Detection of mutational patterns in cell-free DNA of colorectal cancer by custom amplicon sequencing

Simon Herrmann[1,2,3], Tianzuo Zhan[1,2,3], Johannes Betge[1,2,3], Benedikt Rauscher[1,2], Sebastian Belle[3], Tobias Gutting[3], Nadine Schulte[3], Ralf Jesenofsky[3], Nicolai Härtel[3], Timo Gaiser[4], Ralf-Dieter Hofheinz[5], Matthias P. Ebert[3] and Michael Boutros[1,2,6]

1 Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany
2 Department Cell and Molecular Biology, Faculty of Medicine Mannheim, Heidelberg University, Germany
3 Department of Internal Medicine II, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
4 Medical Faculty Mannheim, Institute of Pathology, Heidelberg University, Mannheim, Germany
5 Medical Faculty Mannheim, Interdisciplinary Tumor Centre, Heidelberg University, Mannheim, Germany
6 German Cancer Consortium (DKTK), Heidelberg, Germany

Correspondence

M. Boutros, Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ) and Heidelberg University, Im Neuenheimer Feld 580, Heidelberg D-69120, Germany
Tel: +49 (0) 6221 42-1950
Fax: +49 6221 421959
E-mail: m.boutros@dkfz-heidelberg.de
and
M. P. Ebert, Department of Internal Medicine II, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, Mannheim D-68167, Germany
Tel: +49 (0) 621 383-3284
Fax: +49 621 3833805
E-mail: matthias.ebert@medma.uni-heidelberg.de

Simon Herrmann, Tianzuo Zhan, Johannes Betge, and Benedikt Rauscher contributed equally to this article

Monitoring the mutational patterns of solid tumors during cancer therapy is a major challenge in oncology. Analysis of mutations in cell-free (cf) DNA offers a noninvasive approach to detect mutations that may be prognostic for disease survival or predictive for primary or secondary drug resistance. A main challenge for the application of cfDNA as a diagnostic tool is the diverse mutational landscape of cancer. Here, we developed a flexible end-to-end experimental and bioinformatic workflow to analyze mutations in cfDNA using custom amplicon sequencing. Our approach relies on open-software tools to select primers suitable for multiplex PCR using minimal cfDNA as input. In addition, we developed a robust linear model to identify specific genetic alterations from sequencing data of cfDNA. We used our workflow to design a custom amplicon panel suitable for detection of hotspot mutations relevant for colorectal cancer and analyzed mutations in serial cfDNA samples from a pilot cohort of 34 patients with advanced colorectal cancer. Using our method, we could detect recurrent and patient-specific mutational patterns in the majority of patients. Furthermore, we show that dynamic changes of mutant allele frequencies in cfDNA correlate well with disease progression. Finally, we demonstrate that sequencing of cfDNA can reveal mechanisms of resistance to anti-Epidermal Growth Factor Receptor(EGFR) antibody treatment. Thus, our approach offers a simple and highly customizable method to explore genetic alterations in cfDNA.

1. Introduction

Although survival rates have been increasing over the last decades, the prognosis of patients with advanced stages of gastrointestinal cancer is persistently poor (Jemal *et al.*, 2011). Major challenges include tumor heterogeneity and the clonal expansion of resistant clones during the course of treatment, resulting in cancer progression or relapse (Alizadeh *et al.*, 2015). Monitoring the dynamics of cancer evolution during the course of a treatment may help to early identify the development of drug resistance. However, in patients with advanced disease, obtaining serial tumor biopsies to study tissue-based biomarkers is not feasible in the majority of cases. Most biopsy techniques bear risks due to their invasiveness and are not suitable for morbid patients (Vanderlaan *et al.*, 2014). Besides, a tissue biopsy only represents a small section of a heterogenous and disseminated disease, ignoring cancer biology at other (metastatic) sites. Hence, it may result in sampling errors due to branched evolution of metastatic sites and intratumoral heterogeneity (Burrell *et al.*, 2013; Gerlinger *et al.*, 2012).

It is known for decades that solid tumors can release single tumor cells (Alix-Panabières and Pantel, 2013; Maheswaran and Haber, 2010) or nucleic acids into the bloodstream, including DNA, mRNA, and miRNA (Schwarzenbach *et al.*, 2011; Stroun *et al.*, 2006). Advances in DNA sequencing opened up opportunities to analyze the genetic composition of this circulating tumor DNA (ctDNA) (Wan *et al.*, 2017). In fact, large-scale sequencing projects of ctDNA showed that genetic alterations present in primary tumors can also be found in ctDNA (Strickler *et al.*, 2018; Zill *et al.*, 2018). Furthermore, serial ctDNA analysis proves to be a powerful method to longitudinally monitor therapy response and the emergence of resistant clones (Bettegowda *et al.*, 2014; Siravegna *et al.*, 2015; Thierry *et al.*, 2017). Several methods have been developed to detect mutations in ctDNA, including BEAMing technologies (Diaz *et al.*, 2012; Haselmann *et al.*, 2018; Janku *et al.*, 2015), panel sequencing of cancer-associated genes (Strickler *et al.*, 2018; Zill *et al.*, 2018), or targeted amplicon sequencing (Lebofsky *et al.*, 2015; Tie *et al.*, 2016). While every method offers advantages, a major challenge to all technologies is the divergent mutational landscape of cancer. While some mutations in key oncogenes occur at high frequency across many tumors, most genetic alterations are tissue-specific or even highly variable between cancers of the same tissue origin (Lanman *et al.*, 2015). Current FDA-approved diagnostic tests for cell-free DNA (cfDNA) cover mutational hotspots of large predefined panels of frequently altered genes in cancer (Lanman *et al.*, 2015). While this approach is suitable for application in a broad spectrum of cancers, it is not designed to detect potentially relevant mutations that occur only in specific tumors, such as RNF43 in colorectal (CRC) and pancreatic cancer (Steinhart *et al.*, 2017). Thus, complementary tests are required that can be used to explore a custom-defined set of genetic alterations in cfDNA that are not covered by commercially available tests. These tests are valuable to correlate the occurrence of rare mutations with disease outcome or drug resistance.

Here, we present an end-to-end experimental and bioinformatic workflow that can be used to analyze a custom set of genetic alterations in cfDNA. We designed an amplicon panel that is tailored to the genetic landscape of CRC, covering mutations in oncogenes and tumor suppressors as well as functionally relevant mutations in genes that are less frequently altered. We performed amplicon sequencing with our custom panel in serial cfDNA samples from a pilot cohort of 34 patients with metastatic CRC. Using this approach, we detected patient-specific mutational patterns that reoccurred in serial samples. We also show that the mutant allele frequencies of genetic alterations in cfDNA increase with disease progression.

2. Materials and methods

2.1. Study design and clinical databank

Patients with gastrointestinal cancers across all tumor stages were included in the study. Blood was collected before start of anticancer treatment and every 3–6 months during radiologic restaging or if patients were hospitalized. Simultaneously, the following blood markers were determined: cell counts of erythrocytes, leukocytes, and thrombocytes; AST; ALT; GGT; ALP; bilirubin; albumin; creatinine; INR; LDH; CRP; and the serum tumor markers CEA, CA19-9, and AFP. Clinical characteristics were collected in a prospective database.

Written informed consent was obtained from all patients, and the study was approved by the local ethics board (Ethikkommission II, Medical Faculty Mannheim, Heidelberg University, identifier 2013-640N-MA). The study design is in accordance with the standards proposed by the Declaration of Helsinki.

2.2. Sample collection and plasma DNA extraction

Two 10-mL K_3EDTA tubes were taken during regular blood sampling and before administration of

chemotherapy. Blood samples were immediately stored at 4 °C and processed within 16 h to minimize release of genomic DNA from nontumorous sources (Norton *et al.*, 2013; Risberg *et al.*, 2018; Wong *et al.*, 2013). For isolation of blood plasma, K_3EDTA tubes were centrifuged at 1900 *g* for 10 min at 4 °C. Afterwards, supernatants were centrifuged at 16 000 *g* for 10 min at 4 °C to remove remaining blood cells. Plasma was harvested without mobilizing the cell pellet and stored at −80 °C until needed. For cfDNA isolation, 1–4 mL of frozen plasma was used. Extraction of DNA was performed using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). DNA was dissolved in 50–70 μL nuclease-free water and stored at −20 °C for further use. The amount of DNA was quantified with QUBIT 2.0 (Life Technologies, Carlsbad, CA, USA) using the high-sensitivity assay according to the manufacturer's protocol.

2.3. Design of amplicon primers and multiplex reaction

Positions of frequently mutated hotspots in CRC-associated genes were obtained from the COSMIC database (Forbes *et al.*, 2015). In addition, frequently mutated loci in genes that regulate the metabolism of 5-fluorouracil (*DPYD*, *TYMS*, *TYMP*, *UPP1*) and mismatch repair (*MLH1*, *MLH3*, *MSH6*, *PMS2*) were selected. Nucleotide sequences surrounding the genetic alteration of interest were obtained from the NCBI database with the genome assembly GRCH38 as a reference. Specific primers for regions that cover the selected mutations were designed with Primer3 (Untergasser *et al.*, 2012) with the following settings: primer length of 19–27 bp, GC ratio of 30–63%, and melting temperature of 55–61.5 °C. The universal adapter sequence TCCCTACACGACGCTCTTCCGATCT was added to the 5′ end of each forward primer. To each reverse primer, the sequence AGTTCAGACGTGTGCTCTTCCGATC was added to the 5′ end. The amplicon length varied between 100 and 175 bp. Characteristics and sequences of all primers can be found in Table S1. To develop a multiplex PCR assay, we used the software tool MULTIPLX 2.1 (Kaplinski and Remm, 2015) to identify appropriate primer combinations. To this end, we set the concentration of monovalent salts to 50 mM and the concentration of Mg^{2+} to 1.5 mM. All five possible primer interaction scores that might affect primer compatibility were calculated. To group the 43 primer pairs for multiplex PCR, we used the stringency value 'normal'. The best combination resulted in six different pools with 5–9 primer pairs.

2.4. Nested PCR protocol and library preparation

We developed a custom, nested PCR protocol suited for multiplex reactions. For the first PCR step, at least 500 pg of cfDNA was used per multiplexed reaction. In all cases, we used Q5 Hot Start Polymerase (NEB, Ipswich, MA, USA) to minimize amplification errors. To reduce the formation of primer-dimers, we selected a concentration of 40 nM for each primer pair. The PCR protocol was adapted from a previous study (Ståhlberg *et al.*, 2017). In brief, after an initial denaturation step (3 min, 98 °C), 18 PCR cycles were run with the following settings: 10 s of denaturation at 98 °C, 6 min of annealing at 62 °C, and 30 s of elongation at 72 °C. In the end, a 10-min elongation step at 72 °C was included. The PCR products were semi-automatically purified with a Biomek FXP (Beckman Coulter) using AMPure XP beads (Beckman Coulter, Munich, Germany) with a bead-to-sample ratio of 0.76 and eluted in 25 μL of nuclease-free water. By this step, fragments larger than 120 bp were separated from primer-dimers. In the second PCR step, the same settings were applied except for the following modifications: First, we used 15 μL of DNA products from the first PCR step as input. Second, primers consisting of Illumina universal adapter sequence and a unique combination of the TruSeq DNA HT indexes (forward primer: AATGATACGGCGAC-CACCGAGATCTACAC-(Index)-ACACTCTTTCCC-TACACGACGCTCTTCCGATCT; reverse primer: CAAGCAGAAGACGGCATACGAGAT-(Index)-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC) were used at a concentration of 300 nM. Third, the annealing temperature was increased to 72 °C and the annealing time was reduced to 15 s. After the second PCR step, all six multiplexed PCRs of each patient sample were pooled. Ninety micro litre of the mixed PCR products was then cleaned with AMPure XP beads using a bead-to-sample ratio of 0.67 and eluted in 20 μL of nuclease-free water.

2.5. Amplicon sequencing

All patient samples were pooled to a single library with a final concentration of 4 nM. The library was sequenced on an Illumina MiSeq using the MiSeq Reagent Kit V2 (Illumina Inc., San Diego, CA, USA) (150 bp single-end sequencing) with 5% PhiX as spike-in. Between 30 and 50 patient samples were sequenced in one MiSeq run.

2.6. Analysis of sequencing data and variant calling

First, a quality report was generated for each sequenced sample using 'FASTQC' (Andrew, 2010).

Quality reports were examined manually to ensure sufficient sample quality. Subsequently, the 'TRIMMO-MATIC' software version 0.36 (Bolger *et al.*, 2014) was used to remove sequenced base pairs with a quality score of < 15 (PHRED33) at both ends of the reads to avoid false-positive mutations due to sequencing errors. In addition, a sliding window trimming was performed cutting the read once the average quality within a window of size 4 bp was detected to be < 15. Reads with a length of < 70 base pairs were removed. Next, reads were mapped to the human genome GRCh38 using the 'bwa mem' algorithm of the 'BWA' alignment software version 0.7.15-r1140 (Li and Durbin, 2010) with default parameters. 'SAM-TOOLS' version 1.4 (Li *et al.*, 2009) was used to remove reads that could not be mapped to genome. In addition, reads mapped with a quality of < 13 were excluded. To determine variants, the 'mpileup' algorithm implemented in the 'SAMTOOLS' software was applied. The resulting variant calls were summarized and quantified using 'bam-readcount' (https://github.com/genome/bam-readcount), counting only variants at positions with a sequencing quality of at least 30.

2.7. Analysis of sequencing variants

The allele frequency was determined for each detected variant by dividing the number of reads containing the variant by the total number of reads mapping to that position. Subsequently, variants were mapped to the amplicon panel. To this end, the genomic coordinates were determined for each amplicon using the 'BLASTn' algorithm as implemented in the 'BLAST+' software package (Altschul *et al.*, 1990). Here, each amplicon sequence was mapped to the human genome GRCh38 requiring 100% sequence identity. Sequencing variants were then matched to the custom amplicon panel by genomic coordinates. Correctness of the variant mapping and matching was assessed by taking advantage of prior knowledge about mutations in the *NRAS*, *KRAS,* and *BRAF* oncogenes that had previously been determined using Sanger sequencing. In order to distinguish true mutations from false-positive mutations introduced by, for example, PCR or sequencing errors, a model-based approach was applied. Assuming that (a) the majority of variants are caused by PCR or sequencing errors, that (b) PCR errors do on average occur at the same frequency at each position of an amplicon, and that (c) sequencing errors at a specific genomic position are equally likely in each sample. A robust linear model of the form

$$\log\left(\frac{a}{1 - a}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

was fit for each amplicon, where a is the allele frequency of a variant, X_1 is the variant's position on the amplicon, and X_2 represents the sequenced patient sample. The resulting fit represents a noise model. A variant was considered a true mutation if its model residual exceeded the median of all residuals by at least three standard deviations, indicating that its presence cannot be explained by the estimated noise present in the data. In addition, we required a minimum allele frequency of at least 0.5% for a variant to be considered a true clinically relevant mutation. Next, the remaining mutations were annotated using the COSMIC database of somatic mutations in cancer (Forbes *et al.*, 2015). Mutations that were not listed in COSMIC were excluded from further analysis. In addition, COSMIC mutations marked as SNPs were excluded.

2.8. Statistics

All *P*-values reported in this study were computed using a two-sample Wilcoxon rank sum test as implemented in the R statistical programming language (R Core Team, 2016). As metrics for the relationships between quantitative variables, both the parametric Pearson correlation coefficient and the rank-based Spearman correlation coefficient are reported in all cases.

2.9. Data and software availability

Documented computer code to reproduce analyses presented in this study is available from GitHub at https://github.com/boutroslab/Supplemental-Material/tree/master/Herrmann%26Zhan%26Betge%26Rauscher_2018.

3. Results

3.1. cfDNA levels correlate with clinical parameters in CRC patients

To study the value of cfDNA as a biomarker in gastrointestinal cancers, we prospectively collected serial blood samples from cancer patients at a tertiary university hospital in Germany (Mannheim Liquid Biopsy Unit—MALIBU, University Hospital Mannheim, Heidelberg University) (Fig. 1A). For patients undergoing palliative chemotherapy, we obtained blood samples
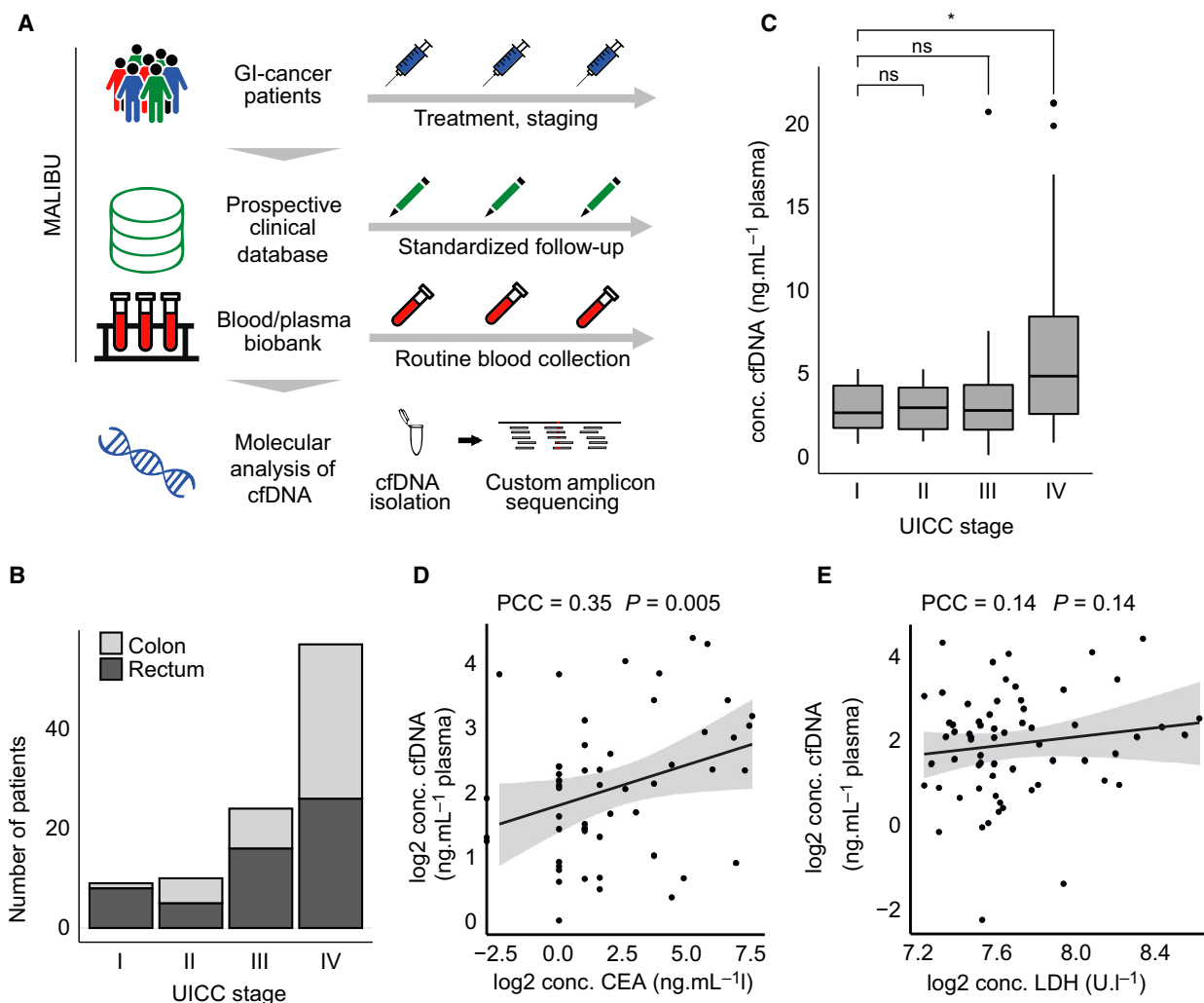
Fig. 1. Standardized collection of blood plasma and clinical data from patients with colorectal cancer reveals stage-dependent levels of cell-free DNA (cfDNA). (A) The Mannheim Liquid Biopsy Unit (MALIBU) prospectively collects blood and plasma, as well as clinical data from patients with gastrointestinal tumors. Blood is collected at regular time intervals during the course of treatment and follow-up. The plasma is used for isolation of cfDNA and subsequently for custom amplicon sequencing. (B) Clinical characteristics (location and UICC/AJCC stage) of patients with colorectal cancer were included in the MALIBU biobank and database. (C) Patients with stage IV colorectal cancer have significantly higher cfDNA levels than nonmetastasized patients. cfDNA was isolated from 1-4 mL plasma of each patient and quantified with QUBIT. * p < 0.05, two-sided *t*-test. (D, E) Association of cfDNA levels with levels of tumor markers. cfDNA levels are significantly associated with CEA levels ($P = 0.005$) (D), but not with LDH levels (E).

prior to start of treatment and in parallel to each radiologic assessment of therapy response. In all cases, blood sampling was performed before administration of anticancer drugs. For patients undergoing surgical removal of the cancer, we collected blood samples prior and at multiple time points after resection. In parallel, we documented relevant clinical parameters and blood markers in a prospective database. Plasma from each blood sample was isolated within 16 h after sampling and stored at −80 °C until further use. Most samples stored in the biobank were derived from

patients with CRC. In total, we collected blood samples from 104 CRC patients across all UICC stages, but predominantly from patients with metastatic CRC (Fig. 1B). To correlate levels of cfDNA with clinical parameters, we measured cfDNA concentrations from blood samples of all CRC patients. Comparison of cfDNA levels demonstrated significantly elevated levels in patients with metastatic compared to nonmetastatic tumors (Fig. 1C). In contrast, no significant difference was found between cfDNA levels of patients with UICC stage I to III cancers. Next, we analyzed

the correlation of cfDNA levels with two biomarkers commonly used in gastrointestinal oncology, carcinoembryonic antigen (CEA), and lactate dehydrogenase (LDH) (Fig. 1D,E). We observed a weak, but significant ($P = 0.005$; Pearson's correlation test) correlation of cfDNA concentration with CEA levels. In contrast, no correlation was found with LDH levels. This finding suggests that the cfDNA concentration may be a biological marker that occurs independently of tumor cell necrosis.

3.2. A pipeline for the design and analysis of custom amplicon panels for cfDNA

To analyze mutational patterns in cfDNA, we developed a pipeline for the design of custom amplicon panels. We selected primers that cover mutational hotspots in oncogenes and tumor suppressors of CRC based on the COSMIC database (Forbes *et al.*, 2015). In addition, we included primers that amplify frequently mutated loci in genes that affect metabolism of 5-fluorouracil (Jennings *et al.*, 2012; Ooyama *et al.*, 2007; Pullarkat *et al.*, 2001) or DNA mismatch repair (Li, 2008). These mutations occur at a lower frequency and are not commonly covered by commercial amplicon

panels. An overview of all genes can be found in Fig. 2A. We then designed primer pairs that bind 50–75 bp up- and downstream of the mutations of interest and calculated optimal multiplexes (see Section 2 for details). Our custom panel contained a total of 43 primer pairs, which were distributed to six multiplex reactions containing 5–9 primer pairs each (see Table S1 for primer sequences and associated mutations). We tested different amounts of cfDNA as input and found that 500 pg per multiplex reaction was sufficient for our nested PCR protocol (see overview in Fig. 2B and Section 2). The purified and pooled amplicons were sequenced on an Illumina MiSeq by 150 bp single-end sequencing. In parallel, we developed a custom bioinformatic pipeline for the analysis of sequencing results (Fig. 3A). This pipeline comprises four consecutive steps. First, in the quality control step, reads with low sequencing quality are discarded. In the second step, all nucleotide variants across all reads of the same amplicon are detected and quantified. To discriminate PCR or sequencing artifacts from true genetic variants, we developed a robust linear regression model that we applied to each individual amplicon. Alterations identified as true variants by the model were then matched to the COSMIC database and discarded if not covered.
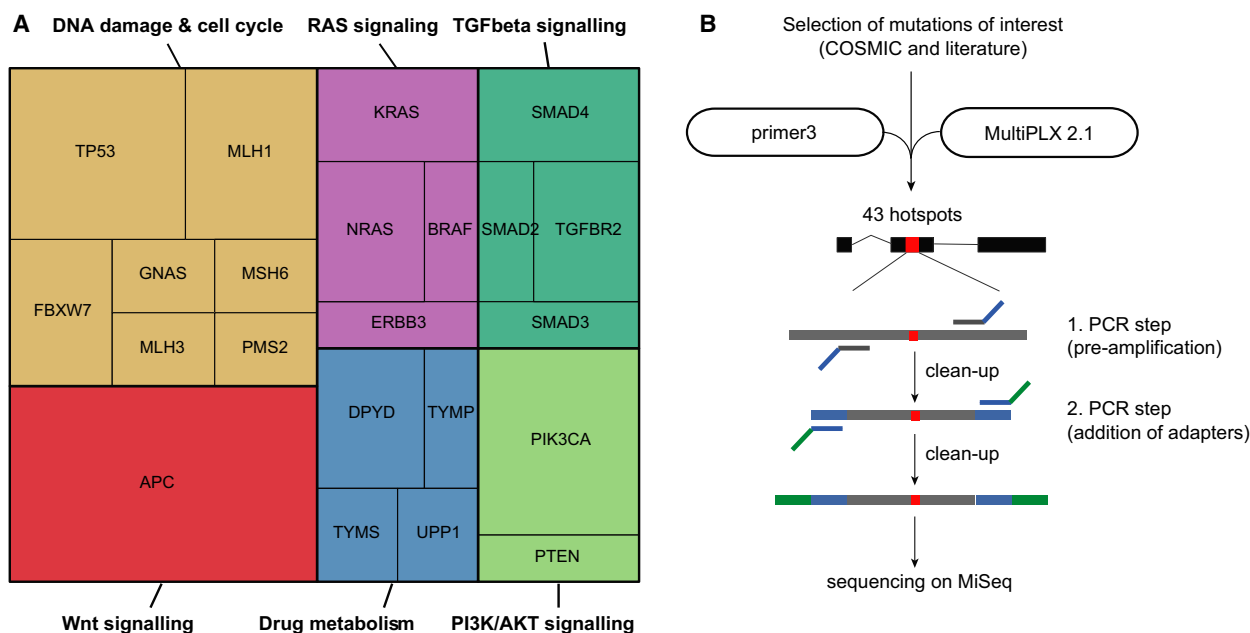
Fig. 2. Detection of recurrent mutations in cfDNA with a custom amplicon sequencing workflow. (A) Overview of genes included in the custom amplicon panel. A total number of 43 amplicons were designed to detect hotspot mutations in 18 genes frequently altered in colorectal cancer, including genes related to DNA damage, cell cycle, RAS signaling, TGF beta signaling, Wnt signaling, PI3K/AKT signaling, as well as four genes related to drug metabolism. (B) Experimental workflow of custom amplicon sequencing. Mutational hotspots of interest were selected in COSMIC database and literature, amplicon primers were then designed with Primer3, and multiplexes were determined with MULTIPLX 2.1. A two-step nested PCR workflow was established, including pre-amplification and addition of sequencing adapters before PCR clean-up and subsequent sequencing on MiSeq (Illumina).
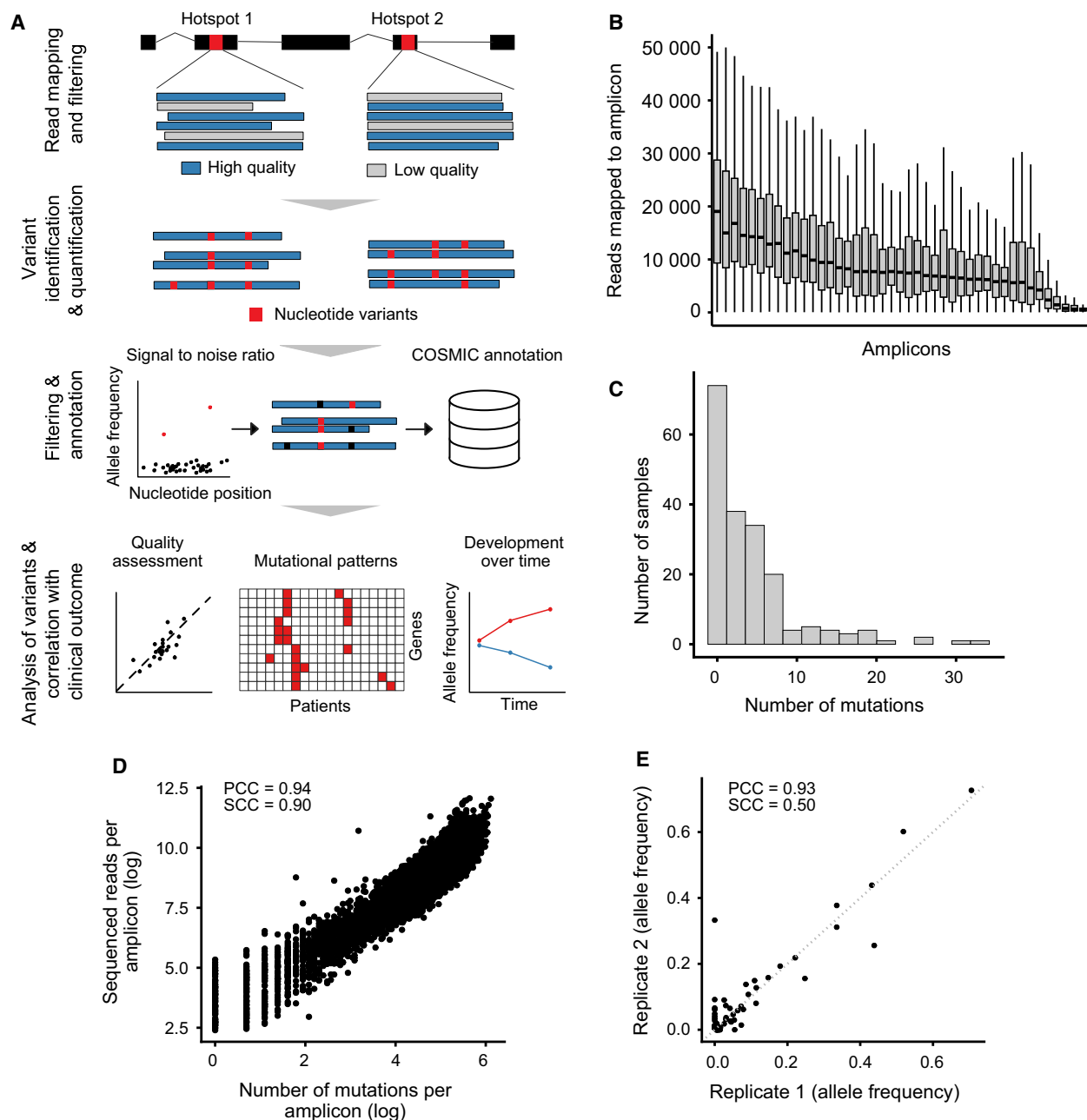
Fig. 3. An automated workflow for bioinformatic analysis of amplicon sequencing data from cfDNA. (A) Bioinformatic analysis workflow. Reads generated with Illumina MiSeq were mapped to the human reference genome, and low-quality reads were excluded before variant identification by comparison of reads and reference. Allele frequency was calculated, and variants were filtered by signal-to-noise analysis. Subsequently, variants were annotated with metadata from public databases. Importantly, variants not present in the COSMIC database were filtered out. With this set of mutations, further analyses regarding mutational patterns or the development of mutations over time were performed. (B) Distribution of reads on amplicons. The number of reads mapped to each amplicon region is shown for all 43 amplicons. Data from all cfDNA samples were used for this analysis. (C) Number of mutations per sample. Only mutations with positive signal-to-noise ratio and present in COSMIC database were considered true somatic mutations. (D) Association between number of mutations found per amplicon and sequencing depth. Raw mutations were used for this analysis without any filter steps. (E) Correlation of replicates. Two independent replicates were analyzed, starting from cfDNA isolation from the plasma. Shown are the allele frequencies of detected mutations in replicate 1 vs. replicates 2 of 19 analyzed patients.

By this bioinformatic approach, we assured that only highly confident alterations previously found in cancer tissues are reported.

3.3. Functional validation of the amplicon sequencing assay

We tested our amplicon panel in genomic DNA of cancer tissue and cell lines, as well as cfDNA of patients with CRC. First, we evaluated the performance of individual amplicons with our panel. Mapping of sequencing reads to the amplicons showed a continuous distribution of sequencing depth between amplicons (median 7576, minimum 503, maximum 20 809) (Fig. 3B). We found that the number of mutations identified varied between the different samples (Fig. 3C). Also, we noticed a correlation between number of sequenced reads and number of mutations per amplicon (Fig. 3D). Next, we tested the sensitivity of our method to detect specific alterations. To this end, we mixed sheared genomic DNA (fragment size 200–300 bp) from two CRC cell lines (HT29, HCT116) in different ratios. Both cell lines have distinct, monoallelic genetic alterations that can be detected with our amplicon panel. As shown in Fig. S1A, we could reliably identify mutant alleles of HCT116 at a concentration of 1.25% in the total genomic DNA mix, which corresponds to a mutant allele frequency of approximately 0.6%.

We then used our amplicon panel on genomic DNA from formalin-fixed paraffin-embedded primary CRC tissues (Fig. S1B). We could detect all RAS mutations previously found by Sanger sequencing, except for two KRAS mutations in Exon 4, which were not covered by the panel. Finally, we tested the reproducibility of our experimental and analysis pipeline by performing amplicon sequencing on independent biological replicates of 19 cfDNA samples, starting from cfDNA isolation from plasma. We show that mutations that occurred at an allele frequency > 0.5% could be detected as indicated by a strong correlation between mutant allele frequencies of the same genetic alterations from independent replicates (Fig. 3E). In summary, our quality control experiments indicate that our custom amplicon sequencing method is functional in both genomic and cfDNA. Furthermore, we identified thresholds for mutant allele frequencies that enable the detection of genetic alterations with high confidence.

3.4. Mutational patterns in cfDNA are patient-specific and highly recurrent

Next, we used our custom amplicon panel to analyze mutational patterns in a cohort of 34 patients with

Table 1. Patient, treatment, and tumor baseline characteristics.[a] 5-FU, 5-fluorouracil or capecitabine; FOLFOX, 5-FU + oxaliplatin; CAPOX, capecitabine + oxaliplatin; FOLFIRI, 5-FU + irinotecan; FOLFIRINOX, 5-FU + oxaliplatin + irinotecan

Parameter	All CRC, $n = 104$	mCRC cohort, $n = 34$
Gender		
Male	74 (71%)	25 (74%)
Female	30 (29%)	9 (26%)
Age (years)	65 (34–88)	65 (36–88)
Stage		
I	9 (9%)	
II	10 (10%)	
III	25 (24%)	3 (9%)[b]
IV	60 (58%)	31 (91%)
Tumor location		
Colon cancer	45 (43%)	15 (44%)
Rectal cancer	58 (56%)	19 (56%)
Both entities	1 (1%)	
Therapy setting		
Surveillance	24 (23%)	
Adjuvant/neoadjuvant	21 (20%)	
Palliative	59 (57%)	34 (100%)
Metastases (only stage IV)		
Hepatic	47 (78%)	24 (71%)
Pulmonary	28 (47%)	18 (53%)
Peritoneal carcinomatosis	14 (23%)	4 (12%)
Pleural carcinomatosis	5 (8%)	2 (6%)
Osseous	4 (7%)	1 (3%)
Other	5 (8%)	0 (0%)
Therapy regimen		
5-FU	17 (16%)	6 (18%)
FOLFOX/CAPOX	18 (17%)	4 (12%)
FOLFIRI	37 (36%)	22 (65%)
FOLFOXIRI	3 (3%)	2 (6%)
Irinotecan	2 (2%)	0 (0%)
No chemotherapy	25 (24%)	0 (0%)
With anti-VEGF-antibody	25 (24%)	16 (47%)
With anti-EGFR-antibody	27 (26%)	13 (38%)

[a] n is shown for categorical variables with percentage in parentheses. For continuous variables, median is shown with range in parentheses. [b] These three patients all had an inoperable relapse of their primaries.

advanced CRC, for which serial plasma samples were available. The patients received a chemotherapy backbone consisting of 5-fluorouracil with or without oxaliplatin or irinotecan. In addition, most patients also received treatment with antibodies targeting Vascular Endothelial Growth Factor (VEGF) or Epidermal Growth Factor Receptor (EGFR). Detailed patient characteristics can be found in Table 1. Within the cohort, 14 patients had a radiologic stable disease or partial remission while the others had disease progression between two consecutive blood sampling time points. Amplicon sequencing of cfDNA demonstrated
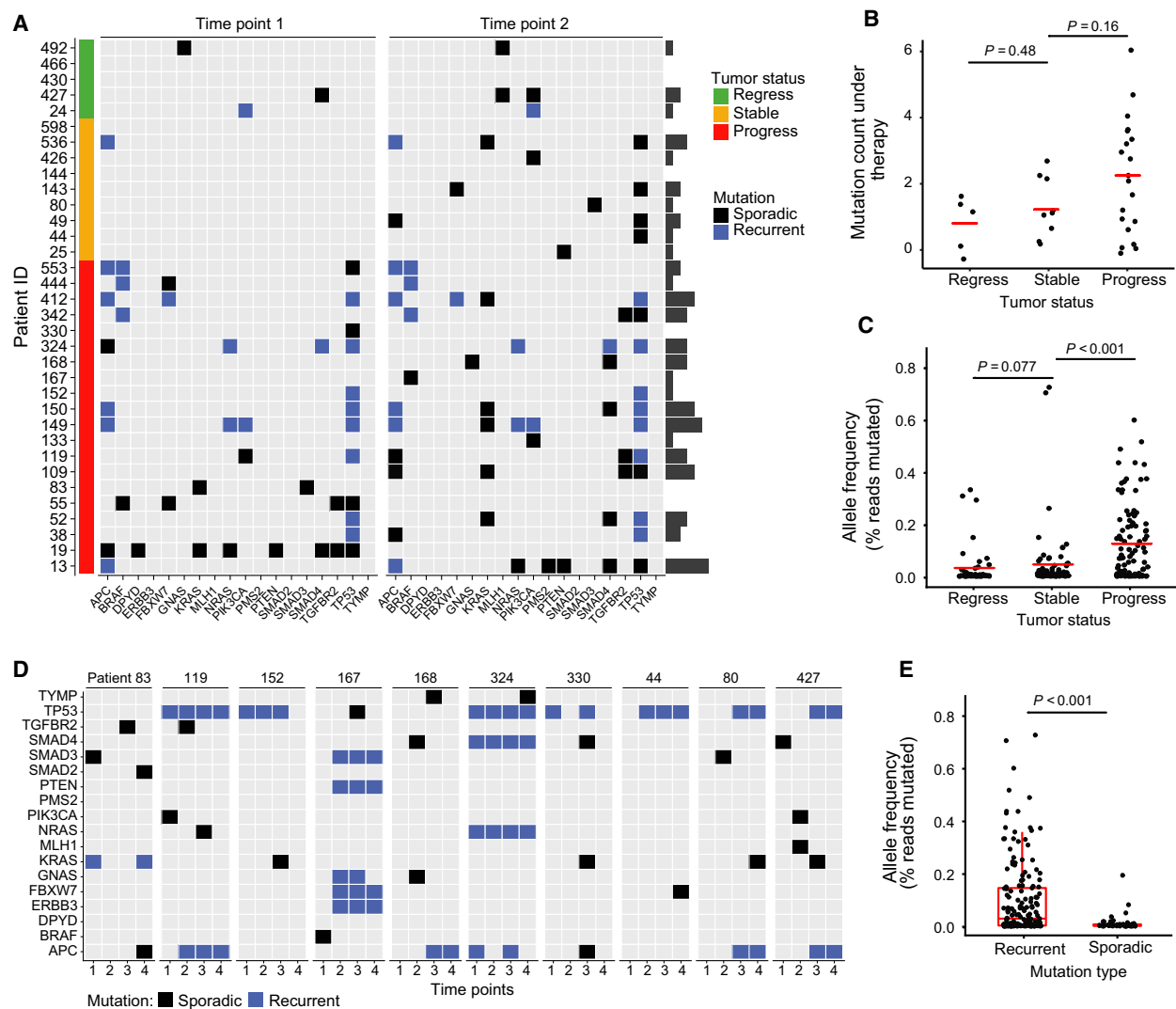
Fig. 4. Mutational patterns in cfDNA are patient-specific and highly recurrent. (A) Heatmap of all mutations detected in cfDNA at two subsequent time points during the course of treatment. Patients' disease was categorized as regressing, stable, or progressive according to radiologic imaging between the two time points assessed, several mutations were recurrent in both samples, while others were only noted in either time point 1 or 2 (sporadic). (B) Mutation count under therapy. Tumor status was categorized as regressive, stable, or progressive by radiologic imaging at the time the sample was taken. (C) Mutant allele frequency of detected alterations according to tumor status. Tumor status was categorized as regressive, stable, or progressive by radiologic imaging at the time the cfDNA sample was taken. (D) Development of mutational patterns during the course of treatment. Four cfDNA samples were taken from each patient at different time points during the course of treatment, and mutations were analyzed. Several mutations were reoccurring between different time points, while others appeared only at one time point (sporadic). (E) Comparison of mutant allele frequencies of recurrent and sporadic mutations.

that 2.41 alterations could be detected on average per sample (Fig. 4A). These alterations were either recurrent (present in serial samples) or sporadic. Alterations occurred at an allele frequency between 0.5% and 96.7% (median 2.25%). Most mutations were found in amplicons of TP53. Overall, we observed a trend toward a higher number of alterations in patients with disease progression (Fig. 4B). Also, we found that mutant allele frequencies were significantly higher ($P < 0.001$) in patients with progressive disease

compared to those with stable disease or radiologic therapy response (Fig. 4C), consistent with previous findings (Forshew *et al.*, 2012; Gray *et al.*, 2015; Newman *et al.*, 2014). Using APC as an example case, we observed that the distribution of mutations in cfDNA is very similar to the mutational spectrum found in primary CRC tissue (Fig. S2) (Forbes *et al.*, 2017).

For a small group of patients, we were able to obtain up to four consecutive blood samples during the course of the therapy, corresponding to a

treatment period of approximately 1 year. For all patients in this cohort, we detected both recurrent and sporadic mutations (Fig. 4D). Recurrent mutations were more frequent than sporadic mutations and showed a higher mutant allele frequency (Fig. 4E). Furthermore, we found the pattern of recurrent mutations to be distinct for each patient and conserved throughout multiple blood samples.

3.5. Liquid biopsy allows monitoring the clinical course of CRC patients and reveals resistance mechanisms

Next, we aimed to analyze how the spectrum of mutations found in serial cfDNA samples corresponds to the course of treatment in CRC patients. To this end, we correlated the clinical course with the mutant allele frequency of specific alterations in cfDNA (Fig. 5). All four presented cases were initially treated with 5-fluorouracil, folinic acid, irinotecan (FOLFIRI), and cetuximab and were in stable disease or had tumor remissions by the time of the first liquid biopsy. Patient 119 had achieved tumor remission and therefore received maintenance treatment with 5-FU and cetuximab. Two novel mutations occurred in the cfDNA upon radiologic disease progression. The allele frequencies of those mutations closely matched the radiologic extent of disease during the following course of treatment, as did the CEA levels (Fig. 5A). Accordingly, in patients 324 (Fig. 5B), 24 (Fig. 5C), and 553 (Fig. 5D), allele frequencies increased upon tumor progression, while regression or stable disease was associated with stable or decreasing allele frequencies. Interestingly, liquid biopsies revealed an NRAS mutation (patient 324) and a BRAF mutation (patient 553) not identified by Sanger sequencing in the patients' primary tumors. CEA levels were not congruent with radiologic disease progression and allele frequencies of mutations in cfDNA in patient 553. Hence, in all presented cases, we observed that an increase in allele frequency is closely associated with radiologic progression of the disease. Furthermore, and in contrast to the traditional tumor marker CEA, our amplicon sequencing assay for cfDNA analysis could also reveal biological insights into the mechanisms of therapy resistance to targeted therapies.

4. Discussion

In the present study, we describe a flexible experimental and bioinformatic method to analyze custom-defined gene loci in cfDNA by multiplexed amplicon sequencing. We applied our method on serial cfDNA

samples from a pilot cohort of 34 CRC patients to detect recurrent and patient-specific mutational patterns in key cancer genes.

Over the past decade, significant progress has been made in the analysis of cfDNA from cancer patients. Two recent studies have comprehensively dissected the mutational landscape of cfDNA in large cohorts of patients with CRC (Strickler *et al.*, 2018) and other advanced tumors (Zill *et al.*, 2018). Both studies demonstrate that the composition of mutations is highly similar between cfDNA and the corresponding primary tumors. Therefore, detection and analysis of cfDNA have been proposed as a biomarker to monitor tumor progression, treatment response, and disease recurrence (Bettegowda *et al.*, 2014; Newman *et al.*, 2014; Tie *et al.*, 2016). Until now, many different approaches have been developed to characterize cfDNA, including the detection of specific mutations (Haselmann *et al.*, 2018) or the sequencing of a broad panel of genes (Goodall *et al.*, 2017; Quigley *et al.*, 2017). Technologies based on both approaches have been approved by the FDA, that is Guardant360 or cobas EGFR Mutation Test V2. A major limitation to their broad clinical application is the diversity of the genetic landscape of cancers from different tissues. Particularly, mutations that occur only in selected cancer types and at medium frequencies can represent potentially druggable conditions, but are not covered by commercial gene panels. Examples include mutations in RNF43 (Steinhart *et al.*, 2017) or PMS2 (Khagi *et al.*, 2017), which could mark clonal subtypes that may make the tumor responsive to specific agents, such as checkpoint inhibitors. Hence, a main challenge is to customize the analysis of cfDNA to specific tumor types or defined sets of mutations for exploratory analysis. In this regard, our study provides a methodological framework that enables the design and bioinformatic analysis of custom amplicon panels using open-source software tools. We used our method to design a gene panel tailored for colorectal cancer, including amplicons that cover hotspot mutation sites in oncogenes and tumor suppressor genes, as well as rare mutations in genes relevant for drug metabolism.

We used our custom amplicon panel to analyze sequential cfDNA samples of a pilot cohort of patients with advanced CRC. In accordance with previous studies, we found that most genetic alterations could be found in genes that are known to be highly mutated, such as *APC* or *TP53*. Furthermore, we demonstrate that alterations in specific genes, such as BRAF or KRAS (Diaz *et al.*, 2012), occur *de novo* in cfDNA during treatment failure, thereby revealing genetic alterations that can potentially mediate drug
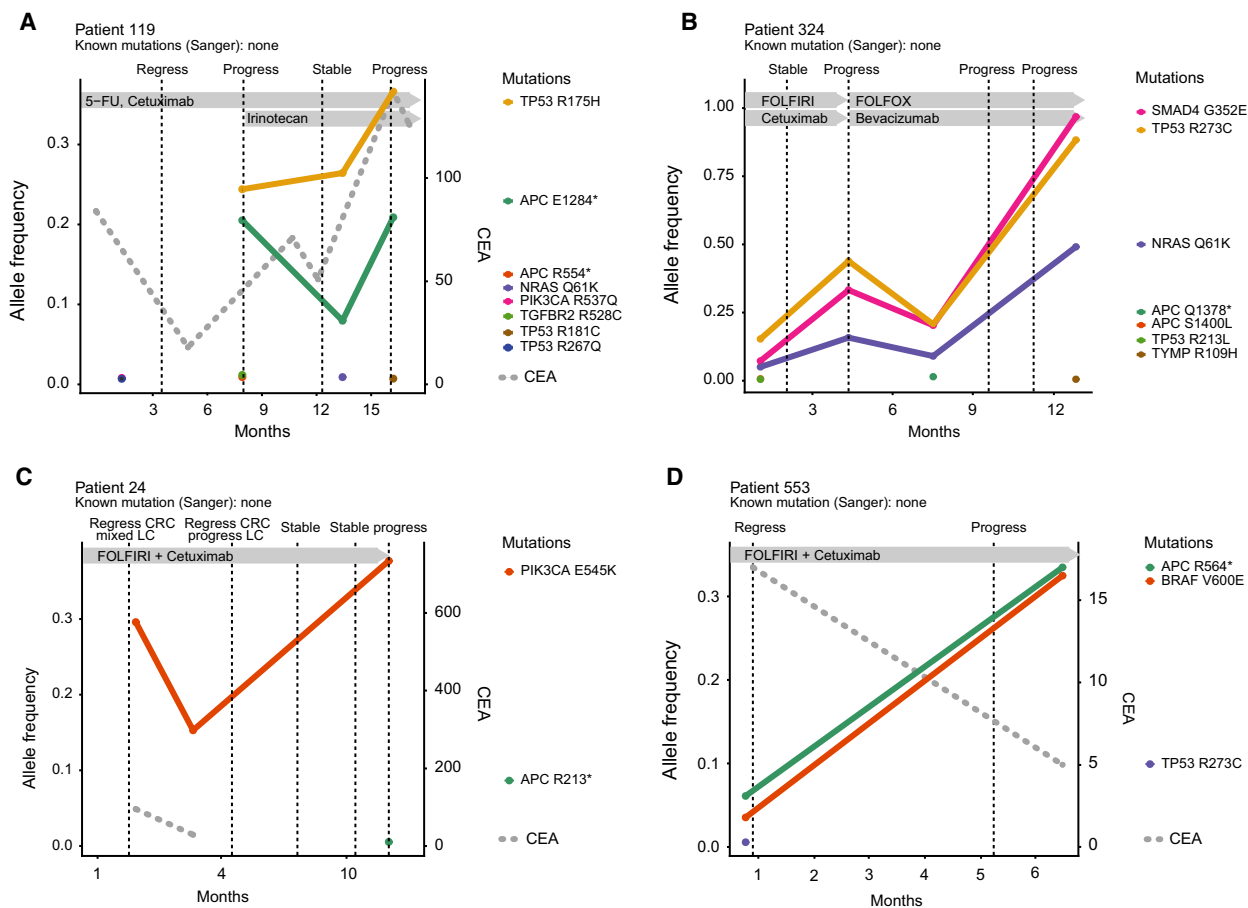
Fig. 5. Amplicon sequencing of cfDNA allows monitoring the clinical course of colorectal cancer patients. (A) Patient 119 had received FOLFIRI and cetuximab for metastasized rectal cancer. The patient was in a stable situation and underwent maintenance treatment with 5-FU and cetuximab by the time of the first liquid biopsy. When the patient experienced radiologic disease progression, treatment was re-escalated to FOLFIRI and cetuximab, which led to stable disease in the next restaging and progressive disease another 3 months later. Accordingly, novel mutations were discovered by the time of first progression that decreased in allele frequency during stable disease and increased upon radiologic disease progression. (B) Patient 324 had a local recurrence of rectal cancer and hepatic metastases. Under palliative combination treatment with FOLFIRI and cetuximab, the patient had a stable disease by the time of liquid biopsy 1. When he experienced disease progression, liquid biopsy 2 showed increased allele frequencies of several mutations, including an NRAS mutation that had not been detected by Sanger sequencing of the primary tumor. The treatment was switched to FOLFOX + bevacizumab, leading to a decrease in allele frequency, before another radiologic disease progression and, accordingly, rising allele frequencies were observed. CEA levels were not available for this patient. (C) Patient 24 had both a colon cancer with hepatic metastases and a synchronous non-small-cell lung cancer. Each tumor manifestation was histologically proven. The colorectal cancer was RAS wild-type according to Sanger sequencing, and the lung cancer was RAS-mutated. Nevertheless, the patient received FOLFIRI and cetuximab. During the course of treatment, a PIK3CA mutation that was also identified by amplicon sequencing in material from the colon cancer sank in allele frequency before radiologic imaging proved regression of the colorectal cancer (and progression of the lung cancer). Upon progress of the colon cancer, the mutant allele frequency increased accordingly. (D) Patient 553 was treated for metachronous liver metastases of RAS/RAF wild-type rectal cancer with FOLFIRI and cetuximab. After initial tumor regression, allele frequency of BRAF V600E and APC R564* mutations rose significantly after radiologic tumor progression, while other mutations, likely from a different tumor subclone, remained on low levels.

resistance. In addition, our data show that the mutant allele frequency correlates with disease progression. Interestingly, we observed that the occurrence of individual mutations remained stable across the course of the therapy. These distinct mutations could be part of a stem mutational signature which is conserved across cfDNA, primary tumors, and metastasis and has been described for a case of breast cancer (Murtaza *et al.*, 2015). Our results indicate that such a signature can be found in many patients with CRC. This application might be of particular interest for patients who do not have elevated tumor markers in spite of a high disease

burden, which in case of CRC accounts for 30% of patients (Liu *et al.*, 2014). However, we are aware that our amplicon sequencing-based method works best in gastrointestinal cancers that have a limited set of frequently mutated genetic loci, as it is the case for CRC or pancreatic adenocarcinoma. For these cancer types, compact amplicon panels can be designed that cover most hotspot positions. For other cancers, such as gastric or hepatocellular cancer, the mutational load is more equally distributed to many loci, which requires the design of larger amplicon panels for sufficient coverage. This will, however, lead to an increase in cost and technical complexity. An additional goal of our custom amplicon panel was the detection of genetic alterations in genes that are not commonly covered by commercial panels. These include genes involved in 5-fluorouracil metabolism or regulators of DNA repair. Analysis of cfDNA in our cohort shows that mutations in these genes are rarely found in cfDNA. However, we could detect the *de novo* occurrence of a mutation leading to a premature stop codon in *PMS2* and novel mutations in *TYMP* and *DYPD*. The functional relevance of these mutations is currently unknown. Therefore, a major challenge for the future is the functional annotation of these rare mutations, in order to understand their impact on tumor biology and drug resistance. These insights will enable the profound interpretation of sequencing results from cfDNA in cancer patients.

Although our method enables the flexible design of custom amplicon panels, it still has several limitations. First, calculation of mutant allele frequencies is biased by the presence of genomic DNA from nontumorous cells, which can be increased by unwanted cell lysis. To minimize the extent of contamination by genomic DNA, we selected storage and sample processing conditions that prevent cell lysis, based on findings from other studies (Norton *et al.*, 2013; Risberg *et al.*, 2018; Wong *et al.*, 2013). However, since the time to sample processing varied between 1 and 16 h in our cohort, we are unable to exclude that the measured cfDNA concentrations might be slightly biased by the different storage times. Furthermore, since amplicon sequencing is based on pre-amplification by PCR, results can be influenced by PCR artifacts and sequencing errors. An effective method to reduce these technical errors is the use of unique molecular identifiers (UMI). For the analysis of cfDNA, methods have been developed to introduce UMI during first PCR amplification cycles by using barcoded primers. These approaches, for instance Safe-SeqS (Kinde *et al.*, 2011) or SiMSen-Seq (Ståhlberg *et al.*, 2017), have been shown to suppress PCR errors and enable more precise counting of

mutant allele copies. A drawback of using UMIs is that a vastly higher sequencing depth is required to discriminate between different UMI families, which will increase the costs of the method. However, as sequencing costs will further decrease in the near future, the integration of UMIs in our approach will be feasible. Another limitation of our methods is that it is currently unable to detect copy number alterations or gene fusions, which can be analyzed by other methods (Strickler *et al.*, 2018; Zill *et al.*, 2018) and provide valuable insights into tumor biology (Stover *et al.*, 2018; Weiss *et al.*, 2017). Lastly, the prediction of primer pairs that will function in multiplex assays is limited with the currently available bioinformatic tools and we had to discard several primer pairs because they did not yield a functional amplicon. In another case, we were unable to differentiate between mutations and PCR errors due to a poly C region within the MSH6 amplicon region. Therefore, in-depth testing and experimental validation of primer pairs will be necessary when establishing novel custom amplicon panels. In spite of these limitations, we could show by multiple quality control experiments that our method can reliably detect cancer mutations in cfDNA.

5. Conclusions

In summary, we describe a flexible method to analyze mutations in cfDNA by custom amplicon panels. We show that this method can be used to detect recurrent mutational patterns as well as *de novo* mutations in cfDNA in a pilot cohort of patients with metastatic CRC, which enables the monitoring of therapy response and development of drug resistance. Finally, we show that our method can be applied to explore novel mutations at genetic loci that are not covered by commercial amplicon panels. The integration of UMI for error suppression and the validation of the method in a larger, clinically defined cohort of cancer patients will be necessary to further improve our workflow.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

All sequencing data generated for this study are available from the European Genome-phenome Archive under the accession number EGAS00001003382. Documented computer code to reproduce the figures of this manuscript is available from GitHub at https://github.com/boutroslab/Supplemental-Material/tree/master/Herrmann%26Zhan%26Betge%26Rauscher_2018.

Author contributions

SH, TZ, and JB established methods and carried out the experiments. SH, TZ, JB, SB, TG, NS, RJ, NH, TG, and R-DH collected samples and clinical data. BR performed bioinformatic analysis. MPE and MB provided funding and supervised the project. SH, TZ, JB, and MB wrote the manuscript text.

References

Alix-Panabières C and Pantel K (2013) Circulating tumor cells: liquid biopsy of cancer. *Clin Chem* **59**, 110–118.

Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, Caldas C, Califano A, Doherty M, Elsner M *et al.* (2015) Toward understanding and exploiting tumor heterogeneity. *Nat Med* **21**, 846–853.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Andrew S (2010) A quality control tool for high throughput sequence data. http://wwwBioinformaticsBabrahamAcUk/Projects/Fastqc/

Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM *et al.* (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra24.

Bolger AM, Lohse M and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Burrell RA, McGranahan N, Bartek J and Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345.

Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA *et al.* (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540.

Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45**, D777–D783.

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805–D811.

Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D *et al.* (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, 136ra68.

Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892.

Goodall J, Mateo J, Yuan W, Mossop H, Porta N, Miranda S, Perez-Lopez R, Dolling D, Robinson DR, Sandhu S *et al.* (2017) Circulating cell-free DNA to guide prostate cancer treatment with PARP inhibition. *Cancer Discov* **7**, 1006–1017.

Gray ES, Rizos H, Reid AL, Boyd SC, Pereira MR, Lo J, Tembe V, Freeman J, Lee JH, Scolyer RA *et al.* (2015) Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. *Oncotarget* **6**, 42008–42018.

Haselmann V, Gebhardt C, Brechtel I, Duda A, Czerwinski C, Sucker A, Holland-Letz T, Utikal J, Schadendorf D and Neumaier M (2018) Liquid profiling of circulating tumor DNA in plasma of melanoma patients for companion diagnostics and monitoring of BRAF inhibitor therapy. *Clin Chem* **64**, 830–842.

Janku F, Angenendt P, Tsimberidou AM, Fu S, Naing A, Falchook GS, Hong DS, Holley VR, Cabrilo G, Wheler JJ *et al.* (2015) Actionable mutations in plasma cell-free DNA in patients with advanced cancers referred for experimental targeted therapies. *Oncotarget* **6**, 12809–12821.

Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D (2011) Global cancer statistics. *CA Cancer J Clin* **61**, 69–90.

Jennings BA, Kwok CS, Willis G, Matthews V, Wawruch P and Loke YK (2012) Functional polymorphisms of folate metabolism and response to chemotherapy for colorectal cancer, a systematic review and meta-analysis. *Pharmacogenet Genomics* **22**, 290–304.

Kaplinski L and Remm M (2015) MultiPLX: automatic grouping and evaluation of PCR primers. *Methods Mol Biol* **1275**, 127–142.

Khagi Y, Kurzrock R and Patel SP (2017) Next generation predictive biomarkers for immune checkpoint inhibition. *Cancer Metastasis Rev* **36**, 179–190.

Kinde I, Wu J, Papadopoulos N, Kinzler KW and Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci USA* **108**, 9530–9535.

Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, Collisson EA, Divers SG, Hoon DS, Kopetz ES *et al.* (2015) Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS One* **10**, e0140712.

Lebofsky R, Decraene C, Bernard V, Kamal M, Blin A, Leroy Q, Rio Frio T, Pierron G, Callens C, Bieche I *et al.* (2015) Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol* **9**, 783–790.

Li G-M (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**, 85–98.

Li H and Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liu Z, Zhang Y, Niu Y, Li K, Liu X, Chen H and Gao C (2014) A systematic review and meta-analysis of diagnostic and prognostic serum biomarkers of colorectal cancer. *PLoS One* **9**, e103910.

Maheswaran S and Haber DA (2010) Circulating tumor cells: a window into cancer biology and metastasis. *Curr Opin Genet Dev* **20**, 96–99.

Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, Chin SF, Tsui DWY, Marass F, Gale D *et al.* (2015) Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760.

Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE *et al.* (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548–554.

Norton SE, Lechner JM, Williams T and Fernando MR (2013) A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR. *Clin Biochem* **46**, 1561–1565.

Ooyama A, Okayama Y, Takechi T, Sugimoto Y, Oka T and Fukushima M (2007) Genome-wide screening of loci associated with drug resistance to 5-fluorouracil-based drugs. *Cancer Sci* **98**, 577–583.

Pullarkat ST, Stoehlmacher J, Ghaderi V, Xiong YP, Ingles SA, Sherrod A, Warren R, Tsao-Wei D, Groshen S and Lenz HJ (2001) Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *Pharmacogenomics J* **1**, 65–70.

Quigley D, Alumkal JJ, Wyatt AW, Kothari V, Foye A, Lloyd P, Aggarwal R, Kim W, Lu E, Schwartzman J *et al.* (2017) Analysis of circulating cell-free DnA identifies multiclonal heterogeneity of BRCA2 reversion mutations associated with resistance to PARP inhibitors. *Cancer Discov* **7**, 999–1005.

R Core Team, (2016) R: A language and environment for statistical computing. R Foundation for Statistical ComputingVienna, Austria. http://www.r-project.org/

Risberg B, Tsui DWY, Biggs H, Martin R-V, de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A *et al.* (2018) Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patients. *J Mol Diagn* **20**, 883–892.

Schwarzenbach H, Hoon DSB and Pantel K (2011) Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* **11**, 426–437.

Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, Ponzetti A, Cremolini C, Amatu A, Lauricella C *et al.* (2015) Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 827.

Ståhlberg A, Krzyzanowski PM, Egyud M, Filges S, Stein L and Godfrey TE (2017) Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nat Protoc* **12**, 664–682.

Steinhart Z, Pavlovic Z, Chandrashekhar M, Hart T, Wang X, Zhang X, Robitaille M, Brown KR, Jaksani S, Overmeer R *et al.* (2017) Genome-wide CRISPR screens reveal a Wnt-FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nat Med* **23**, 60–68.

Stover DG, Parsons HA, Ha G, Freeman SS, Barry WT, Guo H, Choudhury AD, Gydush G, Reed SC, Rhoades J *et al.* (2018) 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.

Strickler JH, Loree JM, Ahronian LG, Parikh AR, Niedzwiecki D, Pereira AAL, McKinney M, Korn WM, Atreya CE, Banks KC *et al.* (2018) Genomic landscape of cell-free DNA in patients with colorectal cancer. *Cancer Discov* **8**, 164–173.

Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, Rossier A, Chen XQ and Anker P (2006) The origin and mechanism of circulating DNA. *Ann N Y Acad Sci* **906**, 161–168.

Thierry AR, Pastor B, Jiang ZQ, Katsiampoura AD, Parseghian C, Loree JM, Overman MJ, Sanchez C,

Messaoudi SE, Ychou M *et al.* (2017) Circulating DNA demonstrates convergent evolution and common resistance mechanisms during treatment of colorectal cancer. *Clin Cancer Res* **23**, 4578–4591.

Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, Silliman N, Tacey M, Wong HL, Christie M *et al.* (2016) Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* **8**, 346ra92.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Res* **40**, 1–12.

Vanderlaan PA, Yamaguchi N, Folch E, Boucher DH, Kent MS, Gangadharan SP, Majid A, Goldstein MA, Huberman MS, Kocher ON *et al.* (2014) Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. *Lung Cancer* **84**, 39–44.

Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R and Rosenfeld N (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* **17**, 223–238.

Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, Quan W Jr, Sangal A, Khemka V, Waypa J *et al.* (2017) Tumor cell-free DNA copy number instability predicts therapeutic response to immunotherapy. *Clin Cancer Res* **23**, 5074–5081.

Wong D, Moturi S, Angkachatchai V, Mueller R, DeSantis G, van den Boom D and Ehrich M (2013) Optimizing blood collection, transport and storage conditions for cell free DNA increases access to prenatal testing. *Clin Biochem* **46**, 1099–1104.

Zill OA, Banks KC, Fairclough SR, Mortimer SA, Vowles JV, Mokhtari R, Gandara DR, Mack PC, Odegaard JI, Nagy RJ *et al.* (2018) The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin Cancer Res* **24**, 3528–3538.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Sensitivity of the amplicon sequencing assay.

Fig. S2. Distribution of APC mutations in primary colorectal cancer and cfDNA.

Table S1. List and characteristics of primers in the amplicon panel.