treatment with lapatinib[12,13], a splicing mutation in *GAS6* (growth arrest-specific 6) in the same patient; and a resistance-conferring mutation in *EGFR* (epidermal growth factor receptor; T790M) following treatment with gefitinib[14]. These results establish proof of principle that exome-wide analysis of circulating tumour DNA could complement current invasive biopsy approaches to identify mutations associated with acquired drug resistance in advanced cancers. Serial analysis of cancer genomes in plasma constitutes a new paradigm for the study of clonal evolution in human cancers.

Serial sampling of the tumour genome is required to identify the mutational mechanisms underlying drug resistance[2]. Serial tumour biopsies are invasive and often unattainable. Tumours are heterogeneous and continuously evolve, and even if several biopsies are obtained, these are limited both spatially and temporally. Analysis of isolated circulating tumour cells (CTCs) has been proposed, but circulating tumour DNA (ctDNA) is more accessible and easier to process[15]. Previous studies of tumour mutations in plasma have analysed individual loci, genes or structural variants to quantify tumour burden and to detect previously-characterized resistance-conferring mutations[1,6,16–18]. Genome-wide sequencing of plasma samples is used in prenatal diagnostics, demonstrating comprehensive coverage of the genome[19]. More recently, genome-wide sequencing of plasma DNA has been demonstrated as a potential tool for detection of disease or analysis of tumour burden in patients with advanced cancers[5,7]. These studies established that plasma DNA contains representation of the entire tumour genome[7], mixing together variants originating from multiple independent tumours[5]. This suggests that deeper sequencing of plasma DNA, applied to selected samples with high tumour burden in blood, may allow assessment of clonal heterogeneity and selection. In this study, we applied exome sequencing of ctDNA as a platform for non-invasive analysis of tumour evolution during systemic cancer treatment (Fig. 1).

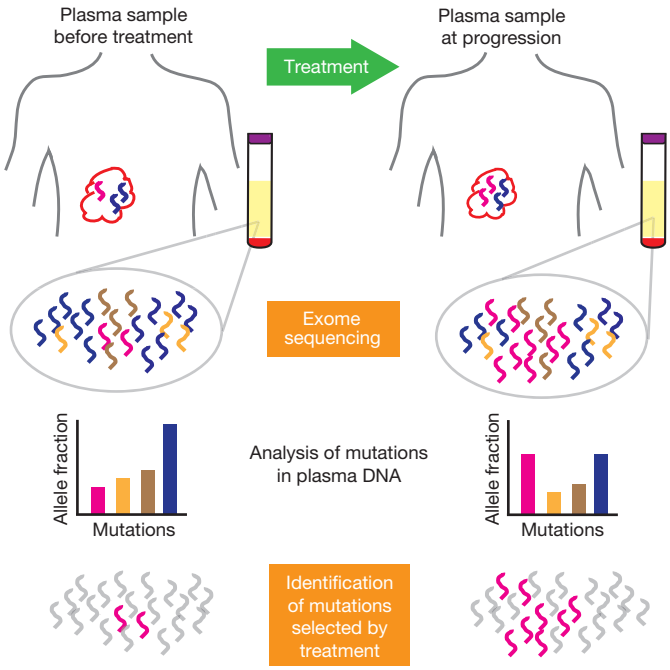

Figure 1 | Identification of treatment-associated mutational changes from exome sequencing of serial plasma samples. Overview of the study design: plasma was collected before treatment and at multiple time-points during treatment and follow-up of advanced cancer patients. Exome sequencing was performed on circulating DNA from plasma at selected time-points, separated by periods of treatment, and germline DNA. Mutations were identified across the plasma samples, and their abundance (allele fraction) at different time-points compared, generating lists of mutations that showed a significant increase in abundance, which may indicate underlying selection pressures associated with specific treatments. These lists contained mutations known to promote tumour growth and drug resistance, but also mutations of unknown significance. Accumulating such data across large cohorts could identify genes or pathways with recurrent mutations.

[1]Cancer Research UK Cambridge Institute and University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [2]Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [3]Illumina, Inc., Chesterford Research Park, Little Chesterford CB10 1XL, UK. [4]Department of Haematology-Oncology, National University Cancer Institute, National University Health System, 5 Lower Kent Ridge Road, Tower block level 7, 119074 Singapore. [5]Cancer Science Institute, National University of Singapore, Centre for Translational Medicine, 14 Medical Drive, #12-01, 117599 Singapore. [6]Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.
*These authors contributed equally to this work.

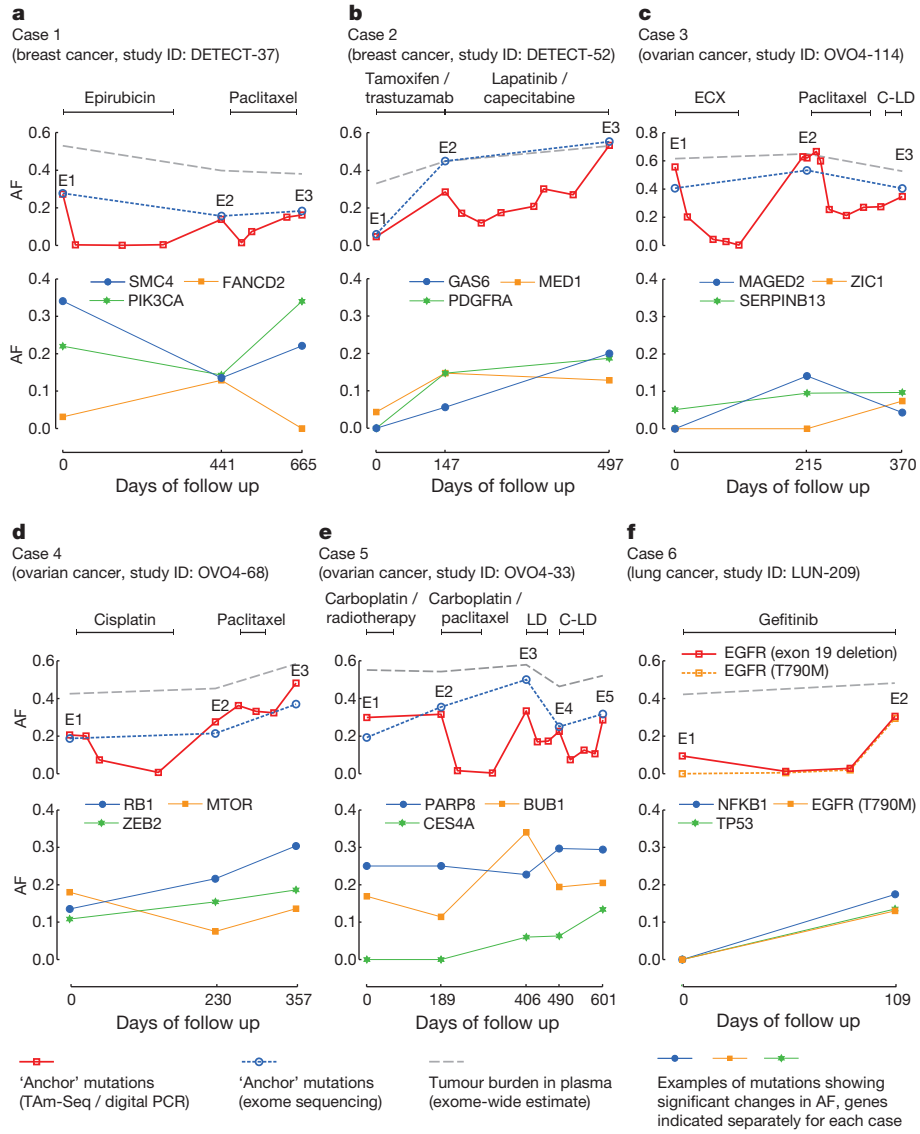

Figure 2 | Mutations showing evidence of genomic tumour evolution. All panels (**a**–**f**) are made up of an upper and a lower subpanel. Upper subpanels, time courses for allele fractions (AF; data points) of 'anchor' mutations used for initial quantification of ctDNA levels, and the fractional concentration of tumour DNA (tumour burden; grey dashed lines). 'Anchor' mutations were measured using digital PCR or TAm-Seq[6] for all available plasma samples, and using exome sequencing at selected time points indicated by E1, E2, E3 (and E4 and E5 for case 5). Tumour burden was estimated from exome data (an adaptation of genome-wide aggregated allelic loss[7]). In **a**, AF was averaged over six mutations measured in parallel using digital PCR. In **b**, a single mutation in *ATM* (predicted amino acid change I2948F) was measured by TAm-Seq. In **c**, **d** and **e**, a single mutation in *TP53* was measured by digital PCR for each case (R175H, K132N and R175H, respectively). In **f**, digital PCR was used to measure abundance of a deletion in exon 19 of *EGFR* (not quantified in exome sequencing data) and the EGFR T790M mutation. Lower subpanels, AF in exome data for selected mutations (blue, green and orange datapoints, see key) for each of the cases. Additional details are listed in Table 1, and a full list of mutations that showed a significant increase in abundance is included in Supplementary Tables 2–7. ECX, epirubicin, cisplatin and capecitabine; C-LD, carboplatin and liposomal doxorubicin; LD, liposomal doxorubicin.

We performed whole exome sequencing of plasma DNA in six patients with advanced cancers (Supplementary Table 1): two with breast cancer (cases 1 and 2), three with ovarian cancer (cases 3–5), and one with non-small-cell lung cancer (NSCLC, case 6). Exome sequencing was performed on multiple plasma samples from each patient separated by consecutive lines of therapy, spanning up to 665 days of clinical follow up (range 109–665 days, median 433 days). The ability to detect genomic events using redundant sequencing is dependent on the allele fraction (AF) of the mutant alleles in the samples analysed (ratio of mutant reads to depth of coverage at that locus), the sequencing depth, and the background noise rates of sequencing. Levels of ctDNA were previously quantified in these patients using digital PCR and tagged-amplicon deep sequencing[6] (TAm-Seq; Fig. 2, upper subpanels), allowing us to focus on samples with a high mutant AF in plasma, in which genomic changes related to the tumour could be identified even at relatively modest depth of sequencing. Comparison of AF measured using exome sequencing, digital PCR and TAm-Seq showed a high degree of concordance (correlation coefficient 0.8, $P < 0.0001$; Supplementary Fig. 1). Using as little as 2.3 ng of DNA (4%–20% of the DNA extracted from 2.0–2.2 ml of plasma), and an average of 169 million reads of sequencing per sample, we analysed the coding exons of all protein-coding genes at an average unique coverage depth ranging from 31-fold to 160-fold across 19 plasma samples (Supplementary Table 2). Consistent with previous reports[5,7], we observed copy number aberrations (CNAs, both gains and losses) in plasma samples in all patients across the whole genome (Supplementary Figs 2–7). These were strongly modulated by the fraction of tumour DNA in plasma and were particularly prominent in plasma samples in which mutant AF exceeded 50%.

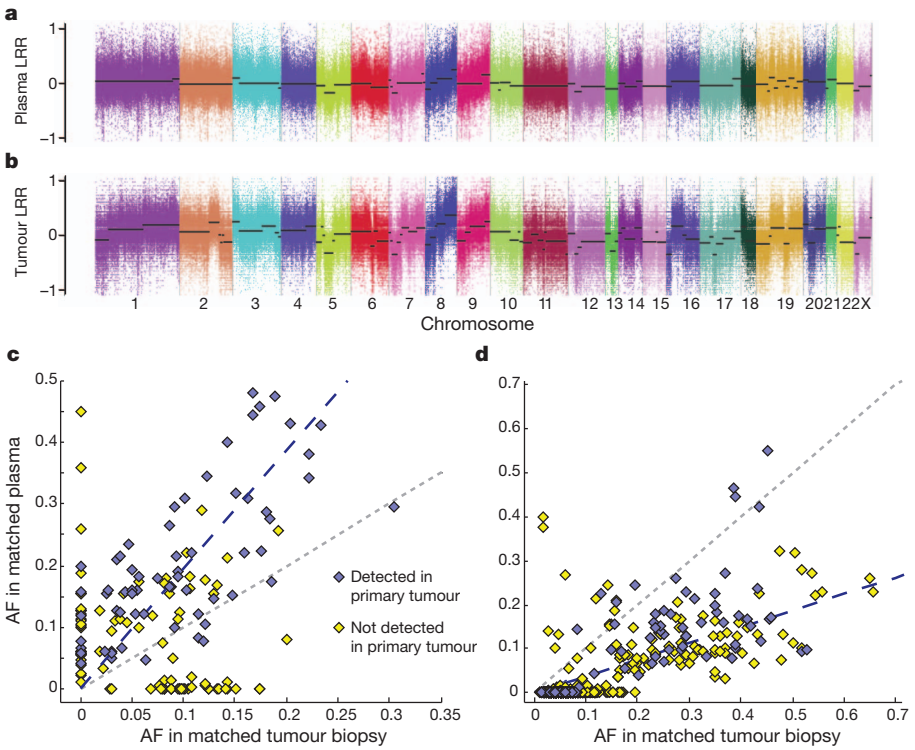

Figure 3 | Genome-wide concordance between plasma DNA and tumour DNA. **a**, **b**, Sequencing data were used to assess CNAs in the plasma sample (**a**) and in the synchronous metastatic tumour biopsy (**b**) from case 4. Panels show log *R* ratio (LRR), calculated on the basis of exome data, between plasma DNA and normal DNA (**a**) and between tumour and normal DNA (**b**). **c**, AF of mutations identified in exome data from plasma or metastatic biopsy for case 1. Grey dotted line shows equality. Blue dashed line has a slope of 1.93, indicating the median of the AF ratio for mutations found in both samples. Key applies to **c** and **d**. **d**, As **c** but for case 4, blue dashed line has a slope of 0.37.

For two cases, sequencing data were also available from metastatic tumour biopsies, collected at the same time as plasma samples (case 1 sample E1, and case 4 sample E2), and from tumour samples collected at the patients' initial presentation, 9 and 4.5 years earlier. CNAs were concordant between plasma and metastasis DNA in both patients (Fig. 3a, b, and Supplementary Fig. 7). Mutations identified in sequencing data[20–23] from the plasma or metastatic biopsy were compared (Supplementary Information). In case 1 with breast cancer, 151 mutations were identified in either the plasma or the synchronous biopsy. Of these, 93 mutations were found in both, and mutant AFs for these were higher in the plasma sample compared to the metastatic biopsy. The correlation coefficient of mutant AFs was positive (0.71) for mutations that were also found in the primary tumour, but negative (−0.22) for other mutations (Fig. 3c). In case 4 with ovarian cancer, 895 mutations were identified in either plasma or the tumour biopsy. For 172 mutations found in both, AFs were positively correlated (0.72) and were higher in the metastatic biopsy, which also contained 686 'private' mutations with AF < 0.2 that were not found in either the plasma or the earlier tumour sample (Fig. 3d).

To identify changes in the mutation profiles of the tumours, we compared the abundance of somatic mutations found in plasma before and after each course of systemic treatment. For each patient, we examined a conservative list of mutations, including all mutations that were called in any of the plasma samples with a Bonferroni-corrected binomial probability of <0.05 assuming a background sequencing error rate of 0.1%. For each mutation and course of treatment (spanned by a pair of plasma samples), a *P*-value for a possible change in mutant AF was calculated as the binomial probability of obtaining the observed number of mutant reads, given the sequencing depth and the observed abundance in the paired time-point, normalized by the fractional concentration of tumour-derived DNA in the plasma (based on genome-wide aggregated allelic loss[5], Supplementary Table 3). Overall, 364 non-synonymous mutations passed with false discovery rate of <10% for significant changes in normalized abundance, ranging from 15 to 121 for each case (median 49). These include mutations in well-known cancer genes, genes linked to drug resistance and drug metabolism, and genes not previously associated with carcinogenesis or therapy resistance (Supplementary Tables 4–9). Selected examples are shown in Table 1 and Fig. 2.

We highlight here five examples. In case 1 with breast cancer, a strong increase was observed in the abundance of an activating mutation in *PIK3CA* following treatment with paclitaxel (Fig. 2a and Table 1). This mutation has been shown to promote resistance to paclitaxel in mammary epithelial cells[8]. In case 2, a patient with an oestrogen-receptor (ER)-positive, HER2-positive breast cancer, treatment with tamoxifen in combination with trastuzumab led to an increase in abundance of a nonsense mutation near the carboxy terminus of *MED1*, an ER co-activator that has been shown to be involved in tamoxifen resistance[10,11]. After further treatment of this patient with lapatinib in combination with capecitabine, we observed an increase in abundance of a splicing mutation in *GAS6*, the ligand for the tyrosine kinase receptor AXL (Fig. 2b, Table 1). Activation of the AXL kinase pathway has been shown to cause resistance to tyrosine kinase inhibitors in NSCLC[13] and resistance to lapatinib in ER-positive, HER2-positive breast cancer cell lines[12]. In case 4 with ovarian cancer, following treatment with cisplatin, we observed increase in abundance of a truncating mutation in the tumour-suppressor *RB1* (Fig. 2d, Table 1), predicted to inactivate the RB1 protein (Supplementary Fig. 8). In the matched metastasis biopsy obtained after treatment, the mutation was found in 95% of sequencing reads (59 of 62), with apparent loss of heterozygosity at 13q containing the *RB1* gene (Fig. 3a, b). Loss of *RB1* has been linked with chemotherapy response[9]. Case 6 was a NSCLC patient with an activating mutation in *EGFR* who was treated with gefitinib but progressed on treatment. Analysis by digital PCR detected the EGFR T790M mutation in plasma at progression, but not at the start of treatment. This mutation inhibits binding of

Table 1 | Selected mutations whose mutant AF significantly increased following treatment

Patient	Cancer type	Gene	Effect	Potential biological interest	Associated treatment	Mutant AF in plasma	
						Before	After
Case 1	Breast	*PIK3CA*	E545K	PI-3-kinase. p.E545K mutation associated with chemoresistance in mammary epithelial cells[8].	Paclitaxel	14%	34%
Case 1	Breast	*BMI1*	S324Y	BMI1 polycomb ring finger oncogene. Associated with chemoresistance[25].	Paclitaxel	3%	12%
Case 1	Breast	*SMC4*	I1000S	Structural maintenance of chromosomes 4. Downregulated in taxane resistant cell lines[26].	Paclitaxel	14%	22%
Case 1	Breast	*FANCD2*	G56V	Fanconi anaemia complementation group D2. Chromatin dynamics and DNA crosslink repair[27].	Epirubicin	3%	13%
Case 2	Breast	*MED1*	S1179X	Mediator complex subunit 1. Co-activator of ER with functional role in tamoxifen resistance[10,11].	Tamoxifen/trastuzumab	4%	15%
Case 2	Breast	*ATM*	I2948F	Ataxia telangiectasia mutated.	Tamoxifen/trastuzumab	6%	45%
Case 2	Breast	*PDGFRA*	D714E	Platelet-derived growth factor alpha. Cell surface tyrosine kinase receptor.	Tamoxifen/trastuzumab	0%	15%
Case 2	Breast	*GAS6*	Splicing	Growth arrest-specific 6. Ligand for AXL, overexpression associated with TKI resistance[12,13].	Lapatinib/capecitabine	6%	30%
Case 2	Breast	*TP63*	Splicing / S551G	Tumour protein p63.	Lapatinib/capecitabine	4%	20%
Case 4	Ovarian	*RB1*	E580X	Retinoblastoma 1. Loss of RB1 associated with EMT and drug resistance[9].	Cisplatin	14%	22%
Case 4	Ovarian	*ZEB2*	Y663C	Zinc finger E-box binding homeobox 2. Overexpression associated with cisplatin resistance in ovarian cancer[28].	Cisplatin	11%	15%
Case 4	Ovarian	*MTOR*	K1655N	Mechanistic target of rapamycin. Activating mutations in mTOR confers resistance to antimicrotubule agents[29].	Paclitaxel	8%	14%
Case 5	Ovarian	*CES4A*	P55S	Carboxylesterase 4A. Hydrolysis or transesterification of various xenobiotics.	Carboplatin/paclitaxel Carboplatin/liposomal doxorubicin	0% 6%	6% 13%
Case 5	Ovarian	*BUB1*	M889K	Mitotic checkpoint serine/threonine-protein kinase.	Carboplatin/paclitaxel	11%	34%
Case 5	Ovarian	*PARP8*	P81T	Poly [ADP-ribose] polymerase family, member 8.	Liposomal doxorubicin	23%	30%
Case 6	Lung	*EGFR*	T790M	Epidermal growth factor receptor. Established to cause gefitinib resistance by inhibiting drug binding[14].	Gefitinib	0%	13%
Case 6	Lung	*TP53*	Y163C	Tumour protein p53[30].	Gefitinib	0%	14%
Case 6	Lung	*NFKB1*	G489V	Nuclear factor κB[30].	Gefitinib	0%	17%

Potential biological role and associations with drug resistance described in literature are highlighted. The ''Effect'' column lists predicted change in amino acid sequence.

gefitinib to EGFR and has been established as the main driver of acquired resistance to gefitinib[14]. Unbiased analysis of plasma DNA by exome sequencing identified selection for this mutation amongst genomic changes that occurred following therapy (Fig. 2f, Table 1).

In this proof of principle study, we demonstrate that exome analysis of plasma ctDNA represents a novel paradigm for non-invasive characterization of tumour evolution. Our data, together with recent reports[5,7], show that CNAs and somatic mutations identified in ctDNA are widely representative of the tumour genome and provide an alternative method of tumour sampling that can overcome limitations of repeated biopsies. Cell-free DNA fragments from multiple lesions in the same individual all mix together in the peripheral blood[5], therefore ctDNA is likely to contain a wider representation of the genomes from multiple metastatic sites, whereas mutations present in a single biopsy or minor sub-clone may be missed. This strengthens the case for the use of ctDNA as a biomarker for monitoring tumour burden or for the analysis of hotspot mutation regions[1,6,16,17], but also indicates that tracking different mutations for assessment of tumour heterogeneity and clonal evolution is now possible. Our data identified a subset of genes that were positively selected following treatment, many of which have been previously associated with drug resistance. Other changes may represent 'passenger' mutations or false-positives, but some are likely to contribute to resistance to therapy. Accumulating data across a large number of cases could identify new genes or pathways that are frequently mutated following specific treatment types, and help refine analysis algorithms.

The approach we describe here may be broadly applicable to a large fraction of advanced cancers, where the median mutation burden in plasma (before start of treatment) is 5%–10% (refs 6, 16, 24). Analysis of acquired drug resistance is of particular utility in advanced or metastatic cancers, which is the target population for nearly all early phase clinical trials. Improvements in sequencing and associated technologies may enable similar analysis in cases with a lower tumour burden in plasma. At present, this non-invasive approach for characterizing cancer exomes in plasma is readily applicable to patients with high systemic tumour burden, enabling detailed and comprehensive evaluation of clonal genomic evolution associated with treatment response and resistance.

METHODS SUMMARY

Patients and samples. Cases 1–5 were recruited as part of prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference nos 07/Q0106/63, 08/H0306/61 and 07/Q0106/63). Case 6 was recruited as part of the 'Hydroxychloroquine and gefitinib to treat lung cancer' study (NCT00809237) at the National University Health System, Singapore, approved by the National Healthcare Group NHG IRB—DSRB 2008/00196. Written informed consent was obtained from patients, and serial blood samples were collected at intervals of ≥3 weeks.

Extraction and sequencing of plasma DNA. DNA was extracted from plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions. Barcoded sequencing libraries were prepared using a commercially available kit (ThruPLEX-FD, Rubicon Genomics). Pooled libraries were enriched for the exome using hybridization (TruSeq Exome Enrichment Kit, Illumina), quantified using quantitative PCR and pooled in 1:1 ratio for paired-end sequencing on a HiSeq2500 (Illumina).

Variant calling and analysis. Sequencing data were demultiplexed and aligned to the hg19 genome using BWA[20]. Pileup files for properly paired reads with mapping quality ≥60 were generated using samtools[22]. AFs were calculated for all Q30 bases. A mutation was called if ≥4 mutant reads were found in plasma with ≥1 read on each strand, and no mutant reads were observed in germline DNA or in a prior plasma sample with ≥10-fold coverage. For comparison between consecutive plasma samples in a patient, we calculated the binomial probability of obtaining the observed AF (or greater) if the abundance of the mutant allele, normalized by tumour load in plasma (based on a modified genome-wide aggregated allelic loss method[5]), had remained constant between the two samples.

Full Methods and any associated references are available in the online version of the paper.

Received 5 October 2012; accepted 11 March 2013.
Published online 7 April 2013.

1. Diaz, L. A. Jr *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486,** 537–540 (2012).

2. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368,** 842–851 (2013).
3. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
4. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–399 (2012).
5. Chan, K. C. *et al.* Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin. Chem.* **59,** 211–224 (2013).
6. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4,** 136ra168 (2012).
7. Leary, R. J. *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci. Transl. Med.* **4,** 162ra154 (2012).
8. Isakoff, S. J. *et al.* Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Res.* **65,** 10992–11000 (2005).
9. Knudsen, E. S. & Knudsen, K. E. Tailoring to RB: tumour suppressor status and therapeutic response. *Nature Rev. Cancer* **8,** 714–724 (2008).
10. Cui, J. *et al.* Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res.* **72,** 5625–5634 (2012).
11. Nagalingam, A. *et al.* Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis* **33,** 918–930 (2012).
12. Liu, L. *et al.* Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer Res.* **69,** 6871–6878 (2009).
13. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nature Genet.* **44,** 852–860 (2012).
14. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2,** e73 (2005).
15. Punnoose, E. A. *et al.* Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. *Clin. Cancer Res.* **18,** 2391–2401 (2012).
16. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nature Med.* **14,** 985–990 (2008).
17. McBride, D. J. *et al.* Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosom. Cancer* **49,** 1062–1069 (2010).
18. Yung, T. K. *et al.* Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin. Cancer Res.* **15,** 2076–2084 (2009).
19. Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2,** 61ra91 (2010).
20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
21. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
22. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
23. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
24. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl Acad. Sci. USA* **102,** 16368–16373 (2005).
25. Siddique, H. R. & Saleem, M. Role of BMI1, a stem cell factor, in cancer recurrence and chemoresistance: preclinical and clinical evidences. *Stem Cells* **30,** 372–378 (2012).
26. Chang, H. *et al.* Identification of genes associated with chemosensitivity to SAHA/taxane combination treatment in taxane-resistant breast cancer cells. *Breast Cancer Res. Treat.* **125,** 55–63 (2011).
27. Sato, K. *et al.* Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J.* **31,** 3524–3536 (2012).
28. Haslehurst, A. M. *et al.* EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer. *BMC Cancer* **12,** 91 (2012).
29. VanderWeele, D. J., Zhou, R. & Rudin, C. M. Akt up-regulation increases resistance to microtubule-directed chemotherapeutic agents through mammalian target of rapamycin. *Mol. Cancer Ther.* **3,** 1605–1613 (2004).
30. Wu, C. C., Yu, C. T., Chang, G. C., Lai, J. M. & Hsu, S. L. Aurora-A promotes gefitinib resistance via a NF-κB signaling pathway in p53 knockdown lung cancer cells. *Biochem. Biophys. Res. Commun.* **405,** 168–172 (2011).

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

Acknowledgements We thank J. Langmore and K. Solomon (Rubicon Genomics) for early access to library preparation products. We thank L. Jones, S. Richardson, C. Hodgkin and H. Biggs for recruiting patients into the DETECT and CTCR-OVO4 studies, all medical and ancillary staff in the breast and gynaecological cancer clinic and patients for consenting to participate. We thank the Human Research Tissue Bank at Addenbrooke's Hospital which is supported by the NIHR Cambridge Biomedical Research Centre. We thank the Cancer Science Institute, National University of Singapore, and the Hematology-Oncology Research Group, National University Health System, Singapore for their support. We acknowledge the support of Cancer Research UK, the University of Cambridge, National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge Experimental Cancer Medicine Centre, Hutchison Whampoa Limited, and the National Medical Research Council, Singapore. S.-J.D. is supported by an Australian NHMRC/RG Menzies Early Career Fellowship that is administered through the Peter MacCallum Cancer Centre, Victoria, Australia.

Author Contributions M.M., S.-J.D., T.F., D.W.Y.T., D.G., J.D.B., C.C. and N.R. designed the study. M.M., D.W.Y.T. and T.F. developed methods. S.-J.D., C.P., A.S.C.W., T.M.C., J.D.B. and C.C. designed and conducted the prospective clinical studies. M.M., S.-J.D., D.W.Y.T., D.G., T.F. and A.M.P. generated data. Z.K., S.H. and D.B. contributed sequencing data. M.M., F.M. and N.R. analysed sequencing data. S.-F.C. and J.H. contributed to experiments and data analysis. M.M., S.-J.D., D.W.Y.T., T.M.C., J.D.B., C.C. and N.R. interpreted data. M.M. and N.R. wrote the paper with assistance from S.-J.D., D.W.Y.T., C.C., J.D.B. and other authors. All authors approved the final manuscript. J.D.B., C.C. and N.R. are the project co-leaders and joint senior authors.

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. Correspondence and requests for materials should be addressed to J.D.B. (james.brenton@cruk.cam.ac.uk), C.C. (carlos.caldas@cruk.cam.ac.uk) or N.R. (nitzan.rosenfeld@cruk.cam.ac.uk).

METHODS

Sample collection. Cases 1–5: patients were recruited as part of prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by local research ethics committee (REC reference nos 07/Q0106/63, 08/H0306/61 and 07/Q0106/63). Written informed consent was obtained from the patients. Serial blood samples were collected in EDTA tubes at intervals of ≥3 weeks, and centrifuged within 1 h at 820*g* for 10 min to separate the plasma from the peripheral blood cells. The plasma was then further centrifuged at 20,000*g* for 10 min to pellet any remaining cells. The plasma was then stored at −80°C until DNA extraction.

Case 6: this patient was recruited as part of the 'Hydroxychloroquine and gefitinib to treat lung cancer' study (NCT00809237) at the National University Health System, Singapore, approved by the National Healthcare Group NHG IRB-DSRB 2008/00196. Blood was collected in CPT tubes (BD Vacutainer) before gefitinib was started, and at monthly intervals while the patient was on treatment, until disease progression. Blood collected was spun within 1 h at 1,500*g* for 20 min, and the plasma fraction was frozen at −80 °C. Thawed samples were recentrifuged at 20,000*g* for 10 min to further separate any cellular portions.

Extraction of plasma DNA. DNA was extracted from aliquots of plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions (see Supplementary Table 1 for volumes used). DNA was eluted into buffer AVE, eluted twice through each column to maximize yield, and stored at −20 °C.

Extraction of normal and tumour DNA. DNA from tumour sections was extracted using DNeasy tissue or DNA Allprep kits (Qiagen) according to manufacturer's instructions. Matched germline DNA was derived from normal peripheral blood leucocytes. After the collection of plasma from each blood sample, the remaining layer of normal peripheral blood lymphocytes ('buffy coat') was removed. This layer was either subjected to red cell lysis using a red cell lysis buffer (155 mM NH_4Cl, 10 mM $KHCO_3$ and 0.1 mM EDTA pH 7.4) and DNA extracted using a standard phenol-chloroform extraction protocol; or frozen at −80 °C before extraction using QIAamp DNA mini kit (Qiagen).

Sequencing of plasma DNA. Concentration of DNA for each plasma sample was determined using digital PCR, with an assay targeting *RPP30* for case 2, *TP53* for cases 3–5 and *EGFR* for case 6. For case 1, DNA concentration and 'anchor' mutation AF were calculated by averaging results from six assays targeting *PIK3CA*, *MET*, *IQCA1*, *CD1A*, *KIAA0406* and *ZFYVE21*. Libraries were generated using a commercially available kit for fragmented DNA (ThruPLEX-FD, Rubicon Genomics). 2.3–40 ng of DNA (Supplementary Table 2) was used to generate a sequencing library using manufacturer's protocols. Separate unique molecular identifiers were used for each sample. 30 μl of the library volume was obtained for each sample. 2–5 plasma DNA libraries from each patient were made and pooled together for exome capture using hybridization (TruSeq Exome Enrichment Kit, Illumina). Pools were concentrated using vacuum (Eppendorf Vacuum Concentrator) and prepared to 40 μl volume. Exome enrichment was performed following manufacturer's protocols. Enriched libraries were quantified using quantitative PCR and pooled in 1:1 ratio for paired-end next generation sequencing on HiSeq2500 (Illumina).

Sequencing of normal and tumour DNA. Sequence data for tumour and germline samples for case 1 have been reported previously. In brief, genomic libraries from tumour and matched normal tissue were prepared using the standard Illumina paired-end sample preparation kit according to the manufacturer's instructions. DNA fragments of 300 bp in size were sequenced using paired-end 100 bp reads on a HiSeq2000 (Illumina) achieving a depth of >30×. Germline samples for cases 2–6 and tumour sample for case 4 were sheared using Covaris and exome sequenced as described above.

Digital PCR. The principle of microfluidic digital PCR and its use for quantification of tumour DNA has been described previously[6,18]. Assays were designed based on TaqMan chemistry. All digital PCR analysis was carried out on the BioMark system using 12.765 Digital Arrays (Fluidigm) following manufacturer's instructions and protocol. Briefly, 3.5 μl from the eluted DNA was heated to 95 °C for 1 min and placed on ice, then mixed with TaqMan Universal PCR Master Mix (Applied Biosystems) and sample loading buffer (Fluidigm) into a final reaction volume of 10 μl and loaded into each panel of the chip. The reaction mix was then automatically partitioned into 765 reaction chambers. The numbers of starting template DNA molecules were calculated using Poisson statistics based on the number of positive amplifications[6,18].

Analysis of sequencing data. Sequencing reads were demultiplexed allowing zero mismatches in barcodes. Paired-end alignment to the hg19 genome was performed using BWA version 0.5.9 for all exome sequencing data including germline samples, plasma samples and tumour metastasis where generated[20]. PCR duplicates were marked using Picard. Local realignment was performed using Genome Analysis Tool Kit (GATK)[21]. Pileup files were generated for the genomic regions targeted by exome enrichment using samtools v0.1.17[22]. For plasma samples, properly paired reads with mapping quality ≥60 were used to generate the pileup. AFs for each single-base locus were calculated for all bases with phred quality ≥30. For germline DNA, an additional pileup file was generated (using a mapping quality cut-off of ≥1 and without any base quality cut-offs) and was used as reference for calling somatic variants. A mutation was called if no mutant reads for an allele were observed in germline DNA at a locus that was covered at least 10 fold, and if at least 4 reads supporting the mutant were found in the plasma data with at least 1 read on each strand (forward and reverse). At loci with <10-fold coverage in normal DNA and no mutant reads, mutations were called in plasma if a prior plasma sample showed no evidence of a mutation and was covered adequately (10 fold or more). All mutations were annotated for genes and function as well as repeated genomic regions using ANNOVAR[23].

AF was defined as the number of high quality reads supporting a mutation as a fraction of the total number of high quality reads covering the locus. For each patient, AF and number of reads for any mutations called with the above parameters were identified in all plasma samples. A binomial probability of obtaining the observed number of reads given depth in each plasma sample was calculated. The minimum of these probability values was corrected using Bonferroni correction for 62 million × n hypotheses tested, where n was the number of plasma samples sequenced (3 samples for cases 1–4, 5 samples for case 5 and 2 samples for case 6). Mutations with corrected P-values under 0.05 were retained for further analysis in plasma samples.

Estimation of CNAs. To assess CNAs, plasma DNA and tumour sequencing data were compared to germline DNA data at single nucleotide polymorphisms (SNPs) covered within the targeted exome region. The SNPs were identified from the publicly available 1000 Genomes Project data.

Depth information was normalized by dividing the depth of each SNP by the median depth across all SNPs. The log R ratio (LRR) was computed as the base-10 logarithm of the sample depth (metastasis or plasma) divided by the depth of the normal. Each chromosome was segmented by an iterative process that considered non-overlapping blocks of 1,000 data points. Points lying at least 1.5 standard deviations away from the median LRR for the block were removed from the mean LRR computation. If the difference in mean LRR between two consecutive blocks was less than 0.12, the blocks were merged into a single segment whose mean LRR was re-computed using points from both blocks.

Segmentation of B allele frequency (BAF) plots was similarly performed, considering windows of 1,000 data points and starting new segments if the difference in median frequency was greater than 4%. Blocks whose median frequency was within 8% of the median chromosome frequency in the normal sample were considered consistent with the BAF of the normal sample.

Comparison of mutations between plasma and tumour. For tumour/plasma comparison presented for cases 1 and 4, we identified all mutations called in data from synchronous plasma and metastatic tumour samples, as described above. We retained all mutations adequately covered in both samples (minimum 50 reads in plasma, minimum 10 reads in synchronous tumour whole genome data for case 1, minimum 50 reads in synchronous tumour exome data for case 4). We further discarded all mutations with no coverage in archived tumour samples obtained earlier (9 years earlier for case 1, and 4.5 years earlier for case 4).

Identification of mutations that changed in representation over treatment. To estimate systemic tumour burden, we calculated fractional concentration of ctDNA in blood using an adaptation of genome-wide aggregated allelic loss[5]. AFs of SNPs from the 1000 Genomes Project were obtained for germline and plasma data. SNPs with $0 < AF < 1$ in germline DNA were identified. SNPs where the minor AF in the germline data deviated from heterozygosity were identified using a binomial probability of obtaining the observed number of minor allele reads given depth in germline DNA and expected AF of 0.5. SNPs with probability <0.25 were discarded from further analysis.

Of the remaining SNPs, significant deviation from heterozygosity in any of the sequenced plasma samples, determined by a binomial distribution using sequencing depth and expected AF of 0.5, was used to identify loss of heterozygosity (LOH). SNPs with a probability <0.01 in any of the sequenced plasma samples were retained for estimation of tumour burden as described previously[5]. Fractional ctDNA burden was calculated as follows:

1 – [(sum of reads in the lost alleles)/(sum of reads in the retained alleles)]

AFs for all mutations were normalized by the estimated tumour burden. For any comparison between two consecutive plasma samples in a patient, we calculated the binomial probability for the observed difference in AF assuming no difference in normalized abundance. For a comparison between (for example) E1 and E2, we calculated the probability of obtaining the observed number of mutant reads or greater in E2 if normalized abundance in E2 had remained the same as in E1; this probability was multiplied by the probability of the observed number of mutant reads or less in E1 if the normalized abundance in E1 was the same as observed in E2. Where no mutant reads were obtained in the E1, only the reverse direction was used for this analysis. Changes in representation with a false discovery rate of 10% or lower, which were exonic non-synonymous or splicing mutations, were retained and are presented in Supplementary Tables 2–7.