

## CANCER

# Enhanced detection of circulating tumor DNA by fragment size analysis

Florent Mouliere<sup>1,2\*†</sup>, Dineika Chandrananda<sup>1,2\*</sup>, Anna M. Piskorz<sup>1,2\*</sup>, Elizabeth K. Moore<sup>1,2,3\*</sup>, James Morris<sup>1,2</sup>, Lise Barlebo Ahlbom<sup>4,5</sup>, Richard Mair<sup>1,2,6</sup>, Teodora Goranova<sup>1,2</sup>, Francesco Marassi<sup>1,2,7,8</sup>, Katrin Heider<sup>1,2</sup>, Jonathan C. M. Wan<sup>1,2</sup>, Anna Supernat<sup>1,2,9</sup>, Irena Hudecová<sup>1,2</sup>, Ioannis Gounaris<sup>1,2,3</sup>, Susana Ros<sup>1,2</sup>, Mercedes Jimenez-Linan<sup>2,3</sup>, Javier Garcia-Corbacho<sup>10</sup>, Keval Patel<sup>1,2</sup>, Olga Østrup<sup>5</sup>, Suzanne Murphy<sup>1,2</sup>, Matthew D. Eldridge<sup>1,2</sup>, Davina Gale<sup>1,2</sup>, Grant D. Stewart<sup>2,3,11</sup>, Johanna Burge<sup>2,11</sup>, Wendy N. Cooper<sup>1,2</sup>, Michiel S. van der Heijden<sup>12,13</sup>, Charles E. Massie<sup>1,2,14</sup>, Colin Watts<sup>15</sup>, Pippa Corrie<sup>3</sup>, Simon Pacey<sup>3,14</sup>, Kevin M. Brindle<sup>1,2,16</sup>, Richard D. Baird<sup>17</sup>, Morten Mau-Sørensen<sup>4</sup>, Christine A. Parkinson<sup>1,2,3,18,19</sup>, Christopher G. Smith<sup>1,2</sup>, James D. Brenton<sup>1,2,3,18,19‡§</sup>, Nitzan Rosenfeld<sup>1,2‡§</sup>

Copyright © 2018  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.  
Government Works

Existing methods to improve detection of circulating tumor DNA (ctDNA) have focused on genomic alterations but have rarely considered the biological properties of plasma cell-free DNA (cfDNA). We hypothesized that differences in fragment lengths of circulating DNA could be exploited to enhance sensitivity for detecting the presence of ctDNA and for noninvasive genomic analysis of cancer. We surveyed ctDNA fragment sizes in 344 plasma samples from 200 patients with cancer using low-pass whole-genome sequencing (0.4x). To establish the size distribution of mutant ctDNA, tumor-guided personalized deep sequencing was performed in 19 patients. We detected enrichment of ctDNA in fragment sizes between 90 and 150 bp and developed methods for in vitro and in silico size selection of these fragments. Selecting fragments between 90 and 150 bp improved detection of tumor DNA, with more than twofold median enrichment in >95% of cases and more than fourfold enrichment in >10% of cases. Analysis of size-selected cfDNA identified clinically actionable mutations and copy number alterations that were otherwise not detected. Identification of plasma samples from patients with advanced cancer was improved by predictive models integrating fragment length and copy number analysis of cfDNA, with area under the curve (AUC) >0.99 compared to AUC <0.80 without fragmentation features. Increased identification of cfDNA from patients with glioma, renal, and pancreatic cancer was achieved with AUC > 0.91 compared to AUC < 0.5 without fragmentation features. Fragment size analysis and selective sequencing of specific fragment sizes can boost ctDNA detection and could complement or provide an alternative to deeper sequencing of cfDNA.

<sup>1</sup>Cancer Research UK Cambridge Institute, University of Cambridge, CB2 0RE Cambridge, UK. <sup>2</sup>Cancer Research UK Major Centre—Cambridge, Cancer Research UK Cambridge Institute, CB2 0RE Cambridge, UK. <sup>3</sup>Cambridge University Hospitals NHS Foundation Trust, CB2 0QQ Cambridge, UK. <sup>4</sup>Department of Oncology, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. <sup>5</sup>Centre for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. <sup>6</sup>Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, CB2 0QQ Cambridge, UK. <sup>7</sup>Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. <sup>8</sup>Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. <sup>9</sup>Department of Medical Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, 80-211 Gdańsk, Poland. <sup>10</sup>Clinical Trials Unit, Clinic Institute of Haematological and Oncological Diseases, Hospital Clinic de Barcelona, 170 08036 Barcelona, Spain. <sup>11</sup>Academic Urology Group, Department of Surgery, University of Cambridge, CB2 0QQ Cambridge, UK. <sup>12</sup>Division of Molecular Carcinogenesis, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. <sup>13</sup>Department of Medical Oncology, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. <sup>14</sup>Department of Oncology, University of Cambridge, CB2 0XZ Cambridge, UK. <sup>15</sup>Institute of Cancer Genomics Science, University of Birmingham, B15 2TT Birmingham, UK. <sup>16</sup>Department of Biochemistry, University of Cambridge, CB2 1QW Cambridge, UK. <sup>17</sup>Early Phase Clinical Trials and Breast Cancer Research Teams, Cancer Research UK Cambridge Centre, CB2 0QQ Cambridge, UK. <sup>18</sup>Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, CB2 0XZ Cambridge, UK. <sup>19</sup>NIHR Cambridge Biomedical Research Centre, CB2 0QQ Cambridge, UK.

\*These authors contributed equally to this work as co-first authors.

†Present address: Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Pathology, Cancer Center Amsterdam, de Boelelaan 1117, 1081 HV Amsterdam, Netherlands.

‡These authors contributed equally to this work as co-senior authors.

§Corresponding author. Email: james.brenton@cruk.cam.ac.uk (J.D.B.); nitzan.rosenfeld@cruk.cam.ac.uk (N.R.)

## INTRODUCTION

Blood plasma of patients with cancer contains circulating tumor DNA (ctDNA), but this valuable source of information is diluted by much larger quantities of DNA of noncancerous origins, such that ctDNA usually represents only a small fraction of the total cell-free DNA (cfDNA) (1, 2). High-depth targeted sequencing of selected genomic regions can be used to detect low amounts of ctDNA, but broader analysis with methods such as whole-exome sequencing (WES) and shallow whole-genome sequencing (sWGS) is only generally informative when ctDNA content is ~10% or greater (3–5). The concentration of ctDNA can exceed 10% of the total cfDNA in patients with advanced-stage cancers (6–8), but is much lower in patients with low tumor burden (9–12) and in patients with some cancer types such as gliomas and renal cancers (6). Current strategies to improve ctDNA detection rely on increasing depth of sequencing coupled with various error correction methods (2, 13, 14). However, approaches that focus only on genomic alterations do not take advantage of the potential differences in chromatin organization or fragment sizes of ctDNA (15–17). Results of ever-deeper sequencing are also confounded by the likelihood of false-positive results from detection of mutations from noncancerous cells, clonal expansions in normal epithelia, or clonal hematopoiesis of indeterminate potential (CHIP) (13, 18, 19).

The cell of origin and the mechanism of cfDNA release into blood can mark cfDNA with specific fragmentation signatures, potentially

providing precise information about cell type, gene expression, cell physiology or pathology, or action of treatment (15, 16, 20). cfDNA fragments commonly show a prominent mode at 167 bp, suggesting release from apoptotic caspase-dependent cleavage (Fig. 1A) (21–24). Circulating fetal DNA has been shown to be shorter than maternal DNA in plasma, and these size differences have been used to improve sensitivity of noninvasive prenatal diagnosis (22, 25–27). The size distribution of tumor-derived cfDNA has only been investigated in a few studies, encompassing a small number of cancer types and patients, and showed conflicting results (28–33). A limitation of previous studies is that determining the specific sizes of tumor-derived DNA fragments requires detailed characterization of matched tumor-derived alterations (30, 33), and the broader understanding and implications of potential biological differences have not previously been explored.

We hypothesized that we could improve the sensitivity for noninvasive cancer genomics by selective sequencing of ctDNA fragments and by leveraging differences in the biology that determine DNA fragmentation. To test this, we established a pan-cancer catalog of cfDNA fragmentation features in plasma samples from patients with different cancer types and healthy individuals to identify biological features enriched in tumor-derived DNA. We developed methods for selecting specific sizes of cfDNA fragments before sequencing and investigated the impact of combining cfDNA size selection with genome-wide sequencing to improve the detection of ctDNA and the identification of clinically actionable genomic alterations.

## RESULTS

### Surveying the fragmentation features of tumor cfDNA

We generated a catalog of cfDNA fragmentation features (Fig. 1A) in 344 plasma samples from 200 patients with 18 different cancer types and additional 65 plasma samples from healthy controls (Fig. 1B, fig. S1, and tables S1 and S2). The size distribution of cfDNA fragments in patients with cancer differed in the size ranges of 90 to 150 bp, 180 to 220 bp, and 250 to 320 bp compared to healthy individuals (Fig. 1B and fig. S2). cfDNA fragment sizes in plasma of healthy individuals and in plasma of patients with late-stage glioma, renal, pancreatic, and bladder cancers were significantly longer than in other late-stage cancer types including breast, ovarian, lung, melanoma, colorectal, and cholangiocarcinoma (Kruskal-Wallis,  $P < 0.001$ ; Fig. 1C). Sorting the 18 cancer types according to the proportion of cfDNA fragments in the size range of 20 to 150 bp resulted in an order very similar to that obtained by Bettegowda *et al.* (6) based on the concentrations of ctDNA measured by individual mutation assays (Fig. 1D). In contrast to previous reports (6, 34), this sorting was performed without any analysis or prior knowledge of the presence of mutations or somatic copy number alterations (SCNAs) yet allowed the investigation of ctDNA content in different cancers.

### Sizing up mutant ctDNA

We determined the size profile of mutant ctDNA in plasma using two high-specificity approaches. First, we inferred the specific size profile of ctDNA and nontumor cfDNA with sWGS from the plasma of mice bearing human ovarian cancer xenografts (Fig. 2A). We observed a shift in ctDNA fragment sizes to less than 167 bp (Fig. 2B). Second, the size profile of mutant ctDNA was determined in plasma from 19 patients with cancer, using deep sequencing with patient-

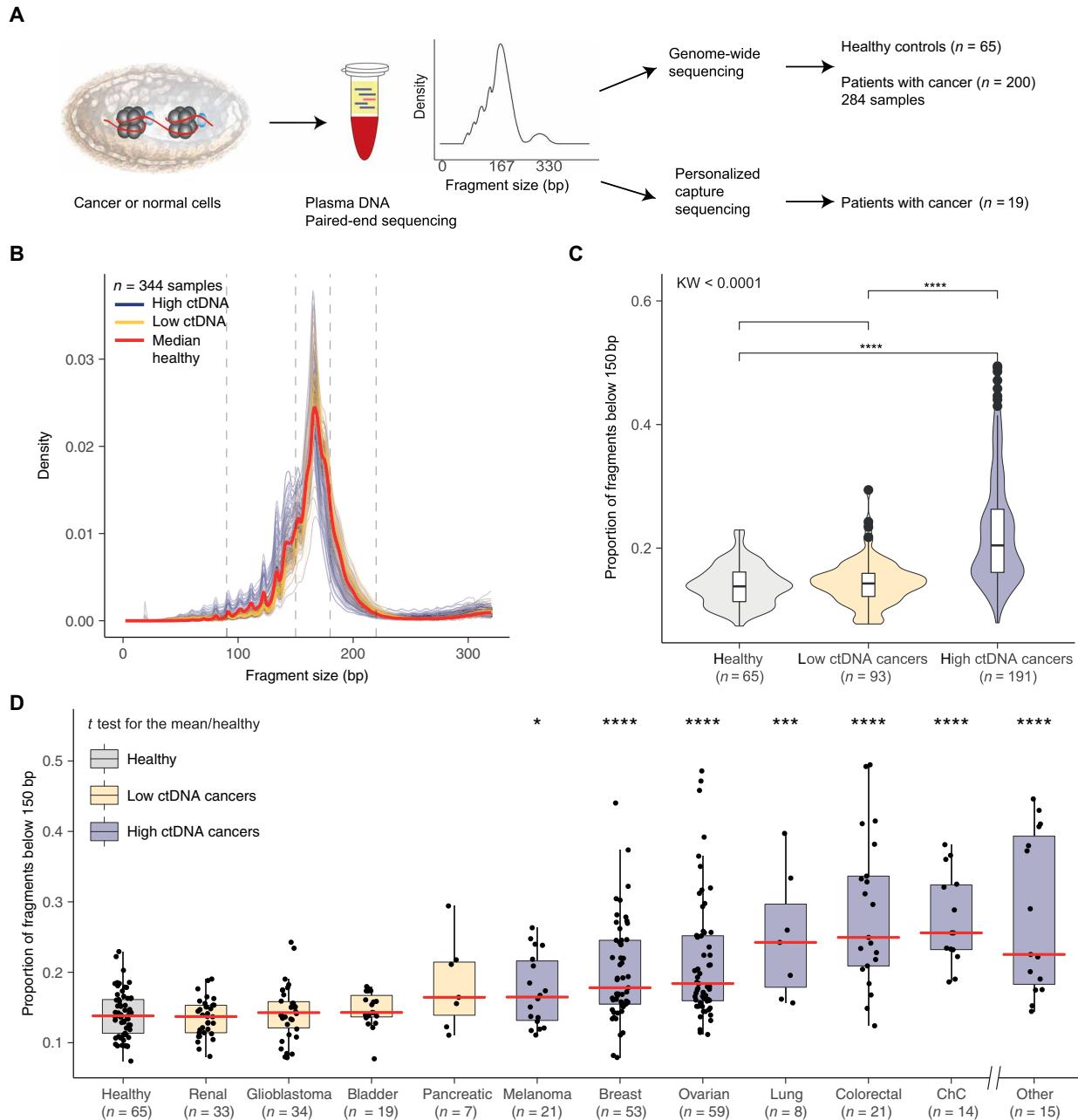
specific hybrid-capture panels developed from whole-exome profiling of matched tumor samples (Fig. 2C). By sequencing hundreds of mutations at a depth of  $>300\times$  in cfDNA, we obtained allele-specific reads from mutant and normal DNA. Enrichment of DNA fragments carrying tumor-mutated alleles was observed in fragments between ~20 and 40 bp shorter than nucleosomal DNA sizes (multiples of 167 bp; Fig. 2D). We determined that mutant ctDNA is generally more fragmented than nonmutant cfDNA, with a maximum enrichment of ctDNA in fragments between 90 and 150 bp (fig. S3), as well as enrichment in the size range of 250 to 320 bp. These data also indicated that mutant DNA in plasma of patients with advanced cancer (before treatment) is consistently shorter than predicted mononucleosomal and dinucleosomal DNA fragment lengths (Fig. 2D).

### Selecting tumor-derived DNA fragments

We evaluated whether the shorter cfDNA fragments in plasma can be harnessed to improve ctDNA detection. We determined the feasibility of selective sequencing of shorter fragments using in vitro size selection with a bench-top microfluidic device followed by sWGS in 48 plasma samples from 35 patients with high-grade serous ovarian cancer (HGSOC; Fig. 3A and figs. S4 and S5). We assessed the accuracy and quality of the size selection with the plasma from 20 healthy individuals (Fig. 3B and fig. S6). We also explored the utility of in silico size selection of fragmented DNA using read-pair positioning from unprocessed sWGS data (Fig. 3A). In silico size selection was performed once reads were aligned to the genome reference, by selecting the paired-end reads that corresponded to the fragment lengths in a 90- to 150-bp size range. Figure 3 (C to E) shows the effect of in vitro size selection for one HGSOC case (see all five samples in figs. S7 and S8). First, we identified SCNAs in plasma cfDNA before treatment, when the concentration of ctDNA was high (Fig. 3C). Only a small number of focal SCNAs were observed in the subsequent plasma sample collected 3 weeks after initiation of chemotherapy (without size selection; Fig. 3D). In vitro size selection of the same posttreatment plasma sample showed a median increase of  $6.4\times$  in the amplitude of detectable SCNAs without size selection. Selective sequencing of shorter fragments in this sample resulted in the detection of multiple other SCNAs that were not observed without size selection (Fig. 3E) and a genome-wide copy number profile that was similar to that obtained before treatment when ctDNA concentrations were four times higher, with additional copy number alterations identified in this sample despite the lower initial concentration of ctDNA (Fig. 3C). In silico size selection also enriched ctDNA but to a lower extent than using in vitro size selection (fig. S7). We concluded that selecting short DNA fragments in plasma can enrich tumor content on a genome-wide scale.

### Quantifying the impact of size selection

To quantitatively assess the enrichment after size selection on a genome-wide scale, we developed a metric from sWGS data ( $<0.4\times$  coverage) called t-MAD (trimmed median absolute deviation from copy number neutrality; see Fig. 4A). All sWGS data were downsampled to 10 million sequencing reads for comparison. To define the detection threshold, we measured the t-MAD score for sWGS data from 65 plasma samples from 46 healthy individuals and took the maximal value (median, 0.01; range, 0.004 to 0.015). We compared t-MAD to the mutant allele fraction (MAF) in high ctDNA cancer types as assessed by digital polymerase chain reaction (dPCR)



**Fig. 1. Survey of plasma DNA fragmentation with genome-wide sequencing on a pan-cancer scale.** (A) The size profile of cfDNA can be determined by paired-end sequencing of plasma samples and reflects its organization around the nucleosome. cfDNA is released into the blood circulation by various means, each of which leaves a signature on the DNA fragment sizes. We inferred the size profile of cfDNA by analyzing with sWGS ( $n = 344$  plasma samples from 65 healthy controls and 200 patients with cancer) and the size profile of mutant ctDNA by personalized capture sequencing ( $n = 19$  plasma samples). (B) Fragment size distributions of 344 plasma samples from 200 patients with cancer. Samples are split into two groups based on the previous literature (6), with orange representing samples from patients with cancer types previously observed to have low amounts of ctDNA (renal, bladder, pancreatic, and glioma) and blue representing samples from patients with cancer types previously observed to have higher amounts of ctDNA (breast, melanoma, ovarian, lung, colorectal, cholangiocarcinoma, and others; see table S1). (C) Proportion of cfDNA fragments below 150 bp in those samples, grouped into cancer types as defined in (B). The Kruskal-Wallis (KW) test for difference in size distributions indicated a significant difference between the group of samples from cancer types releasing high amounts of ctDNA and the group of samples from cancer types releasing low amounts, as well as the group of samples from healthy individuals. (D) Proportion of cfDNA fragments below 150 bp by cancer type (all samples). Cancer types represented by fewer than four individuals are grouped in the “other” category. Red lines indicate the median proportion for each cancer type. ChC, cholangiocarcinoma. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ .

or WES in 97 samples. We observed a high correlation (Pearson correlation,  $r = 0.80$ ) between t-MAD and MAF (Fig. 4B) for samples with t-MAD greater than the detection threshold (0.015) or

with MAF > 0.025. Figure S9 shows that the slope of t-MAD versus MAF fit lines differed between cancer types (range, 0.17 to 1.12), likely reflecting differences in the extent of SCNAs. We estimated the

**Fig. 2. Determining the size profile of mutant ctDNA with animal models and personalized capture sequencing.** (A) A mouse model with xenografted human tumor cells enabled the discrimination of DNA fragments released by cancer cells (reads aligning to the human genome) from the DNA released by healthy cells (reads aligning to the mouse genome), with the use of sWGS. (B) Fragment size distribution from the plasma extracted from a mouse xenografted with a human ovarian tumor, showing ctDNA originating from tumor cells (red) and cfDNA from noncancerous cells (blue). Two vertical dashed lines indicate 145 and 167 bp. The fraction of reads shorter than 150 bp is indicated. (C) Design of personalized hybrid-capture sequencing panels developed to specifically determine the size profiles of mutant DNA and non-mutant DNA in plasma from 19 patients with late-stage cancers. Capture panels included somatic mutations identified in tumor tissue by WES. A mean of 165 mutations per patient was then analyzed from matched plasma samples. Reads were aligned and separated into fragments carrying either the reference or the mutant sequence. Fragment sizes for paired-end reads were calculated. (D) Size profiles of mutant DNA and nonmutant DNA in plasma from 19 patients with late-stage cancers were determined by tumor-guided capture sequencing. The fraction of reads shorter than 150 bp is indicated.

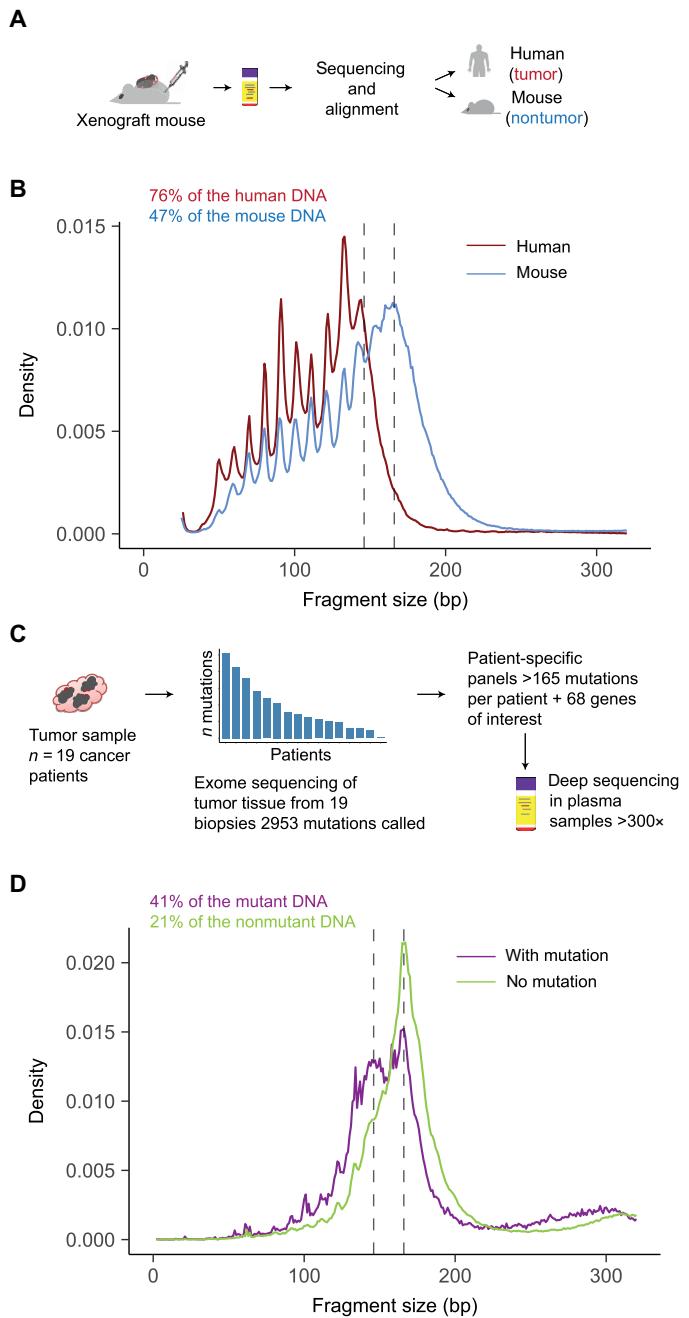
sensitivity of t-MAD for detecting low amounts of ctDNA using a spike-in dilution of DNA from a patient with a *TP53* mutation into DNA from a pool of seven healthy individuals (fig. S10), which confirmed that the t-MAD score was linear with ctDNA fraction down to MAF of ~0.01. In addition, t-MAD scores greater than the detection threshold (0.015) for samples were present even in samples with MAF as low as 0.004. t-MAD was also strongly correlated with tumor volume determined by RECIST1.1 (Pearson correlation,  $r = 0.6$ ;  $P < 0.0001$ ;  $n = 35$ ; fig. S11).

Using t-MAD, we detected ctDNA from 69% (130 of 189) of the samples from cancer types where ctDNA concentrations were shown to be high (Fig. 4C). From cancer types for which ctDNA concentrations are suspected to be low (glioma, renal, bladder, and pancreatic), we detected ctDNA in 17% (10 of 57) of the cases (Fig. 4C). We used in silico size selection of the DNA fragments between 90 and 150 bp from the high ctDNA cancers ( $n = 189$ ) and healthy controls ( $n = 65$ ) to improve the sensitivity for detecting t-MAD (Fig. 4D). Receiver operating characteristic (ROC) analysis comparing the t-MAD score for the samples revealed an area under the curve (AUC) of 0.90 after in silico size selection, against an AUC of 0.69 without size selection (Fig. 4D).

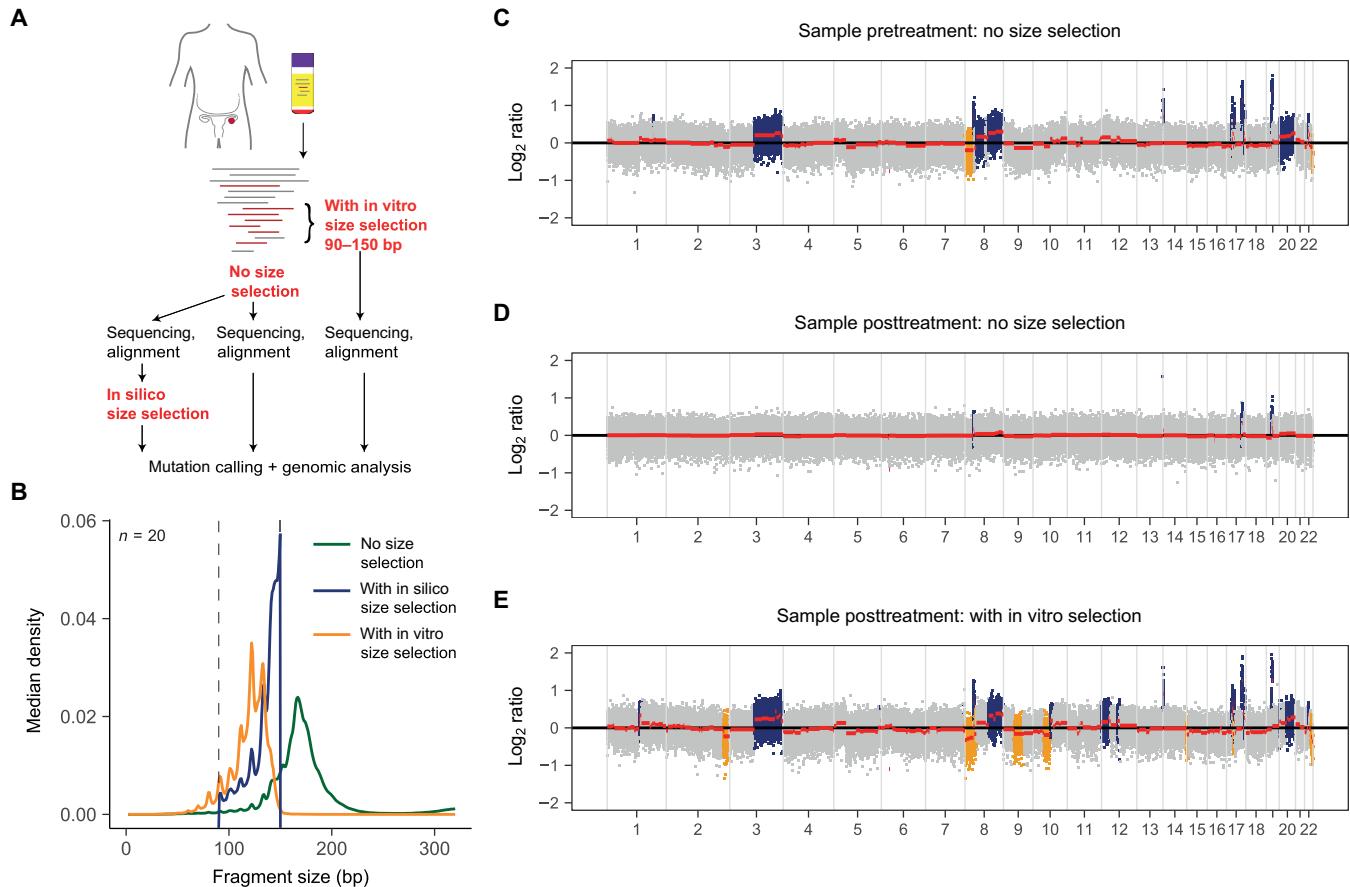
We explored whether size-selected sequencing could improve the detection of response or disease progression. We used sWGS of longitudinal plasma samples from six patients with cancer (Fig. 4, E and F) and in silico size selection of the cfDNA fragments between 90 and 150 bp. In two patients, size-selected samples indicated tumor progression 60 and 87 days before detection by imaging or unselected t-MAD analysis (Fig. 4, E and F). Other longitudinal samples exhibited improvements in the detection of ctDNA with t-MAD and size selection (Fig. 4F).

### Identifying more clinically relevant genomic alterations with size selection

We next tested whether size selection could increase the sensitivity for detecting cancer genomic alterations in cfDNA. To test effects on copy number aberrations, we studied 35 patients with HGSOC as the archetypal copy number–driven cancer (35). t-MAD was used to quantify the enrichment of ctDNA with in vitro size selection in 48 plasma samples, including samples collected before and after initiation of chemotherapy treatment. In vitro size selection resulted



in an increase in the calculated t-MAD score from the sWGS data for 47 of 48 of the plasma samples (98%;  $t$  test,  $P = 0.06$ ) with a mean of 2.5 and median of 2.1-fold increase (Fig. 5A and table S3). We compared the t-MAD scores against those obtained by sWGS for the plasma samples from healthy individuals. Thirty-nine of the 48 size-selected HGSOC plasma samples (82%) had a t-MAD score greater than the highest t-MAD value determined in the in vitro size-selected healthy plasma samples (Fig. 5A and figs. S6 and S12), compared to 24 of 48 without size selection (50%). ROC analysis comparing the t-MAD score for the samples from patients with cancer (pre- and posttreatment initiation,  $n = 48$ ) and healthy controls ( $n = 46$ ) revealed an AUC of 0.97 after in vitro size selection, with maximal sensitivity and specificity of 90 and 98%, respectively.



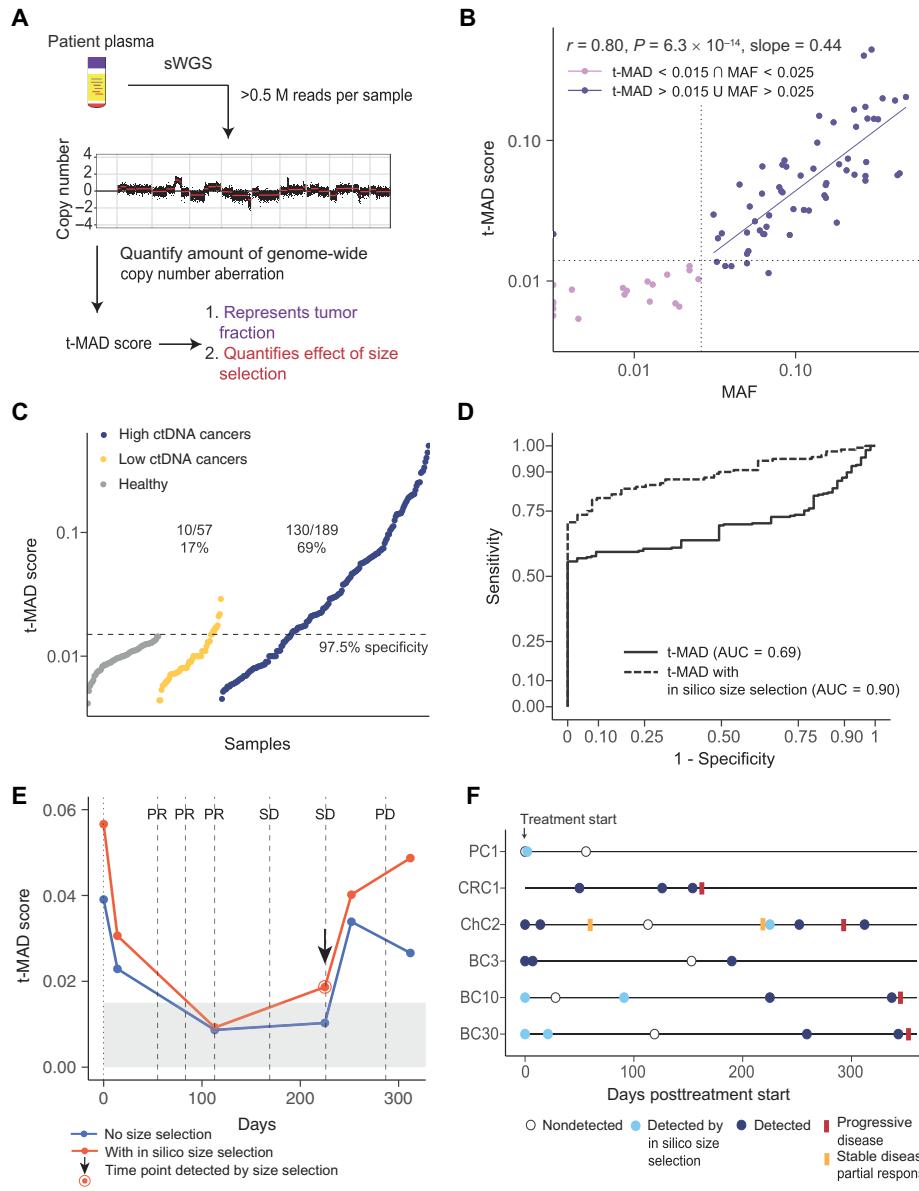
**Fig. 3. Enhancing the tumor fraction from plasma sequencing with size selection.** (A) Plasma samples collected from patients with ovarian cancer were analyzed in parallel without size selection or using either in silico or in vitro size selection. (B) Accuracy of the in vitro and in silico size selection determined on a cohort of 20 healthy controls. The size distribution before size selection is shown in green, after in silico size selection (with sharp cutoff at 90 and 150 bp) in blue and after in vitro size selection in orange. Vertical lines indicate 90 and 150 bp. (C) SCNA analysis with sWGS from plasma DNA of a patient with ovarian cancer collected before initiation of treatment, when ctDNA MAF was 0.271 for a *TP53* mutation as determined by tagged-amplicon deep sequencing (TAm-Seq). Inferred amplifications are shown in blue and deletions in orange. Copy number neutral regions are shown in gray. (D) SCNA analysis of a plasma sample from the same patient as in (C), collected 3 weeks after treatment start. The MAF for the *TP53* mutation at this time point was 0.068, and sWGS revealed only limited evidence of copy number alterations (before size selection). (E) Analysis of the same plasma sample as in (D) after in vitro size selection of fragments between 90 and 150 bp in length. The MAF for the *TP53* mutation increased to 0.402 after in vitro size selection, and SCNAs were apparent by sWGS. More SCNAs were detected in comparison to (C) and (D) (for example, in chr2, chr9, and chr10). SCNAs were also detected in this sample after in silico size selection (fig. S7).

This was superior to detection by sWGS without size selection (AUC, 0.64; Fig. 5B).

We then determined whether this improved sensitivity resulted in the detection of SCNAs with potential clinical value. Across the genome, t-MAD scores evaluating SCNAs were higher after size selection in 33 of 35 (94%) patients with HGSOC, and the magnitude of copy number ( $\log_2$  ratio) values significantly increased after in vitro size selection ( $t$  test for the means,  $P = 0.003$ ; Fig. 5C). We compared the relative copy number values for 15 genes frequently altered in HGSOC (table S4). Analysis of plasma cfDNA after size selection revealed a large number of SCNAs that were not observed in the same samples without size selection (Fig. 5D), including amplifications in key genes such as *NF1*, *TERT*, and *MYC* (fig. S13).

We also tested whether similar enrichment was seen for substitutions to exclude the possibility that size selection might only increase the sensitivity for sWGS analysis. We performed WES of plasma cfDNA from 23 patients with seven cancer types (fig. S1). We used the WES data to compare the size distributions of fragments carrying

mutant or nonmutant alleles (Fig. 6A) and to test whether size selection could identify additional mutations. We first selected six patients with HGSOC and performed WES of plasma DNA with and without in vitro size selection in the range of 90 to 150 bp, analyzing time points before and after initiation of treatment (36). In addition, in silico size selection for the same range of fragment sizes was performed (Fig. 6A). Analysis of the MAF of SNVs revealed statistically significant enrichment of the tumor fraction with both in vitro size selection (mean, 4.19-fold; median, 4.27-fold increase;  $t$  test,  $P < 0.001$ ) and in silico size selection (mean, 2.20-fold; median, 2.25-fold increase;  $t$  test,  $P < 0.001$ ; Fig. 6A and fig. S14). Three weeks after initiation of treatment, ctDNA fractions are often lower (36), and therefore, we further analyzed posttreatment plasma samples using TAm-Seq (37). We observed enrichment of MAFs by in vitro size selection between 0.9 and 11 times (mean, 2.1 times; median, 1.5 times), with one outlier sample exhibiting a relative enrichment of 118 times compared to the same samples without size selection (fig. S15).



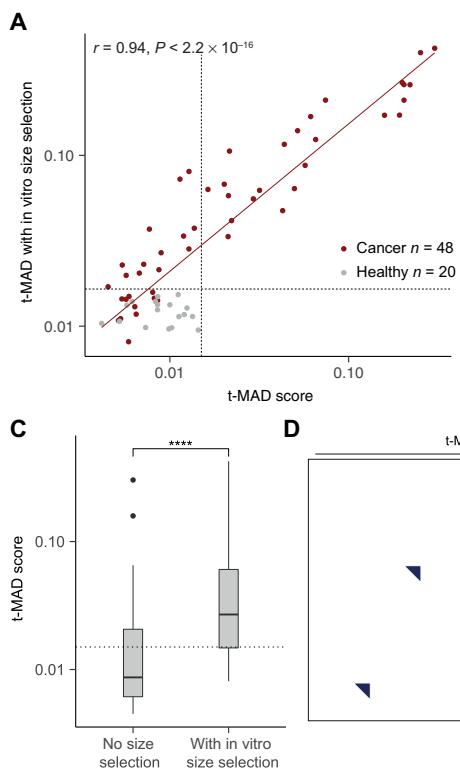
**Fig. 4. Quantifying the ctDNA enrichment by sWGS with in silico size selection and t-MAD.** (A) Workflow to quantify tumor fraction from SCNA as a genome-wide score named t-MAD. (B) Correlation between the MAF of single-nucleotide variants (SNVs) determined by dPCR or hybrid-capture sequencing and t-MAD score determined by sWGS. Data included 97 samples from patients with multiple cancer types with matched MAF measurements and t-MAD scores. Pearson correlation (coefficient  $r$ ) between MAF and t-MAD scores was calculated for all cases with MAF > 0.025 and t-MAD > 0.015. Linear regression indicated a fit with a slope of 0.44 (purple solid line). (C) Comparison of t-MAD scores determined from sWGS between healthy samples and samples collected from patients with cancer types that exhibit low amounts of ctDNA and from patients with cancer types that exhibit high amounts of ctDNA (as in Fig. 1). All samples for which t-MAD could be calculated have been included. (D) ROC analysis comparing the classification of these plasma samples from high ctDNA cancer samples ( $n = 189$ ) and plasma samples from healthy controls ( $n = 65$ ) using t-MAD had an AUC of 0.69 without size selection (black solid curve). After applying in silico size selection to the samples from patients with cancer, we observed an AUC of 0.90 (black dashed curve). (E) Determination of t-MAD from longitudinal plasma samples of a patient with colorectal cancer. t-MAD was analyzed before and after in silico size selection of the DNA fragments between 90 and 150 bp and then compared to the RECIST status for this patient. PR, partial response; SD, stable disease; PD, progressive disease. (F) Application of in silico size selection to six patients with long-term follow-up. t-MAD score was determined before and after in silico size selection of the short DNA fragments. Dark blue circles indicate samples in which ctDNA was detected both with and without in silico size selection. Light blue circles indicate samples where ctDNA was detected only after in silico size selection. Open circles indicate samples where ctDNA was not detected by either analysis. Times when RECIST status was assessed are indicated by a red bar for progression or an orange bar for regression or stable disease. PC, prostate cancer; CRC, colorectal cancer; ChC, cholangiocarcinoma; BC, breast cancer. The numbers correspond to the patients.

Size selection with both in vitro and in silico methods increased the number of mutations detected by WES by an average of 53% compared to no size selection (Fig. 6B). We identified a total of 1023 mutations in the samples without size selection. An additional 260 mutations were detected by in vitro size selection, and an additional 310 mutations were called after in silico size selection (Fig. 6B and table S5). To exclude the possibility that the improved sensitivity for mutation detection was a result of sequencing artifacts, we validated whether new mutations were also detectable in tumor specimens. We used in silico size selection in an independent cohort of 16 patients for whom matched tumor tissue DNA was available (table S6). In silico size selection enriched the MAF for nearly all mutations (2061 of 2133, 97%), with an average increase of MAF of 1.7× (Fig. 6C). For 13 of 16 patients (81%), we identified additional mutations in plasma after in silico size selection. Of these 82 additional mutations, 23 (28%) were confirmed to be present in the matched tumor tissue DNA (Fig. 6D). This included mutations in key cancer genes including *BRAF*, *ARID1A*, and *NF1* (fig. S16).

### Detecting cancer by supervised machine learning combining cfDNA fragmentation and somatic alteration analysis

Although in vitro and in silico size selection increase the sensitivity of detection, they also result in a loss of cfDNA for analysis. In analysis of ctDNA based on genomic signals, potentially informative data are lost because regions of the cancer genome that are not mutated or altered do not contribute to detection (fig. S17). We hypothesized that leveraging other biological properties of the cfDNA fragmentation profile could enhance the detection of ctDNA.

We defined other cfDNA fragmentation features from sWGS data including (i) the proportion of fragments in multiple size ranges, (ii) the ratios of proportions of fragments in different sizes, and (iii) the amplitude of oscillations in fragment size density with 10-bp periodicity (see Materials and Methods and Fig. 7A). These fragmentation features were compared between patients with cancer and healthy individuals (fig. S18),



**Fig. 5. Quantifying the ctDNA enrichment by sWGS with in vitro size selection.** (A) The effect of in vitro size selection on the t-MAD score. For each of 48 plasma samples collected from 35 patients, the t-MAD score was determined from the sWGS after in vitro size selection (y axis) and without size selection (x axis). In vitro size selection increased the t-MAD score for nearly all samples, with a median increase of 2.1-fold (range from 1.1- to 6.4-fold). t-MAD scores determined from sWGS for 46 samples from healthy individuals were all  $<0.015$  both before and after in vitro size selection. (B) ROC analysis comparing the classification of plasma samples from patients with cancer ( $n = 48$ ) and plasma samples from healthy controls ( $n = 46$ ) using t-MAD had an AUC of 0.64 without size selection (green curve). After applying in silico size selection to the samples from the patients and controls, we observed an AUC of 0.78 (blue curve), and after in vitro size selection, an AUC of 0.97 (orange curve). (C) Comparison of t-MAD scores determined from sWGS between matched ovarian cancer samples with and without in vitro size selection. The  $t$  test for the difference in means indicates a significant increase in tumor fraction (measured by t-MAD) with in vitro size selection (\*\*\*\* $P < 0.0001$ ). (D) Detection of SCNAs across 15 genes frequently mutated in recurrent ovarian cancer, measured in plasma samples collected during treatment for 35 patients. Patients were ranked from left to right by increasing tumor fraction as quantified by t-MAD (before in vitro size selection). SCNAs were labeled as detected for a gene if the mean  $\log_2$  ratio in that region was greater than 0.05. Empty squares represent copy number neutral regions, bottom left triangles in light blue indicate that SCNAs were detected without size selection, and top right triangles in dark blue represent SCNAs detected after in vitro size selection.

and the feature representing the proportion (P) of fragments between 20 and 150 bp exhibited the highest AUC (0.819). Principal components analysis (PCA) of the samples represented by t-MAD and fragmentation features showed a separation between healthy samples and samples from patients with cancer and identified fragment features that were aligned (in PCA) with t-MAD scores (Fig. 7B).

We next explored the potential of fragmentation features to enhance the detection of tumor DNA in plasma samples. A predictive analysis was performed using the t-MAD score and nine fragmentation features across 304 samples (239 from patients with cancer and 65 from healthy controls; Fig. 7C, fig. S19, and table S2). The nine fragmentation features determined from sWGS included five features based on the proportion (P) of fragments in defined size ranges: P(20 to 150), P(100 to 150), P(160 to 180), P(180 to 220),

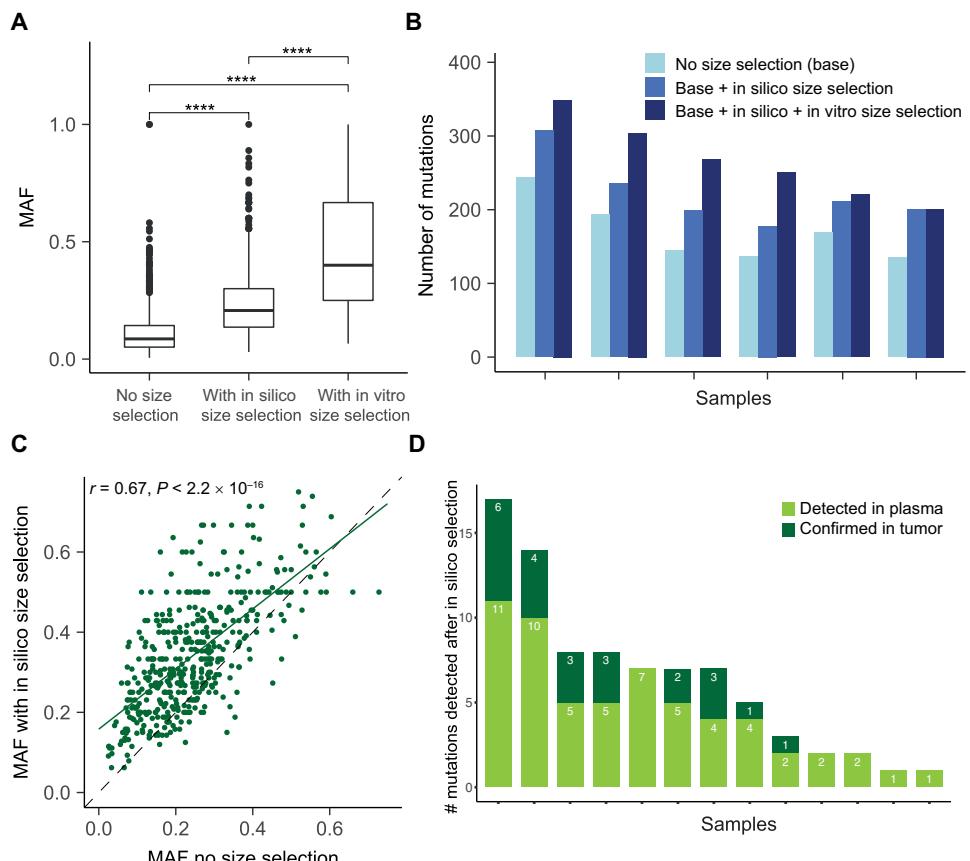
and P(250 to 320); three features based on ratios of those proportions: P(20 to 150)/P(160 to 180), P(100 to 150)/P(163 to 169), and P(20 to 150)/P(180 to 220); and a further feature based on the amplitude of the oscillations having 10-bp periodicity observed below 150 bp.

Variable selection and the classification of samples as “healthy” or “cancer” were performed using logistic regression (LR) and random forest (RF) models trained on 153 samples and validated on two datasets of 94 and 83 independent samples (Fig. 7C). The best feature set for the LR model included t-MAD, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320). The same five variables were independently identified using the RF model (with some differences in their ranking). Figure S20 shows performance metrics for the different algorithms on training set data using cross-validation. Using t-MAD alone in the validation pan-cancer dataset (Fig. 7D and fig. S19), we could distinguish cancer samples from healthy individuals with an AUC of 0.764. Using the LR model improved the classification of the samples to an AUC of 0.908. The RF model (trained on the 153-sample training set) could distinguish cancer from healthy individuals even more accurately in the validation dataset ( $n = 94$ ) with an AUC of 0.994. On the second validation dataset containing low-ctDNA cancer samples ( $n = 83$ ; Fig. 7E), t-MAD alone or the LR performed less well, with AUC values of 0.421 and 0.532, respectively. However, the RF model was still able to distinguish low-ctDNA cancer samples from healthy controls with an AUC of 0.914. At a specificity of 95%, the RF model correctly classified as cancer in 64 of 68 (94%) of the samples from high-ctDNA

cancers (colorectal, cholangiocarcinoma, ovarian, breast, and melanoma) and 37 of 57 (65%) of the samples from low-ctDNA cancers (pancreatic, renal, and glioma; Fig. 7F). In a second iteration of model training, we omitted t-MAD using only the four fragmentation features (fig. S21). The RF model could still distinguish cancer from healthy controls, albeit with slightly reduced AUCs (0.989 for cancer types with high amounts of ctDNA and 0.891 for cancer types with low amounts of ctDNA), suggesting that the cfDNA fragmentation pattern is the most important predictive component.

## DISCUSSION

Our results indicate that exploiting fundamental properties of cfDNA with fragment-specific analyses can allow more sensitive evaluation



**Fig. 6. Improving the detection of somatic alterations by WES in multiple cancer types with size selection.**

(A) Analysis of the MAF of mutations detected by WES in six patients with HGSC without size selection and with either in vitro or in silico size selection. \*\*\*\* $P < 0.0001$ . (B) Comparison of size-selected WES data with nonselected WES data to assess the number of mutations detected in plasma samples from six patients with HGSC. For each patient, the first bar in light blue shows the number of mutations called without size selection, the second bar quantifies the number of mutations called after the addition of those identified with in silico size selection, and the third bar in dark blue shows the number of mutations called after addition of mutations called after in vitro size selection. (C) Patients ( $n = 16$ ) were retrospectively selected from a cohort with different cancer types (colorectal, cholangiocarcinoma, pancreatic, and prostate) enrolled in early-phase clinical trials. Matched tumor tissue DNA was available for each plasma sample, and two patients also had a biopsy collected at relapse. WES was performed on tumor tissue DNA and plasma DNA samples, and in silico size selection was applied to the data. A total of 97% (2061 of 2133) of the shared mutations detected by WES showed higher MAF after in silico size selection. (D) Mutations detected only after in silico selection of WES data from 16 patients [as in (C)] compared to mutations called by WES of the matched tumor tissue. Three of 16 patients had no additional mutations identified after in silico size selection. Of the 82 mutations detected in plasma after in silico size selection, 23 (28%) had low signal in tumor WES data and were not identified in those samples without size selection.

of ctDNA. We based the fragment size selection criteria on a biological observation that ctDNA fragment size distribution is shifted from noncancerous cfDNA. Our work builds on a comprehensive survey of plasma cfDNA fragmentation patterns across 200 patients with multiple cancer types and 65 healthy individuals. We identified features that could determine the presence and amount of ctDNA in plasma samples, without a priori knowledge of somatic aberrations. We caution that this catalog is limited to double-stranded DNA from plasma samples and is subject to potential biases incurred by the DNA extraction and sequencing methods we used. Additional biological effects could contribute to further selective analysis of cfDNA. Other bodily fluids (urine, cerebrospinal fluid, and saliva), different nucleic acids and structures, altered mechanisms of

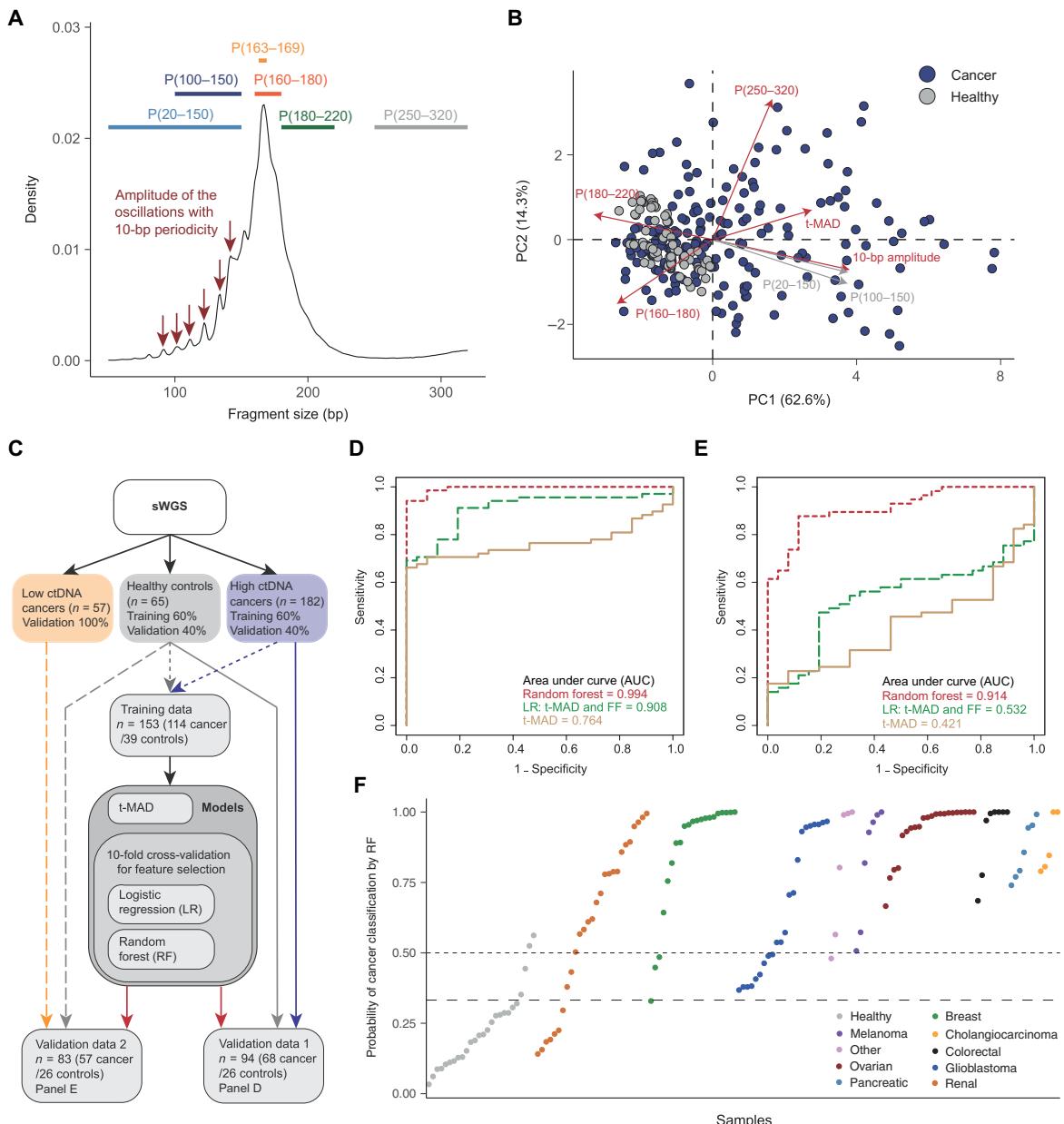
release into circulation, or sample processing methods could exhibit varying fragment size signatures and could offer additional exploitable biological patterns for selective sequencing.

Previous work has reported the size distributions of mutant ctDNA but only considered limited genomic loci, cancer types, or cases (30, 32, 33). We identified the size differences between mutant and nonmutant DNA on a genome-wide and pan-cancer scale. We developed a method to size mutant ctDNA without using high-depth WGS. By sequencing >150 mutations per patient at high depth, we obtained large numbers of reads that could be unequivocally identified as tumor derived and thus determined the size distribution of mutant ctDNA and nonmutant cfDNA in patients with cancer. A potential limitation of our approach is that capture-based sequencing is biased by probe capture efficiency and, therefore, our data may not accurately reflect ctDNA fragments of <100 or >300 bp.

Our work provides strong evidence that the modal size of ctDNA for many cancer types is less than 167 bp, which is the length of DNA wrapped around the chromatosome. In addition, our work also shows that there is enrichment of mutant DNA fragments at sizes greater than 167 bp, notably in the range of 250 to 320 bp. These longer fragments may explain previous observations that longer ctDNA can be detected in the plasma of patients with cancer (29, 32). The origin of these long fragments is still unknown, and their observation could be linked to technical factors. However, it is likely that mechanisms of compaction and release of cfDNA into circulation, which may differ depending on its origin, will be reflected by different fragment sizes (38). Improving the characteriza-

tion of these fragments will be important, especially for future work combining analysis of ctDNA with that of other entities in blood such as microvesicles and tumor-educated platelets (39, 40). Fragment-specific analyses not only increase the sensitivity for detection of rare mutations but could also be used to track modifications in the size distribution of ctDNA. Future work should address whether this approach could be used to elucidate mechanistic effects of treatment on tumor cells, for example, by distinguishing between necrosis and apoptosis based on fragment size (41).

Genome-wide and exome sequencing of plasma DNA at multiple time points during cancer treatment have been proposed as noninvasive means to study cancer evolution and for the identification of possible mechanisms of resistance to treatment (3). However, WGS and WES



**Fig. 7. Enhancing the potential for ctDNA detection by combining SCNA and fragment size features.** (A) Schematic illustrating the selection of different size ranges and features in the distribution of fragment sizes. For each sample, fragmentation features included the proportion (P) of fragments in specific size ranges, the ratio between certain ranges, and a quantification of the amplitude of the 10-bp oscillations in the 90- to 145-bp size range calculated from the periodic “peaks” and “valleys.” (B) PCA comparing cancer and healthy samples using data from t-MAD scores and the fragmentation features. Red arrows indicate features that were selected as informative by the predictive analysis. (C) Workflow for the predictive analysis combining SCNA and fragment size features. sWGS data from 182 plasma samples from patients with cancer types with high amounts of ctDNA (colorectal, cholangiocarcinoma, lung, ovarian, and breast) were split into a training set (60% of samples) and a validation set (validation data 1, together with the healthy individual validation set). A further dataset of sWGS from 57 samples of cancer types exhibiting low amounts of ctDNA (glioma, renal, and pancreatic) was used as validation data 2, together with the healthy individual validation set. Plasma DNA sWGS data from healthy controls were split into a training set (60% of samples) and a validation set (used in both validation data 1 and validation data 2). (D) ROC curves for validation data 1 (samples from patients with cancer with high ctDNA amounts, 68; healthy, 26) for three predictive models built on the pan-cancer training cohort (cancer, 114; healthy, 39). The beige curve represents the ROC curve for classification with t-MAD only, the long-dashed green line represents the LR model combining the top five features based on recursive feature elimination [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed line shows the result for a RF classifier trained on the combination of the same five features, independently chosen for the best RF predictive model. FF, fragment size features. (E) ROC curves for validation data 2 (samples from patients with cancer with low ctDNA amounts, 57; healthy, 26) for the same three classifiers as in (D). The beige curve represents the model using t-MAD only, the long-dashed green curve represents the LR model combining the top five features [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed curve shows the result for a RF classifier trained on the combination of same five predictive features. (F) Plot representing the probability of classification as cancer with the RF model for all samples in both validation datasets. Samples are separated by cancer type and sorted within each by the RF probability of classification as cancer. The horizontal dashed line indicates 50% probability (achieving specificity of 24 of 26, 92.3%), and the long-dashed line indicates 33% probability (achieving specificity of 22 of 26, 84.6%).

approaches are costly and have thus far been applicable only in samples for which the tumor DNA fraction was >5 to 10% (3–5, 42). We demonstrated that we could exploit the differences in fragment lengths using *in vitro* and *in silico* size selection to enrich for tumor content in plasma samples, which improved mutation and SCNA detection in sWGS and WES data. We demonstrated that size selection improved the detection of mutations that are present in plasma at low allelic fractions while maintaining low sequencing depth by sWGS and WES. Size selection can be achieved with simple means and at low cost and is compatible with a wide range of downstream genome-wide and targeted genomic analyses, greatly increasing the potential value and utility of liquid biopsies as well as the cost-effectiveness of cfDNA sequencing.

Size selection can be applied *in silico*, which incurs no added costs, or *in vitro*, which adds a simple and low-cost intermediate step that can be applied to either the extracted DNA or the libraries created from it. This approach, applied prospectively to new studies, could boost the clinical utility of ctDNA detection and analysis and creates an opportunity for reanalysis of large volumes of existing data (4, 34, 43). The limitation of this technique is a potential loss of material and information, because some of the informative fragments may be found in size ranges that are filtered out or deprioritized in the analysis. This may be particularly problematic if only a few copies of the fragments of interest are present in the plasma. Despite potential loss of material, we demonstrated that classification algorithms can learn from cfDNA fragmentation features and SCNA analysis and improve the detection of ctDNA with a cheap sequencing approach. Moreover, the cfDNA fragmentation features alone can be leveraged to classify cancer and healthy samples with a high accuracy [AUC, 0.989 (high ctDNA cancers) and 0.891 (low ctDNA cancers)].

Analysis of fragment sizes could provide improvements in other applications. Introducing fragment size information on each read could enhance mutation-calling algorithms from high-depth sequencing to distinguish tumor-derived mutations from other sources such as somatic variants or background sequencing noise. In addition, cfDNA from patients analyzed with CHIP is likely to be structurally different from ctDNA released during tumor cell proliferation (18, 19). Thus, fragmentation analysis or selective sequencing strategies could be applied to distinguish clinically relevant tumor mutations from those present in clonal expansions of normal cells. This will be critical for the development of cfDNA-based methods for identification of patients with early-stage cancer.

Size selection could also have an impact on the detection of other types of DNA in body fluids or enrichment of signals from circulating bacterial or pathogen DNA and mitochondrial DNA. These DNA fragments are not associated with nucleosomes and are often highly fragmented below 100 bp. Filtering or selection of such fragments may prove to be important in light of the recently established link between the microbiome and treatment efficiency (17, 44). Moreover, recent work highlights a stronger correlation of ctDNA detection with cellular proliferation than with cell death (45). We hypothesize that the mode of the distribution of ctDNA fragment sizes at 145 bp could reflect cfDNA released during cell proliferation, and the fragments at 167 bp may reflect cfDNA released by apoptosis or maturation/turnover of blood cells. The effect of other cancer hallmarks (46) on ctDNA biology, structure, concentration, and release is yet unknown.

In summary, ctDNA fragment size analysis, via size selection and machine learning approaches, boosts noninvasive genomic analysis

of tumor DNA. Size selection of shorter plasma DNA fragments enriches ctDNA and assists in the identification of a greater number of genomic alterations with both targeted and untargeted sequencing at minimal additional cost. Combining cfDNA fragment size analysis and the detection of SCNAs with a nonlinear classification algorithm improved the discrimination between samples from patients with cancer and those from healthy individuals. Because the analysis of fragment sizes is based on the structural properties of ctDNA, size selection could be used with any downstream sequencing applications. Our work could help overcome current limitations of sensitivity for liquid biopsy, supporting expanded clinical and research applications. Our results indicate that exploiting the endogenous biological properties of cfDNA provides an alternative paradigm to deeper sequencing of ctDNA.

## MATERIALS AND METHODS

### Study design

Three hundred forty-four plasma samples from 200 patients with multiple cancer types were collected along with plasma from 65 healthy controls. Among the patients, 172 individuals, and notably the OV04 samples, were recruited through prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference number: 07/Q0106/63; and National Research Ethics Service Committee East of England—Cambridge Central 03/018). Written informed consent was obtained from all patients, and blood samples were collected before and after initiation of treatment with surgery or chemotherapeutic agents. DNA was extracted from 2 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (QIAGEN) or QIASymphony (QIAGEN) according to the manufacturer's instructions. In addition, 28 patients were recruited as part of the Copenhagen Prospective Personalized Oncology (CoPPO) program (PMID reference number: 25046202) at Rigshospitalet, Copenhagen, Denmark, approved by the local research ethics committee. Baseline tumor tissue biopsies were available from all 28 patients, together with rebiopsies collected at relapse from two patients, and matched plasma samples. Brain tumor patients were recruited at Addenbrooke's Hospital, Cambridge, UK as part of the BLING (biopsies of liquids in new gliomas) study (REC reference number: 15/EE/0094). Patients with bladder cancer were recruited at the Netherlands Cancer Institute, Amsterdam, The Netherlands, and approval according to national guidelines was obtained (N13KCM/CFMPB250) (47). Sixty-five plasma samples were obtained from healthy control individuals using a similar collection protocol (Seralab). Plasma samples have not been freeze thawed more than two times to reduce artificial fragmentation of cfDNA. A flowchart of the study is presented in fig. S1.

## SUPPLEMENTARY MATERIALS

[www.scientifictranslationalmedicine.org/cgi/content/full/10/466/eaat4921/DC1](http://www.scientifictranslationalmedicine.org/cgi/content/full/10/466/eaat4921/DC1)

Materials and Methods

Fig. S1. Flowchart summarizing the experiments performed in this study and the sample numbers used at each step.

Fig. S2. Size distribution of cfDNA determined by sWGS for different cancer types.

Fig. S3. Insert size distribution of mutant cfDNA determined with hybrid-capture sequencing for 19 patients.

Fig. S4. DNA fragment size distribution for plasma samples from patients with ovarian cancer.

Fig. S5. Quality control assessed for *in vitro* size selection.

Fig. S6. Quality control assessed for *in vitro* and *in silico* size selection on healthy control samples.

Fig. S7. SCNA analysis of the segmental  $\log_2$  ratio determined after sWGS (<0.4x coverage) for the patient OV04-83.

- Fig. S8. SCNA analysis of the segmental log<sub>2</sub> ratio determined after sWGS (<0.4× coverage) for plasma samples from patients with ovarian cancer (from the OV04 study).
- Fig. S9. MAF and t-MAD score compared for different cancer types.
- Fig. S10. t-MAD score measured on a plasma DNA dilution series.
- Fig. S11. t-MAD scores and fragmentation features compared to tumor volume.
- Fig. S12. Changes to t-MAD after in vitro size selection.
- Fig. S13. SCNA analysis in cfDNA from plasma samples collected at baseline and after treatment for 13 patients with HGSC.
- Fig. S14. MAF for SNVs called by WES with and without size selection.
- Fig. S15. TAm-Seq before and after in vitro size selection.
- Fig. S16. Mutations in clinically relevant genes detected by WES with and without in silico size selection.
- Fig. S17. Size distribution of nonmutant DNA and ctDNA concentration.
- Fig. S18. ROC curve for individual fragmentation features in high ctDNA cancers versus controls.
- Fig. S19. t-MAD score compared with seven fragmentation features.
- Fig. S20. Performance metrics for the two algorithms, LR and RF.
- Fig. S21. LR and RF models using the fragmentation features without t-MAD.
- Table S1. Summary table of the patients and samples included in this study.
- Table S2. Values for nine fragmentation features determined from sWGS data for the samples included in the study.
- Table S3. t-MAD score for the 48 plasma samples of the OV04 cohort before and after in vitro size selection.
- Table S4. Log<sub>2</sub> of the signal ratio observed by sWGS of the plasma samples from the OV04 cohort.
- Table S5. Mutations called by WES of six patients selected from the OV04 cohort.
- Table S6. Mutations called by WES data of the plasma samples from 16 patients from the CoPPO cohort.
- References (48, 49)

## REFERENCES AND NOTES

1. G. Siravagna, S. Marzoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).
2. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
3. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshaw, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108–112 (2013).
4. V. A. Adalsteinsson, G. Ha, S. S. Freeman, A. D. Choudhury, D. G. Stover, H. A. Parsons, G. Gyduch, S. C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. M. Francis, C.-Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M. S. Merrill, R. A. Santiago, E. P. O'Connor, S. H. Jeong, R. Leeson, R. M. Barry, J. F. Kramkowski, Z. Zhang, L. Polacek, J. G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N. M. Oliver, L. Marini, A. G. Waks, L. C. Harshman, S. M. Tolaney, E. M. Van Allen, E. P. Winer, N. U. Lin, M. Nakabayashi, M.-E. Taplin, C. M. Johannessen, L. A. Garraway, T. R. Golub, J. S. Boehm, N. Wagle, G. Getz, J. C. Love, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
5. E. Heitzer, P. Ulz, J. Belic, S. Gutschi, F. Quehenberger, K. Fischereider, T. Benezeder, M. Auer, C. Pischedler, S. Mannweiler, M. Pichler, F. Eisner, M. Haeusler, S. Riethdorf, K. Pantel, H. Samonigg, G. Hoeffer, H. Augustin, J. B. Geigl, M. R. Speicher, Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **5**, 30 (2013).
6. C. Bettogowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravagna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Oliví, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, I.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
7. 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).
8. S.-J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshaw, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
9. 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, Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
10. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmidier, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8**, 346ra92 (2016).
11. A. A. Chaudhuri, J. J. Chabon, A. F. Lovejoy, A. M. Newman, H. Stehr, T. D. Azad, M. S. Khodadoust, M. S. Esfahani, C. L. Liu, L. Zhou, F. Scherer, D. M. Kurtz, C. Say, J. N. Carter, D. J. Merriott, J. C. Dudley, M. S. Binkley, L. Modlin, S. K. Padua, M. F. Gensheimer, R. B. West, J. B. Shrager, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., A. A. Alizadeh, M. Diehn, Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov.* **7**, 1394–1403 (2017).
12. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afshari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettogowda, L. A. Diaz Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
13. I. S. Haque, O. Elemento, Challenges in using ctDNA to achieve early detection of cancer. *bioRxiv* 237578 [preprint]. 21 December 2017. <https://doi.org/10.1101/237578>
14. A. M. Newman, A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V. Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge Jr., J. B. Shrager, B. W. Loo Jr., J. W. Neal, H. A. Wakelee, M. Diehn, A. A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
15. P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G. Pristauz, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48**, 1273–1278 (2016).
16. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
17. P. Burnham, M. S. Kim, S. Agbor-Enoh, H. Luikart, H. A. Valentine, K. K. Khush, I. De Vlaminck, Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).
18. G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhour, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landén, M. Höglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Grönberg, C. M. Hultman, S. A. McCarroll, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
19. Y. Hu, B. Ulrich, J. Supplee, Y. Kuang, P. H. Lizotte, N. B. Feeney, N. M. Guibert, M. M. Awad, K.-K. Wong, P. A. Jänne, C. P. Paweletz, G. R. Oxnard, False-positive plasma genotyping due to clonal hematopoiesis. *Clin. Cancer Res.* **24**, 4437–4443 (2018).
20. A. J. Bronkhorst, J. F. Wentzel, J. Aucamp, E. van Dyk, L. du Plessis, P. J. Pretorius, Characterization of the cell-free DNA released by cultured cancer cells, *Biochim. Biophys. Acta* **1863**, 157–165 (2016).
21. S. Jahr, H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R.-D. Hesch, R. Knippers, DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).
22. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
23. D. Chandrananda, N. P. Thorne, M. Bahlo, High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genomics* **8**, 29 (2015).
24. P. Jiang, Y. M. D. Lo, The long and short of circulating cell-free DNA and the ins and outs of molecular diagnostics. *Trends Genet.* **32**, 360–371 (2016).
25. S. C. Y. Yu, K. C. A. Chan, Y. W. L. Zheng, P. Jiang, G. J. W. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. J. I. Go, J. M. G. van Vugt, R. Minekawa, C. B. M. Oudejans, K. H. Nicolaides, R. W. K. Chiu, Y. M. D. Lo, Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8583–8588 (2014).
26. F. M. F. Lun, N. B. Y. Tsui, K. C. A. Chan, T. Y. Leung, T. K. Lau, P. Charoenwan, K. C. K. Chow, W. Y. W. Lo, C. Wanapirak, T. Sanguansermsri, C. R. Cantor, R. W. K. Chiu, Y. M. D. Lo, Noninvasive prenatal diagnosis of monogenic diseases by digital size

- selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19920–19925 (2008).
27. G. Minarik, G. Repiska, M. Hyblová, E. Nagyova, K. Soltyš, J. Budis, F. Duris, R. Sysák, M. Gerykova Bujalkova, B. Vlkova-Israel, O. Biro, B. Nagy, T. Szemes, Utilization of benchtop next generation sequencing platforms ion torrent PGM and MiSeq in noninvasive prenatal testing for chromosome 21 trisomy and testing of impact of in silico and physical size selection on its analytical performance. *PLOS ONE* **10**, e0144811 (2015).
  28. M. B. Giacoma, G. C. Ruben, K. A. Iczkowski, T. B. Roos, D. M. Porter, G. D. Sorenson, Cell-free DNA in human blood plasma: Length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* **17**, 89–97 (1998).
  29. N. Umetani, A. E. Giuliano, S. H. Hiramatsu, F. Amersi, T. Nakagawa, S. Martino, D. S. B. Hoon, Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J. Clin. Oncol.* **24**, 4270–4276 (2006).
  30. F. Mouliere, B. Robert, E. Arnau Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, High fragmentation characterizes tumour-derived circulating DNA. *PLOS ONE* **6**, e23418 (2011).
  31. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
  32. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1317–E1325 (2015).
  33. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, Fragment length of circulating tumor DNA. *PLOS Genet.* **12**, e1006162 (2016).
  34. O. A. Zill, K. C. Banks, S. R. Fairclough, S. A. Mortimer, J. V. Vowles, R. Mokhtari, D. R. Gandara, P. C. Mack, J. I. Odegaard, R. J. Nagy, A. M. Baca, H. Eltoukhy, D. I. Chudova, R. B. Lanman, A. Talasz, The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin. Cancer Res.* **24**, 3528–3538 (2018).
  35. G. Macintyre, T. E. Goranova, D. De Silva, D. Ennis, A. M. Piskorz, M. Eldridge, D. Sie, L.-A. Lewisley, A. Hanif, C. Wilson, S. Dowson, R. M. Glasspool, M. Lockley, E. Brockbank, A. Montes, A. Walther, S. Sundar, R. Edmondson, G. D. Hall, A. Clamp, C. Gourley, M. Hall, C. Fotopoulos, H. Gabra, J. Paul, A. Supernat, D. Millan, A. Hoyle, G. Bryson, C. Nourse, L. Mincarelli, L. N. Sanchez, B. Ylstra, M. Jimenez-Linan, L. Moore, O. Hofmann, F. Markowitz, I. A. McNeish, J. D. Brenton, Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).
  36. C. A. Parkinson, D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld, J. D. Brenton, Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: A retrospective study. *PLOS Med.* **13**, e1002198 (2016).
  37. T. Forshaw, 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).
  38. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35**, 347–376 (2016).
  39. M. G. Best, N. Sol, B. A. Tannous, P. Wesseling, T. Wurdinger, RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).
  40. M. G. Best, N. Sol, S. G. J. G. In 't Veld, A. Vancura, M. Muller, A.-L. N. Niemeijer, A. V Fejes, L.-A. Tjon Kon Fat, A. E. Huis In 't Veld, C. Leurs, T. Y. Le Large, L. L. Meijer, I. E. Kooi, F. Rustenburg, P. Schellen, H. Verschueren, E. Post, L. E. Wedekind, J. Bracht, M. Esenkbrink, L. Wils, F. Favaro, J. Schoonhoven, J. Tannous, H. Meijers-Heijboer, G. Kazemier, E. Giovannetti, J. C. Reijneveld, S. Idema, J. Killestein, M. Heger, S. C. de Jager, R. T. Urbanus, I. E. Hoefer, G. Pasterkamp, C. Mannhalter, J. Gomez-Arroyo, H.-J. Bogaard, D. P. Noske, W. P. Vandertop, D. van den Broek, B. Ylstra, R. J. A. Nilsson, P. Wesseling, N. Karachaliou, R. Rosell, E. Lee-Lewandrowski, K. B. Lewandrowski, B. A. Tannous, A. J. de Langen, E. F. Smit, M. M. van den Heuvel, T. Wurdinger, Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* **32**, 238–252.e9 (2017).
  41. A. L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmamn-Groth, M. Schneider, T. Muley, M. Thomas, H. Sultmann, Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients. *Sci. Rep.* **6**, 33505 (2016).
  42. J. Belic, M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher, E. Heitzer, Rapid identification of plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. *Clin. Chem.* **61**, 838–849 (2015).
  43. D. G. Stover, H. A. Parsons, G. Ha, S. S. Freeman, W. T. Barry, H. Guo, A. D. Choudhury, G. Gyudush, S. C. Reed, J. Rhoades, D. Rotem, M. E. Hughes, D. A. Dillon, A. H. Partridge, N. Wagle, I. E. Krop, G. Getz, T. R. Golub, J. C. Love, E. P. Winer, S. M. Tolaney, N. U. Lin, V. A. Adalsteinsson, Association of cell-free DNA tumor fraction and somatic copy number alterations with survival in metastatic triple-negative breast cancer. *J. Clin. Oncol.* **36**, 543–553 (2018).
  44. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flamant, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelin, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J. R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).
  45. C. Abbott, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. James, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann; TRACERx consortium; PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
  46. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
  47. K. M. Patel, K. E. van der Vos, C. G. Smith, F. Mouliere, D. Tsui, J. Morris, D. Chandrananda, F. Marass, D. van den Broek, D. E. Neal, V. J. Gnanaprasgam, T. Forshaw, B. W. van Rhijn, C. E. Massie, N. Rosenfeld, M. S. van der Heijden, Association of plasma and urinary mutant DNA with clinical outcomes in muscle invasive bladder cancer. *Sci. Rep.* **7**, 5554 (2017).
  48. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
  49. I. Scheinin, D. Sie, H. Bengtsson, M. A. van de Wiel, A. B. Olshen, H. F. van Thuijl, H. F. van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson, B. Ylstra, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).

**Acknowledgments:** We would like to thank all members of the Rosenfeld Lab and Brenton Lab for their help and constructive discussion, in particular, M. Thompson, A. Ruiz-Valdepanas, J. P. Y. Chan, and A. L. Riediger. We would also like to thank the Cancer Research UK Cambridge Institute core facilities for their support, in particular, the genomics, bioinformatics, and biorepository facilities. Support is also acknowledged from the Cancer Research UK Cambridge Cancer Centre, the Cambridge Experimental Cancer Medicine Centre (ECMC), Cancer Molecular Diagnostics Laboratory (CMDL), and NIHR Biomedical Research Centre (BRC). We would like to acknowledge our patients and caregivers and the help and support of the research nurses, trial staff, and the staff at Addenbrooke's Hospital and Rigshospitalet. In particular, we would like to acknowledge C. Hodgkin, H. Biggs, and K. Hosking. We would like to thank H. Carr and AstraZeneca for support for the CALIBRATE study. **Funding:** We would like to acknowledge the support of the University of Cambridge, Cancer Research UK, and the EPSRC [CRUK grant numbers A11906 (to N.R.), A20240 (to N.R.), A22905 (to J.D.B.), A15601 (to J.D.B.), A25177 (CRUK Cancer Centre Cambridge), A17242 (to K.M.B.), and A16465 (CRUK-EPSRC Imaging Centre in Cambridge and Manchester)]. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 337905. The research was supported by the National Institute for Health Research Cambridge, National Cancer Research Network, Cambridge Experimental Cancer Medicine Centre, and Hutchison Whampoa Limited. This research is also supported by Target Ovarian Cancer and the Medical Research Council through their Joint Clinical Research Training Fellowship for E.K.M. The CALIBRATE study was supported by funding from AstraZeneca. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** F. Mouliere, A.M.P., D.C.,

E.K.M., J.D.B., and N.R. conceptualized and designed the study. F. Mouliere, A.M.P., E.K.M., L.B.A., K.H., C.G.S., J.C.M.W., D.G., R.M., T.G., A.S., I.G., O.Ø., C.A.P., M.M.-S., I.H., K.P., C.E.M., and W.N.C. performed experiments and collected data. F. Mouliere, A.M.P., D.C., E.K.M., and C.G.S. conceptualized the size selection approach. F. Mouliere, A.M.P., and E.K.M. designed and performed in vitro size selection. F. Mouliere and D.C. conceptualized and designed the fragmentation feature analysis, with input from F. Marass and N.R. D.C. conceptualized and designed the t-MAD index with input from F. Mouliere. F. Mouliere and D.C. carried out bioinformatics analysis of SCNA from sWGS. J.M. performed bioinformatics analysis of TAm-Seq. F. Mouliere and L.B.A. designed the tailored captured sequencing and performed WES. F. Mouliere and J.M. performed bioinformatics analysis of the capture sequencing and WES. M.D.E. developed and optimized mutation calling algorithms. R.M., K.M.B., and S.R. designed the animal model. J.G.-C., S.P., R.D.B., M.M.-S., G.D.S., J.B., S.M., P.C., C.W., R.M., and M.S.v.d.H. have collected human samples. M.J.-L. and J.B. performed histopathology revision. F. Mouliere, D.C., A.M.P., E.K.M., J.D.B., and N.R. wrote the manuscript. All authors have critically reviewed the manuscript. F. Mouliere, A.M.P., D.C., J.D.B., and N.R. supervised the study. F. Mouliere coordinated the study. **Competing interests:** N.R., J.D.B., and D.G. are cofounders, shareholders, and officers/consultants of Invata Ltd., a cancer genomics company that commercializes ctDNA analysis. Invata Ltd. had no role in the conceptualization, study design, data collection and analysis, and decision to publish or preparation of the manuscript. J.D.B. received research funding from Aprea and NCI and has received advisory board fees from AstraZeneca. F. Marass and N.R. are co-inventors of patent WO/2016/009224 on "A method for detecting a genetic variant." F. Mouliere, J.C.M.W., K.H.,

C.E.M., C.G.S., N.R., and other authors may be listed as co-inventors on patent application number 1803596.4 on "Improvements in variant detection" and other potential patents describing methods for the analysis of DNA fragments and applications of ctDNA. I.G. is currently an employee of Novartis AG, a relationship that started after all his work contributing to this manuscript had been completed. Novartis had no role in the work presented in this manuscript. Other authors declare that they have no competing interests.

**Data and materials availability:** Sequencing data for this study are deposited in the EGA database (accession number EGAS00001003258). Other data associated with this study are present in the paper or the Supplementary Materials.

Submitted 29 March 2018

Accepted 17 October 2018

Published 7 November 2018

10.1126/scitranslmed.aat4921

**Citation:** F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlbom, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).

# Science Translational Medicine

## Enhanced detection of circulating tumor DNA by fragment size analysis

Florent Mouliere, Dineika Chandrananda, Anna M. Piskorz, Elizabeth K. Moore, James Morris, Lise Barlebo Ahlbom, Richard Mair, Teodora Goranova, Francesco Marass, Katrin Heider, Jonathan C. M. Wan, Anna Supernat, Irena Hudecova, Ioannis Gounaris, Susana Ros, Mercedes Jimenez-Linan, Javier Garcia-Corbacho, Keval Patel, Olga Østrup, Suzanne Murphy, Matthew D. Eldridge, Davina Gale, Grant D. Stewart, Johanna Burge, Wendy N. Cooper, Michiel S. van der Heijden, Charles E. Massie, Colin Watts, Pippa Corrie, Simon Pacey, Kevin M. Brindle, Richard D. Baird, Morten Mau-Sørensen, Christine A. Parkinson, Christopher G. Smith, James D. Brenton and Nitzan Rosenfeld

*Sci Transl Med* **10**, eaat4921.  
DOI: 10.1126/scitranslmed.aat4921

### Sizing up circulating tumor DNA

Unlike solid tumors, which are often hidden deep within a patient's body, a patient's blood is easy to access safely. As a result, liquid biopsy, the analysis of tumor DNA in the blood, is an attractive alternative to conventional biopsy. Unfortunately, tumor DNA molecules are usually vastly outnumbered by the fragments of noncancer DNA in the blood, and detecting them can be a challenge, especially in early stages of cancer. Mouliere *et al.* identified characteristic differences in the size distribution of tumor-derived and noncancer DNA fragments and then used these observations to design a method of tumor DNA detection with greater sensitivity.

#### ARTICLE TOOLS

<http://stm.sciencemag.org/content/10/466/eaat4921>

#### SUPPLEMENTARY MATERIALS

<http://stm.sciencemag.org/content/suppl/2018/11/05/10.466.eaat4921.DC1>

#### RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/10/466/eaav3873.full>  
<http://stm.sciencemag.org/content/scitransmed/9/403/eaan2415.full>  
<http://stm.sciencemag.org/content/scitransmed/8/364/364ra155.full>  
<http://stm.sciencemag.org/content/scitransmed/8/346/346ra92.full>  
<http://stm.sciencemag.org/content/scitransmed/11/504/eaax7392.full>  
<http://stm.sciencemag.org/content/scitransmed/11/507/eaay1984.full>  
<http://stm.sciencemag.org/content/scitransmed/12/524/eaax7533.full>  
<http://science.sciencemag.org/content/sci/369/6499/eabb9601.full>  
<http://science.sciencemag.org/content/sci/368/6490/461.full>  
<http://science.sciencemag.org/content/sci/368/6491/589.full>  
<http://stm.sciencemag.org/content/scitransmed/12/548/eaaz8084.full>  
<http://stm.sciencemag.org/content/scitransmed/12/557/eaaz3738.full>

#### REFERENCES

This article cites 48 articles, 17 of which you can access for free  
<http://stm.sciencemag.org/content/10/466/eaat4921#BIBL>

#### PERMISSIONS

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

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. The title *Science Translational Medicine* is a registered trademark of AAAS.

Copyright © 2018 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works