CANCER GENOMICS

Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA

Tim Forshew,[1]* Muhammed Murtaza,[1,2]* Christine Parkinson,[1,2,3]* Davina Gale,[1]* Dana W. Y. Tsui,[1]* Fiona Kaper,[4†] Sarah-Jane Dawson,[1,2,3] Anna M. Piskorz,[1,2] Mercedes Jimenez-Linan,[3,5] David Bentley,[6] James Hadfield,[1] Andrew P. May,[4] Carlos Caldas,[1,2,3,7] James D. Brenton,[1,2,3,7‡] Nitzan Rosenfeld[1,2‡]

Plasma of cancer patients contains cell-free tumor DNA that carries information on tumor mutations and tumor burden. Individual mutations have been probed using allele-specific assays, but sequencing of entire genes to detect cancer mutations in circulating DNA has not been demonstrated. We developed a method for tagged-amplicon deep sequencing (TAm-Seq) and screened 5995 genomic bases for low-frequency mutations. Using this method, we identified cancer mutations present in circulating DNA at allele frequencies as low as 2%, with sensitivity and specificity of >97%. We identified mutations throughout the tumor suppressor gene *TP53* in circulating DNA from 46 plasma samples of advanced ovarian cancer patients. We demonstrated use of TAm-Seq to noninvasively identify the origin of metastatic relapse in a patient with multiple primary tumors. In another case, we identified in plasma an *EGFR* mutation not found in an initial ovarian biopsy. We further used TAm-Seq to monitor tumor dynamics, and tracked 10 concomitant mutations in plasma of a metastatic breast cancer patient over 16 months. This low-cost, high-throughput method could facilitate analysis of circulating DNA as a noninvasive "liquid biopsy" for personalized cancer genomics.

INTRODUCTION

Circulating cell-free DNA extracted from plasma or other body fluids has potentially transformative applications in cancer management (*1*–*7*). Characterization of tumor mutation profiles is required for informed choice of therapy, given that biological agents target specific pathways and effectiveness may be modulated by specific mutations (*8*–*11*). Yet, mutation profiles in different metastatic clones can differ significantly from each other or from the parent primary tumor (*12*, *13*). Evolutionary changes within the cancer can alter the mutational spectrum of the disease and its responsiveness to therapies, which may necessitate repeat biopsies (*14*–*17*). Biopsies are invasive and costly and only provide a snapshot of mutations present at a given time and location. For some applications, mutation detection in plasma DNA as a "liquid biopsy" could potentially replace invasive biopsies as a means to assess tumor genetic characteristics (*2*–*7*). Sensitive methods for detecting cancer mutations in plasma may find use in early detection screening (*1*), prognosis, monitoring tumor dynamics over time, or detection of minimal residual disease (*3*, *18*, *19*). In high-grade serous ovarian carcinomas (HGSOC), mutations in the tumor suppressor gene *TP53* have been observed in 97% of cases (*20*, *21*), but these are located throughout the gene and are difficult to assay. A cost-effective method that could detect and measure allele frequency (AF) of *TP53* mutations in plasma may be highly applicable as a biomarker for HGSOC (*22*).

Circulating DNA is fragmented to an average length of 140 to 170 base pairs (bp) and is present in only a few thousand amplifiable copies per milliliter of blood, of which only a fraction may be diagnostically relevant (*2*, *3*, *23*–*25*). Recent advances in noninvasive prenatal diagnostics highlight the clinical potential of circulating DNA (*25*–*28*), but also the challenges involved in analysis of circulating tumor DNA (ctDNA), where mutated loci and AFs may be more variable. Various methods have been optimized to detect extremely rare alleles (*1*, *2*, *6*, *7*, *29*–*31*), and can assay for predefined or hotspot mutations. These methods, however, interrogate individual or few loci and have limited ability to identify mutations in genes that lack mutation hotspots, such as the *TP53* and *PTEN* tumor suppressor genes (*32*). In patients with more advanced cancers, ctDNA can comprise as much as 1% to 10% or more of circulating DNA (*2*), presenting an opportunity for more extensive genomic analysis. Targeted resequencing has been recently used to identify mutations in selected genes at AFs as low as 5% (*33*–*35*). However, identifying mutations across sizeable genomic regions spanning entire genes at an AF as low as 2%, or in few nanograms of fragmented template from circulating DNA, has been more challenging.

In response, we describe a tool for noninvasive mutation analysis on the basis of tagged-amplicon deep sequencing (TAm-Seq), which allows amplification and deep sequencing of genomic regions spanning thousands of bases from as little as individual copies of fragmented DNA. We applied this technique for detection of both abundant and

[1]Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [2]Department of Oncology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK. [3]Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [4]Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080, USA. [5]Department of Histopathology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [6]Illumina Cambridge, Chesterford Research Park, Little Chesterford, Cambridge CB10 1XL, UK. [7]Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.
*These authors contributed equally to this work.
†Present address: Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA.
‡To whom correspondence should be addressed. E-mail: nitzan.rosenfeld@cancer.org.uk (N.R.); james.brenton@cancer.org.uk (J.D.B.)

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

TEMPUS EX1010

rare mutations in circulating DNA from blood plasma of ovarian and breast cancer patients. This sequencing approach allowed us to monitor changes in tumor burden by sampling only patient plasma over time. Combined with faster, more accurate sequencing technologies or rare allele amplification strategies, this approach could potentially be used for personalized medicine at point of care.

RESULTS

Targeted deep sequencing of fragmented DNA by TAm-Seq

To amplify and sequence fragmented DNA, we designed primers to generate amplicons that tile regions of interest in short segments of about 150 to 200 bases (Fig. 1A and table S1), incorporating universal

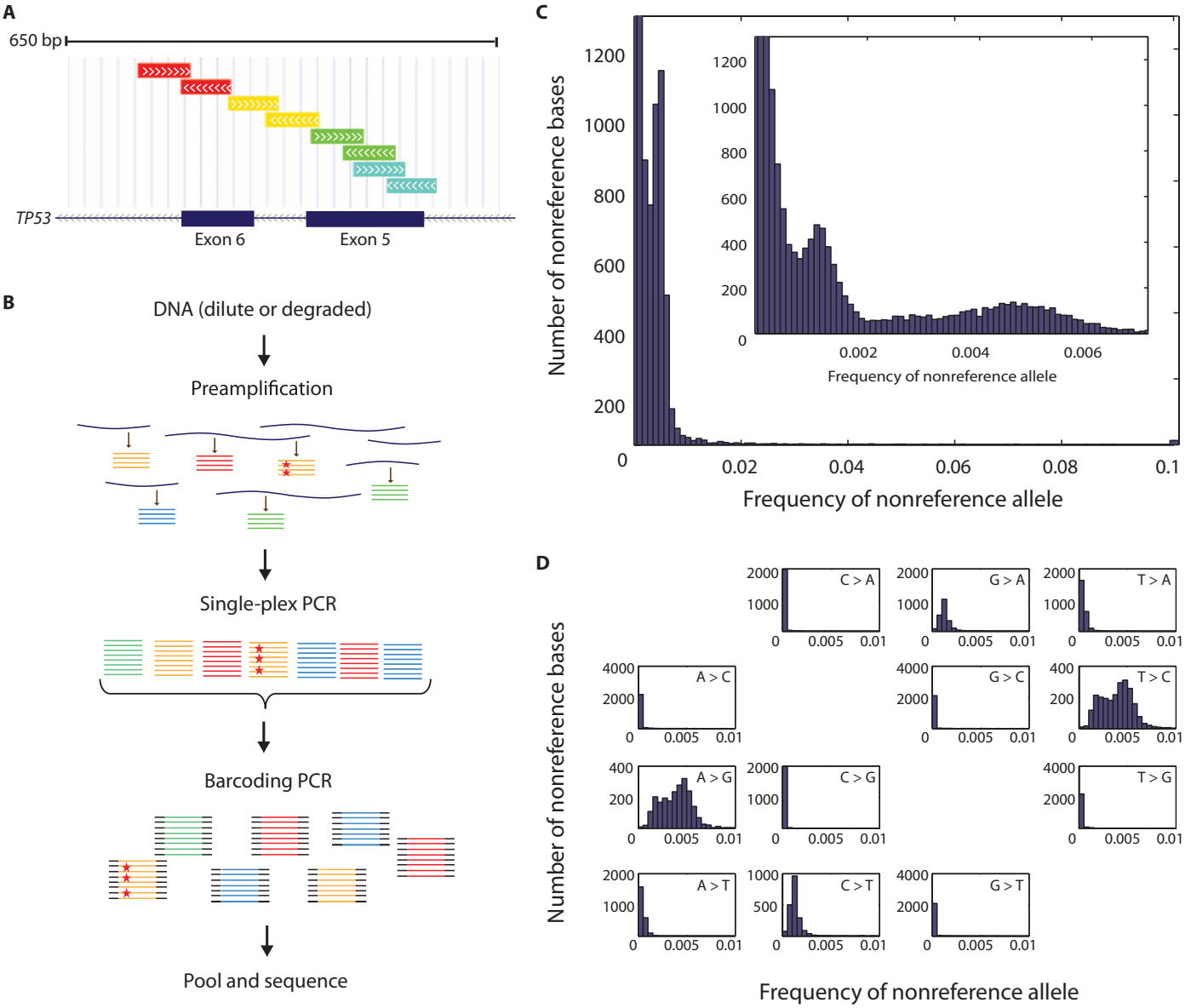

Fig. 1. Overview of tagged amplicon sequencing (TAm-Seq). (**A**) Illustration of amplicon design. Primers were designed to amplify regions of interest in overlapping short amplicons (table S1). Amplicon design is illustrated for a region covering exons 5 to 6 of *TP53*. Colored bars, segmented into forward and reverse reads, show regions covered by different amplicons (excluding primer regions). Sequencing adaptors are attached at either end, such that a single-end read generates separate sets of forward and reverse reads (fig. S1). Because amplicons are mostly shorter than 200 bp, the forward and reverse reads also partially overlap. Figure adapted from University of California, Santa Cruz, Genome Browser (http://genome.ucsc.edu/). (**B**) Workflow overview. Multiple regions were amplified in parallel. An initial preamplification step was performed for 15 cycles using a pool of the target-specific primer pairs to preserve representation of all alleles in the template material. The schematic diagram shows DNA molecules that carry mutations (red stars) being amplified alongside wild-type molecules. Regions of interest in the preamplified material were then selectively amplified in individual (single-plex) PCR, thus excluding nonspecific products. Finally, sequencing adaptors and sample-specific barcodes were attached to the harvested amplicons in a further PCR. (**C**) Distribution of observed nonreference read frequencies, averaged over 47 FFPE samples, across all loci and all nonreference bases. Inset expands the low-frequency range. (**D**) Distribution of the observed background nonreference read frequencies averaged over 47 FFPE samples for the 12 different A/C/G/T base substitutions.

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

adaptors at 5′ ends (fig. S1). Performing single-plex amplification with each of these primer pairs would require dispersing the initial sample into many separate reactions, considerably increasing the probability of sampling errors and allelic loss. Multiplex amplification using a large set of primers could result in nonspecific amplification products and biased coverage. We therefore applied a two-step amplification process: a limited-cycle preamplification step where all primer sets were used together to capture the starting molecules present in the template, followed by individual amplification to purify and select for intended targets (Fig. 1B) (Supplementary Methods). The final concentration of each primer in the preamplification reaction was 50 nM, reducing the potential for interprimer interactions, and 15 cycles of long-extension (4 min) polymerase chain reaction (PCR) were used to remain in the exponential phase of amplification. We used a microfluidic system (Access Array, Fluidigm) to perform parallel single-plex amplification from multiple preamplified samples using multiple primer sets. An additional PCR step attached sequencing adaptors (fig. S1) and tagged each sample by a unique molecular identifier or "barcode" (table S2). Sequencing adaptors were separately attached at either end and the products mixed together, such that single-end sequencing generated separate sets of forward and reverse reads. We performed 100-base single-end sequencing (GAIIx sequencer, Illumina), with an additional 10 cycles using the barcode sequencing primer, generating ~30 million reads per lane. This produced an average read depth of 3250 for each of 96 barcoded samples for 48 amplicons read in two possible orientations.

Validation and sensitivity for mutation identification in ovarian tumor samples

We designed a set of 48 primer pairs to amplify 5995 bases of genomic sequence covering coding regions (exons and exon junctions) of *TP53* and *PTEN*, and selected regions in *EGFR*, *BRAF*, *KRAS*, and *PIK3CA* (table S1) by overlapping short amplicons (Fig. 1A). The sequenced regions cover mutations that account for 38% of all point mutations in the COSMIC database (v55) (*32*). We used TAm-Seq to sequence DNA extracted from 47 formalin-fixed, paraffin-embedded (FFPE) tumor specimens of ovarian cancers (table S3), which were also sequenced for *TP53* by Sanger sequencing (*36*) (Supplementary Methods). DNA extracted from FFPE samples is generally degraded and fragmented as a result of fixation and long-term ambient storage. We amplified DNA from each sample in duplicate, tagging each replicate with a different barcode. Using a single lane of sequencing, we generated 3.5 gigabases of data passing signal purity filters, producing mean read depth of 3200 above Q30 for each of the 9024 expected read groups (48 amplicons × 2 directions × 94 barcoded samples). Background frequencies of nonreference reads were ~0.1% (median, 0.03%; mean, 0.2%; in keeping with Q30 quality threshold applied), yet varied substantially between loci and base substitutions (Fig. 1C) and showed a clear bias toward purine/pyrimidine conservation (Fig. 1D). Sixty-six percent of loci had mean background rate of <0.1%, and 96% of loci had background rate of <0.6%.

The data set interrogated nearly 18,000 possible single-base substitutions for each sample, which introduces a risk of false detection. To control for sporadic PCR errors and reduce false positives, we called point mutations in a sample only if nonreference AFs were above the respective substitution-specific background distribution at a high confidence margin (0.9995 or greater), and ranked high in the list of nonreference AFs, in both replicates (Supplementary Methods). Duplicate sequencing data were obtained for 44 samples, and 43 single-base substitutions were called (table S3). These matched 100% of mutations identified by Sanger sequencing and included three additional mutations at low AFs that were below detection thresholds of Sanger sequencing (fig. S2). The upper bound of AFs that may have been missed was estimated (Supplementary Methods) at <5% for 36 of 44 FFPE samples (82%) and <10% for 42 of 44 samples (95%), with median value of 1.3% and mean value of 2.7%. Mutant AFs were highly reproducible in duplicate samples. For 42 of 43 mutations called, the difference in measured frequency between duplicates was less than 0.08, and the relative difference was 25% or less (Fig. 2A). Mutant AFs correlated significantly with tumor cellularity in the FFPE block (correlation coefficient = 0.422; $P = 0.0049$, t test) (Fig. 2B).

In a separate run, we sequenced libraries prepared from six different diluted mixtures of six FFPE samples, with a different known point mutation in *TP53* in each, to mean read depth of 5600. Of more than 100,000 possible non-SNP (single-nucleotide polymorphism) substitutions, we identified all 33 expected point mutations present at AF >1%, including 6 mutations present at AF <2%, with one false-positive called with AF = 1.9%. Using less stringent parameters (Supplementary Methods), we identified three additional mutations present at AF = 0.6% (Fig. 2C), with no additional false positives. Thus, we obtained 100% sensitivity, identifying mutations at AFs as low as 0.6%. A positive predictive value (PPV) of 100% was calculated for mutations at AF >2%, and a PPV of 90% for mutations identified at AF <2% (Fig. 2D).

Quantitative limitations of mutation detection

When applying TAm-Seq to measure a predefined mutation (as opposed to screening thousands of possible substitutions), the frequency of the mutant allele can be read out directly from the data at the desired locus. False detection is less likely, and criteria for confident mutation detection for a predefined substitution can be less stringent than those described above for de novo mutation identification (Supplementary Methods). The minimal nonreference AFs that could be detected depend on the read depth and background rates of nonreference reads, which vary per locus and substitution type. Minimal detectable frequencies increase when higher confidence margins are used (Supplementary Methods) and had a median value of 0.14% at confidence margin of 0.95 and 0.18% at confidence margin of 0.99 (fig. S3). The minimal detectable frequency would also be limited if a minimal number of reads is applied for confident mutation detection; for example, a minimum of 10 reads implies that sequencing depth of 5000 would be required to detect mutations at AF as low as 0.2%. For alleles present at ~10 or fewer copies in the starting template, reproducibility would also be limited by sampling noise, because these alleles may be over- or underrepresented in any particular reaction.

To characterize the quantitative accuracy of TAm-Seq as applied to circulating DNA, we simulated rare circulating tumor mutations by mixing plasma DNA from two healthy individuals. Using the same set of primers as used for the FFPE experiment, we identified that these two individuals differed at five known SNP loci (table S4). Total amplifiable copies in both plasma DNA samples were determined by digital PCR and mixed to obtain minor AFs ranging from 0.16% to 40% (Supplementary Methods). We sequenced diluted templates containing between 250 and <1 expected copy of the minor allele (table S5). The coefficient of variation (CV) of the observed AFs was equal on average to the inverse square root ($1/\sqrt{n}$) of the expected number of copies of the rare allele (Fig. 3A), which is the theoretical

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

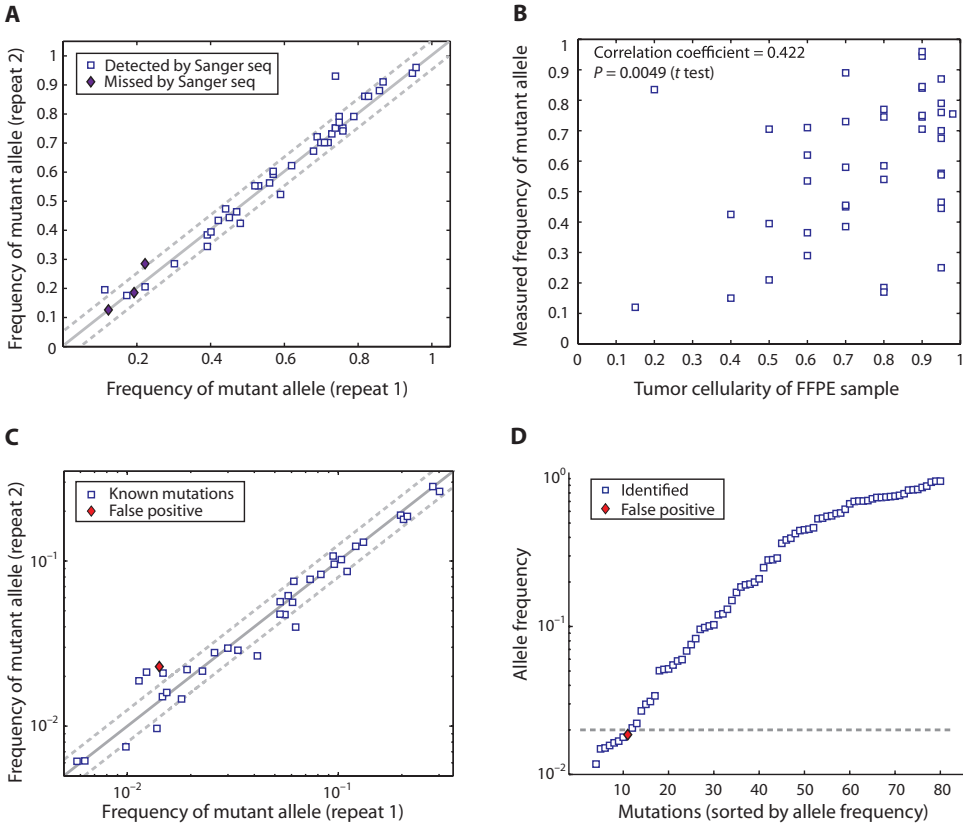

Fig. 2. Identification of mutations in ovarian cancer FFPE samples by TAm-Seq. (**A**) Concordance between duplicate measurements of AFs of mutations identified in fragmented DNA extracted from FFPE samples. The mutation frequency in each library was calculated as the fraction of reads with the mutant (nonreference) base. Solid line indicates equality. Dotted lines indicate a difference in AF of 0.05. (**B**) Correlation of AF with FFPE tumor cellularity. The measured mutant AF (average of both repeats) correlated significantly with the cellularity, estimated from histology (table S3). (**C**) Concordance between duplicate measurements of AFs of mutations identified in a mixture of DNA extracted from different FFPE samples. (**D**) Summary of mutations called in FFPE using TAm-Seq, sorted by increasing AF. Dotted line indicates AF of 2%.

limit of accuracy set by the Poisson distribution for independently segregating molecules. We compared the observed AF to the expected AF for cases where more than six copies of the minor allele were expected. Of 24 such cases, the root mean square (RMS) relative error between the expected and the observed frequency was 14%, with only 2 of 24 cases exhibiting more than 20% discrepancy. For samples with expected minor AF of 0.025, the RMS error was 23% (Fig. 3B).

Noninvasive identification of cancer mutations in plasma circulating DNA

We applied TAm-Seq to directly identify mutations in plasma of cancer patients. We studied a cohort of samples from individuals with HGSOC. These samples were first analyzed for tumor-specific mutations using digital PCR (Supplementary Methods), a method that is highly accurate (*2*, *3*, *7*, *37*) but requires design and validation of a different assay for every mutation screened and relies on previous identification of mutations in tumor samples from the same patients (*2*, *3*). We initially selected for analysis seven cases that had relatively high levels of circulating mutant *TP53* DNA in the plasma (as assessed by digital PCR). Using the equivalent amount of DNA present in 30 to 120 μl of plasma, we performed duplicate preamplification reactions for each sample. For all seven patients, *TP53* tumor mutations were identified in the circulating DNA at frequencies of 4% to 44% (Table 1). In one plasma sample collected from an ovarian cancer patient at relapse, we also identified a de novo mutation in the tyrosine kinase domain of *EGFR* (exon 21), at AF of 6% (patient 27, Table 1). We subsequently validated the presence of this mutation in plasma by performing replicate Sanger sequencing reactions of highly diluted template (Supplementary Methods), and 4 of 91 wells that were successfully Sanger-sequenced contained the *EGFR* mutation (fig. S4). We further validated the presence of this mutation by designing a sequence-specific TaqMan probe targeting this mutation and performing digital PCR (Table 1). The mutation was also identified by TAm-Seq in additional plasma collected from the same individual (sample 16, Table 2). This mutation in *EGFR* was not found in the ovarian mass removed by interval debulking surgery 15 months before the blood sample was collected, although the same sample did contain the concomitant *TP53* mutation found in the same patient's plasma, at AF of 85% (patient 27, table S3). We subsequently used TAm-Seq to sequence seven additional samples collected at the time of initial surgery including deposits in right and left ovaries and omentum. The *EGFR* mutation was detected in the two omental samples above the 0.99 confidence margin (fig. S3) at AF of 0.7%, but was not detected in the six ovarian samples (below the 0.8 confidence margin). Without previous identification in plasma, this mutation would not have been directly identified on screening those samples using high-specificity mutation identification criteria owing to its low AF. In contrast, the *TP53* mutation was identifiable in all biopsy and plasma samples (Fig. 4A). The frequency of mutant alleles in the relapsed tumor could not be directly assessed because a biopsy at relapse was not available.

We validated the TAm-Seq method on a larger panel of plasma samples in which levels of tumor-specific mutations were measured in parallel using patient-specific digital PCR assays. DNA extracted from 62 additional plasma samples collected at different time points from 37 patients with advanced HGSOC was amplified in duplicate (table S6), using DNA present in ~0.15 ml of plasma per reaction (range, 0.06 to 0.2 ml). Amplicon libraries were tagged and pooled together for sequencing with libraries prepared from 24 control samples. This generated an average sequencing depth of 650 for 62 plasma samples, sufficient to detect mutations present at AFs of 1% to 2%. Of >1.5 million possible substitutions, 42 mutations were called using the parameters previously optimized for FFPE analysis (table S6).

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

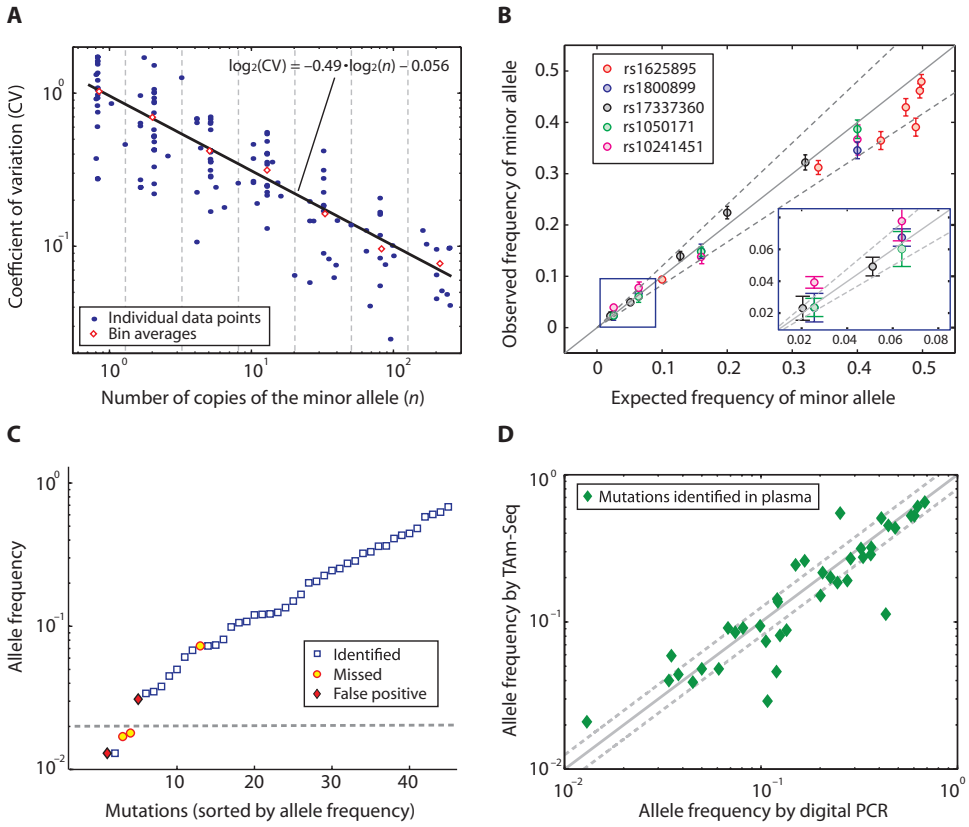

Fig. 3. Noninvasive identification and quantification of cancer mutations in plasma DNA by TAm-Seq. (**A**) Sampling noise in sequencing of sparse DNA using dilutions of plasma DNA from healthy individuals. CV of triplicate AF readings was calculated for each of the five SNPs in each of the mixes, which had varying numbers of copies of the minor allele (n) (blue dots). Bin averages (red diamonds) are the mean CVs calculated for each bin (bin edges denoted by the dotted vertical lines). A linear fit to the $\log_2$ of the mean CV as a function of the $\log_2$ expected copy number was calculated (black line). Two data points, with (n = 100, CV = 0.0064) and (n = 32, CV = 0.0185), were omitted from the figure for enhanced scaling. Three data points with minor allele copies of <0.8 were omitted from the analysis (n = 0.51, CV = 0.62; n = 0.41, CV = 0.86; n = 0.20, CV = 0.99). (**B**) Expected versus observed frequency of rare alleles in a dilution series of circulating DNA. Mean observed frequency was calculated for each of five SNPs for samples, where expected initial number of minor allele copies was greater than 6. Expected frequencies were calculated on the basis of quantification by digital PCR. Dotted lines represent 20% deviation from the expected frequencies. Inset highlights cases with expected minor AF <0.025. (**C**) Mutations identified in 62 plasma samples from patients with advanced HGSOC using TAm-Seq. AFs are based on digital PCR measurement for confirmed mutations (identified or missed by TAm-Seq), and on TAm-Seq for the false positives called using parameters optimized for analysis of FFPE samples. The dashed horizontal line indicates AF of 2%. Mutations detected by digital PCR at AF <1% are not shown. (**D**) AFs measured by TAm-Seq versus digital PCR for mutations identified in plasma DNA.

Table 1. Mutations identified by TAm-Seq in plasma samples from seven ovarian cancer patients. TAm-Seq was used to sequence DNA extracted from plasma of subjects with HGSOC (stage III/IV at diagnosis). Plasma was collected when patients presented with relapse disease, before initiation of chemotherapy. For patient 46, DNA from a formalin-fixed, paraffin-embedded (FFPE) sample was not included in the TAm-Seq set and the mutation was validated in FFPE by Sanger sequencing. CA125 was measured at time of plasma collection. Mean depth of coverage at the mutation locus in the TAm-Seq data was averaged over the repeats (RMS deviation = 850). AF, allele frequency; N, no; Y, yes.

Patient ID	Age at diagnosis	Time elapsed since surgery (months); number of previous lines of chemotherapy	CA125 (U/ml)	Plasma per amplification reaction (μl)	Gene	Mutation and base change (genome build hg19)		Protein change	Detected in FFPE	Mean depth (sequencing reads)	Mean AF using TAm-Seq	Mean AF using digital PCR
8	60	13; 1	2122	50	*TP53*	17:7577120	C>T	p.R273H	Y	5000	0.09	0.10
12	62	27; 3	365	50	*TP53*	17:7577579	G>T	p.Y234*	Y	5000	0.10	0.08
14	58	50; 3	260	120	*TP53*	17:7578212	G>A	p.R213*	Y	5800	0.15	0.12
25	61	9; 1	944	110	*TP53*	17:7578404	A>T	p.C176S	Y	4800	0.04	0.08
27†	68	15; 1	1051	90	*TP53*	17:7578262	C>G	p.R196P	Y	7700	0.06	0.14
					EGFR	7:55259437	G>A	p.R832H	N	5700	0.06	0.05
31	64	12; 1	313	30	*TP53*	17:7578406	C>T	p.R175H	Y	4500	0.44	0.56
46	56	30; 2	1509	30	*TP53*	17:7578406	C>T	p.R175H	Y	4200	0.23	0.30

*Indicates stop codon. †Both a *TP53* and an *EGFR* mutation were identified in this sample (Fig. 4A).

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Table 2. Mutations identified by TAm-Seq in a set of 62 plasma samples from ovarian cancer patients. Forty mutations were identified by TAm-Seq using stringent parameters for mutation calling. Plasma samples described in this table are distinct from those in Table 1, but patients included overlap. Additional data on patients and mutations are provided in table S6.

Sample number	Plasma volume per amplification reaction (μl)	DNA amount per amplification reaction (ng)	Gene	Protein change	Mean depth (sequencing reads)	Mean AF using TAm-Seq	Mean AF using digital PCR
1	70	0.9	*TP53*	p.R273C	640	0.260	0.167
2	160	4.2	*TP53*	p.R248Q	340	0.244	0.150
3	160	5.7	*TP53*	p.R248Q	640	0.507	0.410
4	120	9.9	*TP53*	p.R213X	810	0.059	0.035
5	120	1.4	*TP53*	p.C141Y	680	0.021	0.013
6	120	2.1	*TP53*	p.C141Y	720	0.044	0.038
7	190	17.9	*TP53*	p.I195N	800	0.091	0.081
8	160	14.8	*TP53*	p.R175H	510	0.608	0.627
9	160	10.7	*TP53*	p.R175H	550	0.526	0.604
10	160	6.1	*TP53*	p.R175H	530	0.651	0.682
11	160	4.9	*TP53*	p.R175H	490	0.526	0.581
13	160	2.8	*TP53*	p.C135R	480	0.039	0.045
14	160	2.5	*TP53*	p.C135R	610	0.046	0.120
15	160	3.0	*TP53*	p.C135R	470	0.091	0.068
16†	130	3.7	*TP53*	p.R196P	1070	0.088	0.135
			EGFR	p.R832H	614	0.048	0.050
17	160	4.2	*TP53*	p.C176S	580	0.113	0.432
18	160	4.4	*TP53*	p.C176S	620	0.029	0.108
20	140	5.2	*TP53*	p.R175H	650	0.201	0.226
21	140	3.6	*TP53*	p.R175H	650	0.085	0.074
22	140	4.1	*TP53*	p.R175H	630	0.081	0.125
23	140	3.7	*TP53*	p.R175H	710	0.074	0.106
24	140	7.1	*TP53*	p.R175H	760	0.269	0.286
25	130	3.9	*TP53*	p.R273H	750	0.094	0.099
26	160	5.7	*TP53*	p.R282W	640	0.048	0.061
27	150	3.6	*TP53*	p.C141Y	480	0.321	0.364
29	150	9.5	*TP53*	p.E258K	190	0.548	0.253
31	160	3.6	*TP53*	p.C135Y	620	0.040	0.034
32	140	2.4	*TP53*	p.E56X	1480	0.137	0.122
33	160	13.2	*TP53*	p.K132N	740	0.216	0.206
34	60	5.3	*TP53*	p.K132N	570	0.151	0.201
36	160	5.8	*TP53*	p.K132N	620	0.191	0.275
37	160	9.4	*TP53*	p.K132N	530	0.287	0.362
38	160	10.1	*TP53*	p.K132N	590	0.275	0.331
39	160	16.4	*TP53*	p.K132N	700	0.315	0.323
40	160	19.7	*TP53*	p.K132N	830	0.435	0.482
41	160	15.0	*TP53*	p.K132N	730	0.452	0.445
42	160	8.5	*TP53*	p.K132N	560	0.185	0.245
43	150	3.6	*TP53*	Splicing	680	0.143	0.121
44‡	170	5.2	*TP53*	p.C238R	1543	0.071	0.073

†Both a *TP53* and an *EGFR* mutation were identified in this sample, collected from patient 27 (Table 1), 25 months after initial surgery (Fig. 4A). ‡The amplicon containing the mutation failed amplification in this sample in the initial experiment and was identified successfully in repeat analysis.

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Thirty-nine of these matched mutations detected by digital PCR in those samples (Fig. 3C). Three potential false positives were called, at AF of 3.1%, 1.3%, and 0.7% (the latter in a control sample). Using higher-stringency parameters for mutation identification (Supplementary Methods), we retained only the 39 validated mutations called, with no false positives (Table 2).

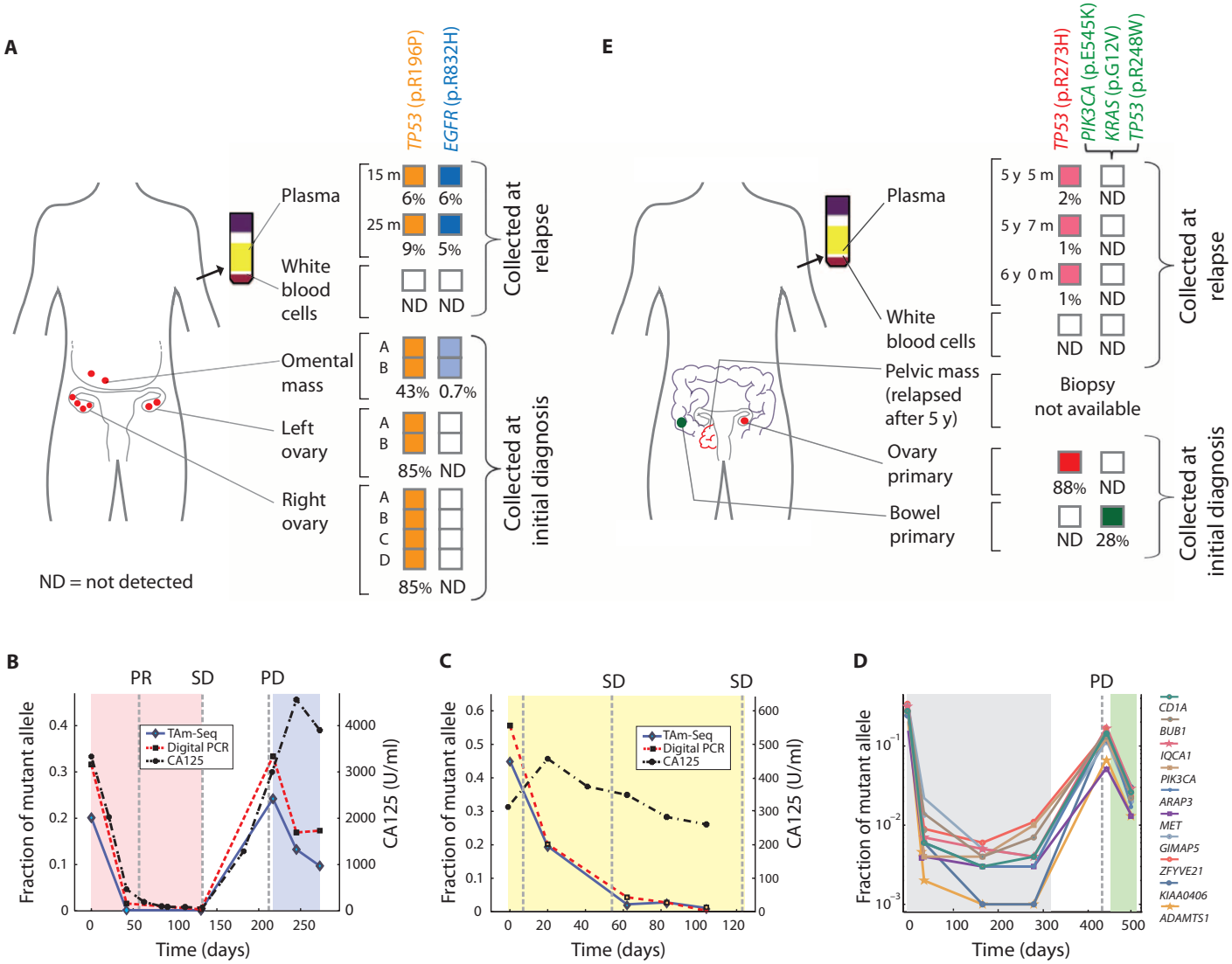

Fig. 4. Clinically relevant applications of plasma DNA sequencing using TAm-Seq. (**A**) Retrospective analysis by TAm-Seq of plasma samples collected during patient follow-up and biopsy specimens collected at initial surgery. We identified a mutation in exon 21 of *EGFR* (dark blue boxes) in two separate plasma samples, collected 15 and 25 months after initial surgery from patient 27 (Tables 1 and 2). This mutation was not directly identified in eight tumor biopsy specimens collected at the time of initial surgery (two from omental mass, two from left ovary, and four from right ovary). Having identified the mutation in the plasma samples, we examined this mutation using the lower-specificity criteria defined for mutation detection (Supplementary Methods) and detected the mutation in the two specimens that had been collected from the omentum at the time of surgery (light blue boxes) but not in the six ovarian specimens. A mutation in *TP53* was identified in all tumor and plasma samples collected from this patient (Tables 1 and 2 and table S3), but not in white blood cells (buffy coat). Percentages indicate mutant AFs. Empty boxes and "ND" indicate samples where a mutation was not identified or detected (below 0.8 confidence margin). (**B**) Monitoring frequency of mutant DNA in plasma of an ovarian cancer patient (patient 46) over time using TAm-Seq and digital PCR. TAm-Seq results are reported as the mean frequency of duplicate analyses. Parallel data are shown for digital PCR and serum CA125. Shaded regions indicate periods of chemotherapy, and vertical dashed lines indicate radiological assessment of patient responses: PR, partial response; SD, stable disease; PD, progressive disease. (**C**) Monitoring frequency of mutant DNA in plasma of an ovarian cancer patient (patient 31) over time. (**D**) Dynamics of 10 tumor-specific mutations in plasma of a breast cancer patient (not included in the other sets of samples analyzed). (**E**) Retrospective analysis of samples from synchronous primary tumors (bowel and ovarian) collected at the time of initial surgery and three plasma samples collected at relapse. In primary tumors from this patient (not included in the other sets of samples analyzed), a *TP53* mutation was identified in the ovarian cancer (red box), and mutations in *PIK3CA*, *KRAS*, and *TP53* were identified in the bowel cancer (green box). At relapse, a biopsy was not performed on the pelvic mass. The *TP53* mutation that was identified in the ovarian primary tumor (p.R273H) was detected in plasma, whereas the bowel-associated mutations were not detected.

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Of 40 point mutations detected at AF >2% by digital PCR, 38 (95%) were identified by TAm-Seq in a single experiment (Fig. 3C). One additional mutation was located in an amplicon that failed in that sample and was identified in repeated analysis; the other was likely missed by TAm-Seq owing to sampling noise, because it was found in one of the duplicate preamplified libraries but not the other (table S6). One of three mutations detected by digital PCR at 1% < AF < 2% was identified by TAm-Seq (Fig. 3C). Eleven additional point mutations detected by digital PCR at AF <1% were not detected by TAm-Seq at these settings. TAm-Seq and digital PCR measurements of AF had excellent agreement, with correlation coefficient of 0.90, increasing to 0.97 when discarding the two strongest outliers (Fig. 3D). Thus, we screened 62 samples across sizeable genomic stretches, using minute amounts of plasma DNA (median, 4 ng), and obtained 97.5% sensitivity with PPV of 100% for identifying mutations at AF >2% in plasma by TAm-Seq. Using parameters optimized for FFPE samples, one potential false positive was called at AF >2%, reducing the PPV to 97.5% (Table 3).

Monitoring levels of ctDNA

Various methods have been suggested to monitor changes in mutation load in plasma. These can have enhanced sensitivity compared to TAm-Seq for tracking individual mutations, but require design of personalized assays (*3*, *18*, *19*). None of these methods have been widely adopted. We therefore applied TAm-Seq as a generic tool to measure changes in the frequency of ctDNA over time. We studied serial plasma samples collected during follow-up and treatment of two patients with relapsed HGSOC, collected during 104 and 273 days of follow-up and treatment, respectively. Frequencies of mutant *TP53* alleles were measured by TAm-Seq and in parallel by digital PCR using a mutation-specific probe. The two methods of quantification had excellent agreement. Mutant AFs in plasma of ovarian cancer patients reflected well the clinical course of the disease compared to the serum marker CA125, showed marked decrease when systemic treatment was initiated, and increased in parallel to disease progression. In the first case (Fig. 4B), a 56-year-old woman with relapsed ovarian cancer (patient 46) was treated with fourth-line carboplatin + paclitaxel chemotherapy for six cycles (pink-shaded region). Radiology showed partial response on mid-treatment computed tomography (CT) scan. End-of-treatment CT showed stable disease. Twelve weeks from the end of her fourth-line treatment, the patient developed progressive disease. The patient then initiated fifth-line chemotherapy with liposomal doxorubicin (purple-shaded region). In the second case (Fig. 4C), a 64-year-old woman with relapsed ovarian cancer (patient 31) was treated with second-line ECX (epirubicin, cisplatin, and capecitabine) chemotherapy for six cycles. Radiology showed stable disease on mid- and end-of-treatment CT scans. The patient then remained off treatment, until she progressed 3 months later.

TAm-Seq can be flexibly adapted to sequence different genomic regions by designing primers to amplify regions of interest. We used this capability to study dynamics of multiple mutations in parallel.

Table 3. Summary of mutations identified in 69 plasma samples of ovarian cancer patients. Samples were analyzed by TAm-Seq and in parallel by digital PCR. Using parameters optimized for plasma DNA, false-positive calls were lost, whereas all confirmed calls were retained, resulting in specificity and PPV of 100%.

First set of plasma samples	
Plasma samples analyzed	7
Point mutations originally detected by digital PCR, using patient-specific assays targeting mutations identified in tumor samples	7
Point mutations identified directly in plasma by TAm-Seq	8
De novo mutations identified by TAm-Seq only, subsequently confirmed by digital PCR	1
Second set of plasma samples	
Plasma samples analyzed	62
Point mutations detected by digital PCR at AF >2%	40
Point mutations with AF >2% (by digital PCR) identified by TAm-Seq	39
Point mutations missed by TAm-Seq due to sampling error	1
Sensitivity of TAm-Seq for identifying mutations at AF >2%	97.5%
PPV of mutations called by TAm-Seq with AF >2%	97.5%*
ctDNA in ovarian cancer	
Advanced ovarian cancer patients in both sets†	38
Patients where TAm-Seq identified cancer mutations	20

*One unconfirmed substitution was called at AF >2% using parameters optimized for FFPE material.

†The first set included 7 patients (Table 1), and the second set included 37 patients (table S6), 6 of whom overlap.

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Whole-genome sequencing of tumor material was used to identify tumor mutations in a patient with metastatic breast cancer undergoing two phases of chemotherapy. Ten mutations were selected, and short amplicons (<120 bp) were designed to cover the mutation loci (table S7). Serial plasma samples were collected over the course of 497 days, both before and after treatment. We performed TAm-Seq in duplicate, using DNA from 0.08 ml of plasma per amplification, and tracked dynamics of all mutations in parallel (Fig. 4D). The patient was treated with single-agent epirubicin (gray-shaded region). After 4 months off treatment, a CT scan showed progressive disease and the patient commenced further treatment with paclitaxel chemotherapy. The 10 mutations followed a common pattern of sharp decline in AF upon onset of therapy and an increase in AF upon disease progression after termination of therapy (Fig. 4D).

Finally, we used TAm-Seq to study plasma from a patient who had a history of two synchronous primary cancers, bowel and ovarian, which were resected simultaneously. After a 5-year remission, a pelvic mass of uncertain origin was detected. A biopsy was considered to guide selection of therapy but was not performed owing to risk of complications and comorbidities. The patient commenced empirically on an ovarian cancer chemotherapy regimen, to which she responded. Retrospective analysis by TAm-Seq of FFPE from the primary tumors collected at initial surgery, and three plasma samples collected serially at the time of relapse (5 years and 5 months, 5 years and 7 months, and 6 years after initial surgery), showed that the patient's plasma at relapse contained the *TP53* (p.R273H) mutation identified in the ovarian primary tumor (exceeding the 0.98, 0.93, and 0.97 confidence margins, respectively), but not the *PIK3CA* (p.E545K), *KRAS* (p.G12V), or *TP53* (p.R248W) mutations identified in the primary bowel cancer (below the 0.8 confidence margin) (Fig. 4E). Had these results been available, uncertainty and treatment delays may have been avoided, as well as the risk of prescribing chemotherapy for an inappropriate tumor site. An alternative possible outcome may have involved a finding of the *PIK3CA* or *KRAS* mutations (present in the primary bowel cancer) in the patient's plasma at the time of relapse. Such a finding, if available to clinicians at the time, may not only have led to alternate chemotherapy being offered but may have also opened the possibility of enrolment into a trial for targeted therapy with mammalian target of rapamycin (mTOR), phosphatidylinositol 3-kinase (PI3K), or mitogen-activated protein kinase kinase (MEK) inhibitors (*11*).

DISCUSSION

Detection of rare mutations in circulating DNA has long been pursued owing to its potentially transformative impact on cancer diagnosis and management. Important progress has been made using sequence-specific assays that target predefined mutations and that detect extremely rare alleles. Assays such as PCR (*6*, *7*), ligation (*5*), and primer extension/mass spectrometry (*27*) can identify specific, predefined mutations in plasma samples. Enhanced detection down to 1 mutant allele in 10,000 or more wild-type alleles can be obtained using a variety of methods, such as peptide nucleic acid and primer extension ("PPEM") (*38*), ligation followed by quantitative PCR ("LigAmp") (*39*), bead-based digital PCR in emulsions ("BEAMing") (*2*, *3*), microfluidic-based (*7*) or droplet-based digital PCR (*40*), or microinsertion/deletion/indel-activated pyrophosphorolysis ("MAP") (*29*). Nonetheless, identification of rare mutations in tumor suppressor genes such as *TP53*, which are widely mutated in cancers but lack a well-defined hotspot region, remains an elusive goal.

In patients with advanced cancers, mutant alleles can reach a sizeable fraction of DNA. For example, Dukes' D colorectal cancers have median 8% mutant AF (*2*). Screening of entire genes for mutations would therefore be useful for some applications, even if analytical selectivity is limited to a few percent. Advances in massively parallel sequencing make new approaches possible. These have largely focused on large-scale analyses, including whole-genome or whole-exome sequencing (*41*). This generates a large amount of data on genomic regions that do not, at present, inform clinical decisions. Moreover, the depth of coverage for clinically significant loci is not sufficient to detect changes that occur at low frequency (<5%). Such approaches have recently been complemented by methods for examination of individual amplicons at great depth (*30*).

The intermediate scale of sequencing is most likely to have immediate impact on clinical genomics. Targeted sequencing has been applied for tumor DNA (*34*, *35*) and cyst fluid (*33*) to detect mutations down to 5% AF, but has not been applied for analysis of circulating tumor nucleic acids. Here, we demonstrate noninvasive identification of mutant alleles in plasma, at AFs as low as 2%, by targeted deep sequencing of circulating DNA. Our TAm-Seq method uses a combination of short amplicons, two-step amplification, sample barcodes, and high-throughput PCR. Because the amplicons are short, this method effectively amplifies even small amounts of fragmented DNA such as are present in circulating DNA. The two-step amplification permits extensive primer multiplexing that enables the amplification and sequencing of sizeable genomic regions by tiling short amplicons without loss of fidelity or efficiency. Duplicate sequencing of each sample is used to avoid false positives stemming from PCR errors. Sample barcodes and high-throughput PCR reduce the per-sample costs to a range where this may be widely applicable. Preparing TAm-Seq libraries for sequencing from 48 samples takes less than 24 hours and involves only few hours of hands-on time. New platforms for massively parallel sequencing allow for fast turnaround times, which make this approach practical in a clinical setting.

The sensitivity presently achieved can provide useful diagnostic information in certain advanced cancers. We studied a cohort of subjects with advanced HGSOC in which the tumor suppressor gene *TP53* is a driver mutation (*20*). Of the 69 plasma samples collected from 38 different individuals with advanced HGSOC, we identified mutations in *TP53* in 46 samples (67%) from 20 of the cases (53%). In contrast, a previous study using a ligase detection reaction with bespoke primers found mutated *TP53* sequences in plasma for only 30% of advanced ovarian cancer patients (*5*), and a study using single-strand conformation polymorphism found no ctDNA in preoperative plasma samples from high-grade serous cancer patients (*42*).

Targeted agents, such as inhibitors of poly(adenosine diphosphate–ribose) polymerase (PARP), or tyrosine kinase inhibitors targeting epidermal growth factor receptor (EGFR), may be applicable for systemic treatment of advanced HGSOC (*8*, *10*, *22*). In a recent study of 203 HGSOC tumors, *EGFR* was found to be the most frequently mutated oncogene and was mutated in nearly 10% of cases (*10*). In one case, we identified in plasma a de novo mutation in the tyrosine kinase domain (exon 21) of *EGFR*, located 26 amino acids upstream of the L858R activating mutation widely documented for lung cancer. In a subset of tumor samples collected from the same patient 15 months earlier, this mutation was detected at AF of 0.7%, but could not have

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

been identified by analysis of those samples alone without previous knowledge of the mutation identified in plasma (Fig. 4A). In a clinical setting, identification of such a mutation could potentially guide treatment with alternative molecularly targeted therapy (*10*). Current clinical recommendations in lung adenocarcinoma suggest mutation assessment in exons 18 to 21 of *EGFR* (a region of ~560 bp) in the tumor tissue to identify patients eligible for treatment with gefitinib or erlotinib (*9*). Using a commercial PCR-based in vitro diagnostic kit (Qiagen), 28 different *EGFR* variants can be assayed (not including the mutation we identified), but the sample needs to be subdivided into seven different reactions. When sample is limited or mutant alleles are rare, this could introduce sampling errors.

Using standard amplification primers tailored to the mutation loci, we also used TAm-Seq to monitor the dynamics of 10 mutations in plasma DNA of a single patient with metastatic breast cancer, using minute amounts of input DNA. Previous studies have followed up to two mutations in any individual patient (*3*, *19*). Tracking multiple mutations can provide insight into clonal evolution and, at the same time, increases the robustness for tumor monitoring by compensating for effects of sampling noise or mutational drift. For example, if a patient has only five copies of a mutant allele per milliliter of plasma (on average), there is a 37% probability that this mutation will not be present in a 0.2-ml sample, and even a perfect assay will fail to detect residual tumor, whereas a method that measures multiple mutations in parallel can have a low likelihood of a false-negative result even if the detection rate for each mutation is less than 50%.

A current limitation of TAm-Seq is the detection limit compared to assays that target individual loci (*2*, *3*, *7*, *40*), which have been shown to detect two to three orders of magnitude lower frequencies. Our approach may be sufficient for analyzing plasma from patients with certain advanced cancers, but further improvement may be necessary before this method can be more widely used in the clinic. Higher read depth or fidelity, additional replicates, or improved algorithms could allow for enhanced mutation detection without change to protocols. An alternative strategy is through rare allele enrichment, for example, by combining TAm-Seq with protocols such as COLD-PCR (co-amplification at lower denaturation temperature PCR) (*31*).

Previously proposed methods for personalized monitoring of tumor dynamics relied on expensive custom-designed probes (*3*) or identification of rearrangements using whole-genome sequencing (*18*, *19*). These have better analytical sensitivity than currently achieved by TAm-Seq, but are difficult to implement on a routine basis. TAm-Seq strikes a balance between sensitivity and ease of use and could facilitate study and application of circulating DNA. Using TAm-Seq, we identified cancer mutations in the plasma of most advanced ovarian cancer patients and tracked dynamics of *TP53* mutations without requiring any specially designed probes. In summary, TAm-Seq is a flexible and cost-effective platform for applications in noninvasive cancer genomics and diagnostics. We have shown that this method can be used for high-throughput sequencing of plasma samples to identify and monitor levels of multiple cancer mutations in circulating DNA. This could also be applied to screen for rare mutations in a variety of heterogeneous sample types such as low-cellularity tumor specimens, cytological samples, or circulating tumor cells (*16*). With further developments, this and derivative methods may be applied in molecular screening for earlier detection or for differential diagnosis of cancer from benign masses. For genetic analysis of FFPE or small biopsy samples, TAm-Seq can be applied as is, as a cost-effective clinical aid.

MATERIALS AND METHODS

Sample collection

FFPE blocks were obtained from the pathology archives at Addenbrooke's Hospital (Cambridge, UK). Plasma samples were collected upon disease relapse, before and during chemotherapy treatment. Sample collection for this study was approved by Cambridgeshire Research Ethics Committee (REC 08/H0306/61 and 07/Q0106/63). Peripheral blood samples were collected into EDTA tubes and centrifuged at 820*g* for 10 min within 1 hour of collection to limit degradation of cell-free DNA and leukocyte lysis. Aliquots (1 ml) of plasma were centrifuged in a benchtop microfuge at 14,000 rpm for 10 min. The supernatant was transferred to sterile 1.5-ml tubes and stored at −80°C before extraction.

Extraction of DNA from FFPE and blood plasma

Paraffin blocks were cut as 8-μm sections on plain glass slides. Targeted regions for sampling were marked on adjacent hematoxylin and eosin sections by the study pathologist and recovered by scrape macrodissection. Between 3 and 20 sections were macrodissected depending on the tissue sample's size. DNA from FFPE sections was extracted with QIAamp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's instructions.

Circulating DNA was extracted from between 0.85 and 2.2 ml of plasma with the QIAamp Circulating Nucleic Acid kit (Qiagen), following the manufacturer's instructions, and with the QIAvac 24 Plus vacuum manifold. Carrier RNA was added to ACL lysis buffer to enhance binding of nucleic acids to the QIAamp membrane with the aim to enhance yields.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/4/136/136ra68/DC1

Methods

Fig. S1. PCR strategy and primer design.

Fig. S2. Sanger traces for mutations identified by tagged-amplicon sequencing.

Fig. S3. Background frequencies and detection limits for base substitutions.

Fig. S4. Replicate dilute Sanger sequencing of a mutation identified in plasma.

Table S1. Target-specific primers.

Table S2. Unique sequencing barcodes.

Table S3. Mutations identified in FFPE samples.

Table S4. SNPs identified in circulating DNA from two plasma control samples.

Table S5. Frequency of SNP alleles in dilution series of DNA from control plasma.

Table S6. Additional data for Table 2 for mutations identified in plasma samples.

Table S7. Mutations and amplicons studied in one breast cancer patient.

REFERENCES AND NOTES

1. E. Gormally, E. Caboux, P. Vineis, P. Hainaut, Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance. *Mutat. Res.* **635**, 105–117 (2007).
2. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz Jr., S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16368–16373 (2005).
3. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
4. H. Schwarzenbach, D. S. Hoon, K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
5. E. M. Swisher, M. Wollan, S. M. Mahtani, J. B. Willner, R. Garcia, B. A. Goff, M. C. King, Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. *Am. J. Obstet. Gynecol.* **193**, 662–667 (2005).

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

6. R. E. Board, A. M. Wardley, J. M. Dixon, A. C. Armstrong, S. Howell, L. Renshaw, E. Donald, A. Greystoke, M. Ranson, A. Hughes, C. Dive, Detection of *PIK3CA* mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Res. Treat.* **120**, 461–467 (2010).
7. T. K. Yung, K. C. Chan, T. S. Mok, J. Tong, K. F. To, Y. M. Lo, 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).
8. S. Banerjee, S. Kaye, The role of targeted therapy in ovarian cancer. *Eur. J. Cancer* **47** (Suppl. 3), S116–S130 (2011).
9. V. L. Keedy, S. Temin, M. R. Somerfield, M. B. Beasley, D. H. Johnson, L. M. McShane, D. T. Milton, J. R. Strawn, H. A. Wakelee, G. Giaccone, American Society of Clinical Oncology provisional clinical opinion: Epidermal growth factor receptor (*EGFR*) mutation testing for patients with advanced non–small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J. Clin. Oncol.* **29**, 2121–2127 (2011).
10. U. A. Matulonis, M. Hirsch, E. Palescandolo, E. Kim, J. Liu, P. van Hummelen, L. MacConaill, R. Drapkin, W. C. Hahn, High throughput interrogation of somatic mutations in high grade serous cancer of the ovary. *PLoS One* **6**, e24433 (2011).
11. J. A. Engelman, L. Chen, X. Tan, K. Crosby, A. R. Guimaraes, R. Upadhyay, M. Maira, K. McNamara, S. A. Perera, Y. Song, L. R. Chirieac, R. Kaur, A. Lightbown, J. Simendinger, T. Li, R. F. Padera, C. García-Echeverría, R. Weissleder, U. Mahmood, L. C. Cantley, K. K. Wong, Effective use of PI3K and MEK inhibitors to treat mutant *Kras* G12D and *PIK3CA* H1047R murine lung cancers. *Nat. Med.* **14**, 1351–1356 (2008).
12. P. J. Campbell, S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. A. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, C. A. Griffin, J. Burton, H. Swerdlow, M. A. Quail, M. R. Stratton, C. Iacobuzio-Donahue, P. A. Futreal, The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
13. S. P. Shah, R. D. Morin, J. Khattra, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R. A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G. A. Taylor, A. E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R. L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M. A. Marra, S. Aparicio, Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
14. M. Inukai, S. Toyooka, S. Ito, H. Asano, S. Ichihara, J. Soh, H. Suehisa, M. Ouchida, K. Aoe, M. Aoe, K. Kiura, N. Shimizu, H. Date, Presence of *epidermal growth factor receptor gene* T790M mutation as a minor clone in non–small cell lung cancer. *Cancer Res.* **66**, 7854–7858 (2006).
15. S. L. Edwards, R. Brough, C. J. Lord, R. Natrajan, R. Vatcheva, D. A. Levine, J. Boyd, J. S. Reis-Filho, A. Ashworth, Resistance to therapy caused by intragenic deletion in *BRCA2*. *Nature* **451**, 1111–1115 (2008).
16. S. Maheswaran, L. V. Sequist, S. Nagrath, L. Ulkus, B. Brannigan, C. V. Collura, E. Inserra, S. Diederichs, A. J. Iafrate, D. W. Bell, S. Digumarthy, A. Muzikansky, D. Irimia, J. Settleman, R. G. Tompkins, T. J. Lynch, M. Toner, D. A. Haber, Detection of mutations in *EGFR* in circulating lung-cancer cells. *N. Engl. J. Med.* **359**, 366–377 (2008).
17. B. Norquist, K. A. Wurz, C. C. Pennil, R. Garcia, J. Gross, W. Sakai, B. Y. Karlan, T. Taniguchi, E. M. Swisher, Secondary somatic mutations restoring *BRCA1/2* predict chemotherapy resistance in hereditary ovarian carcinomas. *J. Clin. Oncol.* **29**, 3008–3015 (2011).
18. R. J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F. M. De La Vega, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., V. E. Velculescu, Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl. Med.* **2**, 20ra14 (2010).
19. D. J. McBride, A. K. Orpana, C. Sotiriou, H. Joensuu, P. J. Stephens, L. J. Mudie, E. Hämäläinen, L. A. Stebbings, L. C. Andersson, A. M. Flanagan, V. Durbecq, M. Ignatiadis, O. Kallioniemi, C. A. Heckman, K. Alitalo, H. Edgren, P. A. Futreal, M. R. Stratton, P. J. Campbell, Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer* **49**, 1062–1069 (2010).
20. A. A. Ahmed, D. Etemadmoghadam, J. Temple, A. G. Lynch, M. Riad, R. Sharma, C. Stewart, S. Fereday, C. Caldas, A. Defazio, D. Bowtell, J. D. Brenton, Driver mutations in *TP53* are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221**, 49–56 (2010).
21. Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
22. R. C. Bast Jr., Molecular approaches to personalizing management of ovarian cancer. *Ann. Oncol.* **22** (Suppl. 8), viii5–viii15 (2011).
23. K. C. Chan, S. F. Leung, S. W. Yeung, A. T. Chan, Y. M. Lo, Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin. Cancer Res.* **14**, 4141–4145 (2008).
24. H. C. Fan, Y. J. Blumenfeld, U. Chitkara, L. Hudgins, S. R. Quake, Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clin. Chem.* **56**, 1279–1286 (2010).
25. Y. M. Lo, K. C. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
26. R. W. Chiu, R. Akolekar, Y. W. Zheng, T. Y. Leung, H. Sun, K. C. Chan, F. M. Lun, A. T. Go, E. T. Lau, W. W. To, W. C. Leung, R. Y. Tang, S. K. Au-Yeung, H. Lam, Y. Y. Kung, X. Zhang, J. M. van Vugt, R. Minekawa, M. H. Tang, J. Wang, C. B. Oudejans, T. K. Lau, K. H. Nicolaides, Y. M. Lo, Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: Large scale validity study. *BMJ* **342**, c7401 (2011).
27. Y. M. Lo, N. B. Tsui, R. W. Chiu, T. K. Lau, T. N. Leung, M. M. Heung, A. Gerovassili, Y. Jin, K. H. Nicolaides, C. R. Cantor, C. Ding, Plasma placental RNA allelic ratio permits non-invasive prenatal chromosomal aneuploidy detection. *Nat. Med.* **13**, 218–223 (2007).
28. Y. M. Lo, F. M. Lun, K. C. Chan, N. B. Tsui, K. C. Chong, T. K. Lau, T. Y. Leung, B. C. Zee, C. R. Cantor, R. W. Chiu, Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13116–13121 (2007).
29. Z. Chen, J. Feng, C. H. Buzin, Q. Liu, L. Weiss, K. Kernstine, G. Somlo, S. S. Sommer, Analysis of cancer mutation signatures in blood by a novel ultra-sensitive assay: Monitoring of therapy or recurrence in non-metastatic breast cancer. *PLoS One* **4**, e7220 (2009).
30. 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).
31. J. Li, L. Wang, H. Mamon, M. H. Kulke, R. Berbeco, G. M. Makrigiorgos, Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat. Med.* **14**, 579–584 (2008).
32. S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, P. A. Futreal, COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
33. J. Wu, H. Matthaei, A. Maitra, M. Dal Molin, L. D. Wood, J. R. Eshleman, M. Goggins, M. I. Canto, R. D. Schulick, B. H. Edil, C. L. Wolfgang, A. P. Klein, L. A. Diaz Jr., P. J. Allen, C. M. Schmidt, K. W. Kinzler, N. Papadopoulos, R. H. Hruban, B. Vogelstein, Recurrent *GNAS* mutations define an unexpected pathway for pancreatic cyst development. *Sci. Transl. Med.* **3**, 92ra66 (2011).
34. O. Harismendy, R. B. Schwab, L. Bao, J. Olson, S. Rozenzhak, S. K. Kotsopoulos, S. Pond, B. Crain, M. S. Chee, K. Messer, D. R. Link, K. A. Frazer, Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biol.* **12**, R124 (2011).
35. N. Wagle, M. F. Berger, M. J. Davis, B. Blumenstiel, M. DeFelice, P. Pochanard, M. Ducar, P. Van Hummelen, L. E. MacConaill, W. C. Hahn, M. Meyerson, S. B. Gabriel, L. A. Garraway, High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.* **2**, 82 (2012).
36. T. Sjöblom, S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. D. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
37. B. Vogelstein, K. W. Kinzler, Digital PCR. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9236–9241 (1999).
38. X. Sun, K. Hung, L. Wu, D. Sidransky, B. Guo, Detection of tumor mutations in the presence of excess amounts of normal DNA. *Nat. Biotechnol.* **20**, 186–189 (2002).
39. C. Shi, S. H. Eshleman, D. Jones, N. Fukushima, L. Hua, A. R. Parker, C. J. Yeo, R. H. Hruban, M. G. Goggins, J. R. Eshleman, LigAmp for sensitive detection of single-nucleotide differences. *Nat. Methods* **1**, 141–147 (2004).
40. L. B. Pinheiro, V. A. Coleman, C. M. Hindson, J. Herrmann, B. J. Hindson, S. Bhat, K. R. Emslie, Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011 (2012).
41. M. Meyerson, S. Gabriel, G. Getz, Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **11**, 685–696 (2010).
42. J. Otsuka, T. Okuda, A. Sekizawa, S. Amemiya, H. Saito, T. Okai, M. Kushima, Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. *Int. J. Gynecol. Cancer* **14**, 459–464 (2004).

Acknowledgments: We thank H. Biggs, C. Hodgkin, S. Richardson, and L. Jones for assistance in sample collection, S. Aldridge for assistance in genomic analysis, and S. Tavaré, B. Davis, and M. Dunning for assistance in data analysis. **Funding:** 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, and Hutchison Whampoa Limited. C.P. was supported in part by the Academy of Medical Sciences, Wellcome Trust, British Heart Foundation, and Arthritis Research UK. **Author contributions:** T.F., M.M., C.P., D.G., D.W.Y.T., C.C., J.D.B., and N.R. designed the study. T.F., M.M., D.W.Y.T., F.K., J.H., A.P.M.,

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

and N.R. developed methods. T.F., D.G., D.W.Y.T., A.M.P., and S.-J.D. collected the data. T.F., M.M., and N.R. analyzed TAm-Seq data. C.P., S.-J.D., C.C., and J.D.B. designed clinical studies and collected samples and clinical data. M.J.-L. performed pathological analysis. D.B. contributed sequencing data. T.F., M.M., C.P., D.G., D.W.Y.T., J.H., A.P.M., J.D.B., and N.R. interpreted the data. T.F., M.M., and N.R. wrote the manuscript with assistance from C.P., D.G., D.W.Y.T., A.P.M., J.D.B., and other authors. All authors approved the final manuscript. **Competing interests:** A.P.M. and F.K. hold equity in Fluidigm and may stand to gain by publication of these findings. D.B. and F.K. hold equity in Illumina and may stand to gain by publication of these findings.

Submitted 24 October 2011
Accepted 18 April 2012
Published 30 May 2012
10.1126/scitranslmed.3003726

Citation: T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68 (2012).

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Science Translational Medicine

Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA

Tim Forshew, Muhammed Murtaza, Christine Parkinson, Davina Gale, Dana W. Y. Tsui, Fiona Kaper, Sarah-Jane Dawson, Anna M. Piskorz, Mercedes Jimenez-Linan, David Bentley, James Hadfield, Andrew P. May, Carlos Caldas, James D. Brenton and Nitzan Rosenfeld

Sci Transl Med **4**, 136ra68136ra68.
DOI: 10.1126/scitranslmed.3003726

Deep Sequencing Tumor DNA in Plasma

Five liters of circulating blood contain millions of copies of the genome, broken into short fragments; in cancer patients, a small fraction is circulating tumor DNA (ctDNA). An even smaller number harbor mutations that affect cancer outcome. Looking for diagnostic answers in circulating DNA is a challenge, but Forshew, Murtaza, and colleagues have risen to the occasion by developing a tagged-amplicon deep sequencing (TAm-Seq) method that can amplify and sequence large genomic regions from even single copies of ctDNA. By sequencing such large regions, the authors were able to identify low-level mutations in the plasma of patients with high-grade serous ovarian carcinomas.

Forshew *et al*. designed primers to amplify 5995 bases that covered select regions of cancer-related genes, including *TP53*, *EGFR*, *BRAF*, and *KRAS*. In plasma obtained from 38 patients with high levels of ctDNA, the authors were able to identify mutations in *TP53* at allelic frequencies of 2% to 65%. In plasma samples from one patient, they also identified a de novo mutation in *EGFR* that had not been detected 15 months prior in the tumor mass itself. Finally, the TAm-Seq approach was used to sequence ctDNA in plasma samples collected from two women with ovarian cancer and one woman with breast cancer at different time points, tracking as many as 10 mutations in parallel. Forshew and coauthors showed that levels of mutant alleles reflected the clinical course of the disease and its treatment—for example, stabilized disease was associated with low allelic frequency, whereas patients at relapse exhibited a rise in frequency.

Through several experiments, the authors were able to show that TAm-Seq is a viable method for sequencing large regions of ctDNA. Although this provides a new way to noninvasively identify gene mutations in our blood, TAm-Seq will need to achieve a more sensitive detection limit (<2% allele frequency) to identify mutations in the plasma of patients with less advanced cancers. Nevertheless, once optimized, this "liquid biopsy" approach will be amenable to personalized genomics, where the level and type of mutations in ctDNA would inform clinical decision-making on an individual basis.

ARTICLE TOOLS http://stm.sciencemag.org/content/4/136/136ra68

SUPPLEMENTARY MATERIALS http://stm.sciencemag.org/content/suppl/2012/05/25/4.136.136ra68.DC1

Use of this article is subject to the Terms of Service

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Translational Medicine* is a registered trademark of AAAS.

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

RELATED CONTENT

http://science.sciencemag.org/content/sci/339/6119/580.full
http://stm.sciencemag.org/content/scitransmed/5/207/207ps14.full
http://stm.sciencemag.org/content/scitransmed/6/224/224ra24.full
http://stm.sciencemag.org/content/scitransmed/6/224/224ra26.full
http://stm.sciencemag.org/content/scitransmed/6/241/241ra77.full
http://stm.sciencemag.org/content/scitransmed/6/243/243fs26.full
http://stm.sciencemag.org/content/scitransmed/7/283/283ra53.full
http://stm.sciencemag.org/content/scitransmed/7/293/293ra104.full
http://stm.sciencemag.org/content/scitransmed/7/302/302ra133.full
http://science.sciencemag.org/content/sci/339/6127/1546.full
http://science.sciencemag.org/content/sci/339/6127/1559.full
http://science.sciencemag.org/content/sci/339/6127/1563.full
http://science.sciencemag.org/content/sci/339/6127/1539.full
http://science.sciencemag.org/content/sci/339/6127/1543.full
http://science.sciencemag.org/content/sci/339/6127/1567.full
http://science.sciencemag.org/content/sci/339/6127/1540.full
http://science.sciencemag.org/content/sci/339/6127/1493.full

REFERENCES

This article cites 42 articles, 16 of which you can access for free
http://stm.sciencemag.org/content/4/136/136ra68#BIBL

PERMISSIONS

http://www.sciencemag.org/help/reprints-and-permissions

Downloaded from http://stm.sciencemag.org/ by guest on December 15, 2017

Use of this article is subject to the Terms of Service

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Translational Medicine* is a registered trademark of AAAS.

www.sciencetranslationalmedicine.org/cgi/content/full/4/136/136ra68/DC1

Supplementary Materials for

Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA

Tim Forshew, Muhammed Murtaza, Christine Parkinson, Davina Gale, Dana W. Y. Tsui, Fiona Kaper, Sarah-Jane Dawson, Anna M. Piskorz, Mercedes Jimenez-Linan, David Bentley, James Hadfield, Andrew P. May, Carlos Caldas, James D. Brenton,* Nitzan Rosenfeld*

*To whom correspondence should be addressed. E-mail: nitzan.rosenfeld@cancer.org.uk (N.R.); james.brenton@cancer.org.uk (J.D.B.)

Published 30 May 2012, *Sci. Transl. Med.* **4**, 136ra68 (2012)
DOI: 10.1126/scitranslmed.3003726

The PDF file includes:

Methods
Fig. S1. PCR strategy and primer design.
Fig. S2. Sanger traces for mutations identified by tagged-amplicon sequencing.
Fig. S3. Background frequencies and detection limits for base substitutions.
Fig. S4. Replicate dilute Sanger sequencing of a mutation identified in plasma.
Table S1. Target-specific primers.
Table S2. Unique sequencing barcodes.
Table S3. Mutations identified in FFPE samples.
Table S4. SNPs identified in circulating DNA from two plasma control samples.
Table S5. Frequency of SNP alleles in dilution series of DNA from control plasma.
Table S6. Additional data for Table 2 for mutations identified in plasma samples.
Table S7. Mutations and amplicons studied in one breast cancer patient.

SUPPLEMENTARY METHODS

Sanger sequencing

Coding sequences of the *TP53* gene (exons 2 to 11) were amplified as described previously (*36*) with the following modifications: PCR reactions were performed in 25 µl, universal primers M13 forward and M13 reverse were tagged onto both primers pairs and used to sequence in both the forward and reverse directions. To sequence exon 7, TP53-7F (CAGGTCTCCCCAAGGCGCAC) was used owing to a poly A tract downstream of the exon 7 forward primer described in (*36*): CATCCTGGCTAACGGTGAAAC. PCR products were sequenced on an ABI 3730 (Applied Biosystem). Mutational analysis was performed using Mutation Surveyor Software version 3.97 (SoftGenetics), using default settings (mutation score=5.00, mutation height=500, overlapping factor=0.20, dropping factor=0.20, SnRatio=1.00, mobility shift=imbedded algorithm). All exons were sequenced for 26 samples. For 12 samples, sequencing was performed for selected exons only.

Digital PCR

The digital PCR method and the use of the BioMark system for digital PCR analysis have been described (*7*, *37*). Samples were mixed with sequence-specific TaqMan probes, Master Mix, and 20× GE Sample Loading Reagent (Fluidigm). This mixture was then added to the 12.765 Digital Array Chip (Fluidigm) where it was partitioned into 765 separate PCR reactions. These were PCR cycled on the Fluidigm BioMark and all positive reaction chambers counted. Amplifiable copies were calculated from the number of positive wells using a Poisson correction.

Estimation of input DNA amounts

DNA amount was calculated as total amplifiable copies, including both mutant and germline alleles, as determined by digital PCR (median amplicon size 84 bp, range 58-177 bp), multiplied by the estimated amount of DNA in 1 haploid genome (3.3 pg). Amplifiable copies in control healthy plasma samples were estimated using the *FTH1* TaqMan assay Hs01694011_s1 (Life Technologies), which has an amplicon size of 180 bp.

Dilution of plasma DNA samples to determine allele quantification accuracy

After estimation of amplifiable copy numbers using digital PCR, both samples were diluted to 100 amplifiable copies per µl. A 5-µl mix was prepared containing 500 amplifiable copies, of which 400 copies were contributed by plasmaDna1 and 100 copies were contributed by plasmaDna2 (table s4). This was serially diluted 2.5× 6 times (reaching 244× dilution) in plasmaDna2. Each of the 7 mixes were then diluted 2.5× in water a further six times until we had an expected 2 total amplifiable copies of DNA. Twenty-eight of these mixes (table s5) were amplified using the described primer set (table S1) in triplicate, barcoded, and sequenced. Minor allele frequencies expected for the dilutions can be calculated using heterozygosity status for each SNP in either sample, an example that applies to 3 of the 5 SNPs (rs1800899, rs1050171 and rs10241451) we studied is given in table s5.

Replicate dilute Sanger sequencing

We performed replicate dilute Sanger Sequencing using the primers EGFR_A_F (CAGCAGGGTCTTCTCTGTTTCA) and EGFR_A_R (GGTGTTTTCACCAGTACGTTCCT). A PCR mastermix was made containing 0.05 U/µl FastStart High Fidelity Enzyme Blend (Roche), 1× FastStart High Fidelity Enzyme Buffer, 200 µM each of dNTPs, 4.5 mM $MgCl_2$, 5% DMSO, 50 nM of each primer, and 1X EvaGreen DNA binding dye (Biotium). Plasma DNA was quantified using digital PCR and diluted such that when partitioned in a 384-well plate, each reaction well would contain on average less than 1 amplifiable template. Forty-five cycles of PCR (50°C 2 min, 70°C 20 min, 95°C 10 min, 45 cycles of 95°C 15 s, 60°C 30 s, 72°C 1 min) were performed on a 7900HT Fast Real-Time PCR machine (Applied Biosystems).

TAm-Seq

Primer design for TAm-Seq

Target-specific primers were designed with universal primer sequences (termed CS1 and CS2) appended at the 5'-end (table S1). Primers were designed using Primer3 with a Tm range of 58-62°C, allowing homopolymer stretches no longer than 3 and keeping primer %GC less than 65%. Primer designs were screened against the hg19 reference genome to prevent cross-product generation, and were screened to prevent primer-dimer interactions and formation of intramolecular secondary structure. Primers were tested by amplifying from 10 ng human genomic DNA (GM17317, Coriell) in 5 µl reaction volumes containing 0.05 U/µl FastStart High Fidelity Enzyme Blend (Roche), 1× FastStart High Fidelity Enzyme Buffer, 200µM each of dNTPs, 4.5 mM MgCl2, 5% DMSO, 50 nM of each primer, and 1× Access Array sample loading solution (Fluidigm) using 35 cycles of amplification: 50°C 2 min, 70°C 20 min, 95°C 10 min, 10 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 2 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 8 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 2 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 8 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 5 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 1 cycle of 72°C for 3 min.

Barcode primers (table S2) comprised either the PE1 or PE2 sequences for Illumina cluster generation, a 10-bp barcode, followed by either CS1 or CS2 adaptor sequences. For example, one pair of barcode primers comprised 5'-PE1-CS1-3'+ 5'-PE2-BC-CS2-3', the second pair comprised 5'-PE1-CS2-3'+ 5'-PE2-BC-CS1-3'.

The sequencing reagent for read1 contained custom sequencing primers designed to anneal to both CS1 and CS2. The sequencing reagent for the index read contained custom sequencing primers designed to anneal to the reverse complements of CS1 and CS2.

Preamplification for TAm-Seq

Preamplification reactions were carried out in 10-µl reaction volumes containing 50 nM of each forward and reverse target-specific primer. Target DNA (1 to 5 µl) was added to a mastermix containing 0.5 U FastStart High Fidelity Enzyme Blend (Roche), 1× FastStart High Fidelity Enzyme Buffer, 200 µM each of dNTPs, 4.5 mM $MgCl_2$, and 5% DMSO. Reactions were subjected to 15 cycles of amplification (95°C 10 min, 15 cycles of 95°C 15 s, 60°C 4 min). Following preamplification, 4 µl Exo-SAP-it (Affymetrix) was added to each reaction and incubated for 15 min at 37°C, then for 15 min at 80°C. The preamplified samples were diluted 5-fold in PCR-grade water prior to amplification on the Access Array IFC system.

Target-specific amplification on the Access Array microfluidic system

Individual primer pairs were loaded into the primer inlets of the Access Array IFC (Fluidigm) at a final concentration of 1 µM (20× the final concentration of 50 nM required for the PCR reaction) to account for the volume ratio of sample and primer chambers (20:1) in the Access Array IFC. One µl of each preamplified sample was added to 4 µl pre-sample mix containing 0.05 U/µl FastStart High Fidelity Enzyme Blend (Roche), 1× FastStart High Fidelity Enzyme Buffer, 200 µM each of dNTPs, 4.5 mM $MgCl_2$, 5% DMSO, and 1× Access Array sample loading solution (Fluidigm). Samples and primers were loaded into the IFC using an IFC-AX controller (Fluidigm). The volume of each sample chamber within the Integrated Fluidic Circuit (IFC) was 33 nl, containing 0.7% of the input reaction volume (5 µl). The IFC was then subjected to thermal cycling using a Biomark system (Fluidigm): 35 cycles of amplification (50°C 2 min, 70°C 20 min, 95°C 10 min, 10 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 2 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 8 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 2 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 8 cycles of 95°C 15 s, 60°C 30 s, 72°C 60 s, 5 cycles of 95°C 15 s, 80°C 30 s, 60°C 30 s, 72°C 60 s, 1 cycle of 72°C for 3 min). Harvesting solution (0.05% Tween-20) was loaded onto the IFC prior to harvesting on an IFC-AX controller. One µl of harvested product was then transferred to a clean PCR plate containing 99 µl PCR-free water.

Sequencing adaptor and barcode primer addition

For each sample, 1 µl of the 100-fold diluted PCR products was added to each of two PCR plates containing 15 µl pre-sample mastermix containing 0.05 U/µl FastStart High Fidelity Enzyme Blend (Roche), 1× FastStart High Fidelity Enzyme Buffer, 200 µM each of dNTPs, 4.5 mM $MgCl_2$, and 5% DMSO. In the first plate, 4 µl of one pair of primers containing an individual 10-base barcode (BC) sequence, and sequence tags for reading in one direction (PE1-BC-CS1 + PE2-CS2) were added to each well. In the second plate, 4 µl of primers containing (PE1-BC-CS2 + PE2-CS1) were added to each well. The corresponding wells in both plates contained primers with the same barcode sequence (e.g. plate 1, well A1 = Barcode FLD0001, plate 2, well A1 = barcode FLD0001). Reaction products in plates were amplified for 15 cycles: 95°C 10 min, 15 cycles of 95°C 15 s, 60°C 30 s, 72°C 4 min, 1 cycle of 72°C for 3 min.

Quantification and clean up of DNA library

After PCR products were barcoded, they were analyzed using Agilent 2100 BioAnalyzer to ensure expected insert size (~180 bp) was obtained. They were then pooled together and purified using AMPure XP beads using a bead to amplicon ratio of 1.8:1. The library was quantified by Agilent BioAnalyzer and subjected to Illumina cluster generation. Single-end sequencing of 100 bases was performed on an Illumina GAIIx sequencer followed by a 10-base indexing (barcode) read, using custom sequencing primers targeted to the CS1 and CS2 tags for both read1 and the index read according to manufacturer's recommendations.

Analysis of sequencing data

De-multiplexing and alignment. Reads generated by the primary Illumina pipeline were demultiplexed using a known list of barcodes (table S2), allowing one base mismatch out of 10 bases. Each set of reads was aligned independently to the hg19 reference genome using bwa-short in the single-end mode. Using expected genomic positions, each set of aligned reads was separated further into its constituent amplicons. A pileup was generated for each amplicon using samtools v1.12a. Using a base quality and a mapping quality cut-off of 30, observed frequencies of non-reference alleles for every sequenced locus across all amplicons and barcodes were calculated.

Mutation identification. For each locus and base, the distribution of non-reference background allele frequencies/reads was modeled as a Normal distribution and as a Poisson distribution, fitting to a Normal distribution and using the mean number of observed non-reference reads in a set of representative barcodes as the parameter for the Poisson distribution. For each barcode, the probability of obtaining the observed frequency/number of reads (or greater) was calculated in both models, and the lower value was retained. Putative substitutions that passed a probability cut-off (confidence margin) of 0.9995 were kept for further analysis as candidate mutations.

For each barcode, all loci that passed the probability cut-off were ranked by observed frequency, corrected for the median frequency (over all samples) for each locus/base. Known SNPs obtained from the 1000 Genomes project and regions covering amplification primers were discarded, leaving 17,934 possible base calls per sample. Indels are not analyzed in the current calling algorithm. A point mutation was called in a sample if it ranked among the 6 loci with highest non-reference frequency in both duplicates, and was represented by at least 10 reads in each. If a mutation was called, it was also discarded from the pool and the process was repeated by testing the 6 highest remaining frequencies. Mutations that occurred at low frequency may have fallen outside the list of top-ranked loci, especially for more noisy samples, and may have therefore been missed. For each sample, we estimated the threshold for mutation calling by the highest background-corrected non-reference allele frequency that did not get into the top-ranked list and may have been missed by the calling algorithm. For mutations called in multiple (overlapping) amplicons, the amplicon showing best concordance between duplicates was retained.

For the sequencing data obtained from a mix of FFPE samples, we performed an additional lower-stringency analysis using a probability cut-off of 0.9995 keeping the 12 loci with highest non-reference frequency to enhance sensitivity. For the data obtained from plasma sequencing, we performed an additional higher-stringency analysis using a probability cut-off of 0.9999 keeping the 6 loci with highest non-reference frequency to enhance sensitivity.

Mutation detection. When applying TAm-Seq to measure a pre-defined mutation (as opposed to screening nearly 18,000 possible substitutions), the frequency of the mutant allele can be read out directly from the data at the desired locus. We used confidence margins of ~0.95 using a Normal distribution model for each examined substitution, and requiring a minimum of 10 representative reads, in at least one amplified library per sample. When a mutation was positively detected, its allele frequency was estimated by averaging the allele frequencies in all replicates after subtracting substitution-specific background frequencies.

SUPPLEMENTARY FIGURES

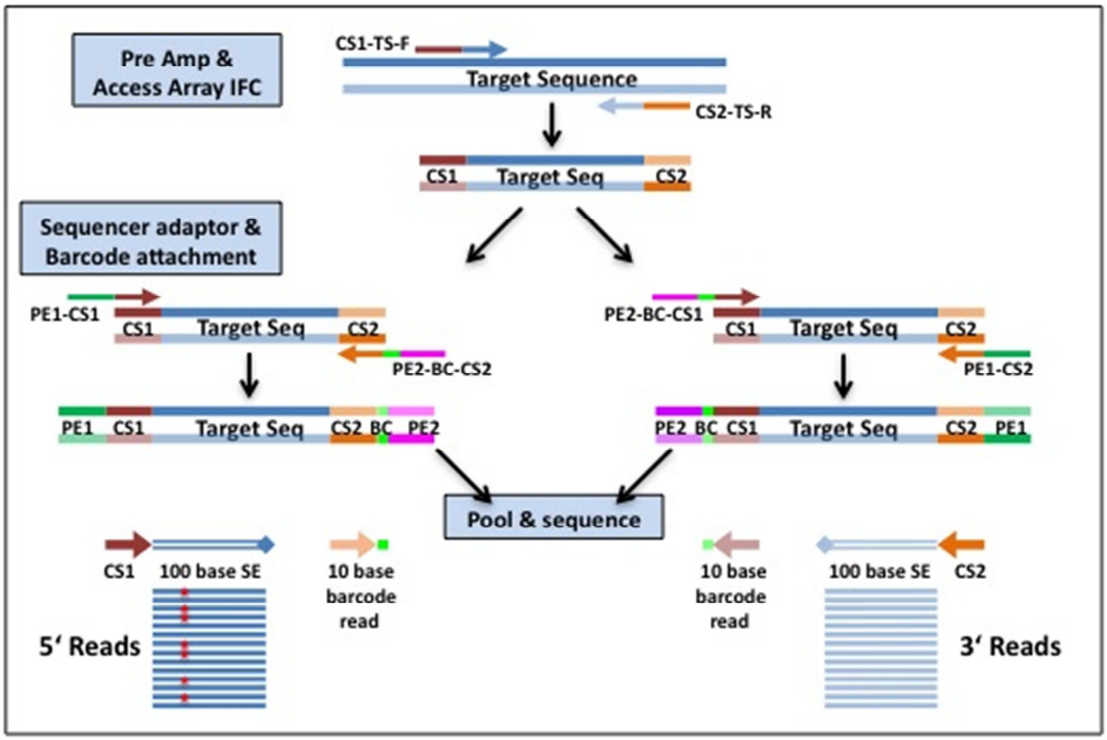

Figure S1: PCR strategy and primer design. Target-specific primers were synthesised with common adaptor sequences at their 5' ends (CS1 and CS2). These were used during the preamplification and single-plex stages to amplify the selected regions. Sequencing platform-specific adaptors and unique sample barcodes were attached during an additional round of PCR. The barcoding primers consist of Illumina sequencer adaptors (PE1 and PE2) at their 3' end, a unique 10 base barcode, and the common CS1 and CS2. Permutations of the universal tags and the sequencing adaptors enabled construction of bidirectional barcoded sequencing libraries. Using custom sequencing primers directed to the two tag sequences, sequence reads representing both strands of the amplicon pool were obtained from a single-end run. Both combinations were used in order to barcode and read amplicons in both directions. All samples were then pooled, cleaned, and then sequenced.

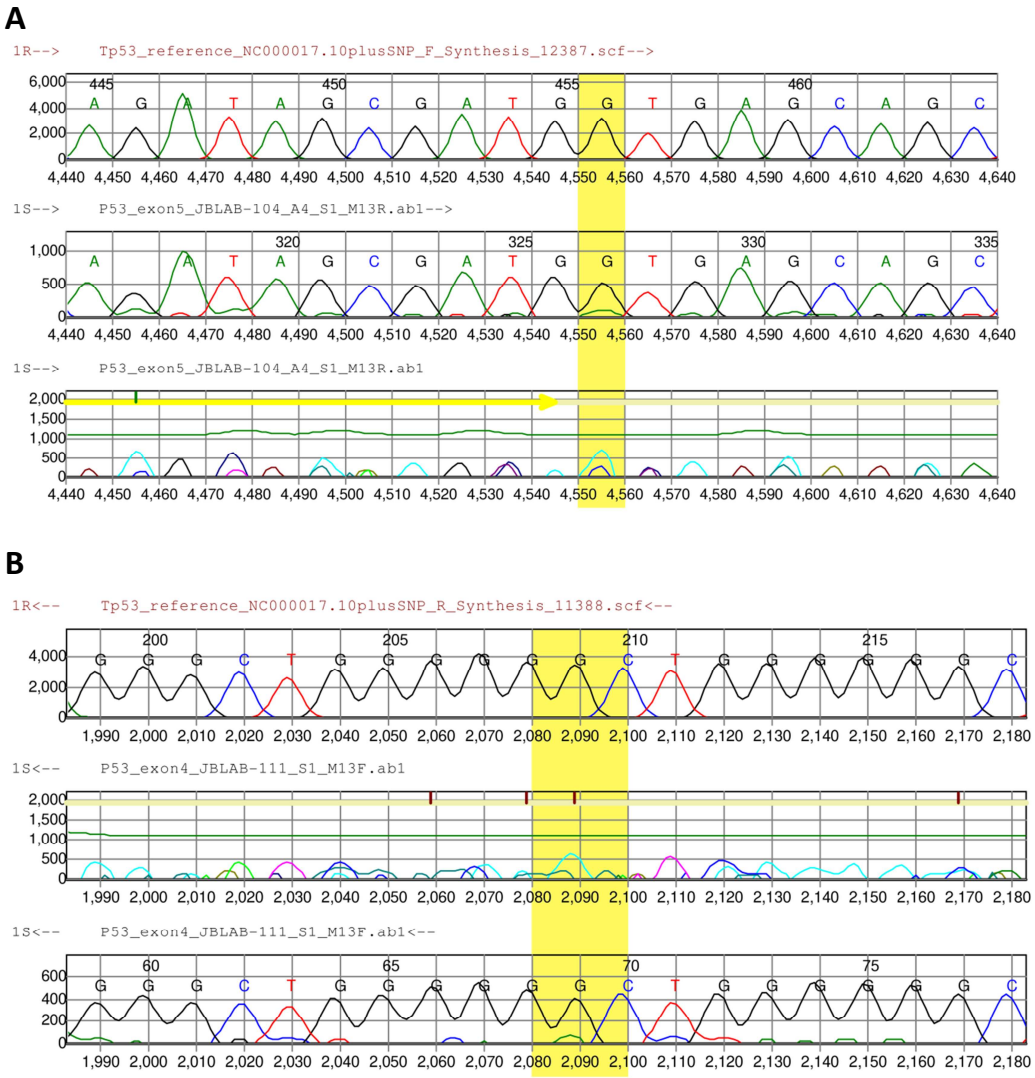

Figure S2: Sanger traces for mutations identified by tagged-amplicon sequencing. Mutational analysis of Sanger traces was performed by Mutation Surveyor Software version 3.97 (SoftGenetics) using default settings. The *TP53* gene was on the reverse strand, therefore mutations seen in the sequencing traces were the reverse complement of the mutations as indicated by the hg19 coordinates. (**A**) Sanger sequencing of sample #104. Highlighted locus shows a mutation that was identified by TAm-Seq (17:7578370C>T), but not Sanger sequencing. The Sanger sequencing trace for the sample (middle panel) shows a weak signal above the reference (top panel), but this was at background noise levels and did not pass detection thresholds (lower panel). (**B**) Sanger sequencing of sample #111. Highlighted locus shows a mutation that was identified by TAm-Seq (17:7579642C>T), but not Sanger sequencing. The Sanger sequencing trace (lower panel) showed a weak signal above the reference (top panel), but this did not pass detection thresholds (middle panel).

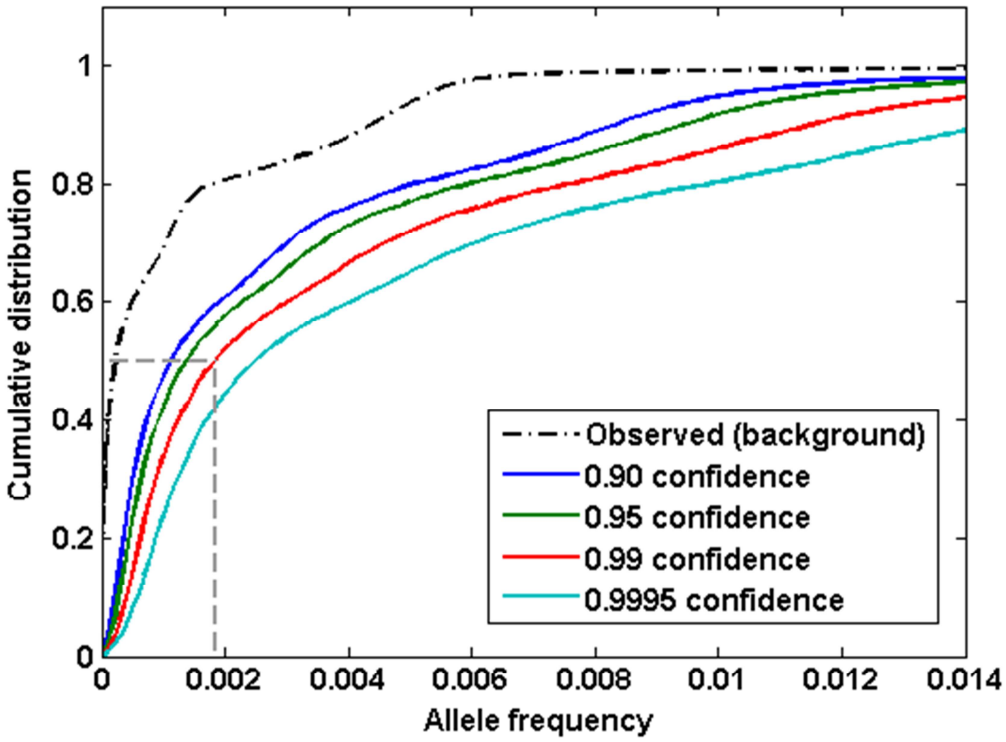

Figure S3: Background frequencies and detection limits for base substitutions. For each one of the possible base substitutions in the sequenced amplicons, we calculated the mean background read rate in the 124 barcoded libraries generated from the set of 62 plasma samples (cumulative distribution shown in black), and the standard deviation. Colored lines show the cumulative distribution of frequencies that would exceed the indicated confidence margins, using a Normal distribution model for each base substitution, given its measured mean and standard deviation. Substitutions that were covered by more than one amplicon or read direction were included multiple times, once for each amplicon/direction in which they were sequenced. Substitutions that had zero reads in all barcodes had undefined confidence limits and were omitted (n = 589). Dashed grey lines indicate the median value for 0.99 confidence margin.

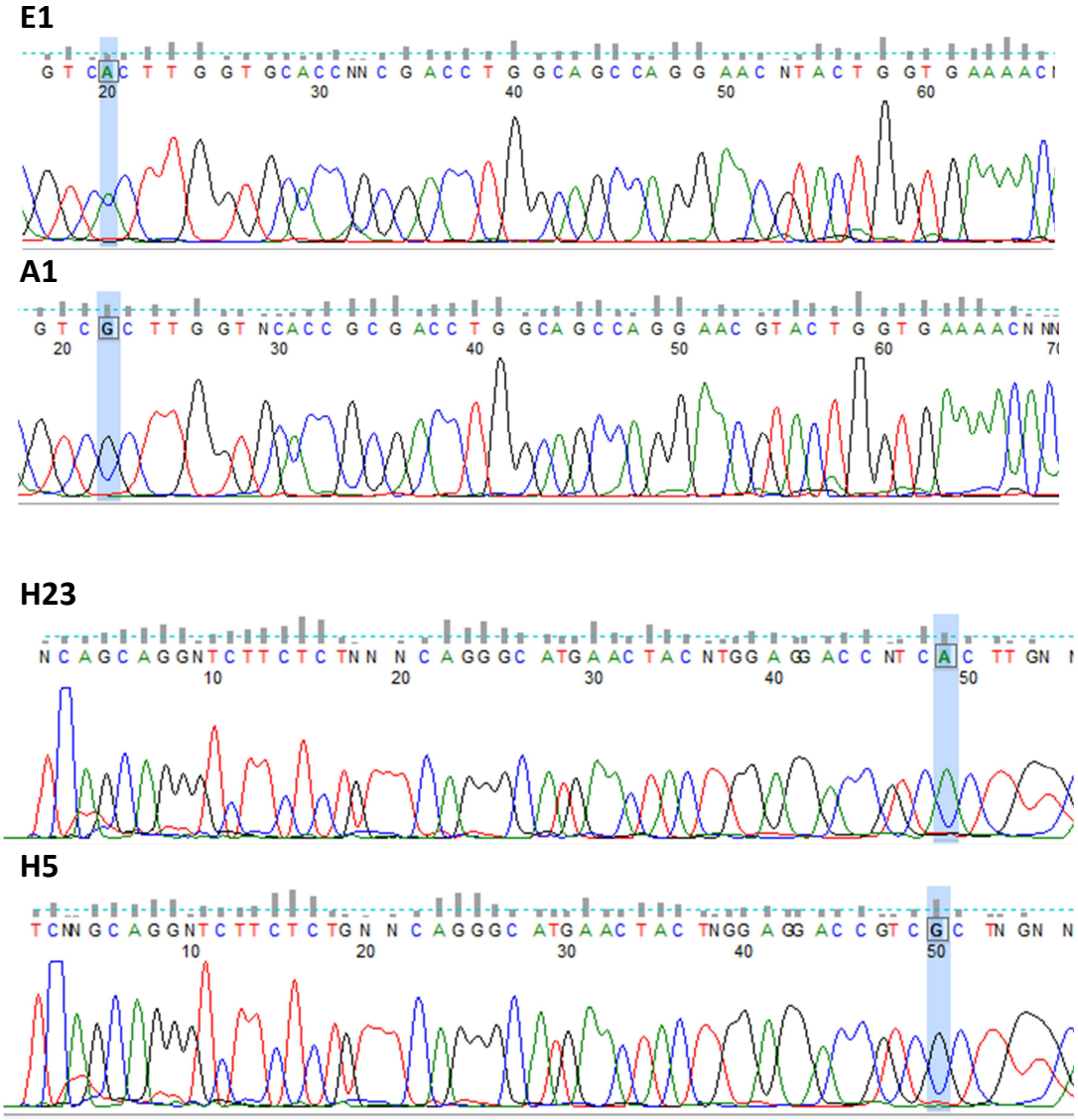

Figure S4: Replicate dilute Sanger sequencing of a mutation identified in plasma. Replicate Sanger sequencing of highly dilute template was used to validate the *EGFR* mutation that was identified *de novo* in plasma and was not found in the corresponding tumour sample from the same patient (Patient #27, Table 1). The panels show examples of mutant and wild type reads in either direction, with the mutated base (chr7:55259437G>A, genome build hg19 coordinates) highlighted in blue. E1 is an example of mutant forward read, and A1 is a matched wild-type. H23 is an example of mutant reverse read, and H5 a matched wild-type.

SUPPLEMENTARY TABLES

Table S1: Target-specific primers. Each target-specific primer consists of a universal 5' end and a target-specific 3' end as follows:

5'-ACACTGACGACATGGTTCTACA-[Target Specific-Forward]-3'.

5'-TACGGTAGCAGAGACTTGGTCT-[Target Specific-Reverse]-3'

Gene name and Amplicon ID	Target Specific Primer – Forward	Target Specific Primer- Reverse	Chr	Amplicon Start	Amplicon End
PIK3CA_E00001077674 *	CAGAGGGGAAAAATATGACAAA	AACAGAGAATCTCCATTTTAGCAC	3	178935943	178936150
PIK3CA_E00001139987	TGAGCAAGAGGCTTTGGAGT	GGTCTTTGCCTGCTGAGAGT	3	178952038	178952227
EGFR_Exon19	TCACAATTGCCAGTTAACGTCT	CCACACAGCAAAGCAGAAAC	7	55242373	55242537
EGFR_E00001601336_1	GCGTCTTCACCTGGAAGGG	CCGGACATAGTCCAGGAGG	7	55248902	55249111
EGFR_E00001601336_2	GCGTGGACAACCCCCAC	GGCTCCTTATCTCCCCTCC	7	55249005	55249213
EGFR_E00001681524_1	GGATGCAGAGCTTCTTCCCA	TTCTCTTCCGCACCCAG	7	55259352	55259542
EGFR_E00001681524_2	GGTCTTCTCTGTTTCAGGGCAT	GCTGACCTAAAGCCACCTCC	7	55259395	55259591
EGFR_E00001631695_1	GTGTCACTCGTAATTAGGTCCA	GGCCTCAGTACAAACTCATTAGC	7	55260366	55260575
EGFR_E00001779947_1	TGTTCATTCATGATCCCACTGC	CCACCAGTCACTCACACTTG	7	55266362	55266571
EGFR_E00001779947_2	TCCCTGCCAGCGAGAT	AGGGATGCAAAGGCCTCA	7	55266461	55266641
EGFR_E00001790701_1	GCCTTCTTTAAGCAATGCCATCTTTAT	CAATGGAAGCACAGACTGCAA	7	55267932	55268134
EGFR_E00001801208_1	CCCCTGCTCCTATAGCCAA	ATGAGGTACTCGTCGGCATC	7	55268806	55268987
EGFR_E00001801208_2	ACTTCTACCGTGCCCTGA	GTTCAAATGAGTAGACACAGCTT	7	55268921	55269101
EGFR_E00001773562_1	TACCCTCCATGAGGCACAC	GGAGAGCTGTAAATTCTGGCTT	7	55269336	55269516
BRAF_E00002324725	TCATAATGCTTGCTCTGATAGGA	CTGATGGGACCCACTCCAT	7	140453108	140453256
PTEN_E00001456562_1	GCAGCTTCTGCCATCTCTCT	TCCGTCTACTCCCACGTTCT	10	89624175	89624372
PTEN_E00001156351_1	TGCTGCATATTTCAGATATTTCTTTCCTTA	ATGAAAACACAACATGAATATAAACATCAAT	10	89653738	89653927
PTEN_E00001156344_1	AATCTGTCTTTTGGTTTTTCTTGATAGT	AATAGTTGTTTTAGAAGATATTTGCAAGC	10	89685172	89685367
PTEN_E00001156337_4	TATATCACTTTTAAACTTTTCTTTTAGTTGTGC	CTCGATAATCTGGATGACTCATTATTGTT	10	89690776	89690940
PTEN_E00001156330_1	TTCTTATTCTGAGGTTATCTTTTTACCAC	TCATTACACCAGTTCGTCCCT	10	89692739	89692919

PTEN_E00001156330_3	TGACCAATGGCTAAGTGAAGATGA	TCCAGGAAGAGGAAAGGAAAAACA	10	89692840	89693048
PTEN_E00001156327_1	TCTTAAATGGCTACGACCCAG	TCCAGATGATTCTTTAACAGGTAGC	10	89711775	89711942
PTEN_E00001156327_4	CAGTCAGAGGCGCTATGTGT	TCTAGATATGGTTAAGAAAACTGTTCCA	10	89711889	89712077
PTEN_E00001156321_1	TGACAGTTTGACAGTTAAAGGCAT	CACACACAGGTAACGGCTGA	10	89717547	89717726
PTEN_E00001156321_2	TGTGGTCTGCCAGCTAAAGG	TCTCCCAATGAAAGTAAAGTACAAACC	10	89717620	89717802
PTEN_E00001156321_4m	TCCACAAACAGAACAAGATGCT	GGCCTTTTCCTTCAAACAGGATT	10	89717748	89717956
PTEN_E00001156315_1m	GCAACAGATAACTCAGATTGCCTT	GTTTCCTCTGGTCCTGGTATGA	10	89720457	89720706
PTEN_E00001156315_5 *	AGGACAAAATGTTTCACTTTTGGGTAA	ACTAGATATTCCTTGTCATTATCTGCAC	10	89720649	89720799
PTEN_E00001156315_7	CCTCAGAAAAAGTAGAAAATGGAAGTC	ACAAGTCAACAACCCCCACA	10	89720706	89720915
PTEN_E00001456541_1	AGATGAGTCATATTTGTGGGTTTTCA	TCTGGATCAGAGTCAGTGGT	10	89724997	89725180
PTEN_E00001456541_2 *	GTAGAGGAGCCGTCAAATCCA	TTCATGGTGTTTTATCCCTCTTGA	10	89725068	89725264
KRAS_E00000936617	GCCTGCTGAAAATGACTGAA	AGAATGGTCCTGCACCAGTAA	12	25398163	25398329
TP53_E00001757276_1	GACCCAAAACCCAAAATGGC	TCCCTGCTTCTGTCTCCTAC	17	7572850	7573030
TP53_E00001728015_1	GGAATCCTATGGCTTTCCAACC	CCCCCTCCTCTGTTGCTG	17	7573859	7574054
TP53_00001404886_13	TCTGTATCAGGCAAAGTCATAGAA	GCCTCAAAGACAATGGCTCC	17	7576584	7576734
TP53_E00001789298_1	AGAAAACGGCATTTTGAGTGT	AAGGGTGCAGTTATGCCTCA	17	7576786	7576983
TP53_E00001789298_2	CTGGTGTTGTTGGGCAGT	ATCTCCGCAAGAAAGGGGAG	17	7576908	7577075
TP53_E00001789298_3	TGTCCTGCTTGCTTACCTCG	GCCTCTTGCTTCTCTTTTCCT	17	7577003	7577187
TP53_E00001665758_1	GGGGTCAGAGGCAAGCAG	CTTGGGCCTGTGTTATCTCC	17	7577432	7577631
TP53_E00001255919_1	GAGAAAGCCCCCCTACTGC	AGCATCTTATCCGAGTGGAAGG	17	7578091	7578274
TP53_E00001255919_3	TCCAAATACTCCACACGCAAA	GCTGCCCCCACCATGAG	17	7578229	7578406
TP53_E00001255919_5	AGCTGCTCACCATCGCTA	CCAACTGGCCAAGACCT	17	7578361	7578525
TP53_E00001255919_6	TGTGCTGTGACTGCTTGTAG	TGCCCTGACTTTCAACTCTGT	17	7578425	7578594
TP53_E00001612188_1	ATACGGCCAGGCATTGAAGT	CCTCCTGGCCCCTGTC	17	7579260	7579421
TP53_E00001612188_2	GGAAACCGTAGCTGCCCTG	AAGACCCAGGTCCAGATGAA	17	7579359	7579520
TP53_E00002359670	CAGCCTCTGGCATTCTGG	CCTGGTCCTCTGACTGCTCT	17	7579479	7579626
TP53_E00002419584	TCAAATCATCCATTGCTTGG	CCATGGGACTGACTTTCTGC	17	7579557	7579754
TP53_E00001596491_1	TTTCGCTTCCCACAGGTCTC	CAGCCAGACTGCCTTCCG	17	7579758	7579940

*Each of these 3 primer pairs potentially amplifies an additional non-target region, owing to large regions of homology and the short amplicon size. This was taken into account in data analysis.

Table S2: Unique sequencing barcodes. Platform-specific adaptors and barcodes are attached through PCR following the single-plex amplification step. The primers consisted of the PE1 and PE2 sequences for Illumina cluster generation, a 10-bp barcode, and the CS1 and CS2 adaptors, used in pairs: PE1-CS1 with PE2-BC-CS2, and PE1-CS2 with PE2-BC-CS1.

PE1: AATGATACGGCGACCACCGAGATCT.

PE2: CAAGCAGAAGACGGCATACGAGAT.

Barcode name	Barcode sequence				
FLD0001	GTATCGTCGT	FLD0060	GCACGTAGCT	FLD0121	CAGAGCTAGT
FLD0002	GTGTATGCGT	FLD0061	TCACGCTATG	FLD0122	CGCAGAGCAT
FLD0003	TGCTCGTAGT	FLD0062	CGTACTACGT	FLD0123	TGTACAGCGA
FLD0004	GTCGTCGTCT	FLD0063	CAGCTGAGTA	FLD0124	ACGTCAGTAT
FLD0005	GTGCGTGTGT	FLD0064	GAGATCAGTC	FLD0125	TCACAGCATA
FLD0006	GCGTCGTGTA	FLD0065	TACTGAGCTG	FLD0126	ACTGCGTGTC
FLD0007	GTCGTGTACT	FLD0066	TAGTAGCGCG	FLD0127	CGATCGACTG
FLD0008	GATGTAGCGT	FLD0067	GACGTCTGCT	FLD0128	GCGAGATGTA
FLD0009	GAGTGATCGT	FLD0068	GTACTCGCGA	FLD0129	CTGATGCAGA
FLD0010	CGCTATCAGT	FLD0069	TCTGAGCGCA	FLD0130	GTGACGTACG
FLD0011	CGCTGTAGTC	FLD0070	TAGACGTGCT	FLD0131	CGACGCTGAT
FLD0012	GCTAGTGAGT	FLD0071	GTGACTCGTC	FLD0132	CTACGATCAG
FLD0013	GAGCTAGTGA	FLD0072	TCGAGTAGCG	FLD0133	GCACTAGACA
FLD0014	CGTGCTGTCA	FLD0073	CGTATGATGT	FLD0134	CTAGCAGATG
FLD0015	GATCGTCTCT	FLD0074	TAGTCTGTCA	FLD0135	CATGATACGC
FLD0016	GTGCTGTCGT	FLD0075	TGTCTCTATC	FLD0136	GCAGCTGTCA
FLD0017	TGAGCGTGCT	FLD0076	CTAGAGTATC	FLD0137	ACGTATCATC
FLD0018	CATGTCGTCA	FLD0077	TATCATGTGC	FLD0138	AGTATCGTAC
FLD0019	TCAGTGTCTC	FLD0078	CATGAGTGTA	FLD0139	GATACACTGA
FLD0020	GTGCTCATGT	FLD0079	TGTCGTCATA	FLD0140	GACTAGTCAG
FLD0021	CGTATCTCGA	FLD0080	TATCTCATGC	FLD0141	GATGACTACG
FLD0022	GTCATGCGTC	FLD0081	TGTGTCACTA	FLD0142	CAGAGAGTCA
FLD0023	CTATGCGATC	FLD0082	TATCGATGCT	FLD0143	TCGATCGACA
FLD0024	TGCTATGCTG	FLD0083	TAGAGTCTGT	FLD0144	ACTGATGTAG
FLD0025	TGTGTGCATG	FLD0084	CATGCATCAT	FLD0145	ACTCGATAGT
FLD0026	GAGTGTCACT	FLD0085	TGATCAGTCA	FLD0146	GACGATCGCA
FLD0027	CTAGTCTCGT	FLD0086	CGTCTATGAT	FLD0147	TCATCATGCG
FLD0028	GAGTGCATCT	FLD0087	GTGATACTGA	FLD0148	ACATGTCTGA
FLD0029	TGCGTAGTCG	FLD0088	CTAGATCTGA	FLD0149	AGTCATCGCA
FLD0030	CTGTGTCGTC	FLD0089	TATCAGTCTG	FLD0150	TAGCATACAG
FLD0031	CTGTAGTGCG	FLD0090	TCAGATGCTA	FLD0151	AGAGTCGCGT
FLD0032	GTGCGCTAGT	FLD0091	TATGTACGTG	FLD0152	TCTACGACAT
FLD0033	TGTGCTCGCA	FLD0092	CTATACAGTG	FLD0153	CACGAGATGA
FLD0034	GATGCGAGCT	FLD0093	TGATACTCTG	FLD0154	ACGCACATAT
FLD0035	CTGTACGTGA	FLD0094	TCAGCGATAT	FLD0155	ACGTGCTCTG
FLD0036	GCGATGATGA	FLD0095	CTACTGATGA	FLD0156	ACGATCACAT
FLD0037	TGTCGAGTCA	FLD0096	GTAGTACACA	FLD0157	AGTGTACTCA
FLD0038	GTCTACTGTC	FLD0097	TGCTACATCA	FLD0158	TGATGTATGT
FLD0039	CAGTCAGAGT	FLD0098	AGTGTGTCTA	FLD0159	GATATATGTC
FLD0040	CGCAGTCTAT	FLD0099	TCATATCGCG	FLD0160	TAGTACTAGA
FLD0041	GTATGAGCAC	FLD0100	TACGTATAGC	FLD0161	TATAGAGATC
FLD0042	CGAGTGCTGT	FLD0101	CAGCTATAGC	FLD0162	TCGATATCTA
FLD0043	TATAGCACGC	FLD0102	TCGATGCGCT	FLD0163	TACATGATAG
FLD0044	TCATGCGCGA	FLD0103	GCACGCGTAT	FLD0164	TGAGATCATA
FLD0045	TATGCGCTGC	FLD0104	GCAGTATGCG	FLD0165	CTACATACTA
FLD0046	TCTCTGTGCA	FLD0105	TGATAGAGAG	FLD0166	ATCAGTGTAT
FLD0047	CTATCGCGTG	FLD0106	GCTACTAGCG	FLD0167	ATCATATCTC
FLD0048	TACGCTGCTG	FLD0107	TGCGAGACGT	FLD0168	AGTAGATCAT
FLD0049	CTGCATGATC	FLD0108	CGATGACAGA	FLD0169	ACATAGTATC
FLD0050	CGCGTATCAT	FLD0109	GACTCATGCT	FLD0170	ATGTATAGTC
FLD0051	GTATCTCTCG	FLD0110	GTCTGATACG	FLD0171	ACAGTCATAT
FLD0052	GCTCATATGC	FLD0111	ACTAGCTGTC	FLD0172	ACATATACGT
FLD0053	CACTATGTCG	FLD0112	GCGTAGACGA	FLD0173	AGCATCTATA
FLD0054	TAGCGCGTAG	FLD0113	CTCAGCAGTG	FLD0174	AGACTATATC
FLD0055	CGTCACAGTA	FLD0114	CAGTCTACAT	FLD0175	CAGCATCTAG
FLD0056	TCGCGTGAGA	FLD0115	TACTGCAGCG	FLD0176	CGAGACGACA
FLD0057	TACATCGCTG	FLD0116	TACACAGTAG	FLD0177	ATCACTCATA
FLD0058	GTGAGAGACA	FLD0117	CACATACAGT	FLD0178	AGCTCTGTGA
FLD0059	GACTGTACGT	FLD0118	CACAGTGATG	FLD0179	ATGTCATGCT
		FLD0119	CGAGCTAGCA	FLD0180	GCTGACAGAG
		FLD0120	GAGACTATGC	FLD0181	ATACAGTCTC

FLD0182	CATAGACGTG
FLD0183	AGAGATATCA
FLD0184	ATGCTGCGCT
FLD0185	AGTCAGACGC
FLD0186	ACGATACACT
FLD0187	AGCGAGTATG
FLD0188	ATCGCTACAT
FLD0189	ATGCTAGAGA
FLD0190	AGCAGTACTC
FLD0191	ATCTAGATCA
FLD0192	ATCGCATAGA
FLD0193	TTGTTGCTGT
FLD0194	GTGTGGTTGT
FLD0195	TAGGTGGAAT
FLD0196	TGTAGGTGGA
FLD0197	TTAGTGGTGA
FLD0198	GTGAAGGTAA
FLD0199	TGTTGTGGTA
FLD0200	GTTGATGAGT
FLD0201	GGTCAGTGTA
FLD0202	GTAATGGAGT
FLD0203	CTCGTTATTC
FLD0204	GGAAGTAAGG
FLD0205	CGGTGTGTGT
FLD0206	CGTCTTCTTA
FLD0207	TGTGAATCTC
FLD0208	CTAATCGTGT
FLD0209	CTCTTAGTTC
FLD0210	GGATAGGATC
FLD0211	GGTGTCTTGT
FLD0212	GATGGTTGTA
FLD0213	CCTCGTTGTT
FLD0214	GGTTGGAGTT
FLD0215	TGGTGTCCGT
FLD0216	CGTTAGCGTA
FLD0217	TACTAGGATC
FLD0218	GTCTCAATGT
FLD0219	GATGAGGTAT
FLD0220	GGTGTTAGTG
FLD0221	CATTCTCTGA
FLD0222	CATCTGGAGT
FLD0223	GAATGGAAGA
FLD0224	GGCTGTGATC
FLD0225	TGGTGCTGGA
FLD0226	TATGGTAAGG
FLD0227	GTTCGATTGT
FLD0228	GGTAGAATGA
FLD0229	TTCTCATCGT
FLD0230	CTCAATCGTA
FLD0231	CGCTAATGTA
FLD0232	GCGTCTGAAT
FLD0233	TTCTGTTGCC
FLD0234	TTGTCCTTGC
FLD0235	CCTGTGTAGA
FLD0236	GATAAGAAGG
FLD0237	CAGGTCACAT
FLD0238	GCCATGTCAT
FLD0239	TCTGCCTATA
FLD0240	CTTAGTTCGC
FLD0241	CGTAATGAGC
FLD0242	TTGCTTAGTC
FLD0243	TCTTGTTCAC
FLD0244	GTGGCTTCGT
FLD0245	TGTTCGATAG
FLD0246	TCATTCAGTG
FLD0247	GTGGAGAGCT
FLD0248	GTAGAAGTGG
FLD0249	TGGAGCATGT
FLD0250	GAAGGAGATA
FLD0251	CGAATGTATG
FLD0252	TCGTGAATGA
FLD0253	GAATAGCTGA
FLD0254	TTGTCACATC
FLD0255	CTGGAGGCTA
FLD0256	TGTCAGCTTA
FLD0257	GTTCTTCGTA
FLD0258	TTACACGTTC
FLD0259	GTAGCCAGTA
FLD0260	TGAGAAGGTA
FLD0261	CCATATGATC
FLD0262	CGATCCTATA
FLD0263	TGACTAGCTT
FLD0264	TAACTCTGCT
FLD0265	TCGAATGTGC
FLD0266	TCGCTGAACA
FLD0267	GCGTTATTGC
FLD0268	GAACTATCAC
FLD0269	TCGAGGTACT
FLD0270	TGCGGATGGT
FLD0271	TTCGAGCTAT
FLD0272	GGTCTGGTGT
FLD0273	CTAAGTCATG
FLD0274	TTGCAGATCA
FLD0275	CTGCGAATGT
FLD0276	CTGTTCTAGC
FLD0277	CACTTGTGTG
FLD0278	TGGATGACAT
FLD0279	GATCCTGAGC
FLD0280	GTCGGTCTGA
FLD0281	TGTTACGATC
FLD0282	GTCTTGGCTC
FLD0283	GGTCGTGCAT
FLD0284	CAGGCTCAGT
FLD0285	TAGCTTCACT
FLD0286	CAGATGTCCT
FLD0287	TTACGCAGTG
FLD0288	TTCGTTCCTG
FLD0289	CACTGCTTGA
FLD0290	TCTAGCGTGG
FLD0291	GCATAATCGC
FLD0292	GTCGTAACAC
FLD0293	GAGATTGCTA
FLD0294	GGACAGATGG
FLD0295	CTTACGTTGC
FLD0296	GTGTTCGGTC
FLD0297	CTCAAGAAGC
FLD0298	TCTCGGATAG
FLD0299	CTCTGGACGA
FLD0300	CGAGCATTGT
FLD0301	CCAAGAAGAA
FLD0302	TCCTTGTTCT
FLD0303	GTAACGATGT
FLD0304	TGGACTCAGA
FLD0305	GGCATCATGC
FLD0306	GTATAACGCT
FLD0307	GCAGATAAGT
FLD0308	GTCGGCTCTA
FLD0309	TTCGATAGCA
FLD0310	GTCTAGCAGG
FLD0311	GGAACACAGG
FLD0312	TGGTTCGCTG
FLD0313	CACATTAGCG
FLD0314	GAAGCGCACT
FLD0315	GCATGCCAGT
FLD0316	GGAGACTGTA
FLD0317	TCGAACTGCA
FLD0318	GAGAGGACAT
FLD0319	GAGCACGGAA
FLD0320	GCTCTAACAT
FLD0321	TGCTGGCTTG
FLD0322	TGCATGGAGC
FLD0323	GTACTAAGAG
FLD0324	GAAGTCAAGC
FLD0325	GCGCATTATG
FLD0326	GTCCAGACAT
FLD0327	GAGACCTCTA
FLD0328	TTGCACTCAG
FLD0329	TGCGGCGATA
FLD0330	AGTTGCTAGT
FLD0331	AGGATTGAGG
FLD0332	CCAGAACAGA
FLD0333	CGTCAAGCAT
FLD0334	TTGTCGAGAC
FLD0335	GACAGGTGAC
FLD0336	CTGACAAGTG
FLD0337	CACGAAGAGC
FLD0338	CATACCTGAT
FLD0339	GACGTGCTTC
FLD0340	ATTGTGGAGT
FLD0341	TCTGGTCTCA
FLD0342	AGGTAAGAGG
FLD0343	TCCTGACAGA
FLD0344	GCACTGTTGC
FLD0345	ACCATGAGTC
FLD0346	AATGCAGTGT
FLD0347	ATATGGTGGA
FLD0348	ACTCAGTTAC
FLD0349	AAGTGCGATG
FLD0350	CCACAGAGTG
FLD0351	AGTGGTGATC
FLD0352	ACTTCTTAGC
FLD0353	GCCACATATA
FLD0354	ACGCAGGAGT
FLD0355	AATATGCTGC
FLD0356	AAGCGTAGAA
FLD0357	GACAGCAAGC
FLD0358	CTGACCGAGA
FLD0359	CGCGACTTGT
FLD0360	CATCAACATG
FLD0361	TGGCTACGCT
FLD0362	ACGCGGACTA
FLD0363	AGAGGTCGGA
FLD0364	AATCGAGCGT
FLD0365	AAGTACACTC
FLD0366	AGCTGAATGA
FLD0367	ATGCCTATCA
FLD0368	ACTGTAGGAC
FLD0369	ATAGCCGTGT
FLD0370	TCACGACGAA
FLD0371	ATCTGTCCAT
FLD0372	ACTTAGAGAG
FLD0373	AGTGGCAGGT
FLD0374	ATGAGGTCGT
FLD0375	AGGAGAAGGA
FLD0376	ACAACTGCAA
FLD0377	ATTAGCGAGT
FLD0378	ACAACGAACA
FLD0379	AGAGCGCCAA
FLD0380	AGGTAGCTCA
FLD0381	AACGCCAAGA
FLD0382	AAGGTATGAG
FLD0383	ATGGAGCACT
FLD0384	ACGGTGCTAG

Table S3: Mutations identified in FFPE samples. Diagnosis: type (FT, fallopian tube; O, ovarian; PP, primary peritoneal), histiotype (END, endometrioid, grade III; HGSOC, high-grade serous ovarian carcinoma; S/CC, mixed serous and clear cell; S/END, mixed serous papillary and endometrioid, high grade), and stage. FFPE source: sample type (B, biopsy; IDS, interval-debulking surgery; PS, primary surgery), sample tissue (FT, fallopian tube; O, ovary; OM, omentum; P, peritoneum; SC, sigmoid colon; U, uterus). Cellularity: estimated from H&E-stained slide of the FFPE block. Identified by Sanger sequencing: I, identified; FAIL, samples where one duplicate had a technical fault; Indel, indel identified (indels were not analyzed by the calling algorithm); M, missed; U, unsequenced region or sample. S-135 and S-165 were from the same patient, but collected at different timepoints. S-140 and S-106 were from the same block, cut in different batches. S-111, S-143, and S-122 had two mutations called each. S-133, S-169, and S-147 had known indels, and did not have any point mutations called. S-119 had no mutations called. S-176, S-177, and S-178 failed harvesting in one of the repeats and calling algorithm was not applied.

Patient ID	Age at diagnosis	Diagnosis: type, histiotype, and stage.			Sample ID	FFPE source		Cellularity	Gene	Mutation and base change (genome build hg19)		cDNA change	Protein change	TAm-Seq Allele frequency (repeat 1)	TAm-Seq Allele frequency (repeat 2)	Identified by Sanger sequencing
1	60	O	HGSOC	III	**S-131**	PS	O	60%	***TP53***	17:7577570	C>T	c.711G>A	p.M237I	0.72	0.70	**I**
2	63	O	HGSOC	III	**S-142**	IDS	O	40%	***TP53***	17:7578203	C>A	c.646G>T	p.V216L	0.42	0.43	**I**
3	54	O	HGSOC	III	**S-150**	IDS	O	70%	***TP53***	17:7577538	C>T	c.743G>A	p.R248Q	0.87	0.91	**I**
4	46	O	HGSOC	I	**S-167**	PS	O	70%	***TP53***	17:7578290	C>T	c.560-1G>A	p.?	0.73	0.73	**I**
5	38	O	HGSOC	III	**S-174**	PS	O	90%	***TP53***	17:7578268	A>G	c.581T>C	p.L194P	0.69	0.72	**I**
6	63	O	HGSOC	III	**S-102**	IDS	FT	60%	***TP53***	17:7578271	T>C	c.578A>G	p.H193R	0.39	0.34	**I**
7	68	O	HGSOC	III	**S-112**	IDS	O	60%	***TP53***	17:7578263	G>A	c.586C>T	p.R196*	0.30	0.28	**I**
8	60	O	HGSOC	III	**S-132**	PS	O	90%	***TP53***	17:7577120	C>T	c.818G>A	p.R273H	0.74	0.75	**I**
9	62	PP	HGSOC	IV	**S-151**	PS	SC	70%	***TP53***	17:7578479	G>A	c.451C>T	p.P151S	0.39	0.38	**I**
10	58	PP	HGSOC	IV	**S-168**	IDS	O	90%	***TP53***	17:7578508	C>T	c.422G>A	p.C141Y	0.82	0.86	**I**
11	55	O	HGSOC	III	**S-104**	IDS	O	15%	***TP53***	17:7578370	C>T	c.559+1G>A	p.?	0.12	0.12	**M**
12	62	O	HGSOC	IV	**S-115**	IDS	O	70%	***TP53***	17:7577579	G>T	c.702C>A	p.Y234*	0.57	0.59	**I**
13	45	O	HGSOC	IV	**S-123**	IDS	O	80%	***TP53***	17:7577574	T>C	c.707A>G	p.Y236C	0.53	0.55	**I**
14	58	O	HGSOC	III	**S-153**	IDS	O	95%	***TP53***	17:7578212	G>A	c.637C>T	p.R213*	0.56	0.56	**I**
15	61	O	HGSOC	IV	**S-116**	IDS	O	98%	***TP53***	17:7577121	G>A	c.817C>T	p.R273C	0.76	0.75	**I**
16	65	O	HGSOC	III	**S-127**	IDS	O	90%	***TP53***	17:7577094	G>A	c.844C>T	p.R282W	0.76	0.74	**I**
17	64	O	HGSOC	III	**S-145**	PS	FT	60%	***TP53***	17:7578176	C>T	c.672+1G>A	p.?	0.52	0.55	**I**
18	63	FT	HGSOC	III	**S-161**	PS	O	95%	***TP53***	17:7578534	C>G	c.396G>C	p.K132N	0.86	0.88	**I**
19	68	PP	HGSOC	III	**S-170**	IDS	U	50%	***TP53***	17:7579312	C>A	c.375G>T	p.T125T	0.22	0.20	**I**

20	67	PP	HGSOC	III	**S-108**	IDS	OM	20%	***TP53***	17:7578431	G>A	c.499C>T	p.Q167*	0.74	0.93	**I**
21	64	O	HGSOC	I	**S-117**	PS	O	95%	***TP53***	17:7574012	C>A	c.1015G>T	p.E339*	0.70	0.70	**I**
22	65	O	HGSOC	III	**S-128**	IDS	O	95%	***TP53***	17:7579521	C>A	c.166G>T	p.E56*	0.68	0.67	**I**
23	54	O	HGSOC	III	**S-138**	IDS	OM	95%	***TP53***	17:7577569	A>C	c.712T>G	p.C238G	0.79	0.79	**I**
24	63	FT	HGSOC	III	**S-146**	PS	O	60%	***TP53***	17:7577559	G>T	c.722C>A	p.S241Y	0.62	0.62	**I**
25	61	PP	HGSOC	III	**S-162**	IDS	OM	40%	***TP53***	17:7578404	A>T	c.526T>A	p.C176S	0.11	0.19	**I**
26	72	PP	HGSOC	III	**S-171**	IDS	OM	50%	***TP53***	17:7578265	A>T	c.584T>A	p.I195N	0.71	0.70	**I**
27	68	O	HGSOC	III	**S-179**	IDS	O	90%	***TP53***	17:7578262	C>G	c.587G>C	p.R196P	0.83	0.86	**I**
28	46	O	HGSOC	III	**S-109**	IDS	O	50%	***TP53***	17:7577548	C>T	c.733G>A	p.G245S	0.40	0.39	**I**
29	82	O	HGSOC	III	**S-118**	IDS	OM	80%	***TP53***	17:7577568	C>G	c.713G>C	p.C238S	0.74	0.75	**I**
30	69	FT	HGSOC	III	**S-130**	PS	O	80%	***TP53***	17:7579389	G>A	c.298C>T	p.Q100*	0.75	0.79	**I**
31	64	O	HGSOC	III	**S-172**	IDS	OM	70%	***TP53***	17:7578406	C>T	c.524G>A	p.R175H	0.48	0.42	**I**
32	71	O	HGSOC	III	**S-101**	IDS	P	90%	***TP53***	17:7578370	C>T	c.559+1G>A	p.?	0.95	0.94	**I**
33	73	O	HGSOC	III	**S-144**	IDS	OM	70%	***TP53***	17:7577557	A>C	c.724T>G	p.C242G	0.44	0.47	**I**
34	54	O	HGSOC	IV	**S-165**	IDS	O	95%	***TP53***	17:7578527	A>G	c.403T>C	p.C135R	0.47	0.46	**I**
					S-135	B	OM	80%	***TP53***	17:7578527	A>G	c.403T>C	p.C135R	0.57	0.60	**I**
35	55	PP	S/END	III	**S-106**	PS	O	90%	***TP53***	17:7574021	C>A	c.1006G>T	p.E336*	0.96	0.96	**I**
					S-140			90%	***TP53***	17:7574021	C>A	c.1006G>T	p.E336*	0.96	0.96	**I**
36	52	O	HGSOC	III	**S-111**	PS	O	95%	***TP53***	17:7578502	A>T	c.428T>A	p.V143E	0.45	0.44	**I**
								95%	***TP53***	17:7579642	C>T	Intron	Intron	0.22	0.28	**M**
37	45	O	END	II	**S-143**	PS	O	95%	***TP53***	17:7577538	C>T	c.743G>A	p.R248Q	0.75	0.77	**I**
								95%	***PTEN***	10:89720778	A>G	c.929A>G	p.D310G	0.59	0.52	**I**
38	55	O	END	IV	**S-122**	IDS	O	80%	***PTEN***	10:89692791	A>G	c.275A>G	p.D92G	0.19	0.18	**M**
								80%	***PTEN***	10:89685287	A>G	c.182A>G	p.H61R	0.17	0.17	**I**
39	54	O	HGSOC	III	**S-119**	PS	OM	80%								**U**
40	71	O	HGSOC	III	**S-133**	IDS	OM	80%								**Indel**
41	66	O	HGSOC	III	**S-169**	IDS	OM	50%								**Indel**
42	68	FT	S/CC	III	**S-147**	IDS	O	90%								**Indel**
43	49	FT	HGSOC	III	**S-176**	PS	O	95%							**FAIL**	
44	73	PP	HGSOC	IV	**S-177**	PS	OM	50%							**FAIL**	
45	44	O	HGSOC	III	**S-178**	IDS	OM	30%							**FAIL**	

Table S4: SNPs identified in circulating DNA from two plasma control samples.

SNP ID	Genomic location (genome build hg19) and genotypes	Normal plasma DNA 1 (“plasmaDna1”)	Normal plasma DNA 2 (“plasmaDna2”)
rs1625895	chr17:7578115 (T/C)	TT	TC
rs1800899	chr17:7576841 (A/G)	AG	AA
rs17337360	chr7:55260440 (C/T)	CC	TT
rs1050171	chr7:55249063 (G/A)	GA	GG
rs10241451	chr7:55248926 (T/C)	TC	TT

Table S5: Frequency of SNP alleles in dilution series of DNA from control plasma.

Total amplifiable copies \ Minor SNP percentage	200.00	80.00	32.00	12.80	5.12	2.05	0.82
2.05							40.00
5.12						40.00	16.00
12.80					40.00	16.00	6.40
32.00				40.00	16.00	6.40	2.56
80.00			40.00	16.00	6.40	2.56	1.02
200.00		40.00	16.00	6.40	2.56	1.02	0.41
500.00	40.00	16.00	6.40	2.56	1.02	0.41	0.16

Minor SNP amplifiable copies

Table S6: Additional data for Table 2 for mutations identified in plasma samples. Patient ID links to patient ID in table S3 (for patients 46 to 56, sequencing of tumour FFPE DNA was not included in table S3). Age: at diagnosis. Time: elapsed since surgery (months); chemotherapy ines: number of previous lines of chemotherapy. Vol: Plasma per amplification reaction (µL). DNA: DNA amount per amplification (in ng) was calculated from the number of total amplifiable copies (both mutant and germline allele) determined by digital PCR. Mean depth: number of sequencing reads covering this locus. AF, TAm-Seq: mean allele frequency using TAm-Seq, with the freq in each replicate given in parenthesis. Mutation call: FP, called using TAm-Seq, but not detected in FFPE DNA, and not called using more stringent calling criteria; M, mutation found in FFPE DNA and detected at allele frequency >2% in the plasma sample by digital PCR, but only found by TAm-Seq in one of the two duplicate amplicon sets; V, validated in FFPE tumour sample from the same patient (or, in the case of *EGFR* mutation in patient 27, using replicate Sanger sequencing of dilute template) and by digital PCR in plasma DNA; V*, sample failed amplification in the first run owing to a technical fault, but was reanalyzed and the mutation validated in FFPE tumour sample from the same patient and by digital PCR in plasma DNA.

Sample number	Patient ID	Age	Time (months); chemotherapy lines	Vol	DNA (ng)	Gene	Location (build hg19) and base change		Protein change	Mean depth	AF, TAm-Seq	Mean AF, digital PCR	Mutation call
1	15	61	35; 4	70	0.9	*TP53*	17:7577121	G>A	p.R273C	640	0.26 (0.227, 0.292)	0.167	V
2	3	54	44; 3	160	4.2	*TP53*	17:7577538	C>T	p.R248Q	340	0.244 (0.245, 0.242)	0.150	V
3	3	54	59; 6	160	5.7	*TP53*	17:7577538	C>T	p.R248Q	640	0.507 (0.49, 0.524)	0.410	V
4	14	58	62; 5	120	9.9	*TP53*	17:7578212	G>A	p.R213X	810	0.059 (0.056, 0.062)	0.035	V
5	10	58	48; 3	120	1.4	*TP53*	17:7578508	C>T	p.C141Y	680	0.021 (0.017, 0.024)	0.013	V
6	10	58	54; 4	120	2.1	*TP53*	17:7578508	C>T	p.C141Y	720	0.044 (0.044, 0.044)	0.038	V
7	26	72	13; 1	190	17.9	*TP53*	17:7578265	A>T	p.I195N	800	0.091 (0.081, 0.101)	0.081	V
8	31	64	21; 2	160	14.8	*TP53*	17:7578406	C>T	p.R175H	510	0.608 (0.614, 0.602)	0.627	V
9	31	64	21; 2	160	10.7	*TP53*	17:7578406	C>T	p.R175H	550	0.526 (0.529, 0.523)	0.604	V
10	31	64	21; 3	160	6.1	*TP53*	17:7578406	C>T	p.R175H	530	0.651 (0.705, 0.597)	0.682	V
11	31	64	22; 3	160	4.9	*TP53*	17:7578406	C>T	p.R175H	490	0.526 (0.548, 0.504)	0.581	V
13	47	54	12; 2	160	2.8	*TP53*	17:7578527	A>G	p.C135R	480	0.039 (0.034, 0.045)	0.045	V
14	47	54	18; 2	160	2.5	*TP53*	17:7578527	A>G	p.C135R	610	0.046 (0.047, 0.045)	0.120	V
15	47	54	20; 3	160	3	*TP53*	17:7578527	A>G	p.C135R	470	0.091 (0.133, 0.05)	0.068	V
16	27	68	25; 3	130	3.7	*TP53*	17:7578262	C>G	p.R196P	1070	0.088 (0.093, 0.084)	0.135	V
16	27	68	25; 3	130	3.7	*EGFR*	7:55259437	G>A	p.R832H	614	0.048 (0.055, 0.068, 0.008, 0.062)[†]	0.050	V

17	25	61	18; 2	160	4.2	*TP53*	17:7578404	A>T	p.C176S	580	0.113 (0.106, 0.12)	0.432	V
18	25	61	24; 3	160	4.4	*TP53*	17:7578404	A>T	p.C176S	620	0.029 (0.024, 0.033)	0.108	V
20	46	56	48; 6	140	5.2	*TP53*	17:7578406	C>T	p.R175H	650	0.201 (0.205, 0.197)	0.226	V
21	46	56	49; 6	140	3.6	*TP53*	17:7578406	C>T	p.R175H	650	0.085 (0.083, 0.087)	0.074	V
22	46	56	50; 6	140	4.1	*TP53*	17:7578406	C>T	p.R175H	630	0.081 (0.053, 0.109)	0.125	V
23	46	56	51; 6	140	3.7	*TP53*	17:7578406	C>T	p.R175H	710	0.074 (0.045, 0.103)	0.106	V
24	46	56	52; 6	140	7.1	*TP53*	17:7578406	C>T	p.R175H	760	0.269 (0.276, 0.261)	0.286	V
25	8	60	12; 1	130	3.9	*TP53*	17:7577120	C>T	p.R273H	750	0.094 (0.082, 0.106)	0.099	V
26	48	47	20; 2	160	5.7	*TP53*	17:7577094	G>A	p.R282W	640	0.048 (0.059, 0.036)	0.061	V
27	55	68	11; 2	150	3.6	*TP53*	17:7578508	C>T	p.C141Y	480	0.321 (0.342, 0.3)	0.364	V
29	50	71	0; 1	150	9.5	*TP53*	17:7577509	C>T	p.E258K	190	0.548 (0.534, 0.563)	0.253	V
31	51	62	0; 1	160	3.6	*TP53*	17:7578526	C>T	p.C135Y	620	0.04 (0.058, 0.021)	0.034	V
32	22	65	12; 1	140	2.4	*TP53*	17:7579521	C>A	p.E56X	1480	0.137 (0.147, 0.127)	0.122	V
33	18	63	48; 3	160	13.2	*TP53*	17:7578534	C>G	p.K132N	740	0.216 (0.238, 0.194)	0.206	V
34	18	63	49; 4	60	5.3	*TP53*	17:7578534	C>G	p.K132N	570	0.151 (0.173, 0.13)	0.201	V
36	18	63	56; 4	160	5.8	*TP53*	17:7578534	C>G	p.K132N	620	0.191 (0.18, 0.202)	0.275	V
37	18	63	57; 5	160	9.4	*TP53*	17:7578534	C>G	p.K132N	530	0.287 (0.245, 0.328)	0.362	V
38	18	63	58; 5	160	10.1	*TP53*	17:7578534	C>G	p.K132N	590	0.275 (0.267, 0.283)	0.331	V
39	18	63	59; 5	160	16.4	*TP53*	17:7578534	C>G	p.K132N	700	0.315 (0.302, 0.328)	0.323	V
40	18	63	60; 5	160	19.7	*TP53*	17:7578534	C>G	p.K132N	830	0.435 (0.439, 0.43)	0.482	V
41	18	63	61; 6	160	15	*TP53*	17:7578534	C>G	p.K132N	730	0.452 (0.461, 0.442)	0.445	V
42	18	63	62; 6	160	8.5	*TP53*	17:7578534	C>G	p.K132N	560	0.185 (0.2, 0.17)	0.245	V
43	56	53	15; 2	150	3.6	*TP53*	17:7578290	C>T	Splicing	680	0.143 (0.166, 0.12)	0.121	V
30	52	55	49; 4	170	5.2	*TP53*	17:7577569	A>G	p.C238R	1543	0.071(0.064, 0.077)	0.073	V*
35	18	63	50; 4	60	2.7	*TP53*	17:7578534	C>G	p.K132N	620	0.015 (0.002, 0.029)	0.073	M
12	53	51	23; 3	160	17	*PTEN*	10:89720799	T>G	Intergenic	1200	0.013 (0.014, 0.012)	-	FP
37	18	63	57; 5	160	9.4	*TP53*	17:7579790	G>A	Intronic	380	0.031 (0.031, 0.03)	-	FP

[†]The mutation locus in the *EGFR* gene was measured four separate times, in two overlapping amplicons in each of the two duplicates

Table S7: Mutations and amplicons studied in one breast cancer patient. To identify multiple mutations, genomic libraries from tumour and matched normal tissues were prepared using the standard IIlumina paired-end sample preparation kit according to the manufacturer's instructions. DNA fragments of 300 bp were sequenced using paired-end 100-bp reads on an Illumina HiSeq 2000 sequencer to 30× genomic coverage. Sequencing reads were aligned to the hg19 reference genome using ELAND (Illumina) and subsequent data analysis was performed using the CASAVA 1.8 analysis pipeline (Illumina) to identify somatic mutations. We selected 10 mutations and designed an amplicon of <120 bp to cover each mutation. We validated these mutations by TAm-Seq in the tumour material and then followed them in serial plasma samples collected from the same patient.

Gene	Genomic location of amplicon (genome build hg19)	Genomic location of mutation (genome build hg19)	Reference allele	Mutant allele
CD1A	chr1:158226710-158226810	chr1:158226762	C	T
BUB1	chr2:111430288-111430374	chr2:111430343	G	A
IQCA1	chr2:237402413-237402499	chr2:237402439	G	A
PIK3CA	chr3:178936048-178936130	chr3:178936091	G	A
ARAP3	chr5:141041694-141041773	chr5:141041739	G	A
MET	chr7:116398593-116398675	chr7:116398635	C	T
GIMAP5	chr7:150439702-150439805	chr7:150439730	C	T
ZFYVE21	chr14:104195437-104195539	chr14:104195486	G	A
KIAA0406	chr20:36641162-36641252	chr20:36641201	C	T
ADAMTS1	chr21:28214239-28214357	chr21:28214274	C	G