

Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer

Isaac Garcia-Murillas^{1†}, Gaia Schiavon^{1,2†#}, Britta Weigelt³, Charlotte Ng³, Sarah Hrebien¹, Rosalind J Cutts¹, Maggie Cheang⁴, Peter Osin², Ashutosh Nerurkar², Iwanka Kozarewa¹, Javier Armisen Garrido¹, Mitch Dowsett^{1,2}, Jorge S Reis-Filho³, Ian E. Smith² and Nicholas C. Turner^{1,2*}

1 The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, Fulham Road, London, SW3 6JB, UK.

2 Breast Unit, Royal Marsden Hospital, Fulham Road, London, SW3 6JJ, UK

3 Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

4 Institute of Cancer Research Clinical Trials and Statistics Unit, The Institute of Cancer Research, 15 Cotswold Road, Surrey, SM2 5NG, UK.

Current address: Translational Science, Oncology iMed, AstraZeneca, Cambridge, UK

*To whom correspondence should be addressed: nicholas.turner@icr.ac.uk

† These authors contributed equally to this work.

One Sentence Summary: Non-invasive mutation tracking in plasma can detect circulating tumor DNA arising from residual micro-metastatic disease and thus identify patients at high-risk of recurrence.

Abstract

The identification of early stage breast cancer patients at high risk of relapse would allow tailoring of adjuvant therapy approaches. We assessed whether analysis of circulating tumor DNA (ctDNA) in plasma can be used to monitor for minimal residual disease (MRD) in breast cancer. In a prospective cohort of 55 early breast cancer patients receiving neoadjuvant chemotherapy, detection of ctDNA in plasma after completion of apparently curative treatment—either at a single post-surgical time-point or with serial follow-up plasma samples—predicted metastatic relapse with high accuracy (hazard ratio 25.1, CI 4.08-130.5, log rank $P < 0.0001$; or hazard ratio 12.0, CI 3.36-43.07, log rank $P < 0.0001$, respectively). Mutation tracking in serial samples increased sensitivity for prediction of relapse, with a median lead-time of 7.9 months over clinical relapse. We further demonstrated that targeted capture sequencing analysis of ctDNA could define the genetic events of MRD, and that MRD sequencing predicted the genetic events of the subsequent metastatic relapse more accurately than sequencing of the primary cancer. Mutation tracking can therefore identify early breast cancer patients at high risk of relapse. Subsequent adjuvant therapeutic interventions could be tailored to the genetic events present in the MRD, a therapeutic approach that could in part combat the challenge posed by intra-tumor genetic heterogeneity.

Introduction

Breast cancer is the most frequently diagnosed cancer worldwide and in women the second most common cause of cancer deaths (1). Approximately 95% of women with breast cancer present with early stage cancer without macroscopic evidence of metastases (2). In many women, however, breast cancer has already metastasized at diagnosis and such micro-metastatic disease can, in time, result in overt metastatic recurrence. The identification of those patients who have residual micro-metastatic disease, or minimal residual disease (MRD), that has not been eradicated by adjuvant systemic therapy and surgery, would allow for the development of clinical trials of adjuvant therapies to prevent relapse focused on those who are at highest risk.

Circulating tumor-derived DNA (ctDNA) can be detected in the plasma and serum of patients with advanced cancer (3), acting as a potential non-invasive source to characterize the somatic genetic features of their tumors (4-7). Limited data are available on whether ctDNA analyses would be applicable to early cancer (8, 9), in part because the low tumor burden of micro-metastatic disease makes detection of ctDNA challenging, as ctDNA is often detectable at a very low level in plasma DNA (10, 11)

Here we assess the potential to detect ctDNA in early-stage, primary breast cancer. We demonstrate that ctDNA detection with personalized digital PCR (dPCR) assays of somatic mutations, can be used to identify MRD, and that such MRD detection with plasma ctDNA can accurately identify patients at risk of cancer relapse. After detection of patient ctDNA with dPCR mutation tracking, we further asked whether high-depth targeted MPS of ctDNA could be used to interrogate the genetics of MRD as a potential route to identify the lethal clone in genetically heterogeneous cancers.

Results

Personalized tumor-specific digital PCR assays for mutation tracking

We investigated the potential utility of ctDNA analysis in early breast cancer in a prospectively accrued cohort of 55 women presenting with early stage breast cancer who received neoadjuvant chemotherapy followed by surgery (Fig. 1). We subjected primary tumor DNA, extracted from a tumor biopsy at diagnosis prior to treatment, to MPS, identifying one or more somatic mutation(s) in 78% (43/55, 95% CI 65%-88%) of cancers, with two or more mutations found in 12 cases.

To track these mutations in plasma DNA and identify the presence of ctDNA, we designed personalized dPCR assays for each somatic mutation identified (table S1). dPCR can accurately quantify mutant DNA at single-molecule sensitivity, even in the presence of vast amounts of wild-type DNA (Fig. 1) (12, 13). MPS and dPCR analysis had a high level of agreement in baseline tumor DNA in the assessment of the mutant allele fractions (Fig. 2A), demonstrating the robust ability to develop dPCR assays for diverse mutations. For a representative number of cases ($n = 9$), estimation of mutation frequency in plasma DNA was highly correlated in replicate dPCR assays ($r^2 = 0.98$) (Fig. 2B). In patients with two mutations identified in the primary tumor, we tracked both mutations in plasma with 96% agreement for present/absent mutation-calling in the same plasma DNA sample ($\kappa = 0.92$; 95% CI 0.77-1.0), emphasizing the reproducible and robust nature of the assays developed.

Tracking mutations in ctDNA to identify MRD and anticipate relapse

We next assessed whether dPCR could be employed for the detection of ctDNA to predict early relapse. The personalized dPCR assays were used to track mutations in

serial plasma samples taken at baseline, post-operatively with the sample taken 2-4 weeks after surgery, and then every 6 months during follow-up (Fig. 2C, fig. S1). Two patients whose tumors harbored the same *PIK3CA* c.3140A>T (p.H1047L) somatic mutation illustrated the potential of mutation tracking. In A310001, who remained disease-free at 30 months post-surgical follow-up, the mutation was undetected in all follow-up plasma samples, suggesting clearance of tumor by neoadjuvant chemotherapy and surgery. In contrast, in A310006 ctDNA was detected in the post-surgery sample, identifying the presence of MRD. At 6.2 months post-surgery there was a dramatic increase in mutation abundance, suggesting increasing disease burden, followed by clinical relapse 8.1 months post-surgery with widespread metastatic disease.

We assessed the potential to predict relapse from the different time points of ctDNA analysis, starting with the baseline plasma sample taken at diagnosis before any treatment. Consistent with previous observations (9), ctDNA was detected in 69% (29/42, 95% CI 53%-82%) of baseline plasma samples (Fig. 3). The level of baseline ctDNA was associated with markers of disease aggressiveness, such as histological grade and estrogen receptor (ER)-negative status, but not primary tumor size (table S2). ctDNA detection at baseline, prior to any treatment, was not predictive of disease-free survival (Fig. 3A), and ctDNA abundance at baseline was not statistically associated with early relapse (median 13.8 vs. 3.13 mutations per ml for early relapse and disease-free survival, respectively) (Fig. 3B).

We next assessed the potential of a single post-surgical sample taken 2-4 weeks after surgery. ctDNA was detected in the single post-operative blood test in 19% (7/37, 95% CI 8-35%) of patients (Fig. 4A), with highly variable mutational load (median of 19.2 copies/ml, range 1.8 to 6284 copies/ml). In these samples, ctDNA

detection was predictive of early relapse (disease-free survival ctDNA detected median 6.5 months vs not-detected median not reached, HR 25.1, 95% CI 4.08-130.5) (Fig. 4A), with a C-index of 0.78. Detection of ctDNA in single post-surgical sample was a significant predictor of early relapse in a multivariable model (table S3). Six patients did not have a sample taken at the 2-4 week post-surgery time-point and were excluded from analysis of this time-point.

We examined whether the detection of ctDNA in serial samples, which we termed “mutation tracking”, could improve relapse prediction compared with a single post-surgery sample (Fig. 4B). Detection of ctDNA in serial samples was predictive of early relapse (disease-free survival ctDNA detected median 13.6 months vs not-detected median not reached, HR 12.0, 95% CI 3.36-43.07), with a C-index of 0.75. Detection of ctDNA by mutation tracking was a significant predictor of early relapse in a multivariable model (table S4).

Sensitivity in a single post-surgical sample was limited by patients who had undetectable ctDNA in the single post-surgical plasma sample and required serial sampling to detect ctDNA. Of the patients who relapsed in the follow-up period, 50% (6/12) had ctDNA detected in the single post-surgical sample and 80% (12/15) had ctDNA detected by mutation tracking. Of the patients who did not relapse, 96% did not have ctDNA detected in either the single post-surgical sample (24/25) or by mutation-tracking (27/28) ($P = 0.0038$ for single post-surgical sample; $P < 0.0001$ for mutation tracking in serial samples, Chi Square test with Yates' correction). One patient with ctDNA detectable post-surgery did not relapse in the follow-up period (A310033) (Fig. 5A). This patient with triple-negative breast cancer had ctDNA detectable in the first post-surgery sample, with a subsequent increase in ctDNA abundance in serial sampling.

Detection of ctDNA had a median 7.9 months (range 0.03-13.6 months) lead time over clinical relapse, identifying the presence of MRD that was not detectable on conventional imaging for patient A310004 (Fig. 5B). ctDNA did not disappear in serial sampling of plasma from patients with early relapse (Fig. 5C). Both assessment of ctDNA in the single post-operative sample and mutation-tracking in serial samples appeared to predict relapse in all the major subtypes of breast cancer (Fig. 6), although mutation-tracking with serial samples was more sensitive in ER⁺ breast cancer.

In the three patients whose metastatic relapse was restricted to the brain, no ctDNA was detectable by dPCR prior to or at relapse. Patients with gliomas frequently have undetectable ctDNA (9), suggesting that the detection of recurrences restricted to the brain may be challenging by means of mutation-tracking. In contrast, extra-cranial MRD was predicted by ctDNA mutation tracking in all patients who relapsed (n=12) (p=0.0022, Chi square test with Yates' correction).

Genomic characterization of MRD by high-depth plasma DNA sequencing

Intra-tumor genetic heterogeneity is found in many solid tumors, including breast cancer (14-16). Adjuvant therapies targeted at the genetic characteristics of the primary cancer may be ineffective if micro-metastatic disease displays different genetic alterations to those found in the primary cancer. Having demonstrated that mutation tracking can anticipate clinical relapse, we set out to establish if high-depth targeted capture MPS of plasma DNA could be employed to interrogate the genetic profile of MRD prior to relapse, and in cancers with intra-tumor genetic heterogeneity assess whether the genetic profile of MRD reflected that of the original primary breast cancer or the subsequent metastatic recurrence.

Once mutation tracking had confirmed the presence of MRD, we sequenced DNA from the primary cancer, from residual primary tumor post-chemotherapy removed at surgery, from plasma DNA taken prior to relapse, and from the subsequent metastasis when biopsied, in 5 patients with a panel targeting all exons of 273 genes recurrently mutated in breast cancer (17). The sequencing strategy was optimized for low DNA inputs using hybrid capture with a custom NimbleGen SeqCap EZ Choice library followed by sequencing on a HiSeq2000 to a mean target depth after duplicate removal of 460x (range 104x-1015x). In all 5 cases, we identified tumor-specific mutations by high-depth sequencing of the plasma DNA taken prior to clinical relapse, demonstrating its widespread applicability in interrogating the genetics of MRD (Fig. 7 and fig S2). Targeted capture MPS also revealed copy number alterations in the plasma samples of two patients, consistent with those documented in the primary tumor and/ or metastatic relapse (fig. S3). Copy number alterations were not identifiable in the remaining cases, likely as ctDNA was present a small fraction of total plasma DNA.

We next addressed whether the repertoire of somatic genetic alterations identified in the analysis of ctDNA arising from MRD would merely recapitulate that of the original primary cancer, or reveal greater diversity reflective of intra-tumoral genetic heterogeneity. In one case, A310006, ctDNA sequencing by MPS revealed no additional genetic events to the *PIK3CA* mutation found in the primary cancer (using AmpliSeq PGM, as there was insufficient DNA for capture sequencing), suggesting homogeneity in the genetics of the MRD for this patient (fig. S2). However, in all other patients we uncovered diversity in the genetics of MRD compared to the primary cancer (Fig. 7 and fig S2). For example, plasma DNA sequencing revealed

substantial divergence of the genetics of the ctDNA arising from RMD compared to the original primary tumor for patient A310012 (fig. S2).

In patient A310003, sequencing of ctDNA revealed the presence of a *PIK3CA* mutation present in both the primary and metastatic lesion; however, the repertoire of somatic mutations found in the plasma (ctDNA arising from MRD) was more similar to that of the subsequently biopsied metastatic relapse than that of the primary cancer (Fig. 7A). In particular, ctDNA sequencing identified an activating *FGFR1* K656E mutation that was not present in the analyzed primary tumor biopsy but was present in the metastasis (Fig. 7A). The *FGFR1* K656E mutation is directly paralogous to the *FGFR3* K650E activating mutation that is frequently found in bladder cancer and in thanatophoric dysplasia type II (18, 19). Similarly, ctDNA sequencing identified loss of an *ESR1* E380Q mutation found in the primary, but not in the metastasis, anticipating loss of estrogen receptor in the metastasis, which was confirmed by immunohistochemistry (fig. S4A). Sequencing of two foci of residual primary tumor post-chemotherapy provided evidence of consistent clonal selection compared to the primary tumor prior to treatment, although the changes did not predict those that were found in the subsequent metastasis (Fig. 7A). When tracked by dPCR, the primary tumor *PIK3CA* mutation remained present in plasma DNA and metastasis, indicating an early clonal event in the cancer.

In patient A310035, sequencing ctDNA prior to relapse predicted acquisition of a *SYNE1* S1244Y mutation in the subsequently biopsied metastasis, as well as enrichment for a *GATA3* frameshift mutation and loss of a *STAT3* mutation (Fig. 7B). In this patient, enrichment for the *SYNE1* mutation was demonstrated in the residual tumor post chemotherapy (fig. S4B). Last, in patient A310004, sequencing of the relapsed tumor revealed an *RBI* R320* somatic mutation that was not detectable

by sequencing of the plasma ctDNA taken 8.1 months prior to relapse (13 months post-surgery) (Fig. 7C). We developed and optimized a multiplex dPCR assay (fig. S4C) to track this mutation, along with the other two mutations in *ANK3* and *XIRP2* that were present in both the primary and recurrence in this patient. dPCR demonstrated that the *RBI* mutation was a late event, only first detectable in a plasma sample taken 16.1 months post-surgery, then expanding in frequency on serial sampling at relapse (Fig. 7C). This suggests that genetic diversity develops in expanding micro-metastatic disease prior to relapse, and that mutation-tracking may have the potential to identify MRD at a point before genetic diversity develops.

Discussion

Here we show that circulating tumor DNA mutation tracking can detect MRD non-invasively and identify earlier which patients are at risk of cancer recurrence. We devised an assay pipeline that uses baseline primary tumor mutations to develop personalized dPCR assays to track the presence of ctDNA in plasma over time. After using dPCR to detect MRD, we showed that high-depth plasma DNA sequencing can help define the repertoire of somatic genetic alterations found in MRD, providing evidence of clonal shifts in response to systemic therapy.

Our data illustrate fundamental principles for the use of ctDNA in the detection of MRD. Driver, and likely clonal, mutations should be tracked in preference to sub-clonal non-driver mutations, which may be lost in the MRD that subsequently repopulates the metastatic recurrence (A310003, A310035, A310012). In our study, serial sampling during follow-up was required for accurate MRD detection and relapse prediction. Our data suggest that the burden of MRD at a single post-surgical time-point soon after completing neoadjuvant chemotherapy is, in some

cases, insufficient for its detection in the plasma DNA; or, owing to lack of proliferation and apoptosis in the MRD, there is no release of ctDNA. These findings contrast with a report where ctDNA detection with BEAMing in a single colorectal cancer sample taken after surgical resection of liver metastases offered high predictive potential, potentially due to the high burden of micro-metastatic disease in this setting (10). Nevertheless, we recommend serial sampling to detect the MRD as it proliferates and expands. Mutation tracking in serial samples may be particularly important in ER-positive breast cancer, to detect ctDNA changes during post-operative endocrine therapy (Figure 6).

We used digital PCR to assay plasma samples in this study. This relatively cost-efficient technology represents, along with BEAMing, the most sensitive techniques currently available for detection of known rare mutations. Alternative techniques include the detection of structural variants (11), although the challenge in advancing this technique to clinical practice is accurate identification of these variants. MPS of plasma is challenged by polymerase error in detecting rare variants, although techniques such as barcoding for error correction (20) may allow MPS to challenge dPCR for sensitivity. Nevertheless, no other technologies have been capable of quantifying ctDNA for early detection of minimal residual disease

Intra-tumor genetic heterogeneity reflecting clonal diversity in cancers (14-16) presents a potential major barrier for successful adjuvant therapy that aims to eradicate micro-metastatic disease and prevent recurrence (21). Adjuvant therapies targeted at the genetic characteristics of the primary cancer may be ineffective if micro-metastatic disease displays different genetic alterations to those found in the primary cancer. Here we show that high-depth targeted capture MPS of ctDNA prior to relapse has the potential to address this challenge by interrogating the genetic

characteristics of MRD to identify the lethal clone that may differ in its repertoire of somatic mutations from the dominant clone in the primary cancer. We detected potentially targetable mutations in the ctDNA that were not in the primary tumor; in some cases, other targetable mutations were lost from primary tumor to ctDNA. Resistance mutations can be detected in plasma many months prior to the development of clinical resistance in the metastatic setting (22, 23) and we demonstrated that this concept can be extended to patients with potential curable micro-metastatic disease.

Targeted therapy has the potential to cure patients in the adjuvant setting with micro-metastatic cancer (24), likely owing to the low disease burden, whereas targeted therapy is almost uniformly unable to cure patients with advanced metastatic disease (25). We demonstrate that the micro-metastatic disease may evolve over time, with late appearance of a *RBI* mutation that might cause resistance to targeted therapy with CDK4/6 inhibitors. This suggests that the approach described here may lead to the detection of MRD prior to the establishment of genetic diversity.

Despite our encouraging findings for improving early treatment of breast cancer patients, the follow-up was short (~2 years), with relatively few patients. Sequencing of the plasma DNA taken at relapse, with the capture 273-gene panel, to 752x depth in one patient with a brain metastasis revealed no detectable mutations. The lack of sensitivity for the detection of metastatic disease restricted to the brain likely suggests that the blood–brain barrier blocks the release of ctDNA into circulation. We were unable to address the potential ctDNA detection in patients who have had primary surgery and no chemotherapy, as our study focused on the ability to detect MRD after neoadjuvant chemotherapy and surgery. In addition, as patients in this series did not have regular imaging studies in the follow-up, a future prospective

study comparing with imaging is required. Our results could potentially be extended by tracking a large number of mutations in each blood sample, in part by greatly extending the targeted sequencing panel to identify multiple mutations in each patient. The theoretical potential for loss of individual driver mutations in metastatic disease could be countered by such an approach, although that is not observed in our series. However, in our study, assaying for three clonal mutations in the post-surgery sample from patient A310004 was unable to detect ctDNA. This suggests that a substantial increase in the number of genetic events tracked per milliliter of plasma would be required.

In conclusion, we have shown that non-invasive mutation tracking in plasma DNA can detect residual micro-metastatic disease, or MRD, which standard treatment has failed to eradicate, and thus identify patients at high-risk of recurrence. Furthermore, we have shown that the genetic events of metastatic disease may differ from those found in the primary tumor. This knowledge will allow for therapeutic interventions tailored to the driver genetic events present in micro-metastases in a new approach that could help combat the challenge posed by intra-tumor genetic heterogeneity.

Methods

Study design

Serial plasma samples were collected from patients with early breast cancer to assess the potential of assays of ctDNA to predict relapse following treatment. Patients (table S5) were recruited from the Royal Marsden Hospital and were treated with standard therapy. Tumor DNA was extracted from the baseline pretreatment tumor biopsy, and sequenced with an amplicon PGM library to identify mutations specific to the tumor.

Digital PCR assays specific to that mutation were developed to track the mutation on circulating tumor DNA, at baseline, and in sequential plasma samples taken post-operatively. The log rank test was used to assess the association between detection of ctDNA and disease free survival (DFS).

Patient Cohort and Sample Collection

Fifty-five patients scheduled to receive neoadjuvant chemotherapy were analyzed from prospective sample collection studies, the ChemoNEAR study (REC Ref No: 11/EE/0063) or the Plasma DNA study (REC Ref No: 10/H0805/50) approved by Research Ethics committees (East of England – Essex and London – Bromley, respectively). Written informed consent was obtained from all participants. Staging investigations were performed at baseline for all node positive and/or cT3/4 patients with CT scan and bone scans, and those with distant metastatic disease were excluded from the study. Patients were treated with standard neoadjuvant chemotherapy (sequential anthracyclines-taxane based chemotherapy in 51 patients, sequential anthracyclines-paclitaxel + carboplatin in 2 patients and docetaxel + cyclophosphamide in 2 patients) with/without trastuzumab depending of HER2 status. After completion of surgery, with or without radiotherapy, patients were treated with adjuvant hormone therapy or trastuzumab as per standard local practice, and followed up in a nurse-led open access follow-up programme. In the series a single patient presented with axillary lymphadenopathy from a cryptic breast primary. This patient had radiotherapy to the breast, axillary and supraclavicular nodes after neoadjuvant chemotherapy, without surgery.

Core biopsies were taken at baseline, at surgery, and where clinically indicated at recurrence. Plasma samples were collected into EDTA K2 tubes at baseline (prior

to chemotherapy), post-surgery (2-4 weeks post-surgery) and every 6-months in follow-up or until relapse, whichever occurred first (Figure 1). Clinicopathological characteristics of the study cohort are available in table S5 while a consort flow diagram of the patients included in the study is available on figure S5. Estrogen receptor (ER), progesterone receptor (PR), and HER2 status were assessed in a single laboratory at the Royal Marsden Histopathology department using standard criteria.

Processing and DNA extraction from tumor samples

Tissue from core biopsies taken at diagnosis, surgery, and recurrence, were formalin-fixed and embedded in paraffin (FFPE). Sections (4-8 x 4µm) were stained with Nuclear Fast Red (NFR) and micro-dissected under a stereomicroscope to achieve >70% tumor cell content. Tumor DNA was isolated using the QIAamp DNA FFPE Tissue Kit (Qiagen) as per manufacturer's instructions. Germline DNA was extracted from buffy coat DNA using DNeasy Blood and Tissue kit (Qiagen) as per manufacturer instructions.

Processing of plasma and extraction of circulating DNA

Blood collected in EDTA K2 tubes was processed within two hours of sample collection and centrifuged at 1600 rpm for 20 min, with plasma stored at -80°C until DNA extraction. DNA was extracted from 2-4 ml of plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions. The DNA was eluted into 50 µl buffer AVE and stored at -20°C.

DNA quantifications from tissue and/or plasma

DNA isolated from tissue or plasma was quantified on a Bio-Rad QX-100 droplet ddPCR using RNase P as the reference gene. One μ l of eluate was added to a digital PCR reaction containing 10 μ l ddPCR Supermix for probes (Bio-Rad) and 1 μ l of TaqMan Copy Number Reference Assay, human, RNase P (Life Technologies) on a total volume of 20 μ l. The reaction was partitioned into \sim 14,000 droplets per sample in a QX-100 droplet generator according to manufacturer's instructions. Emulsified PCR reactions were run on a 96 well plate on a G-Storm GS4 thermal cycler incubating the plates at 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 60 sec, followed by 10 min incubation at 98°C. The temperature ramp increment was 2.5°C/sec for all steps. Plates were read on a Bio-Rad QX-100 droplet reader using QuantaSoft v1.4.0.99 software from Bio-Rad. At least two negative control wells with no DNA were included in every run. The amount of amplifiable RNase P DNA was calculated using the Poisson distribution in QuantaSoft.

Assessment of recovery of mutant DNA extracted from plasma

Genomic DNA (gDNA) was extracted from the *PIK3CA* mutation c.3140A>T GP2d colon adenocarcinoma cell line (European Collection of Cell Cultures (ECACC) Cat no: 95090714), with DNeasy Blood and Tissue Kit (Qiagen) as per manufacturer instructions. DNA was quantified using Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) as per manufacturer instructions. 5 μ g genomic DNA was restriction digested using HindIII endonuclease and the concentration of c.3140A>T mutant DNA copies was calculated using a digital PCR assay on at least 5 replicates, on the QX-100 ddPCR system (Bio-Rad) using the *PIK3CA* c.3140A>T primers and probes described in table S1. A total of 150 copies of the *PIK3CA* c.3140A>T mutation were spiked into 1 ml wild-type (WT) plasma samples and immediately processed as

described above. Of the input *PIK3CA* mutant DNA, 43% 95%CI \pm 2.2% was recovered and analyzed by the digital PCR assay.

Ion PGM sequencing of baseline tumor samples

Sequencing libraries were prepared with a custom Ion AmpliSeq Breast Cancer Panel targeting 14 known breast cancer driver genes (26) (table S6) using the Ion AmpliSeq Library Preparation protocol with 5ng tumor DNA, according to manufacturer's instructions. Following barcoding, libraries were quantified using qPCR, diluted to 100 pM and pooled. Libraries were templated with the Ion PGM template OT2 200 Kit (Life Technologies), and sequenced on a 318 PGM chip using Ion PGM Sequencing 200 Kit v2 (Life Technologies) and 500 flows to a mean depth of x2355. The sequencing resulted in 200,000-650,000 reads per sample. Variant caller v4.0-r73742 with no Hotspot region and configuration "Germ Line Low Stringency" was used for calling variants, and variants not reported as germline in dbSNP were selected as potentially being somatic mutations. Potential somatic mutations were cross-referenced against the 1000 Genomes Project database (www.1000genomes.org), and only variants that did not appear on the 1000 genomes database were taken forward for development of digital PCR assays.

Hybrid capture MPS

In five patients who experienced relapse, DNA extracted from microdissected primary tumor, residual tumor post-chemotherapy removed at surgery, metastatic tumor biopsies where available, plasma DNA samples, and germline lymphocyte DNA to exclude germline polymorphisms, was subjected to targeted capture massively parallel sequencing. Custom oligonucleotides (Roche NimbleGen SeqCap EZ Choice)

were designed for hybridization capture of all protein-coding exons of 273 genes recurrently mutated in breast cancer or involved in DNA repair pathways as previously described (17, 27). Using 12.5 to 50 ng of DNA, barcoded sequence libraries were prepared (NEXTflex barcode adapters, Bioo Scientific), amplified (KAPA Biosystems) and 12 barcoded libraries pooled at equimolar concentrations into a single exon capture reaction as previously described (17, 27, 28). Paired-end 75x75bp sequencing was performed on a single lane of an Illumina HiSeq2000 flow cell. Sequencing reads were aligned to the reference human genome hg19 using the Burrows-Wheeler Aligner (v0.6.2) (29), and local-realignment and base quality recalibration were performed using the Genome Analysis Toolkit (GATK) (30). Duplicates were removed using SAMtools (31). Somatic single nucleotide variants (SNVs) were identified using muTect (32). GATK Haplotype Caller (v3.1.1) (30), and Strelka (33). To minimize potential false positive results obtained with high-depth targeted massively parallel sequencing performed with DNA extracted from formalin-fixed, paraffin-embedded tissues, or plasma only SNVs identified by at least two out of the three callers employed were considered valid as previously described (17). Small insertions and deletions (indels) were identified using GATK Haplotype Caller and VarScan2 (v2.3.6) (34), and gene copy number aberrations were generated using VarScan2, segmented using circular binary segmentation, and gains and losses called using the R package CGHcall. All candidate mutations identified in a given tumor or plasma DNA sample were validated manually in all other plasma DNA or tumor samples of the given patient, and mutations supported by two or more reads were regarded as being present (table S7).

Development of mutation specific digital PCR assays

Digital PCR was performed on a QX-100 droplet digital PCR system (Bio-Rad) using TaqMan chemistry. For each tumor mutation we designed a primer probe combination using Primer 3 plus or Life Technologies' custom Single Nucleotide Polymorphism (SNP) genotyping assays tool (table S1). Primers and probes were designed to avoid reported SNPs. Primers and probes were analyzed for the presence of hairpins, secondary structures or hetero/homo dimers formation. Primers were analyzed for specificity using University of California, Santa Cruz ePCR tool (<http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>). Digital PCR conditions were optimized with a temperature gradient to identify the optimal annealing temperature using either a wild type DNA spiked with a mutant synthetic oligonucleotide, or cell line DNA known to carry the mutation. Following optimization, the baseline tumor DNA sample was analyzed to validate the assay. The complete absence of mutation in the corresponding patients germline lymphocyte DNA, and unmatched plasma samples from metastatic breast cancer patients, was confirmed by dPCR.

Digital PCR analysis of circulating free plasma DNA

Digital PCR was performed on a QX-100 ddPCR system (Bio-Rad) using TaqMan chemistry with primers and probes described in table S1 at a final concentration of 900 nM primers and 250 nM probes. PCR reactions were prepared with ddPCR Supermix for probes (Bio-Rad) and partitioned into a median of 50,000 droplets per sample in a QX-100 droplet generator according to manufacturer's instructions. DNA extracted from 4 ml of plasma was analyzed for the presence of the mutation at each timepoint. Emulsified PCR reactions were run on a 96 well plate on a G-Storm GS4 thermal cycler incubating the plates at 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and specific assay extension temperature (table S1) for 60 sec, followed by

10 min incubation at 98°C. The temperature ramp increment was 2.5°C/sec for all steps. Plates were read on a Bio-Rad QX-100 droplet reader using QuantaSoft v1.4.0.99 software from Bio-Rad to assess the number of droplets positive for mutant DNA, wild type DNA, both or neither. At least two negative control wells with no DNA were included in every run.

Digital PCR analysis

To assess mutation fraction, the concentration of mutant DNA (copies of mutant DNA per droplet) was estimated from the Poisson distribution. Number of mutant copies per droplet $M_{mu} = -\ln(1 - (n_{mu}/n))$, where n_{mu} = number of droplets positive for mutant-FAM probe and n = total number of droplets. The DNA concentration in the reaction was estimated as follows $M_{DNAconc} = -\ln(1 - (n_{DNAconc}/n))$, where $n_{DNAconc}$ = number of droplets positive for mutant-FAM probe and/or Wild Type-VIC probe and n = total number of droplets. The Fraction Mutation = $M_{mu}/M_{DNAconc}$.

To assess the number of mutant copies per ml of plasma, the number of mutant-FAM positive droplets was adjusted for the number of wells run for the sample, the total number of droplets generated, the median volume of a droplet (0.89 pl), and volume equivalent of plasma run, using the following formula:

Mutant copies per ml = (Total number of droplets positive for FAM) x 20,000 x (number of wells run/volume of plasma equivalents) / (total number of droplets generated*0.89)

A mutation was only considered to be present if two or more FAM positive droplets were detected in 4 ml plasma equivalent DNA, with this criterion for a positive test being pre-defined. After taking account of the recovery of 43% of DNA present in plasma (as calculated above) the digital PCR assay would be anticipated to detect tumor specific mutations with 86% probability at an actual concentration of 2 mutations per ml, and 99% probability at an actual concentration of 4 mutations per ml. The presence of increasing ctDNA abundance was defined as the doubling of mutant copies/ml from the nadir or the appearance of ctDNA when previously undetectable. Detection in baseline or post-surgery was defined as detection of ctDNA at the single timepoint. Mutation tracking was defined as detection of ctDNA in any of the post-surgical and serial ctDNA samples.

Statistical analysis

The primary endpoint of the study was to assess disease-free survival in patients with and without detection of ctDNA using univariable survival estimates calculated using the Kaplan-Meier method, and survival differences estimated using the log-rank test. Multivariable Cox regression analyses were used to test the independent prognostic value of mutation tracking and ctDNA, adjusted for tumor size, pathological nodal status and molecular subtypes respectively. Harrell's C-index from the univariable and multivariable survival model was calculated (35). The C-index is a probability of concordance between predicted and observed survival, defined as the probability that risk assignments to members of a random pair are accurately ranked according to their prognosis. C-index values of 0.5 indicate random prediction and higher values indicate increasing prediction accuracy. The associations of median ctDNA level with clinico-pathological markers were assessed using Mann-

Whitney U test or Krustal-Wallis test where appropriate. The odds ratios of ctDNA detection with the standard clinicopathological variables were estimated using univariable logistic regression analysis. All statistical analysis was performed with GraphPad Prism version 6.0, or in R 3.0.1. with R packages *survival*, *Hmisc*, and *rms*. All p values are two sided. Disease-free survival, excluding contralateral invasive breast cancers, was assessed from the date of surgery.

Supplementary Materials

Figure S1. Mutation-tracking by dPCR along a 24-month follow-up of a disease-free patient.

Figure S2. High-depth targeted capture massive parallel sequencing on plasma DNA from two relapsed patients.

Figure S3. Copy number profile in primary tumor and plasma DNA prior to relapse.

Figure S4. Validation and follow-up of MPS on plasma DNA.

Figure S5. CONSORT flow diagram of patients included in this study.

Table S1: Digital PCR assays and mutations analyzed in this study.

Table S2: Clinicopathological factors associated with baseline ctDNA level.

Table S3: Prediction of disease-free survival using a single post-surgery blood sample.

Table S4: Prediction of disease-free survival by mutation-tracking using serial blood samples.

Table S5. Summary of the study cohort.

Table S6: Ion AmpliSeq Breast Cancer driver gene panel

Table S7: Reads from capture MPS of tumor and plasma DNA.

References and Notes

1. A. Jemal *et al.*, Global cancer statistics. *CA Cancer J Clin* **61**, 69-90 (2011).
2. N. Howlader *et al.*, US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status. *J Natl Cancer Inst* **106**, (2014).
3. S. A. Leon, B. Shapiro, D. M. Sklaroff, M. J. Yaros, Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* **37**, 646-650 (1977).
4. T. Forshew *et al.*, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, 136ra168 (2012).
5. S. J. Dawson *et al.*, Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209 (2013).
6. M. Murtaza *et al.*, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108-112 (2013).
7. H. Gevensleben *et al.*, Noninvasive detection of HER2 amplification with plasma DNA digital PCR. *Clin Cancer Res* **19**, 3276-3284 (2013).
8. J. A. Beaver *et al.*, Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res* **20**, 2643-2650 (2014).

9. C. Bettegowda *et al.*, Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014).
10. F. Diehl *et al.*, Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-990 (2008).
11. T. Reinert *et al.*, Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*, (2015).
12. B. Vogelstein, K. W. Kinzler, Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241 (1999).
13. B. J. Hindson *et al.*, High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-8610 (2011).
14. S. P. Shah *et al.*, The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395-399 (2012).
15. N. Navin *et al.*, Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94 (2011).
16. S. Nik-Zainal *et al.*, The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
17. R. Natrajan *et al.*, Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. *J Pathol* **232**, 553-565 (2014).
18. P. M. Lievens, E. Liboi, The thanatophoric dysplasia type II mutation hampers complete maturation of fibroblast growth factor receptor 3 (FGFR3), which activates signal transducer and activator of transcription 1 (STAT1) from the endoplasmic reticulum. *J Biol Chem* **278**, 17344-17349 (2003).
19. D. Cappellen *et al.*, Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nat Genet* **23**, 18-20 (1999).
20. I. Kinde, J. Wu, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 9530-9535 (2011).
21. N. C. Turner, J. S. Reis-Filho, Genetic heterogeneity and cancer drug resistance. *Lancet Oncol* **13**, e178-185 (2012).
22. S. Misale *et al.*, Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532-536 (2012).
23. L. A. Diaz, Jr. *et al.*, The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012).
24. D. Slamon *et al.*, Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med* **365**, 1273-1283 (2011).
25. D. J. Slamon *et al.*, Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**, 783-792 (2001).
26. V. Taly, D. Pekin, A. El Abed, P. Laurent-Puig, Detecting biomarkers with microdroplet technology. *Trends in molecular medicine* **18**, 405-416 (2012).
27. H. H. Won, S. N. Scott, A. R. Brannon, R. H. Shah, M. F. Berger, Detecting somatic genetic alterations in tumor specimens by exon capture and massively parallel sequencing. *J Vis Exp*, e50710 (2013).

28. L. De Mattos-Arruda *et al.*, Establishing the origin of metastatic deposits in the setting of multiple primary malignancies: the role of massively parallel sequencing. *Molecular oncology* **8**, 150-158 (2014).
29. S. Dawood, K. Broglio, A. U. Buzdar, G. N. Hortobagyi, S. H. Giordano, Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. *J Clin Oncol* **28**, 92-98 (2010).
30. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
31. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009).
32. K. Cibulskis *et al.*, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* **31**, 213-219 (2013).
33. C. T. Saunders *et al.*, Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics (Oxford, England)* **28**, 1811-1817 (2012).
34. D. C. Koboldt *et al.*, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* **22**, 568-576 (2012).
35. F. E. Harrell, Jr., K. L. Lee, D. B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* **15**, 361-387 (1996).

Acknowledgments: We thank N. Orr, K. Tomczyk, D. Novo, E. Folkerd, M. Afentakis, K. Sidhu and F. Daley for technical assistance and the Breast Unit Research Team at Royal Marsden for support in enrolment of patients and samples' collection. **Funding:** NIHR funding to the Royal Marsden Biomedical Research Centre, Breakthrough Breast Cancer Research, Cancer Research UK C30746/A16642, the Cridlan Trust and the Breast Cancer Research Foundation. **Authors Contributions:** NCT, IES, and MD designed the trial. NCT, IES, and GS contributed to the recruitment of patients. IG-M, GS, and SH extracted DNA, performed sequencing and dPCR experiments. PO and AN performed pathology and microdissection. BW and IK performed MPS. CN, RJC, MC, and JAG were responsible for bioinformatic analyses. IG-M, GS, JSR-F, and NCT analyzed and interpreted the data. All authors contributed to the writing or review of the manuscript. **Conflicts of interest:** All the authors declare no conflicts of interest. **Data and materials availability:** Targeted capture sequencing data is deposited in NCBI biosample database with accession number SRP058761.

Figures

Figure 1. Personalized digital PCR assays for mutation tracking of circulating tumor DNA in plasma of patients with early breast cancer. The baseline biopsy of patients presenting with early breast cancer was subjected to targeted massively parallel sequencing (MPS) to identify somatic (tumor-specific) mutations. Personalized, patient specific digital PCR (dPCR) assays were developed to detect the mutation in plasma DNA that was extracted from samples taken at baseline, post-surgery, and serially every 6 months during follow-up (mutation tracking). After detection of circulating tumor DNA (ctDNA) with dPCR mutation tracking, plasma samples were subject to high-depth targeted capture MPS to interrogate the repertoire of somatic genetic alterations in detected minimal residual disease. MAF, mutant allele frequency.

Figure 2. Personalized, mutation-specific dPCR accurately quantifies ctDNA and is highly reproducible. (A) Bland-Altman plot of the agreement between mutational frequency assessed by massively parallel sequencing (MPS) and by mutation-specific dPCR on baseline tumor DNA, with 95% CI of agreement -0.13 and 0.14 indicated by dashed lines. Data points from 55 mutation-specific dPCR assays are displayed. (B) Correlation of mutation abundance in repeat assays of mutation-specific dPCR assays in plasma DNA (Pearson's correlation coefficient). Data from 17 mutation-specific assays in 9 different patients. (C) Example of dPCR mutation tracking in two patients with early breast cancer, whose tumors harbored the same tumor *PIK3CA* c.3140A>T (p.H1047L) somatic mutation at the baseline plasma samples. The complete time course for patient A310001 is in fig. S1. In each dPCR plot, green dots represent wild-type DNA (VIC-labeled), blue dots represent mutant DNA (FAM-labeled), brown dots represent droplets containing both wild-type and mutant DNA, and black dots are droplets with no DNA incorporated.

Figure 3. Early relapse is not predicted by analysis of baseline ctDNA. (A) Disease-free survival according to the detection of ctDNA in the baseline plasma sample. *P* value determined by log rank test). (B) Mutant copies per ml of plasma at baseline in patients who relapsed early and who did not relapse during follow up. *P* value determined by Mann Whitney U test. Data in (A and B) are from the same *n* = 42 patients. ctDNA associations with other clinicopathological characteristics are in table S2.

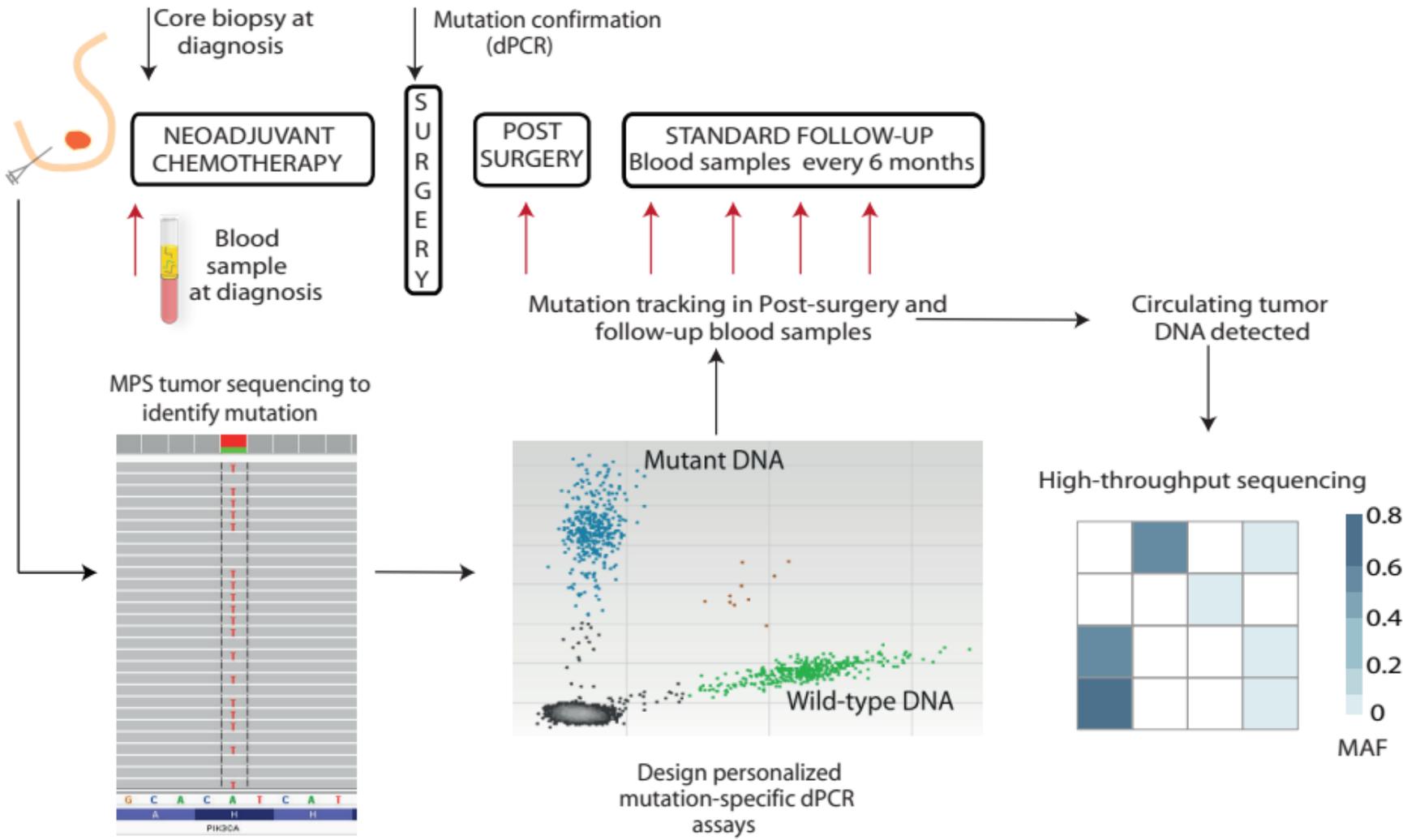
Figure 4. Mutation tracking in serial plasma samples predicts early relapse. (A) Disease-free survival according to the detection of ctDNA in the first post-surgical plasma sample (HR 25.1 95% CI 4.08-130.5). *P* value determined by log rank test. Data are from *n* = 37 patients (B) Disease-free survival according to the detection of ctDNA in serial follow-up samples (HR 12.0 95% CI 3.36-43.07). *P* value determined by log rank test. Data are from *n* = 43 patients [37 of whom are represented in (A)].

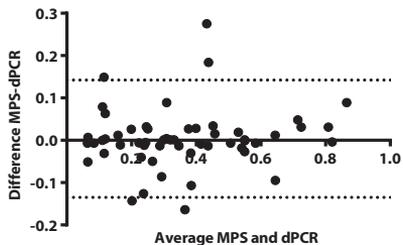
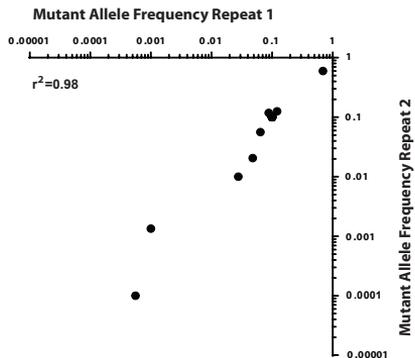
Figure 5. Mutation-tracking in early-relapse and disease-free patients. (A) Mutation tracking profile in 26 patients who are currently disease-free after treatment of primary breast cancer. Mutations remain undetectable in the post-surgical and

follow-up periods in 25/26 patients. The remaining patient (A310033, red), with triple negative disease, had ctDNA detectable post-surgery with an increase in the detectable level of mutational load in follow-up sampling; however, this patient did not have clinical relapse at the time of reporting. **(B)** Mutation tracking detects ctDNA arising from minimal residual disease in patient A310004, with a lead-time of 13.5 months over clinical relapse. Contrast-enhanced spiral CT scan revealed no abnormality at 14 months follow-up, although multiple liver metastases (white arrows) were subsequently detected at 19 months follow-up. M: months; PS: post-surgery. **(C)** Mutation tracking ctDNA profiles prior to relapse, from 12 patients who experienced early relapse after treatment of primary breast cancer. After detecting ctDNA in a post-surgery or follow-up sample, ctDNA was detected in all subsequent samples prior to relapse.

Figure 6. Disease-free survival prediction based on single post-surgery ctDNA and mutation-tracking in serial plasma samples according to tumor subtype. **(A)** ER-positive breast cancers at first post-surgical sample ($n=21$) and serial follow-up samples ($n=24$). **(B)** ER-negative breast cancers at first post-surgical sample ($n=16$) and serial follow-up samples ($n=19$). **(C)** Triple-negative breast cancers (TNBC) at first post-surgical timepoint ($n=11$) and serial follow-up samples ($n=13$). P values determined by log rank test.

Figure 7. High-depth targeted capture MPS of plasma DNA to characterize the genomic landscape of minimal residual disease. For five patients, DNA was extracted from microdissected primary tumor, residual tumor post-chemotherapy removed at surgery, metastatic tumor biopsies where available, plasma DNA samples and germline lymphocyte DNA and subjected to targeted capture massively parallel sequencing using a custom panel targeting the exons of 273 genes recurrently mutated in breast cancer or involved in DNA repair pathways. Two patients are provided in fig. S2. **(A)** (Left) High-depth targeted capture MPS in patient A310003 of baseline primary tumor biopsy, two separate surgery samples post-chemotherapy, plasma DNA samples at 17 and 8 months pre-relapse, and metastatic disease biopsy taken on subsequent relapse. (Right) Validation of MPS findings with dPCR and Sanger sequencing. The *FGFR1* mutation is detectable at a low level 17 months prior to relapse. A diagram illustrates possible clonal selection in this patient through therapy and time. Immunohistochemistry of baseline and metastatic tumor samples are in fig. S5A. **(B)** High-depth targeted capture MPS in patient A310035 of baseline primary tumor biopsy, plasma DNA before relapse, relapse biopsy, and plasma DNA following relapse. Analysis of surgery sample is shown in figure S5B. **(C)** High-depth targeted capture MPS in patient A310004 of baseline primary tumor biopsy, plasma DNA 6.1 months pre-relapse and relapse biopsy (Left). Relapse biopsy revealed a *RBI* c.958C>T (R320*) somatic mutation that was not detectable by sequencing of the plasma. (Right) Validation of MPS findings with digital PCR. *RBI* mutation was a late arising event, that was not detectable 6.1 months prior to relapse, but then rose to relative clonal abundance at relapse ($p<0.0001$, chi-square test for trend). There was a pathological complete response to chemotherapy in the primary tumor. M: months; MAF: mutant allele frequency

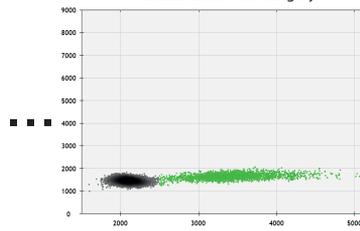
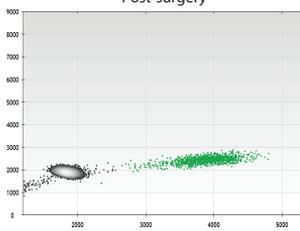
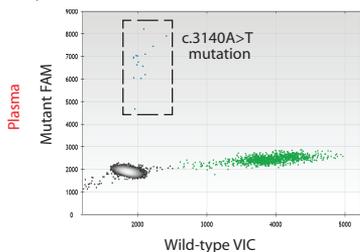


A**B****C**
A310001 *PIK3CA* c.3140A>T disease-free

 Mutant copies
per ml plasma

 12 copies/ml
Baseline

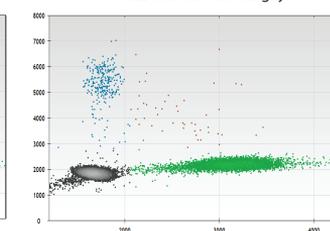
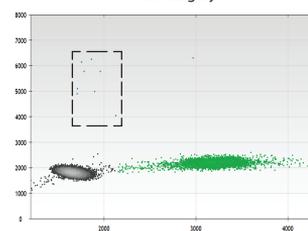
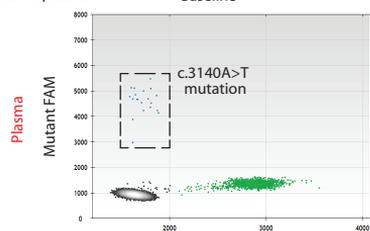
 0 copies/ml
Post-surgery

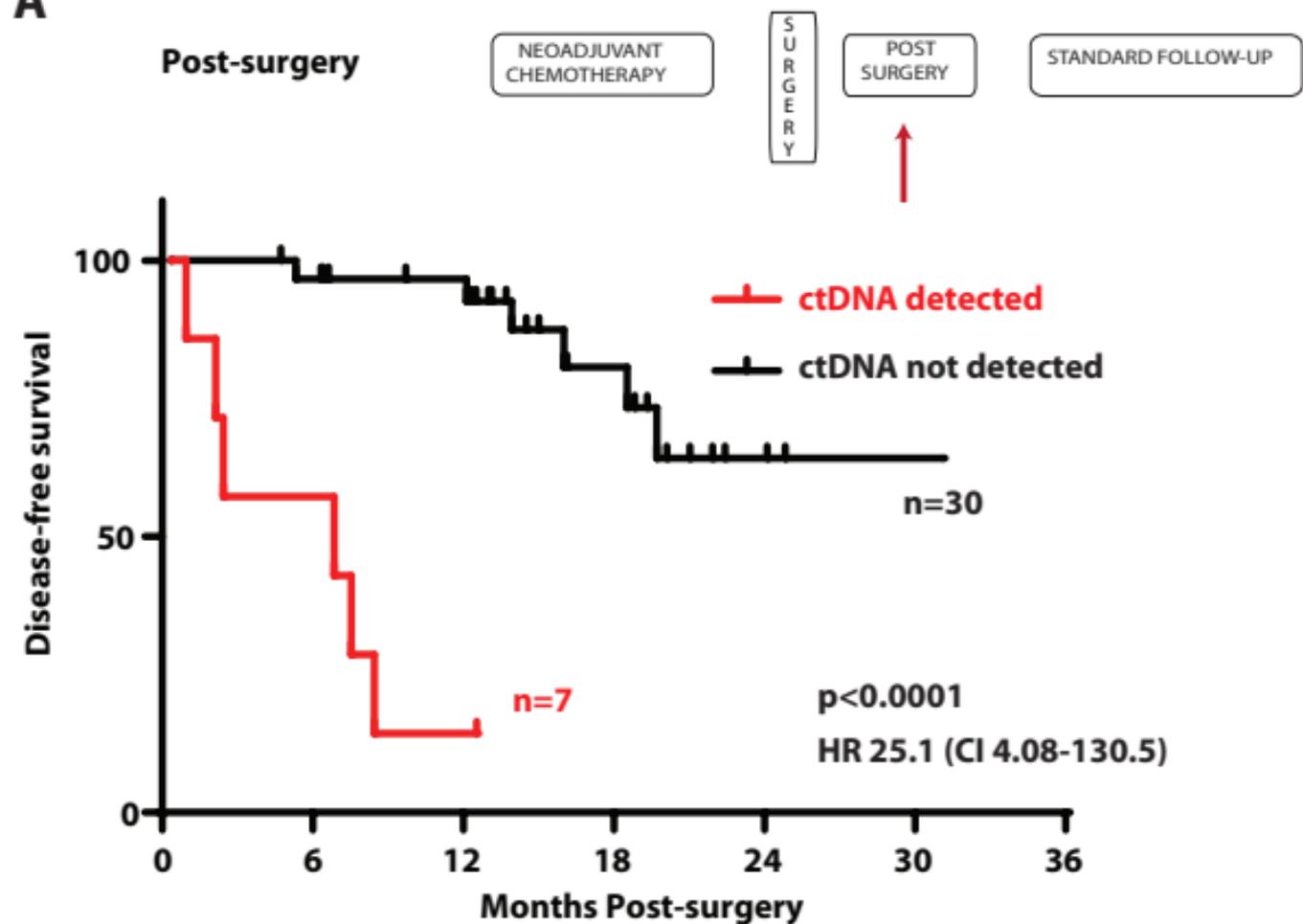
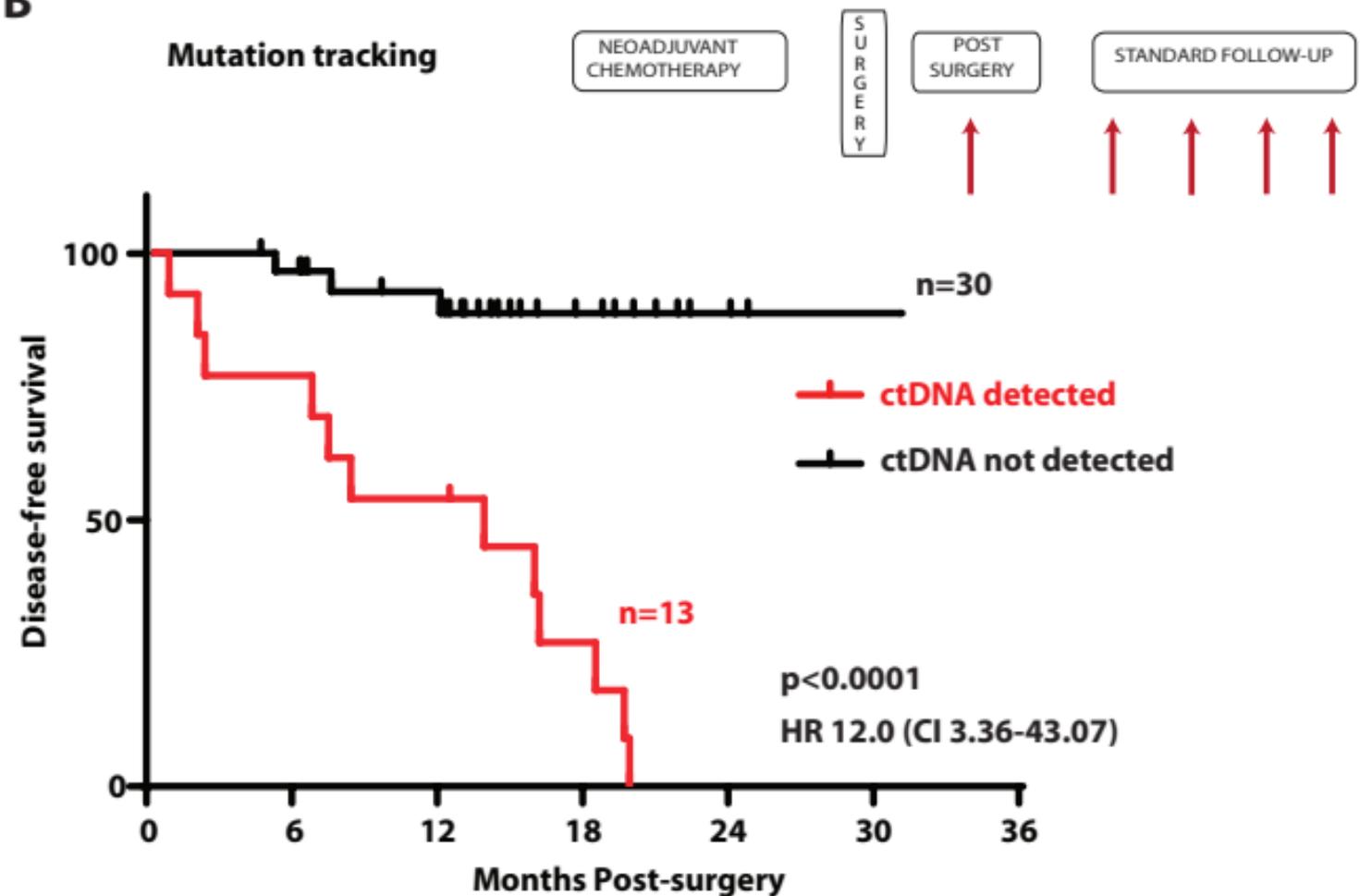
 0 copies/ml
23.2 months Post-surgery

A310006 *PIK3CA* c.3140A>T relapse 8.1 months Post-surgery

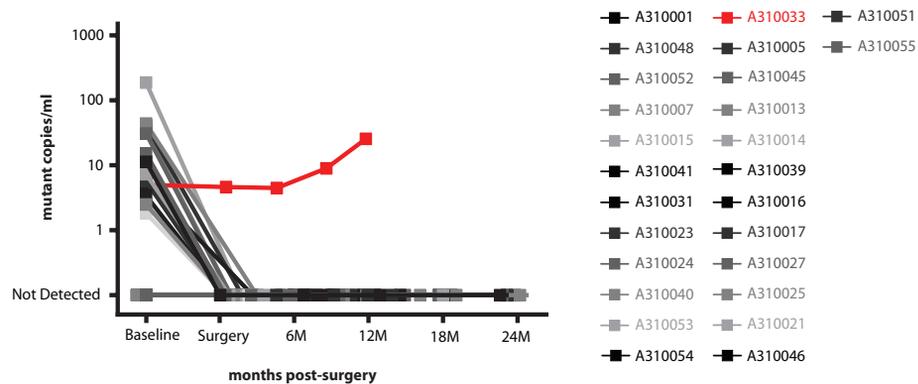
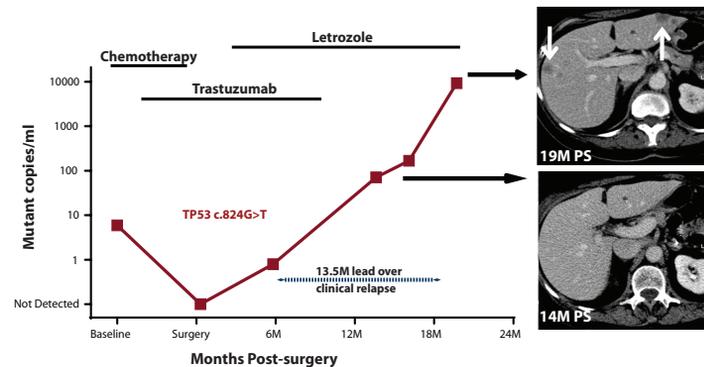
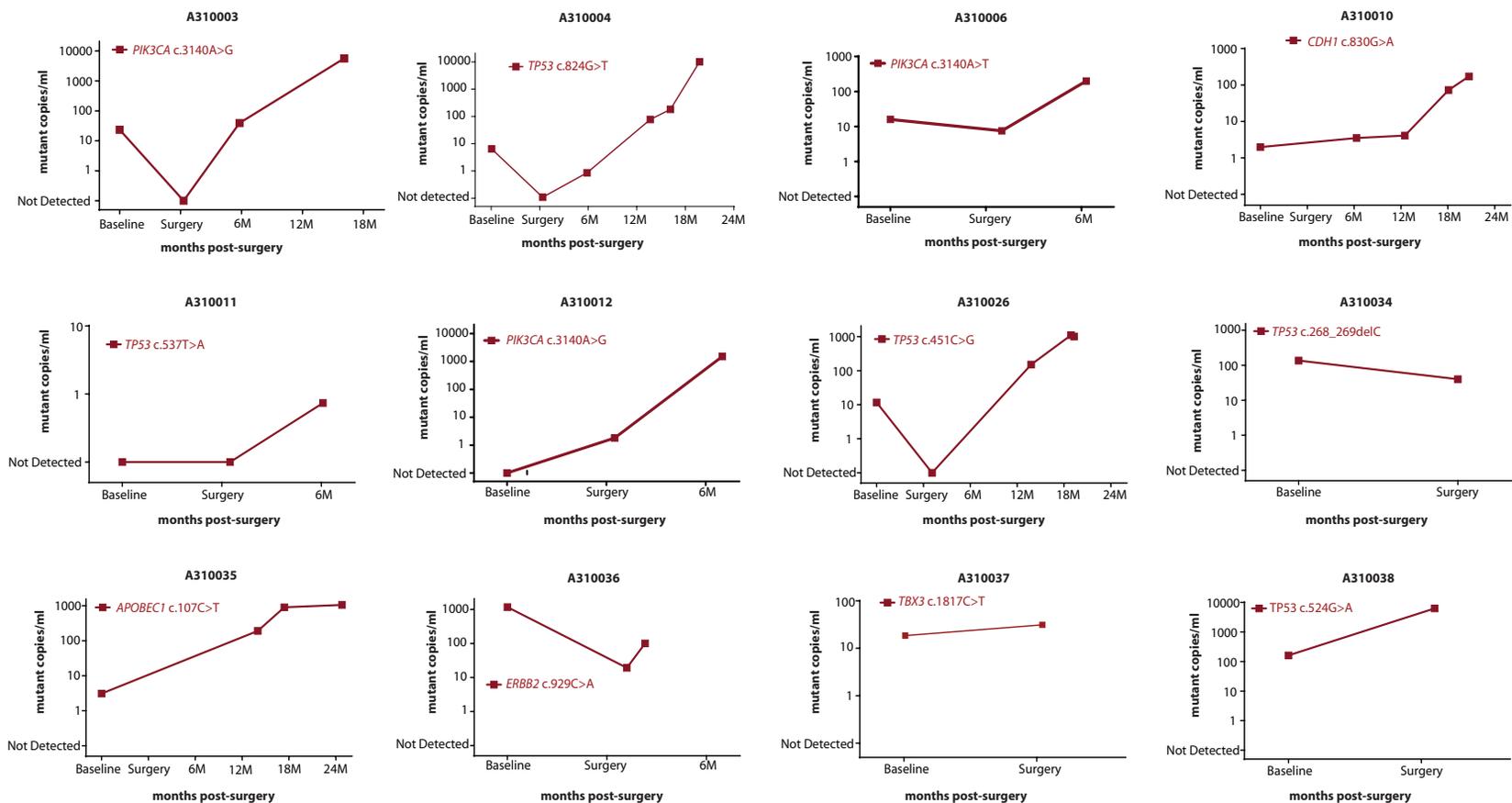
 Mutant copies
per ml plasma

 16 copies/ml
Baseline

 7 copies/ml
Post-surgery

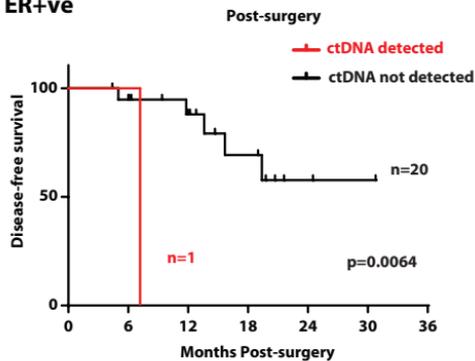
 198 copies/ml
6.2 months Post-surgery


A**B**

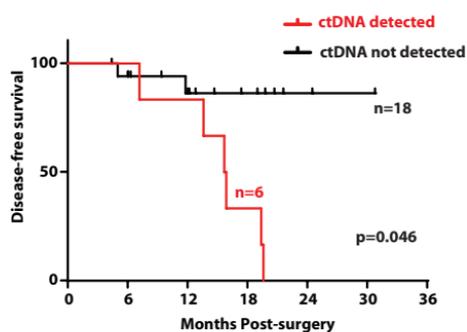
A**B****C**

A

ER+ve

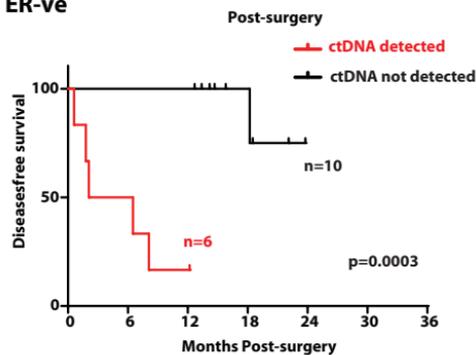


Mutation Tracking

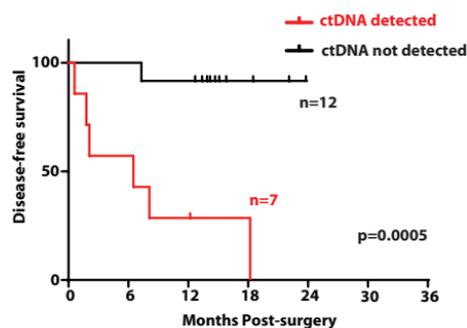


B

ER-ve

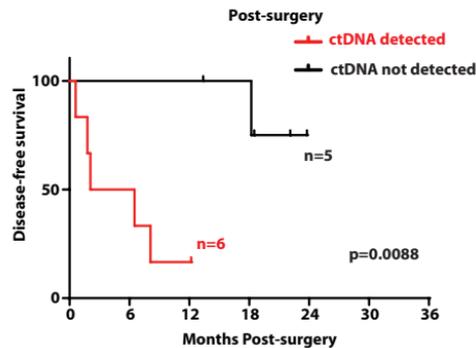


Mutation Tracking

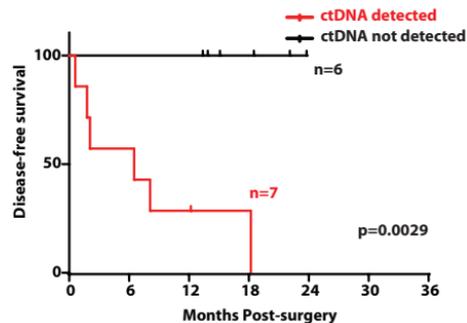


C

TNBC

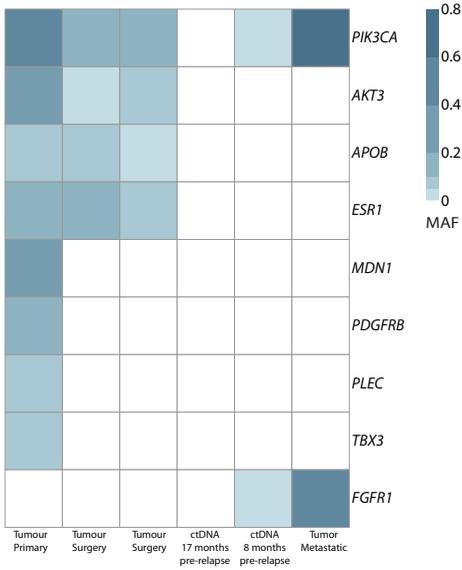


Mutation Tracking

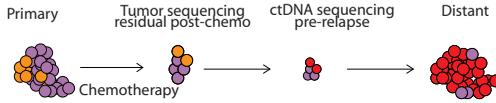
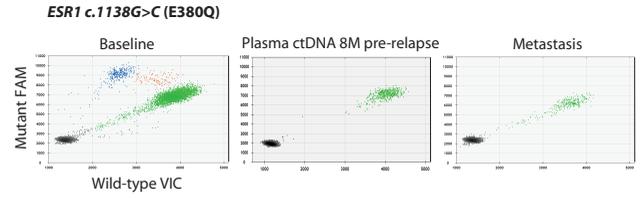
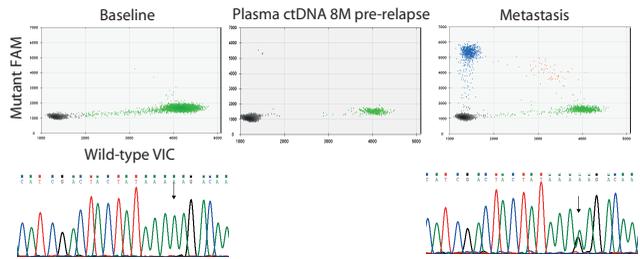


A

A310003



FGFR1 c.1966A>G (K656E)



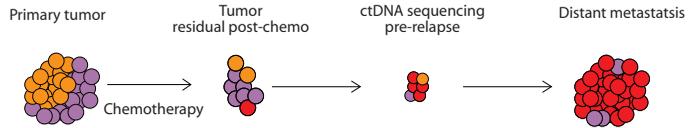
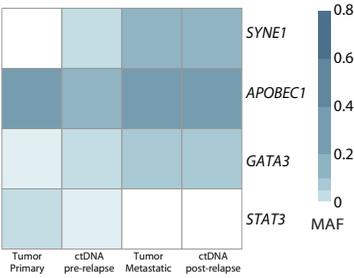
Clonal - PIK3CA c.3140A>G (H1047R)

Subclone - PIK3CA c.3140A>G (H1047R) + ESR1

Subclone - PIK3CA c.3140A>G (H1047R) + FGFR1

B

A310035



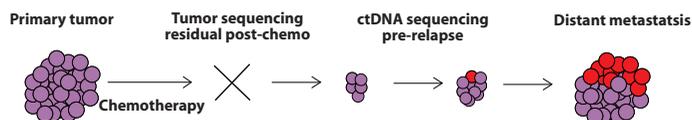
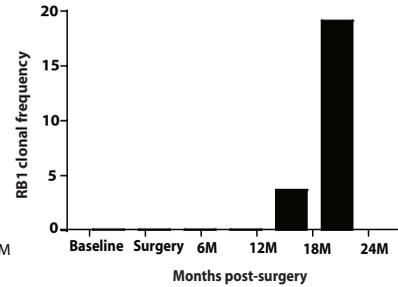
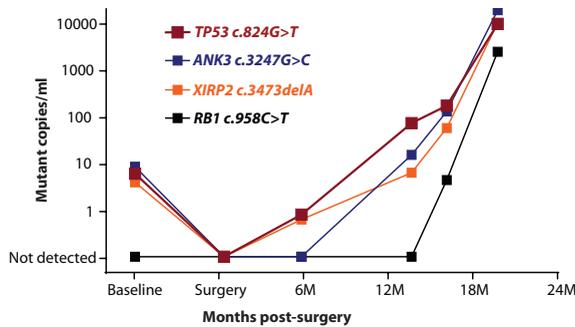
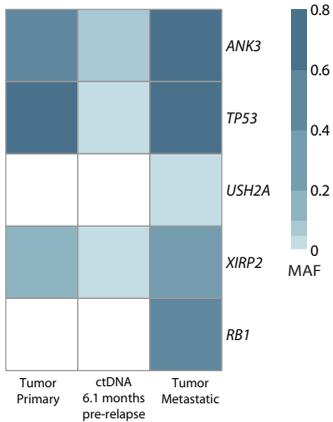
Clonal - APOBEC1 c.107C>T (A36V)

Subclone - APOBEC1 c.107C>T (A36V) + STAT-3 c.530A>G (K177R)

Subclone - APOBEC1 c.107C>T (A36V) + SYNE-1 c.3731C>A (S1244Y) + GATA-3 c.1203_1223 del (S402fs)

C

A310004



TP53 c.824G>T (C275F)

ANK3 c.3247G>C (E1083Q)

XIRP2 c.3473delA (Q1158fs)

Subclone + RB1 c.958C>T (R320*)

SUPPLEMENTARY FIGURES

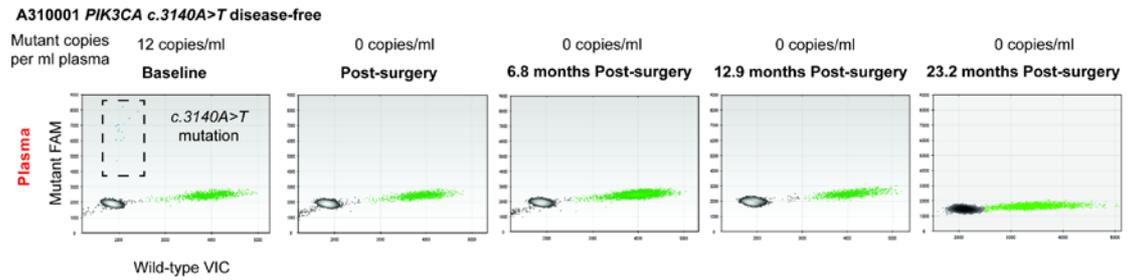


Figure S1. Mutation tracking by dPCR along a 24-month follow-up of a disease-free patient. Complete series of digital PCR plots in chronological order for case A310001 (in Fig. 2C), tracking a *PIK3CA* c.3140A>T (p.H1047L) mutation. In each plot, green dots represent VIC-labeled wild-type DNA and blue dots represent FAM-labeled mutant DNA. Brown dots represent droplets containing both wild-type and mutant DNA, whereas black dots are droplets with no DNA incorporated.

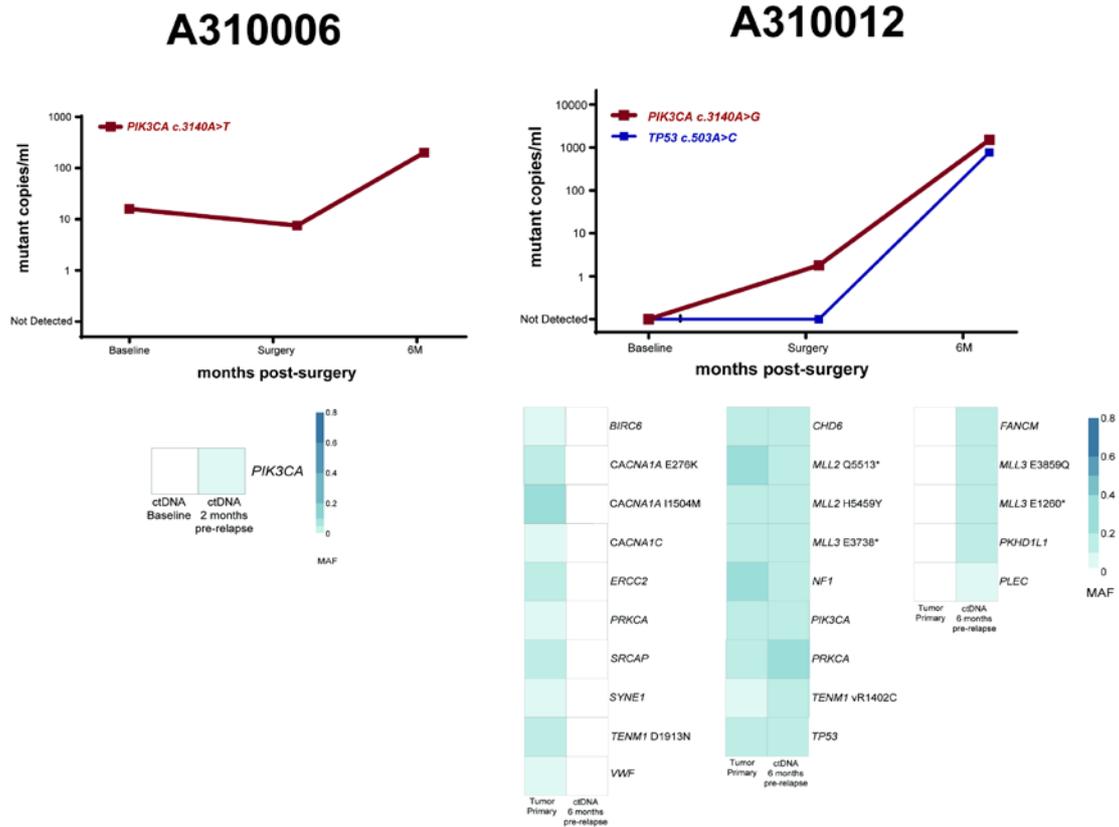
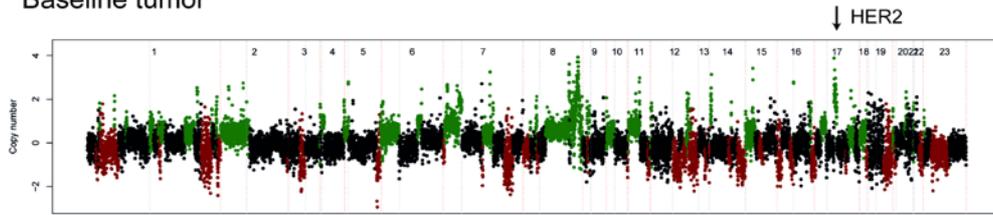


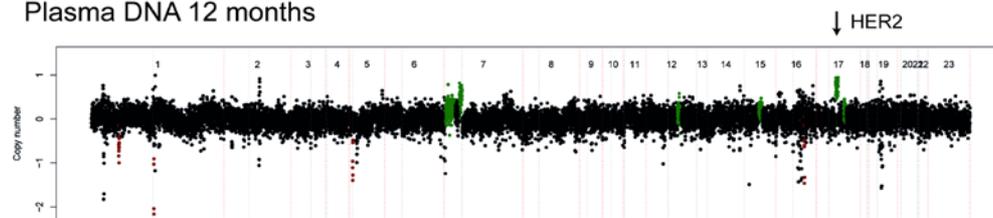
Figure S2. High-depth targeted capture massive parallel sequencing on plasma DNA from two relapsed patients. High-depth targeted capture MPS of primary tumor DNA and plasma DNA in two patients who relapsed with detectable ctDNA levels. Displayed are genes with somatic mutations called by at least two variant callers in one sample (see Methods). In A310006, insufficient DNA was available from the primary cancer, although plasma DNA sequencing revealed no additional mutations other than the *PIK3CA* mutation known to be present in the primary by AmpliSeq PGM sequencing. In A310012, plasma DNA sequencing revealed substantial divergence of the genetics of the ctDNA arising from RMD compared to the original primary tumor. The other three patients are in Fig. 7.

A A310004

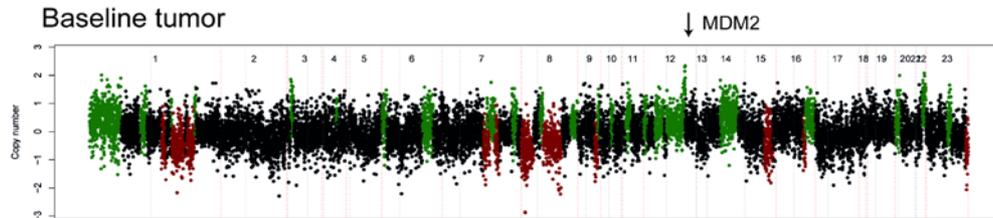
Baseline tumor



Plasma DNA 12 months

**B** A310012

Baseline tumor



Plasma DNA 6 months

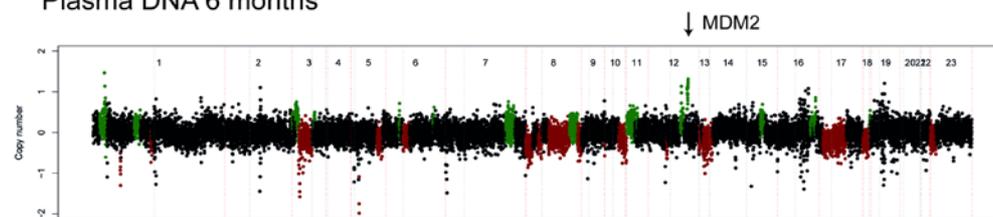


Figure S3. Copy number profile in primary tumor and plasma DNA prior to relapse. Exploratory copy number analysis of high depth capture MPS of ctDNA arising from micro-metastatic disease. Copy number profiles were derived from primary tumor and corresponding plasma DNA samples. Regions of copy number gain are in green, loss in red, neutral in black. The chromosomes are arranged in order from left to right, with red lines demarcating the chromosomes and grey lines demarcating the chromosome arms. **(A)** In known amplified A310004, HER2 amplification was detected in plasma DNA sequencing. Tumor purity in plasma DNA at 12 months was estimated as 9.3%, from frequency of *ANK3* and *TP53* mutations present, assuming heterozygosity and diploid, with the ability to detect HER2 amplification reflecting the likely high level HER2 copy number. **(B)** In A310012 MDM2 amplification was detected in both primary and plasma DNA. Tumor purity in plasma DNA was estimated to be 26%. No copy number events were detectable in plasma of remaining cases reflecting low tumor DNA purity in plasma DNA.

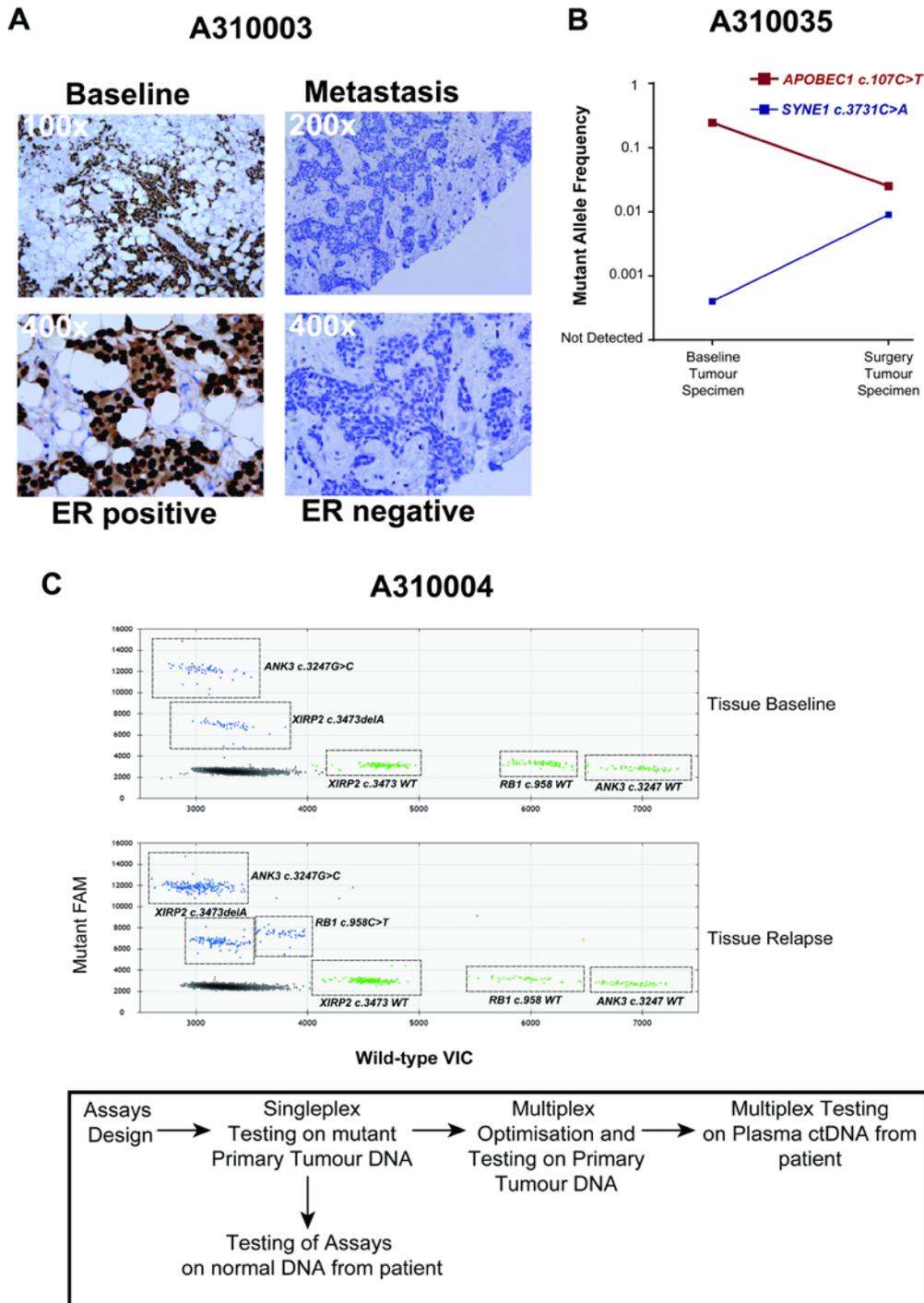


Figure S4. Validation and follow-up of MPS on plasma DNA. MPS data for these three patients are shown in Fig. 7. (A) ER immunohistochemistry staining in patient A310003 at baseline and relapsed disease. (B) Investigation of clonal selection in patient A310035 using dPCR. The *SYNE1* mutation was below the level of discrimination in the baseline tumor, detected only at a level consistent with formalin fixation artifact, but was enriched in the post-chemotherapy resection specimen, and in subsequent follow-up plasma samples. (C) Multiplex dPCR assay used to detect three different mutations in plasma from patient A310004: *ANK3 c.3247G>C*, *XIRP2 c.3473delA*, and *RB1 c.958C>T*. Representative plots are shown for analysis of DNA from primary tissue and metastatic biopsy.

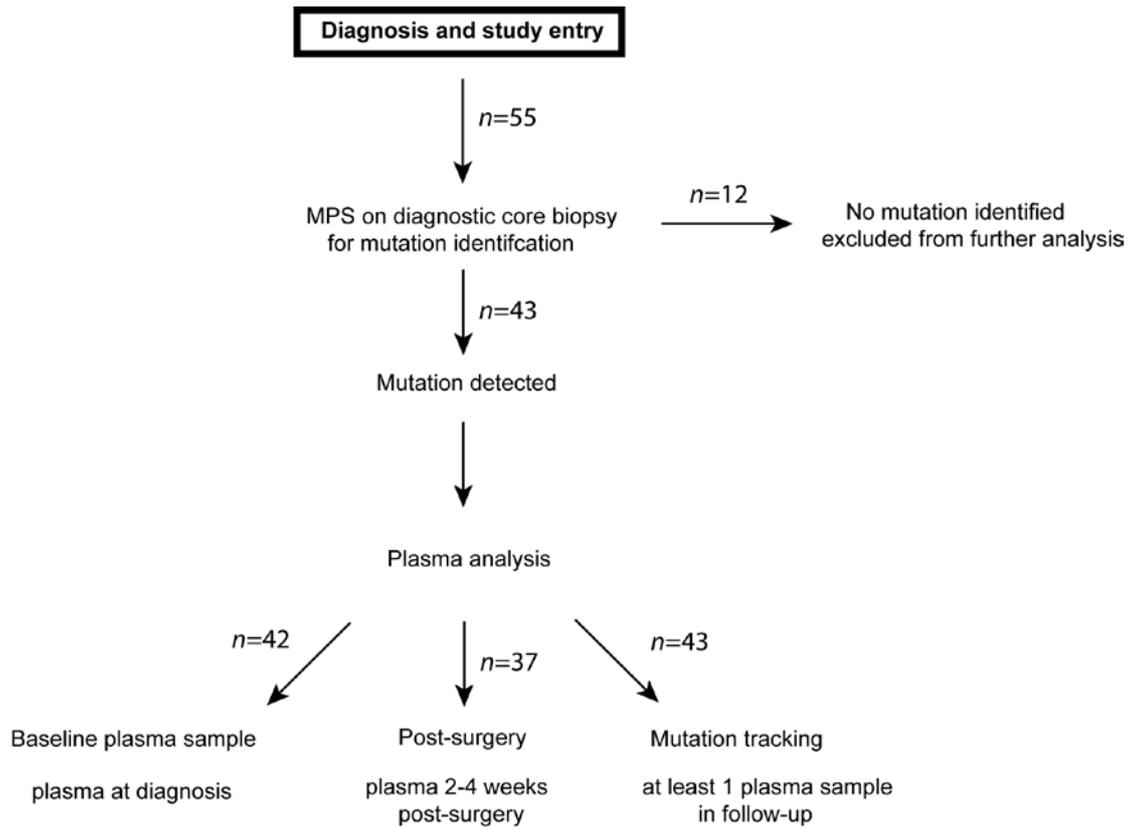


Figure S5. CONSORT flow diagram of patients included in study.

Table S1. Digital PCR assays and mutations analyzed in the study.

Sample	Assay	Seq Primer F	Seq Primer R	WT probe Sequence	5' modification	3' modification	Mutant Probe Sequence	5' modification	3' modification	ddPCR amplicon/extension temp (C)	Amplicon Length (bp)	NCBI Ref Seq
A310052	<i>PIK3CA</i> c.1035T>A	CCCTTTGGGTTATAA ATAGTGCACCTCA	AGCATCAGCATT TGACTTTACCTT ATCA	AACCTACGT GAATGTAA TA	VIC	NFQ MGB	CCTACGTGA AAGTAAATA	6- FAM	NFQ MGB	63	103	NG_012113.2
A310018	<i>PIK3CA</i> c.1258T>C	AAAAGTGTTTGAAA TGTGTTTTATAATTT AGACTAGTGA	CAGATACTAGAG TGCTGTGTAAT CAACA	CAATGGACA GTGTTCC	VIC	NFQ MGB	AATGGACGG TGTTC	6- FAM	NFQ MGB	60	126	NG_012113.2
A310045, A310048	<i>PIK3CA</i> c.1624G>A	GCTCAAAGCAATTT TACACG	TCCATTTTAGCA CTTACCTGTGAC	TCTCTGAAA TCACTGAGC AGGAGAA	VIC	NFQ MGB	TCTCTAAAA TCACTGAGC AGGAGAA	6- FAM	NFQ MGB	63	90	NG_012113.2
A310040, A310047, A310054	<i>PIK3CA</i> c.1633G>A	GCTCAAAGCAATTT TACACG	TCCATTTTAGCA CTTACCTGTGAC	TCTCTGAAA TCACTGAGC AGGAGAA	VIC	NFQ MGB	TCTCTGAAA TCACTAAGC AGGAGAA	6- FAM	NFQ MGB	63	90	NG_012113.2
A310005	<i>PIK3CA</i> c.1636C>A	TCAAAGCAATTTCTA CACCAGATCCT	CTCCATTTTAGC ACTTACCTGTGAC T	AAATCTTTC TCCTGCTCA GTG	VIC	NFQ MGB	AAAATCTTT CTCCTTCTC AGTG	6- FAM	NFQ MGB	60	91	NG_012113.2
A310001, A310006	<i>PIK3CA</i> c.3140A>T	TGAGCAAGAGGCTTT GGAGT	TCAGTCAATGC ATGCTGTTT	AATGATGCA CATCATGGT GGCT	VIC	NFQ MGB	AATGATGCA CTTCATGGT GGCT	6- FAM	NFQ MGB	61.7	115	NG_012113.2
A310003, A310012, A310017, A310021	<i>PIK3CA</i> c.3140A>G	TGAGCAAGAGGCTTT GGAGT	TCAGTCAATGC ATGCTGTTT	AATGATGCA CATCATGGT GGCT	VIC	NFQ MGB	AATGATGCA CGTCATGGT GGCT	6- FAM	NFQ MGB	61.7	115	NG_012113.2
A310034	<i>TP53</i> c.267_268delC	GCACCAGCAGCTCCT ACAC	TGGGAAGGGAC AGAAGATGACA	CGAGCCCC TCCTGG	VIC	NFQ MGB	ACCAGCCCC TCCTGG	6- FAM	NFQ MGB	60	72	NG_017013.2
A310037	<i>TP53</i> c.321C>A	TCCCTTCCCAGAAAA CCTACCA	GCTGTCCAGAA TGCAAGAAG	CGGAAACCG TAGCTGC	VIC	NFQ MGB	ACGGAAACC TTAGCTGC	6- FAM	NFQ MGB	60	66	NG_017013.2
A310007	<i>TP53</i> c.394A>G	GTCTCCTTCTCTTC CTACAGTACT	GCAGGTCTTGGC CAGTTG	CCCTCAACA AGATGTT	VIC	NFQ MGB	CCTCAACGA GATGTT	6- FAM	NFQ MGB	60	69	NG_017013.2
A310016	<i>TP53</i> c.396G>C	GTCTCCTTCTCTTC CTACAGTACT	GCAGGTCTTGGC CAGTTG	CCCTCAACA AGATGTTTT	VIC	NFQ MGB	CCTCAACAA CATGTTTT	6- FAM	NFQ MGB	60	69	NG_017013.2

A3 10 01 7	TP53 c.398T> A	GTCTCCTTCTCTTC CTACAGTACT	GCAGGTCTTGGC CAGTTG	CCTCAACAA GATGTTTTG	VIC	NFQ MGB	CTCAACAAG AAGTTTTG	6- FAM	NFQ MGB	60	69	NG_0 1701 3.2
A3 10 02 6	TP53 c.451C> G	GCCCTGTGCAGTGT G	CATGGCGGGAC GC	TTGATTCCA CACCCCGC	VIC	NFQ MGB	ATTCCACAG CCCCGC	6- FAM	NFQ MGB	60	59	NG_0 1701 3.2
A3 10 03 9, A3 10 04 5	TP53 c.452C> A	TGTGCAGCTGTGGT TGAT	CATGGCGGGAC GC	TCCACACCC CCGCC	VIC	NFQ MGB	CCACACACC CGCCC	6- FAM	NFQ MGB	56	54	NG_0 1701 3.2
A3 10 01 3	TP53 c.477_4 78delC	TGGGTTGATCCACA CCCC	ATGTGCTGTGAC TGCTTGTA	CGGCATGG CCA	VIC	NFQ MGB	CGGCATGG CCA	6- FAM	NFQ MGB	60	69	NG_0 1701 3.2
A3 10 01 2	TP53 c.503A> C	GGCCATGGCCATCT ACA	GGCAGCCCTCA CAAC	AGTCACAGC ACATGAC	VIC	NFQ MGB	CAGTCACAG CCCATGAC	6- FAM	NFQ MGB	60	57	NG_0 1701 3.2
A3 10 03 0, A3 10 03 8, A3 10 04 3, A3 10 05 3	TP53 c.524G> A	ACAGCACATGACGGA GGTT	CTGCTCACCATC GCTATCTGA	TGAGGCGGT GCCC	VIC	NFQ MGB	TGAGGCACT GCCC	6- FAM	NFQ MGB	52	70	NG_0 1701 3.2
A3 10 02 4	TP53 c.528C> A	AGCACATGACGGAGG TTGT	CTGCTCACCATC GCTATCTGA	AGGGCTGC CCCCA	VIC	NFQ MGB	AGGGCTGA CCCCA	6- FAM	NFQ MGB	60	68	NG_0 1701 3.2
A3 10 01 1	TP53 c.537T> A	AGCACATGACGGAGG TTGT	CTGCTCACCATC GCTATCTGA	CCCACATG AGCGCT	VIC	NFQ MGB	CCACCAAGA GCGCT	6- FAM	NFQ MGB	60	68	NG_0 1701 3.2
A3 10 01 5	TP53 c.586C> T	TGGCCCTCCTCAGC AT	CATCCAAATACT CCACACGCAAAT T	CCTTCCACT CGGATAAG	VIC	NFQ MGB	CCTTCCACT CAGATAAG	6- FAM	NFQ MGB	60	60	NG_0 1701 3.2
A3 10 02 1	TP53 c.712T> G	CTGACTGTACCACCA TCCACTAC	ATGGGCCTCGG TTCATG	CAGGAAC TG TTACACATG TA	VIC	NFQ MGB	CAGGAAC TG TTACCCATG TA	6- FAM	NFQ MGB	60	74	NG_0 1701 3.2
A3 10 04 6	TP53 c.713G> T	TGGCTCTGACTGTAC CACCAT	GATGGGCCTCCG GTTCAT	ACAAC TACA TGTGGTAAC AG	VIC	NFQ MGB	ACAAC TACA TGTGTGTAAC AG	6- FAM	NFQ MGB	60	72	NG_0 1701 3.2
A3 10 03 3	TP53 c.742C> T	AACTACATGTGTAAC AGTTCCTGCAT	GAGTCTCCAGT GTGATGATGGT	TGGGCCTCC GGTTCA	VIC	NFQ MGB	TGGGCCTCC AGTTCA	6- FAM	NFQ MGB	60	77	NG_0 1701 3.2
A3 10 04 3	TP53 c.743G> A	AACTACATGTGTAAC AGTTCCTGCAT	CCAGTGTGATGA TGGTGAGGAT	CATGAACCG GAGGCC	VIC	NFQ MGB	CATGAACCA GAGGCC	6- FAM	NFQ MGB	60	70	NG_0 1701 3.2
A3 10 05 5	TP53 c.796G> C	CCTTTGCTTCTCTT TTCCATCCT	CCAGGACAGGCA CAAACAC	AAGCTGTTTC CGTCCCAGT AG	VIC	NFQ MGB	CTGTTCCGT CGCAGTAG	6- FAM	NFQ MGB	60	73	NG_0 1701 3.2
A3 10 02 7	TP53 c.801_8 02delG	CCTTTGCTTCTCTT TTCCATCCT	CCAGGACAGGCA CAAACAC	CAAAGCTGT TCCGTCCC	VIC	NFQ MGB	CAAAGCTGT TCGTCCC	6- FAM	NFQ MGB	60	84	NG_0 1701 3.2
A3 10 00 4	TP53 c.824G> T	TGGGACGGAACAGCT TTGAG	CTGTGCCCGGT CTCT	CGGTGTTTG TGCCTG	VIC	NFQ MGB	TGCGTGTTT TTGCCTG	6- FAM	NFQ MGB	60	60	NG_0 1701 3.2

A3 10 00 01	<i>TP53</i> c.832C> G	AGCTTTGAGGTGCGT GTTTG	TGCGGAGATTCT CTTCTCTGT	TCTCCAGG ACAGGCA	VIC	NFQ MGB	CTCCAGCA CAGGCA	6- FAM	NFQ MGB	60	67	NG_0 1701 3.2
A3 10 03 01	<i>TP53</i> c.841G> C	GCTTTGAGGTGCGTG TTTGTG	TGCGGAGATTCT CTTCTCTGT	CCTGGGAGA GACCG	VIC	NFQ MGB	CCTGGGAGA CACCG	6- FAM	NFQ MGB	60	67	NG_0 1701 3.2
A3 10 02 05	<i>TP53</i> c.853G> A	TGCCTGTCTGGGAG AGA	GGTGAGGCTCCC CTTCTTG	CGCACAGAG GAAGAG	VIC	NFQ MGB	CGCACAAG GAAGAG	6- FAM	NFQ MGB	58	67	NG_0 1701 3.2
A3 10 00 04	<i>ANK3</i> c.3247G >C	GGTCCATGAGAGGAA AAGAGAGAGA	ACTGATGCTCCT TCCAAGTTTAC	CTTCGAAGT GAAAATG	VIC	NFQ MGB	CTTCGAAGT CAAAATG	6- FAM	NFQ MGB	60	75	NG_0 2991 7.1
A3 10 03 05	<i>APOBEC 1</i> c.107C> T	TCTATGACCCAGAG AACTTCGT	TCTATGCCCACT TGATTTCGTA	AGCAGACAG GCCTCTT	VIC	NFQ MGB	AGCAGACAG ACCTCTT	6- FAM	NFQ MGB	60	62	NC_0 0007 2.6
A3 10 01 04	<i>CDH1</i> c.1946_ 1947ins T	CTCCCTGTGCTCAT CAITTTCTTTT	CCACCTAAGG CCATCTTTGG	CAGCCCAAG AATCTATCA T	VIC	NFQ MGB	AGCCCAAGA ATCTTATCA T	6- FAM	NFQ MGB	60	86	NG_0 0802 1.1
A3 10 01 00	<i>CDH1</i> c.830G> A	CCCCCTGTTGGTGT TTTATTATGA	TGGCAATGCGTT CTCTATCCA	ACCTTCAGC CATCCTG	VIC	NFQ MGB	CACCTTCAG TCATCCTG	6- FAM	NFQ MGB	60	82	NG_0 0802 1.1
A3 10 03 06	<i>ERBB2</i> c.929C> A	GGCTACATGTTCTG ATCTCCTTAG	TGCTGTCACCTC TTGGTTGTG	ACGTGGGAT CCTGCACC	VIC	NFQ MGB	CGTGGGATA CTGCACC	6- FAM	NFQ MGB	60	95	NG_0 0750 3.1
A3 10 00 03	<i>ESR1</i> c.1138G >C	GCTTTGTGGATTGA CCCT	AGTAGCTCCCT GGGTGC	TCTAGAATG TGCTTGGCT AGAGATC	VIC	NFQ MGB	TCTACAATG TGCTTGGCT AGAGATC	6- FAM	NFQ MGB	61. 7	112	NG_0 0849 3.1
A3 10 00 03	<i>FGFR1</i> c.1966A >G	GGGACATTCAACACA TCGACTAC	ACAGGGCGCCT TGTC	ACGTTGGTT GTCTTTTTA TA	VIC	NFQ MGB	CGTTGGTTG TCTCTTTAT A	6- FAM	NFQ MGB	60	66	NG_0 0772 9.1
A3 10 05 01	<i>GATA3</i> c.1002_ 1003del GG	GACCACCACAACCAC ACTCT	AGCCACAGGCA TTGCA	AGGGTCCCC ATTGGCAT	VIC	NFQ MGB	CAGGGTCCA TTGGCAT	6- FAM	NFQ MGB	60	63	NG_0 1585 9.1
A3 10 02 03	<i>GATA3</i> c.1667_ 1670del CGAAA	GACTATGAAGAAGGA AGGCATCCA	CCTCCAGTGAGT CATGCACCTTTT	CTAGACATT TTTCGGTTT CT	VIC	NFQ MGB	TTGCTAGAC ATTGTTTCT	6- FAM	NFQ MGB	58	84	NG_0 1585 9.1
A3 10 02 03	<i>GATA3</i> c.1672A >T	GGAAGGCATCCAGAC CAGAAAC	CCTCCAGTGAGT CATGCACCTTTT	TGCTAGACA TTTTTCG	VIC	NFQ MGB	TGCTAGACA TATTTTCG	6- FAM	NFQ MGB	60	77	NG_0 1585 9.1
A3 10 04 01	<i>GATA3</i> c.991_9 93insG	CAGACACCACAACC ACACT	CCCACAGGCATT GCAGACA	CATTGGCAT TCCTCCTCC A	VIC	NFQ MGB	TGGCATTCC TCCTCCA	6- FAM	NFQ MGB	60	66	NG_0 1585 9.1
A3 10 02 00	<i>KMT2C (MLL3)</i> c.2670_ 2671del T	TGTTGACTTTTCCCA ATCTGTTACATAGG	CCTCGCCCCGAC AGT	TTTCTGGA AATCCAG	VIC	NFQ MGB	TTTCTGGA ATCCAG	6- FAM	NFQ MGB	60	88	NG_0 3394 8.1
A3 10 00 04	<i>RB1</i> c.958C> T	GTCAGTGACTTTTTT CTTTCAAGTTGA	CATGATCCAAA ATAATCTTGAT CTAGATCT	AAATTTCTT CGTATCGTT TAGAA	VIC	NFQ MGB	AAATTTCTT CGTATCATT TAGAA	6- FAM	NFQ MGB	60	103	NG_0 0900 9.1
A3 10 03 05	<i>SYNE-1</i> c.3731C >A	GGGACTGTGTCAGT ACAAAGAAAT	AGCTTGTTCTG GACTTCTTTAGA G	TCTTCGAGA GAATTTT	VIC	NFQ MGB	TTCTTCGAG ATAATTTT	6- FAM	NFQ MGB	60	82	NG_0 1285 5.1
A3 10 03 06	<i>SYNE-1</i> c.4065C >G	GGGAGCGATTTGAAA CAAAACAAAGA	CTGCTAAACTC AAGAAGCGTTCA T	CTGTTTGAA AAAGGTATC TT	VIC	NFQ MGB	CTGTTTGAA AAAGCTATC TT	6- FAM	NFQ MGB	60	87	NG_0 1285 5.1

A3 10 03 7	<i>TBX3</i> c.1817C >T	CCCCTTTCGGAAGCC TGTTTC	GGCTGCCGAGA GGAG	ACCCCTACA CGTACATG	VIC	NFQ MGB	TTACCCCTA CATGTACAT G	6- FAM	NFQ MGB	60	73	NG_0 0831 5.1
A3 10 00 4	<i>XIRP2</i> c.3473C A>C	TGAAACAGCAGTCAA ATTGCAAAC	AAAAGAAAACA TGCTGTACGAAC ATCC	CCACCTTGG ATCTC	VIC	NFQ MGB	CCACCTGGA TCTC	6- FAM	NFQ MGB	60	76	NC_0 0000 2.12

Table S2. Clinicopathological factors associated with baseline ctDNA level. *P* values for median ctDNA levels were determined using Kruskal-Wallis rank sum test. *P* values for the percentage of ctDNA were determined using Chi-square test. Clinical tumor size: T1/2 < 50 mm versus T3/4 ≥ 50 mm or locally advanced at presentation. Clinical node positive: cytologically confirmed lymph node involvement at presentation. ER, estrogen receptor expression by immunohistochemistry. HER2 status: HER2+ by immunohistochemistry or in situ hybridization. TNBC, triple negative breast cancer (negative for ER, HER, and progesterone receptor). Subsequent pCR: subsequent achievement of pathological complete response.

	Median ctDNA level copies per ml	<i>P</i> value	ctDNA detection (%)	<i>P</i> value
Clinical tumor size				
T1/2	3.878	0.42	64	0.42
T3/4	4.964		77	
Clinical node positive				
Yes	5.514	0.11	86	0.015
No	0.615		50	
Histological grade				
3	5.947	0.03	79	0.015
2	0		42	
ER status				
Positive	1.597	0.011	54	0.016
Negative	14.61		89	
HER2 status				
Positive	5.355	0.88	71	1
Negative	3.72		68	
Subtype				
TNBC	15.95	0.027	92	0.033
HER2+	5.355		71	
ER+HER2-	0		47	
Subsequent pCR				
Yes	4.679	0.49	75	0.65
No	4.345		67	

Table S3. Prediction of disease free survival using a single post-surgery blood sample. A multivariable Cox model was applied, adjusting for clinicopathological factors including molecular subtypes, clinical tumor size and pathological lymph nodal status. Clinical tumor size was the size of tumor at presentation (T1/2 vs T3/4). Macrometastases on axillary surgery/sentinel node were evaluated after neoadjuvant chemotherapy vs pathological ($n = 36$). One patient was excluded due to missing data. The Cox model had a C-index of 0.795.

Covariates	Multivariable Cox regression model		P value
	HR	95% CI	
ctDNA detection post-surgery			
Yes vs. No	21.1	2.5-177.47	0.005
Subtypes			
TNBC vs. HER2 ⁺ vs. ER ⁺ /HER2 ⁻	1.3	0.56-3.0	0.55
Clinical tumor size			
T3/4 vs. T1/2	0.79	0.13-4.9	0.8
Pathological node-positive post-chemo			
Positive vs. Negative	1.9	0.56-6.5	0.31

Table S4. Prediction of disease-free survival by mutation-tracking using serial blood samples. A multivariable Cox model was applied, adjusting for clinicopathological factors including molecular subtypes, clinical tumor size, and lymph nodal status. Clinical tumor size was the size of tumor at presentation (T1/2 vs T3/4 Macrometastases on axillary surgery/sentinel node were evaluated after neoadjuvant chemotherapy vs pathological ($n = 42$). One patient was excluded due to missing data. The Cox model had a C-index of 0.802.

Covariates	Multivariable Cox Regression model		
	HR	95% CI	p value
Mutation Tracking			
Yes vs. No	9.6	2.4-38.9	0.001
Subtypes			
Triple negative subtype vs. HER2+ vs ER+/HER2-	1.6	0.82-3.0	0.17
Clinical tumor size			
T3/4 vs T1/2	1.3	0.42-4.1	0.64
Pathological node positive post chemo			
Positive vs. Negative	1.1	0.35-3.5	0.86

Table S5. Summary of the study cohort. Summary of the study cohort, illustrating the distribution of the main clinico-pathological factors in overall population, disease-free and early relapse groups. Abbreviations: cT, clinical tumor size; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; mo, months; N/A, not available; NAC, neoadjuvant chemotherapy; NST, non-special type; PgR, progesterone receptor; pathCR, pathological complete response; y, years. pathCR, pathological complete response indicating absence of residual invasive cancer in breast and axillary lymph nodes.

	All patients (%)	Disease free (%)	Relapse (%)
N	55	40 (73)	15 (27)
Median age (range), y	50 (25–76)	50 (25–76)	50 (32–65)
Pathology			
IDC	43 (78)	33 (82.5)	10 (66.5)
ILC	4 (7)	4 (10)	0 (0)
Mixed IDC/ILC	6 (11)	2 (5)	4 (27)
Metaplastic	1 (2)	1 (2.5)	0 (0)
Neuroendocrine	1 (2)	0 (0)	1 (6.5)
Histologic grade			
1	0 (0)	0 (0)	0 (0)
2	17 (31)	13 (32.5)	4 (27)
3	37 (67)	26 (65)	11 (73)
N/A	1 (2)	1 (2.5)	0 (0)
Receptor status			
ER/PgR+ HER2-	20 (36.5)	16 (40)	4 (27)
ER/PgR- HER2+	8 (14.5)	7 (17.5)	1 (6)
ER/PgR+ HER2+	13 (23.5)	9 (22.5)	4 (27)
Triple negative	14 (25.5)	8 (20)	6 (40)
Clinical size at presentation (cT)			
1	6 (11)	6 (15)	0 (0)
2	33 (60)	25 (62.5)	8 (53.5)
3	5 (9)	3 (7.5)	2 (13)
4	10 (18)	5 (12.5)	5 (33.5)
N/A	1 (2)	1 (2.5)	0 (0)
Nodal status at presentation			
Negative	27 (49)	22 (55)	5 (33.5)
Positive	28 (51)	18 (45)	10 (66.5)
Pathological response to NAC			
No pathCR	41 (74.5)	29 (72.5)	12 (80)
pathCR	13 (23.5)	10 (25)	3 (20)
N/A	1 (2)	1 (2.5)	0 (0)
Median follow-up (mo)			
From study entry	21.4	20.8	24.8
From surgery	13.9	14.5	8.1

Table S6. Ion AmpliSeq Breast Cancer driver gene panel.

<i>AKT1 ex4</i>	<i>MAP3K1</i>
<i>BRAF ex15</i>	<i>PIK3CA ex5, 8, 9, 20</i>
<i>CDH1</i>	<i>PIK3R1 ex4, 6</i>
<i>GATA3</i>	<i>PTEN</i>
<i>KIT ex11</i>	<i>RUNX1 ex1, 2, 3</i>
<i>KRAS ex2</i>	<i>SF3B1 ex14, 15</i>
<i>MAP2K4</i>	<i>TP53 ex5, 6, 7, 8</i>

Table S7: Reads from capture MPS of tumor and plasma DNA. Chr, chromosome; AA, aminoacid; Ref, reference nucleotide; Alt, alternative nucleotide; Pos, human genome reference position.

TUMOR_SAMPLE	Mutant reads	All reads	Effect	AA	Gene	Chr	POS	REF	ALT
A310003 Tumor Primary	23	107	NON_SYN ONYMOUS _CODING	D4 55 N	<i>AKT3</i>	1	24366 8628	C	T
A310003 Tumor Primary	61	101	NON_SYN ONYMOUS _CODING	H1 04 7R	<i>PIK3CA</i>	3	17895 2085	A	G
A310003 Tumor Primary	47	233	NON_SYN ONYMOUS _CODING	V5 1I	<i>PDGFR B</i>	5	14951 5331	C	T
A310003 Tumor Primary	35	104	NON_SYN ONYMOUS _CODING	A3 24 4V	<i>MDN1</i>	6	90405 364	G	A
A310003 Tumor Primary	8	130	NON_SYN ONYMOUS _CODING	P4 57 L	<i>TBX3</i>	12	11511 2370	G	A
A310003 Tumor Primary	6	58	NON_SYN ONYMOUS _CODING	I3 32 9T	<i>APOB</i>	2	21229 754	A	G
A310003 Tumor Primary	8	54	NON_SYN ONYMOUS _CODING	E3 80 Q	<i>ESR1</i>	6	15233 2832	G	C
A310003 Tumor Primary	13	234	NON_SYN ONYMOUS _CODING	G1 35 7A	<i>PLEC</i>	8	14500 1675	C	G
A310003 Tumor Surgery 1	24	130	NON_SYN ONYMOUS _CODING	H1 04 7R	<i>PIK3CA</i>	3	17895 2085	A	G
A310003 Tumor Surgery 1	8	105	NON_SYN ONYMOUS _CODING	I3 32 9T	<i>APOB</i>	2	21229 754	A	G
A310003 Tumor Surgery 1	14	113	NON_SYN ONYMOUS _CODING	E3 80 Q	<i>ESR1</i>	6	15233 2832	G	C
A310003 Tumor Surgery 1	5	128	NON_SYN ONYMOUS _CODING	D4 55 N	<i>AKT3</i>	1	24366 8628	C	T
A310003 Tumor Surgery 2	28	147	NON_SYN ONYMOUS _CODING	H1 04 7R	<i>PIK3CA</i>	3	17895 2085	A	G
A310003 Tumor Surgery 2	9	110	NON_SYN ONYMOUS _CODING	D4 55 N	<i>AKT3</i>	1	24366 8628	C	T
A310003 Tumor	12	123	NON_SYN ONYMOUS	E3 80	<i>ESR1</i>	6	15233 2832	G	C

Surgery 2			_CODING	Q					
A310003 Tumor Surgery 2	4	109	NON_SYN ONYMOUS _CODING	I3 32 9T	<i>APOB</i>	2	21229 754	A	G
A310003 cfDNA 8 months pre- relapse	6	239	NON_SYN ONYMOUS _CODING	K6 87 E	<i>FGFR1</i>	8	38272 308	T	C
A310003 cfDNA 8 months pre- relapse	9	231	NON_SYN ONYMOUS _CODING	H1 04 7R	<i>PIK3CA</i>	3	17895 2085	A	G
A310003 Tumor Metastatic	52	113	NON_SYN ONYMOUS _CODING	G1 80 9D	<i>FLG</i>	1	15228 1936	C	T
A310003 Tumor Metastatic	378	499	NON_SYN ONYMOUS _CODING	H1 04 7R	<i>PIK3CA</i>	3	17895 2085	A	G
A310003 Tumor Metastatic	315	704	NON_SYN ONYMOUS _CODING	K6 87 E	<i>FGFR1</i>	8	38272 308	T	C
A310035 Tumor Primary	2	400	FRAME_S HIFT	S4 02 fs	<i>GATA3</i>	10	81158 55	TCC TCCC TGA GCC ACA TCT CG	T
A310035 Tumor Primary	46	200	NON_SYN ONYMOUS _CODING	A3 6V	<i>APOBE C1</i>	12	78053 69	G	A
A310035 Tumor Primary	7	237	NON_SYN ONYMOUS _CODING	K1 77 R	<i>STAT3</i>	17	40490 769	T	C
A310035 cfDNA pre- relapse	39	389	NON_SYN ONYMOUS _CODING	A3 6V	<i>APOBE C1</i>	12	78053 69	G	A
A310035 cfDNA pre- relapse	13	335	NON_SYN ONYMOUS _CODING	S1 24 4Y	<i>SYNE1</i>	6	15276 5652	G	T
A310035 cfDNA pre- relapse	3	417	FRAME_S HIFT	S4 02 fs	<i>GATA3</i>	10	81158 55	TCC TCCC TGA GCC ACA TCT CG	T
A310035 cfDNA pre- relapse	2	569	NON_SYN ONYMOUS _CODING	K1 77 R	<i>STAT3</i>	17	40490 769	T	C
A310035	94	277	NON_SYN	A3	<i>APOBE</i>	12	78053	G	A

Tumor Metastatic			ONYMOUS_CODING	6V	<i>C1</i>		69		
A310035 Tumor Metastatic	10	59	NON_SYN ONYMOUS_CODING	S1 24 4Y	<i>SYNE1</i>	6	15276 5652	G	T
A310035 Tumor Metastatic	36	520	FRAME_S HIFT	S4 02 fs	<i>GATA3</i>	10	81158 55	TCC TCCC TGA GCC ACA TCT CG	T
A310035 cfDNA post-relapse	86	393	NON_SYN ONYMOUS_CODING	A3 6V	<i>APOBE C1</i>	12	78053 69	G	A
A310035 cfDNA post-relapse	63	494	NON_SYN ONYMOUS_CODING	S1 24 4Y	<i>SYNE1</i>	6	15276 5652	G	T
A310035 cfDNA post-relapse	26	461	FRAME_S HIFT	S4 02 fs	<i>GATA3</i>	10	81158 55	TCC TCCC TGA GCC ACA TCT CG	T
A310004 Tumor Primary	20	39	NON_SYN ONYMOUS_CODING	E1 08 3Q	<i>ANK3</i>	10	61865 744	C	G
A310004 Tumor Primary	8	14	NON_SYN ONYMOUS_CODING	C2 75 F	<i>TP53</i>	17	75771 14	C	A
A310004 Tumor Primary	3	23	FRAME_S HIFT	Q1 15 8fs	<i>XIRP2</i>	2	16810 1374	CA	C
A310004 cfDNA 6.1 months pre-relapse	19	315	NON_SYN ONYMOUS_CODING	E1 08 3Q	<i>ANK3</i>	10	61865 744	C	G
A310004 cfDNA 6.1 months pre-relapse	9	272	NON_SYN ONYMOUS_CODING	C2 75 F	<i>TP53</i>	17	75771 14	C	A
A310004 cfDNA 6.1 months pre-relapse	4	258	FRAME_S HIFT	Q1 15 8fs	<i>XIRP2</i>	2	16810 1374	CA	C
A310004 Tumor	94	127	NON_SYN ONYMOUS	E1 08	<i>ANK3</i>	10	61865 744	C	G

Metastatic			_CODING	3Q					
A310004 Tumor Metastatic	49	100	STOP_GAI NED	R3 20 *	<i>RB1</i>	13	48941 648	C	T
A310004 Tumor Metastatic	302	382	NON_SYN ONYMOUS _CODING	C2 75 F	<i>TP53</i>	17	75771 14	C	A
A310004 Tumor Metastatic	11	384	STOP_GAI NED	S3 3*	<i>USH2A</i>	1	21659 5581	G	C
A310004 Tumor Metastatic	53	149	FRAME_S HIFT	Q1 15 8fs	<i>XIRP2</i>	2	16810 1374	CA	C
A310006 cfDNA 2 months pre- relapse	11	321	NON_SYN ONYMOUS _CODING	H1 04 7L	<i>PIK3CA</i>	3	17895 2085	A	T
A310012 Tumor Primary	25	137	NON_SYN ONYMOUS _CODING	S1 00 2C	<i>FLG</i>	1	15228 4357	G	C
A310012 Tumor Primary	22	112	STOP_GAI NED	Q5 51 3*	<i>KMT2D</i>	12	49415 640	G	A
A310012 Tumor Primary	18	71	STOP_GAI NED	Q5 43 *	<i>NF1</i>	17	29546 122	C	T
A310012 Tumor Primary	18	101	STOP_GAI NED	E3 62 *	<i>PRKCA</i>	17	64731 634	G	T
A310012 Tumor Primary	29	138	NON_SYN ONYMOUS _CODING	I1 50 8 M	<i>CACNA 1A</i>	19	13368 242	G	C
A310012 Tumor Primary	26	138	NON_SYN ONYMOUS _CODING	E2 76 K	<i>CACNA 1A</i>	19	13470 572	C	T
A310012 Tumor Primary	14	86	NON_SYN ONYMOUS _CODING	D1 91 3N	<i>TENM1</i>	X	12351 9866	C	T
A310012 Tumor Primary	7	128	NON_SYN ONYMOUS _CODING	S1 97 3C	<i>CACNA 1C</i>	12	27949 97	C	G
A310012 Tumor Primary	6	164	NON_SYN ONYMOUS _CODING	S7 58 C	<i>VWF</i>	12	61558 97	G	C
A310012 Tumor Primary	18	130	NON_SYN ONYMOUS _CODING	H5 45 9Y	<i>KMT2D</i>	12	49416 100	G	A
A310012 Tumor Primary	13	115	NON_SYN ONYMOUS _CODING	L5 78 V	<i>SRCAP</i>	16	30723 395	C	G
A310012	21	108	NON_SYN	H1	<i>TP53</i>	17	75784	T	G

Tumor Primary			ONYMOUS_CODING	68P			27		
A310012 Tumor Primary	11	123	NON_SYN ONYMOUS_CODING	F20L	<i>PRKCA</i>	17	64299029	C	G
A310012 Tumor Primary	14	142	NON_SYN ONYMOUS_CODING	S111C	<i>ERCC2</i>	19	45871916	G	C
A310012 Tumor Primary	7	93	NON_SYN ONYMOUS_CODING	S37L	<i>BIRC6</i>	2	32582339	C	T
A310012 Tumor Primary	8	62	NON_SYN ONYMOUS_CODING	Q631H	<i>CHD6</i>	20	40116413	C	G
A310012 Tumor Primary	7	53	NON_SYN ONYMOUS_CODING	H1047R	<i>PIK3CA</i>	3	178952085	A	G
A310012 Tumor Primary	7	96	NON_SYN ONYMOUS_CODING	D3988H	<i>SYNE1</i>	6	152668310	C	G
A310012 Tumor Primary	15	91	STOP_GAINED	E3738*	<i>KMT2C</i>	7	151859450	C	A
A310012 Tumor Primary	8	128	NON_SYN ONYMOUS_CODING	R1402C	<i>TENM1</i>	X	123556389	G	A
A310012 cfDNA 6 months pre-relapse	36	281	STOP_GAINED	Q5513*	<i>KMT2D</i>	12	49415640	G	A
A310012 cfDNA 6 months pre-relapse	55	328	STOP_GAINED	Q543*	<i>NF1</i>	17	29546122	C	T
A310012 cfDNA 6 months pre-relapse	83	343	STOP_GAINED	E362*	<i>PRKCA</i>	17	64731634	G	T
A310012 cfDNA 6 months pre-relapse	50	302	NON_SYN ONYMOUS_CODING	E3859Q	<i>KMT2C</i>	7	151856043	C	G
A310012 cfDNA 6 months pre-relapse	39	274	NON_SYN ONYMOUS_CODING	R1402C	<i>TENM1</i>	X	123556389	G	A
A310012 cfDNA 6	31	254	NON_SYN ONYMOUS	H545	<i>KMT2D</i>	12	49416100	G	A

months pre-relapse			_CODING	9Y					
A310012 cfDNA 6 months pre-relapse	42	306	NON_SYNONYMOUS_CODING	M1561I	<i>FANCM</i>	14	45656994	G	C
A310012 cfDNA 6 months pre-relapse	27	254	NON_SYNONYMOUS_CODING	H168P	<i>TP53</i>	17	7578427	T	G
A310012 cfDNA 6 months pre-relapse	38	325	NON_SYNONYMOUS_CODING	Q631H	<i>CHD6</i>	20	40116413	C	G
A310012 cfDNA 6 months pre-relapse	43	329	NON_SYNONYMOUS_CODING	H1047R	<i>PIK3CA</i>	3	178952085	A	G
A310012 cfDNA 6 months pre-relapse	33	294	STOP_GAINED	E3738*	<i>KMT2C</i>	7	151859450	C	A
A310012 cfDNA 6 months pre-relapse	35	321	STOP_GAINED	E1260*	<i>KMT2C</i>	7	151904448	C	A
A310012 cfDNA 6 months pre-relapse	37	303	NON_SYNONYMOUS_CODING	G45E	<i>PKHD1L1</i>	8	110376836	G	A
A310012 cfDNA 6 months pre-relapse	11	160	NON_SYNONYMOUS_CODING	S3143C	<i>PLEC</i>	8	144994972	G	C