



OPEN

## Comparative analysis of nuclear and mitochondrial DNA from tissue and liquid biopsies of colorectal cancer patients

Anna Haupts<sup>1</sup>✉, Anne Vogel<sup>1</sup>, Sebastian Foersch<sup>1</sup>, Monika Hartmann<sup>2</sup>, Annett Maderer<sup>2</sup>, Nicolas Wachter<sup>3</sup>, Tobias Huber<sup>3</sup>, Werner Kneist<sup>3,4</sup>, Wilfried Roth<sup>1</sup>, Hauke Lang<sup>3</sup>, Markus Moehler<sup>2</sup> & Nils Hartmann<sup>1</sup>✉

The current standard for molecular profiling of colorectal cancer (CRC) is using resected or biopsied tissue specimens. However, they are limited regarding sampling frequency, representation of tumor heterogeneity, and sampling can expose patients to adverse side effects. The analysis of cell-free DNA (cfDNA) from blood plasma, which is part of a liquid biopsy, is minimally invasive and in principle enables detection of all tumor-specific mutations. Here, we analyzed cfDNA originating from nucleus and mitochondria and investigated their characteristics and mutation status in a cohort of 18 CRC patients and 10 healthy controls using targeted next-generation sequencing (NGS) and digital PCR. Longitudinal analyses of nuclear cfDNA level and size during chemotherapy revealed a decreasing cfDNA content and a shift from short to long fragments, indicating an appropriate therapy response, while shortened cfDNAs and increased cfDNA content corresponded with tumor recurrence. Comparative NGS analysis of nuclear tissue and plasma DNA demonstrated a good patient-level concordance and cfDNA revealed additional variants in three of the cases. Analysis of mitochondrial cfDNA surprisingly revealed a higher plasma copy number in healthy subjects than in CRC patients. These results highlight the potential clinical utility of liquid biopsies in routine diagnostics and surveillance of CRC patients as complementation to tissue biopsies or as an attractive alternative in cases where tissue biopsies are risky or the quantity/quality does not allow testing.

### Abbreviations

cfDNA	Cell-free DNA
cf-mtDNA	Cell-free mitochondrial DNA
cf-nDNA	Cell-free nuclear DNA
CNV	Copy number variation
CRC	Colorectal cancer
ctDNA	Circulating tumor DNA
CTx	Chemotherapy
dPCR	Digital PCR
FFPE	Formalin-fixed paraffin-embedded
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
NGS	Next-generation sequencing
SNV	Single-nucleotide variant
VAF	Variant allele frequency

<sup>1</sup>Institute of Pathology, University Medical Center JGU Mainz, Langenbeckstraße 1, 55131 Mainz, Germany. <sup>2</sup>Department of Internal Medicine I, University Medical Center JGU Mainz, Langenbeckstraße 1, 55131 Mainz, Germany. <sup>3</sup>Department of General, Visceral and Transplantation Surgery, University Medical Center JGU Mainz, Langenbeckstraße 1, 55131 Mainz, Germany. <sup>4</sup>Present address: Department of General and Visceral Surgery, St. Georg Hospital Eisenach gGmbH, Mühlhäuser Straße 94, 99817 Eisenach, Germany. ✉email: ahaupts@uni-mainz.de; Nils.Hartmann@unimedizin-mainz.de

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the fourth most common cause of cancer death worldwide<sup>1,2</sup>. The current standard for characterizing colorectal carcinomas is histopathological examination of the biopsy or resection tissue specimens with subsequent determination of molecular markers in formalin-fixed paraffin-embedded (FFPE) DNA. However, analysis of tumor tissue samples are invasive and do not provide full representation of both spatial and temporal tumor heterogeneity<sup>3</sup>. In contrast, analysis of circulating tumor-derived DNA from blood plasma, which is part of a liquid biopsy, enables minimally invasive detection of in principle all tumor-specific alterations and their dynamic changes. Several studies demonstrated good concordance between profiling molecular alterations in FFPE DNA and matched circulating tumor DNA (ctDNA) from CRC patients, and in addition revealed that plasma analysis could be more informative than corresponding analysis of tissue biopsies<sup>4–7</sup>. The majority of cell-free DNA (cfDNA) is released by apoptotic cells with a modal size of 166 bp, resulting in a “ladder” pattern depending on nuclease action during apoptosis<sup>8,9</sup>. Therefore, cfDNA is highly fragmented and was shown to have an even shorter fragment size in the circulation of cancer patients compared to those derived from non-neoplastic cells<sup>10–14</sup>. Furthermore, quantitative analysis of cfDNA levels have been shown to yield important prognostic value, for example, an increase of concentration correlated with tumor stages and overall survival in CRC patients<sup>6,15</sup>. Hence, analysis of cfDNA level and fragmentation pattern could potentially serve to distinguish between cancer patients and healthy subjects or to follow the course of disease. However, quantitative and qualitative analyses of cfDNA are very challenging and therefore are still rarely used in routine diagnostic of CRC.

Moreover, the proportion of an individual's tumor-derived ctDNA of total cfDNA is often small (<1% according to some studies) but tumor fraction can vary between cancer types and even between patients with histologically identical cancers<sup>6,16</sup>. In general, metastatic patients yield a higher amount of tumor-derived plasma DNA but even patients with advanced stage diseases can harbor only a low amount of ctDNA<sup>17,18</sup>. Furthermore, cfDNA is also released by non-neoplastic cells, resulting in a high wildtype background<sup>16,18</sup>. Thus, analysis of the limited number of ctDNA molecules requires highly sensitive molecular approaches. A high level of sensitivity is achieved by PCR-based techniques, such as digital PCR (dPCR), which rely on detection of specific known mutations using primers that are complementary to the mutant sequences<sup>19</sup>. Next-generation sequencing (NGS)-based techniques also have been implemented to simultaneously investigate multiple molecular alterations to enable genome-wide analysis or targeted sequencing of a subset of genes<sup>20</sup>. However, genomic profiling from ctDNA analysis has difficulties and standards for core quality metrics, such as coverage and sensitivity, have to be defined with regard to the interpretation of variants found near to the limit of detection<sup>21</sup>.

Besides cell nuclei, mitochondria have their own circular genome and thus contribute to total cfDNA content in blood. A single cell contains up to several thousand copies of mitochondrial DNA (mtDNA) opposed to two copies of nuclear DNA (nDNA)<sup>22</sup>. Hence, examination of cell-free mitochondrial DNA (cf-mtDNA) theoretically could offer a higher level of sensitivity than analyzing cell-free nuclear DNA (cf-nDNA). In addition, mtDNA has a high mutation rate and in CRC and other cancers underlying molecular alterations have been reported<sup>23–25</sup>. Thus, evaluation of cf-mtDNA as potential biomarker is interesting for liquid biopsies because high copy number could enable detection of even small amounts of ctDNA and their molecular alterations. Furthermore, previous studies have shown that cf-mtDNA content and fragmentation pattern differentiate between cancer patients and healthy individuals, thus also potentially serving as indicative marker of disease<sup>8,26,27</sup>. However, cf-mtDNA has been not fully characterized yet and an efficient approach for comprehensive analysis is still lacking.

Here, we report a proof-of-principle for evaluation of liquid biopsy workflows for potential application in clinical routine of CRC patients. We examined cfDNA from nuclear and mitochondrial origin, involving determination of cfDNA levels and fragment sizes from healthy controls and CRC patients at baseline and also during chemotherapy, as well as analyses of molecular alterations and copy numbers. We validated the analytical sensitivity of the applied molecular approaches (NGS and dPCR) using reference material with known mutations and investigated paired tissue and plasma mutation analysis concerning sequence variants and copy number variations (CNVs), and identified analytical factors that hamper concordance between both.

## Results

**Patient characteristics.** The 18 patients included in this study had non-metastatic (n=9, 50%), locally advanced (n=3, 16.7%) or metastatic (n=6, 33.3%) CRC. Table 1 provides an overview of patients' clinicopathological features. Overall, 33.3% (n=6) of the primary tumors were localized in the colon (sigmoid colon: n=3, ascending colon: n=1, descending colon: n=1, cecum: n=1) and 66.7% (n=12) in the rectum. According to this, 22.2% (4/6) had a left-sided and 11.1% (2/6) had a right-sided primary colon tumor. The most frequent metastatic sites were lymph nodes followed by liver. Almost all patients (17/18) had microsatellite-stable tumors (94.4%). More than half of the patients (11/18) received radiochemotherapy (n=3, 16.7%) or chemotherapy (n=8, 44.4%). The pretherapeutic CEA was determined in 15/18 patients with a median baseline level of 3.30 ng/mL.

**cfDNA level and fragment size depend on the presence of tumor and treatment.** To study the value of quantitative and qualitative analysis of cfDNA as a biomarker in CRC, we measured levels and fragment sizes in plasma samples from 17 previously untreated CRC patients (one metastatic patient of the cohort was previously treated and thus dropped out) and 10 healthy controls by using Agilent Bioanalyzer instrument (Fig. 1a–c). Thereby, CRC patients were considered as entirety (n=17) as well as divided into different patient groups: non-metastatic (n=9), locally advanced (n=3) and metastatic (n=5). Median pretherapeutic cfDNA level of all CRC patients at baseline blood draw was 1.58 ng/mL plasma and median pretherapeutic fragment size was 174 bp. Thus, cfDNA level of CRC patients (1.58 ng/mL plasma) was as expected about fourfold higher compared to the median cfDNA level of healthy subjects (0.40 ng/mL plasma,  $P < 0.001$ ). Besides, median cfDNA

Characteristics	Patients, n (%)
<b>Median age (years)</b>	
60 (range: 28–87)	18 (100%)
<b>Gender</b>	
Male	12 (66.7%)
Female	6 (33.3%)
<b>ECOG</b>	
0	14 (77.7%)
1	3 (16.7%)
2	1 (5.6%)
<b>Primary tumor site<sup>‡</sup></b>	
Right-sided	2 (11.1%)
Left-sided	4 (22.2%)
Rectum	12 (66.7%)
<b>Status of resection of primary tumor site</b>	
R0	17 (94.4%)
R1	1 (5.6%)
<b>Tumor size</b>	
T1	1 (5.6%)
T2	3 (16.7%)
T3	14 (77.7%)
<b>Nodal stage</b>	
N0	9 (50.0%)
N1	2 (11.1%)
N2	6 (33.3%)
Nx	1 (5.6%)
<b>Distant metastases</b>	
M0	12 (66.7%)
M1	6 (33.3%)
<b>Stage<sup>§</sup></b>	
I	2 (11.1%)
IIA	8 (44.4%)
IIIA	1 (5.6%)
IIIC	1 (5.6%)
IVA	4 (22.2%)
IVB	2 (11.1%)
<b>Histological differentiation</b>	
G2	13 (72.2%)
G3	1 (5.6%)
Unknown	4 (22.2%)
<b>Lymphovascular invasion</b>	
L0	11 (61.1%)
L1	7 (38.9%)
<b>Vein invasion</b>	
V0	14 (77.7%)
V1	4 (22.2%)
<b>Perineural invasion</b>	
Pn0	14 (77.7%)
Pn1	3 (16.7%)
Unknown	1 (5.6%)
<b>Metastatic sites<sup>†</sup></b>	
Lymph nodes	8 (50%)
Liver	6 (37.5%)
Lung	2 (12.5%)
<b>Microsatellite instability</b>	
MS-stable	17 (94.4%)
MSI-high	1 (5.6%)
Continued	

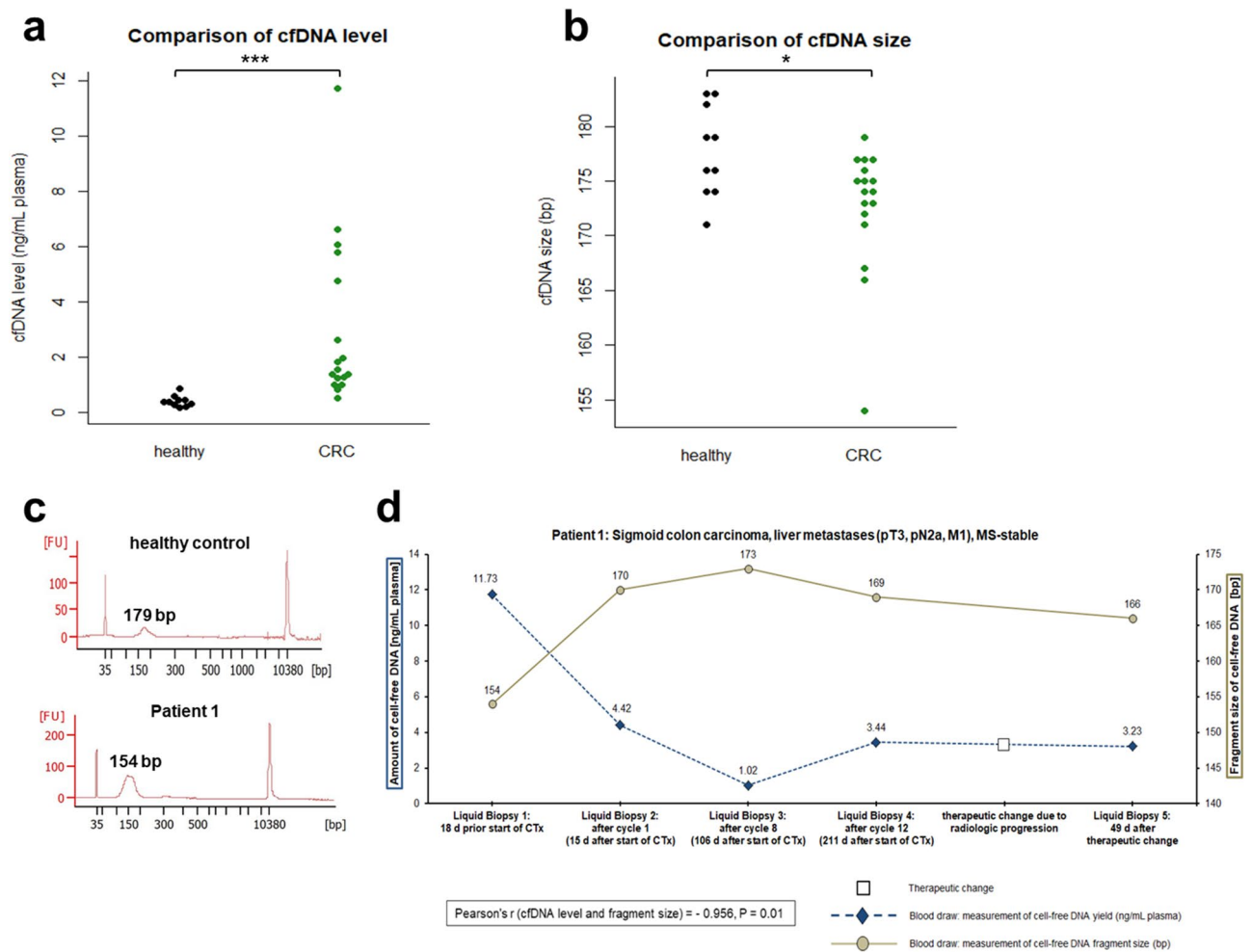
Characteristics	Patients, n (%)
<b>Therapy setting</b>	
Radiochemotherapy	3 (16.7%)
Chemotherapy	8 (44.4%)
Tumor after-care	7 (38.9%)
<b>Baseline CEA</b>	
Negative (< 5 ng/mL)	10 (55.5%)
Positive (≥ 5 ng/mL)	5 (27.8%)
Unknown	3 (16.7%)
<b>Baseline CEA, ng/mL*</b>	
3.30 (IQR: 2.00–9.50)	15 (83.3%)

**Table 1.** Clinicopathological characteristics of CRC patient cohort. *UICC* Union for International Cancer Control, *ECOG* Eastern Cooperative Oncology Group performance status, *CEA* carcinoembryonic antigen, *IQR* interquartile range. <sup>‡</sup>Colon and rectum carcinomas were defined according to European standards of *UICC*. Right-sided colon is defined as a proximal colon region from the cecum and ascending colon to the right transverse colon. Left-sided colon is defined as a distal colon region from the left transverse colon to the descending and sigmoid colon, not including the rectum. <sup>§</sup>Patients were categorized according to *UICC* classification (based on TNM: stage of tumor, metastases and lymph nodes) into CRC patient groups as follows: non-metastatic implies Tumor size = T1 or T2, Nodal stage = N0, Metastases stage = M0, *UICC* stage I or II; locally advanced implies Tumor size = T2 or T3, Nodal stage ≥ N1, Metastases stage = M0, *UICC* stage III; and metastatic implies Tumor size = T3, Nodal stage ≥ N1, Metastases stage = M1, *UICC* stage IV. Patient 19 was a special case due to the occurrence of simultaneous carcinoma of rectum (T1) and descending colon (T3), and therefore was assigned to patient group of locally advanced cancer. <sup>†</sup>Metastatic sites involve repeated counting of patients because of multiple metastases. \*Data are presented as the median (interquartile range).

levels of the respective CRC patient groups (non-metastatic: 1.84 ng/mL plasma,  $P < 0.001$ ; locally advanced: 1.29 ng/mL plasma,  $P < 0.05$ ; metastatic: 1.41 ng/mL plasma,  $P < 0.001$ ) were also significantly higher than in the healthy control group. Furthermore, the median cfDNA fragment size of all CRC patients (174 bp,  $P < 0.05$ ) as well as those of each CRC patient group (non-metastatic: 175 bp,  $P = 0.187$ ; locally advanced: 173 bp,  $P = 0.106$ ; metastatic: 167 bp,  $P < 0.05$ ) was shorter than the median cfDNA fragment size of healthy individuals (178 bp). Therefore, consideration of cfDNA level and fragment size could give an indication of the presence of a disease, and thus might be clinically relevant for diagnosis of CRC.

Next, we examined whether there is a correlation between level and fragment size of cfDNA and the course of disease as well as response to therapy. To track changes of these parameters during therapy, we collected serial blood samples from a metastatic CRC patient undergoing palliative chemotherapy (Fig. 1d). We obtained liquid biopsies prior to start of treatment and at another four time points during therapy (liquid biopsies 1–5: prior chemotherapy, after cycle 1, after cycle 8, after cycle 12, after therapeutic switch). Patient 1 was initially treated with 5-fluorouracil (5-FU), folinic acid, irinotecan (FOLFIRI), and the anti-*EGFR*-antibody cetuximab for metastasized CRC<sup>28</sup> which led to a good therapy response in the next restaging after 3 months. At the same time, longitudinal cfDNA profiling from prior start of therapy until 3-month follow-up after cycle 8 (liquid biopsies 1–3) showed a decreasing cfDNA yield by approximately the factor 11 (range 1.02–11.73 ng/mL plasma), and in contrast the fragment size of cfDNA increased by 19 bp (range 154–173 bp). These observations together suggest that tumor load was shrinking. During further first-line treatment, the patient showed radiologic disease progression. Simultaneously, we detected a threefold increase of cfDNA yield (range 1.02–3.44 ng/mL plasma) and a decrease of cfDNA fragment size by 4 bp (range 169–173 bp) in liquid biopsy 4, indicating a progression of disease as well. The disease progression of patient 1 led to a therapeutic switch to 5-fluorouracil, folinic acid, oxaliplatin (FOLFOX), and the anti-*VEGF*-antibody bevacizumab. After changing therapy regimen, liquid biopsy 5 showed very similar cfDNA level and size as it was before the switch. Overall, we found a significant negative correlation between cfDNA levels and fragment sizes (Pearson's correlation coefficient  $r = -0.956$ ,  $P = 0.01$ ), thus a high cfDNA level is correlated to a short fragment size. Reported cfDNA levels and sizes as well as their changes during therapy suggest that quantitative and qualitative analyses of cfDNA give insights into the course of disease and therefore could be used for monitoring therapy response of CRC patients in clinical setting.

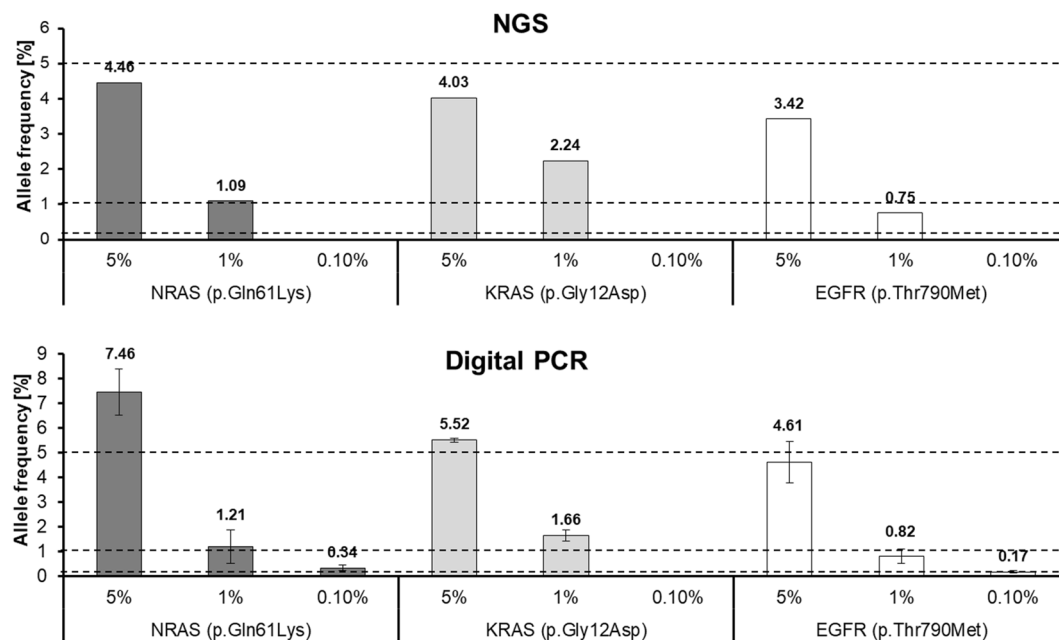
**dPCR enables up to ten times more sensitive variant detection than NGS.** Considering the implementation of plasma variant detection into clinical practice, it is of high importance to define standards for the molecular approaches and to validate the assays. In particular, setting detection limits is required due to the occurrence of variants predominantly at low allele frequencies to ensure the report of only true-positive variants. Evaluation of the detection limits of the applied NGS workflow (Archer Reveal ctDNA 28 Concordance Kit) and dPCR analyses (TaqMan assays) was performed by using a reference standard consisting of cell line DNA fragmented to an average size of 160 bp resembling cfDNA, with mutations (*NRAS* p.Gln61Lys, *KRAS* p.Gly12Asp, *EGFR* p.Thr790Met) at predefined allele frequencies (5%, 1%, 0.1%) (Fig. 2). NGS analysis of the reference material enabled detection of all three mutations at the expected 5% and 1% allele frequency, but none of these variants were detectable at 0.1% allele frequency. In dPCR analyses, all three variants were also detect-



**Figure 1.** cfDNA levels and fragment sizes in CRC patients and healthy individuals. Amount in ng/mL plasma (a) and fragment size in bp (b) of cfDNA extracted from healthy individuals ( $n = 10$ ) and CRC patients ( $n = 17$ ). Data are presented as swarmplots and median values and significance was therefore tested using Mann–Whitney U test. cfDNA level and size were measured by using Agilent High Sensitivity DNA Kit for Bioanalyzer instrument. Representative examples of an electropherogram (fluorescence units FU plotted against fragment size in bp) from a healthy donor and metastatic patient 1 are provided (c). Median cfDNA levels were significantly higher and median cfDNA fragment sizes were shorter in CRC patients than in healthy individuals. Mann–Whitney U test:  $*P < 0.05$ ,  $***P < 0.001$ . (d) Monitoring cfDNA yield (ng/mL plasma) and cfDNA fragment size (bp) of patient 1 prior and during chemotherapy (liquid biopsies 1–5) revealed a significantly strong negative correlation between cfDNA levels and fragment sizes (Pearson's correlation coefficient  $r = -0.956$ ,  $P = 0.01$ ). When patient 1 showed good response to first-line chemotherapeutic treatment (FOLFIRI + cetuximab), liquid biopsies 1–3 showed a decrease of cfDNA levels and an increase of cfDNA fragment sizes. Due to radiologic disease progression in the course of further treatment, therapy regimen changed to FOLFOX + bevacizumab, while liquid biopsy 4 simultaneously revealed an increase of cfDNA yield and a decrease of cfDNA fragment size.

able at the given allele frequency of 5% and 1%, and even two of these variants at 0.1% allele frequency were reproducibly measured. Altogether, NGS approach and dPCR enable variant detection up to 1% allele frequency (3/3 mutations), and dPCR additionally up to 0.1% allele frequency (2/3 mutations).

**Detection of sequence and structural variants in nuclear cfDNA is possible but input amount and read depth are limiting factors.** Next, we have compared NGS detection of sequence variants and CNVs from plasma to the current standard of analyzing tissue biopsies to examine the reliability of the used cfDNA workflow and to define quality thresholds for the applied molecular approaches. NGS screening of the tissue samples from 15 CRC patients of the cohort (non-metastatic:  $n = 7$ , locally advanced:  $n = 2$ , metastatic:  $n = 6$ ) revealed at least one molecular alteration in eight (53.3%), and two or more in seven (46.7%) of the patients, including in total 25 molecular alterations in the FFPE DNA (Fig. 3a). Of these, 15 were missense SNVs (60.0%), 3 were deletions (12.0%), 3 were nonsense SNVs (12.0%), 3 were splice site alterations (12.0%), and 1 was a structural variant, more precisely a CNV (4.0%). Altogether, the most frequently altered gene was *TP53*



**Figure 2.** Detection limits of NGS and dPCR. Validation of applied molecular approaches was performed by using 5 ng of the Multiplex cfDNA Reference Standard Set (HD780) for analyzing with NGS (Archer Reveal ctDNA 28 Concordance Kit,  $n=1$ ) and dPCR TaqMan assays ( $n=3$ , mean  $\pm$  SD). The detection of three variants (NRAS p.Gln61Lys, KRAS p.Gly12Asp, EGFR p.Thr790Met) at given allele frequencies (5%, 1%, 0.1%) was evaluated. Dashed lines indicate expected allele frequencies provided by the manufacturer of the reference standard and columns represent frequencies measured with NGS and dPCR. NGS and dPCR enabled variant detection at 5% and 1% allele frequency, and dPCR assays additionally detected variants at allele frequencies of 0.1%.

(64.0%) and the mutation patterns were consistent with typically mutated genes observed in CRC, previously reported by others<sup>29–33</sup>. However, comparative NGS analysis of plasma DNA exhibited 12 of the 25 molecular alterations from tissue DNA, which resulted in a variant-level concordance of 48.0% (Fig. 3a). Although NGS analysis of the cfDNA reference standard has previously determined a detection limit of 1%, known variants from tissue DNA which occur below 1% allele frequency and in addition show appropriate quality metrics in cfDNA NGS analysis were also considered genuine. Average variant allele frequency (VAF) of the tissue-plasma concordant variants from FFPE DNA samples was 37.8% (range 2.0–64.2%), and 14.5% (range 0.6–59.9%) for cfDNA samples (Table 2). Moreover, in three patients (20%) cfDNA analysis revealed additionally and therefore heterogenous variants in *ERBB2*, *KIT* and *TP53* that were not detected in FFPE DNA, but all of them have uncertain significance (Fig. 3a). Furthermore, tissue-plasma concordant patients were defined as having at least one concordant variant in plasma that matched the tumor tissue results, and thus 8 of the 15 CRC patients were tissue-plasma concordant, resulting in a patient-level concordance of 53.3%. Considering the patient-level concordance with regard to the respective CRC patient groups (non-metastatic, locally advanced, metastatic), patient-level concordance reached 83.3% (5/6) in the metastatic patient group, while among the two patients with locally advanced disease none of the tissue variants were detectable in matched plasma samples and in the non-metastatic patient group concordance was solely 42.9% (3/7). This indicates an impeded variant detection in patients with lower tumor burden. Besides, patients lacking detection of tissue variants in matched cfDNA ( $n=7$ , 46.7%) were primarily non-metastatic patients (4/7) or had locally advanced diseases (2/7). Overall, there was only one case of a metastatic patient (1/7) lacking a concordant variant from tissue in the respective plasma sample, arising presumably due to the technical fact that this patient had nearly the lowest cfDNA input amount (1.6 ng) of whole NGS analysis. Therefore, the sample consequently did not fulfill the quality metrics (coverage, on-target reads) for a reliable evaluation of variants. Thus, variant detection is critical below certain quality thresholds and leads to tissue-plasma discordance. In addition, cfDNA input amount showed overall a significantly strong positive correlation with coverage (Spearman's correlation coefficient  $r=0.868$ ,  $P<0.001$ ; Fig. 3b) and percentage of on-target reads (Spearman's correlation coefficient  $r=0.836$ ,  $P<0.001$ ; Fig. 3c). Furthermore, average target coverage was higher in tissue samples (average: 274) than in corresponding plasma samples (average: 101), as well as the average percentage of on-target reads (FFPE DNA: 77.8%, cfDNA: 34.4%). Accordingly, NGS analysis of cfDNA provided best results in metastatic patients and revealed that a low cfDNA input amount limits detection of molecular alterations due to lower values of coverage and on-target reads.

Apart from sequence variants, NGS analysis revealed a structural variant of the oncogene *ERBB2* (also called *HER2*) in both FFPE DNA and cfDNA of metastatic patient 10 (Fig. 3a). To further validate the NGS results of CNV status of patient 10, we performed dPCR analysis with a copy number TaqMan assay detecting this structural variant. Herein, Log<sub>2</sub>-Ratios from NGS approach and dPCR analyses showed good concordance (Fig. 4a).

Consistently with both approaches, *ERBB2* amplification could be detected in tissue samples of primary tumor and liver metastasis, and in cfDNA before chemotherapeutic treatment, but not after chemotherapy. This is in accordance with the detected amount and fragment size of cfDNA from patient 10, as cfDNA level decreased by almost 20 times and fragment size increased by 10 bp from the time before chemotherapy towards the end of it (Fig. 4b). In relation to this, detected SNVs in *TP53* gene of patient 10 (Fig. 3a) were also solely detectable in cfDNA before chemotherapy, altogether indicating an appropriate therapy response. Overall, the results of orthogonal validation of NGS approach by dPCR showed a strong positive correlation (Spearman's correlation coefficient:  $r = 0.949$ ,  $P = 0.05$ ; Supplementary Fig. S1), demonstrating the robust ability to detect CNVs in cfDNA by using both NGS and dPCR.

**Variant detection in mitochondrial cfDNA is difficult because of lower plasma copy number in CRC patients than in healthy controls.** Addressing the issue of low input amount and the necessity of higher sensitivity for cfDNA analysis, we analyzed mitochondrial DNA (mtDNA). Here, we compared variant detection in tissue and plasma mtDNA and examined properties of cell-free nuclear DNA (cf-nDNA) and cell-free mitochondrial DNA (cf-mtDNA) in CRC patients and healthy controls with regard to implement cf-mtDNA as a potential biomarker for CRC. Sequencing of mtDNA from tumor tissue ( $n = 5$ ) revealed eight tumor-specific molecular alterations with at least one variant in each patient and an average VAF of 26.2% (Fig. 5a). Of these eight mtDNA variants present in tumor specimens, only two (variant-level concordance: 25%) were found in matched cf-mtDNA from plasma samples, resulting in a patient-level concordance of only 20% and therefore showing even a lower concordance than NGS analysis of nuclear cfDNA. Moreover, cf-mtDNA revealed 31 additional variants (Supplementary Table S3) which were not detected in the corresponding tissues of the patients and thus are of unknown tissue origin. These results indicate that detection of tumor-specific mitochondrial variants in cfDNA is possible, but apparently only in patients with high tumor burden, such as metastatic patient 10 as the only case of detectable tissue mtDNA variants in matched plasma cf-mtDNA (Fig. 5a). Patient 10 previously showed a high cfDNA level (Fig. 4b) and a high frequency of tumor-specific sequence variants (Fig. 3a) and CNVs (Fig. 4a), indicating a high tumor load in plasma at time of blood sampling. Furthermore, patient 10 had the highest input amount of cfDNA and this resulted in the highest percentage of reads owning unique molecular identifiers, which are significantly correlated to each other (Spearman's correlation coefficient  $r = 0.975$ ,  $P < 0.01$ ; Supplementary Fig. S2). This demonstrates, as already described for sequencing of cf-nDNA (Fig. 3b,c), the technical influence of cfDNA input amount on variant detection from plasma. Overall, variant detection in mitochondrial cfDNA revealed unexpected difficulties, and thus showed no considerable advantages over cfDNA analysis from nucleus.

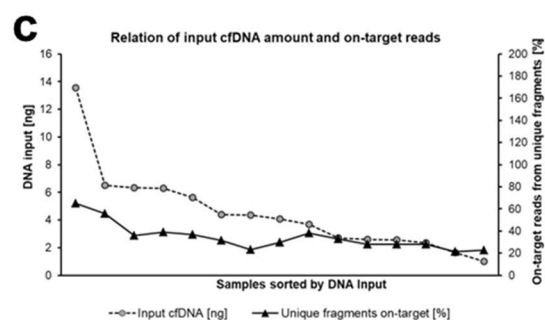
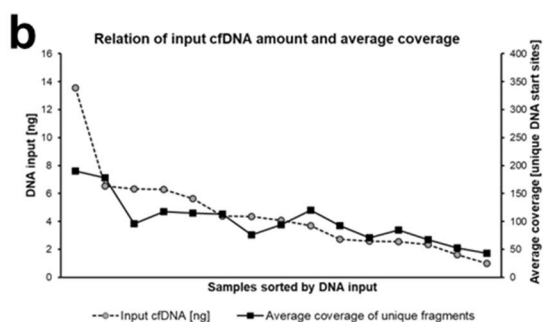
The challenging detection of tumor-specific mitochondrial tissue variants in corresponding plasma samples was very surprising due to the actually higher copy number of mtDNA. To verify the reasons for these difficulties, we compared the copy numbers of mtDNA from tumor and adjacent normal tissue and from pretherapeutic plasma and buffy coat samples from in total six patients of the CRC study cohort. An overview of all individual copy numbers is provided in Supplementary Table S5. Comparison of average mtDNA copy number (mtDNA copies per 1 ng of total DNA; Fig. 5b) calculated from all mitochondrial assays of the six CRC patients revealed that mtDNA has a significantly lower copy number in plasma (mean 125 copies) than in the buffy coat (mean 2949 copies,  $P < 0.001$ ), normal (mean 3982 copies,  $P < 0.001$ ) and tumor tissue (mean 3979 copies,  $P < 0.001$ ). This resulted in an average mtDNA copy number over 30 times higher in tumor tissue compared to plasma. Furthermore, we investigated mitochondrial and nuclear plasma copy numbers (copy number per milliliter of plasma) from CRC patients and healthy controls (Fig. 5c). Although average mitochondrial plasma copy number in CRC patients was significantly higher than nuclear plasma copy number (cf-mtDNA: mean 684 copies/mL plasma; cf-nDNA: mean 93 copies/mL plasma;  $P < 0.001$ ), the mitochondrial plasma concentration was in total surprisingly lower in CRC patients than in healthy individuals (CRC: mean 684 copies/mL plasma, healthy: mean 1081 copies/mL plasma;  $P < 0.05$ ), therefore reducing the intended effect of higher sensitivity for NGS approach. Moreover, mitochondrial plasma concentration was approximately twice as high when running the assay with the shortest amplicon size (*MT-CYB* 63 bp) compared to the assay with the longest amplicon size (*MT-CO2* 187 bp) (Fig. 5c). This resulted in an overall decrease of 45.5% in healthy individuals (decrease 648 copies/mL plasma) and 38.9% in CRC patients (decrease 293 copies/mL plasma), indicating the existence of a crucial proportion of short fragments in plasma mtDNA. Overall, the detection of mitochondrial variants in cfDNA using NGS approach might be affected for various technical and biological reasons. First, there generally is a much lower mitochondrial copy number in plasma than in tissue. Secondly, mitochondrial copy number is lower in CRC patients than in healthy individuals. Lastly, our results indicate that cf-mtDNA might be even shorter than cfDNA from nucleus.

## Discussion

In this study, we assessed the clinical value of several liquid biopsy applications for potential use in routine diagnostic and surveillance of CRC patients. An initial objective of the current study was the measurement of baseline cfDNA levels and fragment sizes. Consistent with previous results, the present study revealed higher cfDNA levels in cancer patients than in healthy individuals that additionally varied in the different clinicopathological groups of CRC patients<sup>9,15,18,34</sup>. These differences between healthy subjects and cancer patients as well as the variabilities among the CRC patient groups could, apart from the presence of a tumor, result from biologic and physiologic factors that affect the release mechanism and the clearance of cfDNA<sup>35–37</sup>. Furthermore, some studies have shown noticeable differences in tumor-derived ctDNA levels between patients even with the same cancer type and disease stage, resulting in an individually varying fraction of ctDNA from total cfDNA<sup>6,11,38</sup>. The mechanism of cfDNA release also has an impact on cfDNA fragment size, as presented by the “ladder” pattern of

**a**

Gene	COSMIC	Clinical significance (ClinVar)	HGVS c.	HGVS p.	Patients															
					non-metastatic					locally advanced	metastatic									
					8	11	12	13	14	16	18	17	19	1	2	5	6	7	10	
TP53	COSM45392	Uncertain significance (VUS)	NM_000546.5:c.936del	NP_000537.3:p.Ser313Alafs*32																
	COSM10704	Likely pathogenic	NM_000546.5:c.844C>T	NP_000537.3:p.Arg282Trp	X															
	COSM13745	Pathogenic	NM_000546.5:c.994-1G>T	N/A																
	COSM10654	Pathogenic	NM_000546.5:c.637C>T	NP_000537.3:p.Arg213*									X							
	COSM13586	Likely pathogenic	NM_000546.5:c.672+1G>T	N/A																
	COSM13747	Uncertain significance (VUS)	NM_000546.5:c.1146del	NP_000537.3:p.Lys382Asnfs*40							X									
	COSM10727	Pathogenic	NM_000546.5:c.438G>A	NP_000537.3:p.Trp146*																
	COSM69405	Likely pathogenic	NM_000546.5:c.375+1G>T	N/A																
	COSM45035	Uncertain significance (VUS)	NM_000546.5:c.761T>G	NP_000537.3:p.Ile254Ser																
	COSM6530	Uncertain significance (VUS)	NM_000546.5:c.723del	NP_000537.3:p.Cys242Alafs*5																
	COSM10867	Likely pathogenic	NM_000546.5:c.797G>A	NP_000537.3:p.Gly266Glu							X									
	COSM10659	Likely pathogenic	NM_000546.5:c.817C>T	NP_000537.3:p.Arg273Cys							X									
	COSM6914682	Likely pathogenic	NM_000546.5:c.325T>A	NP_000537.3:p.Phe109Ile										X						
	BRAF	COSM460	Pathogenic	NM_004333.4:c.1406G>C	NP_004324.2:p.Gly469Ala															
COSM476		Pathogenic	NM_004333.4:c.1799T>A	NP_004324.2:p.Val600Glu	X															
COSM27639		Likely pathogenic	NM_004333.4:c.1780G>A	NP_004324.2:p.Asp594Asn		X														
KRAS	COSM521	Pathogenic	NM_004985.4:c.35G>A	NP_004976.2:p.Gly12Asp						X										
	COSM532	Pathogenic	NM_004985.4:c.38G>A	NP_004976.2:p.Gly13Asp						X			X							
NRAS	COSM566	Pathogenic	NM_002524.4:c.35G>T	NP_002515.1:p.Gly12Val								X								
PIK3CA	COSM763	Likely pathogenic	NM_006218.3:c.1633G>A	NP_006209.2:p.Glu545Lys																
ERBB2	N/A	Uncertain significance (VUS)	NM_004448.3:c.2317G>A	NP_004439.2:p.Val773Met																X
	N/A	N/A	ERBB2 copy number GAIN (amplification)	N/A																
KIT	N/A	Uncertain significance (VUS)	NM_000222.2:c.1708del	NP_000213.1:p.Tyr570Thrfs*2		X														





◀**Figure 3.** Comparative NGS analysis of nuclear DNA from tissue and plasma. (a) Results of NGS analysis of 15 CRC patients using Archer Reveal ctDNA 28 Concordance Kit. We subsequently excluded some patients (patient 4, 9, 15) of the CRC cohort (see “Materials and methods” section) and therefore these sample IDs are missing in the depiction. Patients were categorized as non-metastatic (n = 7), locally advanced (n = 2), and metastatic (n = 6). Black boxes indicate variants identified in FFPE tumor samples, gray boxes indicate variants identified in cfDNA from plasma and crosses indicate lacking variants. Variant allele frequencies (VAFs) are listed in Table 2. Analyzed tissue and plasma samples were except from patient 2 taken before treatment while both samples from patient 2 were taken after radiochemotherapy. Sequence variants were classified according to the recommendations of the American College of Medical Genetics and Genomics (ACMG). Overall, 25 molecular alterations were detected in FFPE DNA of which 12 variants (variant-level concordance: 48.0%) were also detectable in cfDNA of eight patients, resulting in a patient-level concordance of 53.3% (8/15). Moreover, cfDNA analysis showed additionally variants in three patients (20%; patient 10, 11 and 16). (b) Average coverage from unique start sites was much higher in tissue samples than in corresponding plasma samples. Furthermore, average coverage and cfDNA input amount showed a significantly strong positive correlation (Spearman’s correlation coefficient  $r = 0.868$ ,  $P < 0.001$ ). (c) FFPE DNA showed a much higher percentage of on-target reads from unique fragments than cfDNA. Besides, percentage of on-target reads and cfDNA input amount showed a significantly strong positive correlation (Spearman’s correlation coefficient  $r = 0.836$ ,  $P < 0.001$ ). Databases for evaluation of variants: COSMIC<sup>67</sup>: <https://cancer.sanger.ac.uk/cosmic>, ClinVar<sup>69</sup>: <http://www.ncbi.nlm.nih.gov/clinvar>, CIViC<sup>68</sup>: <https://civicdb.org/home>; c.: coding DNA reference sequence, p.: protein reference sequence, VUS: variant of uncertain significance, N/A: not available.

cfDNA because of its apoptotic origin<sup>8,9</sup>. The present results corroborate previous findings that length distribution of cfDNA in cancer patients exhibits a more fragmented pattern than in healthy subjects<sup>13,39,40</sup>. Interestingly, longitudinal analysis of plasma samples from metastatic patient 1 demonstrated very short fragments before initiation of chemotherapy with a shift to longer fragments during chemotherapy. This emergence of longer wildtype cfDNA fragments was accompanied by a decreasing cfDNA level during therapy, indicating a shrinking tumoral bulk and an appropriate therapy response. It has been demonstrated, that cfDNA levels in CRC patients correlate with changes in tumor burden and treatment response<sup>5,41</sup>. Moreover, several studies have shown that cfDNA levels often provide a more sensitive and specific indication of real-time tumor burden in response to therapy than standard serum tumor markers in current clinical use (e.g. CEA)<sup>5,42</sup>. However, most of the studies only considered cfDNA levels and not longitudinal analysis of cfDNA size, although it is difficult to assess therapy response solely based on one parameter. In contrast, our study demonstrates that measurements of both, cfDNA level and size, can serve as reliable markers in clinical practice of CRC patients, not only to discriminate between healthy individuals and cancer patients (alone or in combination with blood protein tumor markers), but also to assess therapy response when performed as serial cfDNA level and size measurements to complement the information available from radiologic imaging.

In the present study, we observed a good patient-level concordance (53.3%) from NGS analysis of nuclear tissue and plasma DNA. This value is consistent with previously reported NGS detection rates in CRC patients (39–85.9%)<sup>43–46</sup>. However, in almost half of the patients (predominantly non-metastatic patients) we could not detect any of the tissue variants in matched plasma cfDNA, indicating that tissue analysis is particularly essential in non-metastatic CRC patients compared to metastatic patients. There are several possible biological and technical explanations for failing detection of variants in cfDNA and thus generating false-negative results. Firstly, it may be assumed that the number of available mutant molecules in plasma of the patients was too low for detection. In a landmark study of 640 patients with different types and stages of cancer, including CRC, quantification of individual tumor-associated mutations in each patient revealed that stage I patients had solely less than 10 mutant copies per 5 mL plasma, while patients with advanced cancers (stage IV) had more than 100 mutant copies per 5 mL plasma<sup>6</sup>. Related to these insights, the non-metastatic CRC patients of our study cohort who failed the detection of variants in cfDNA probably provided only a small fraction of copies carrying the mutation because they were at lower tumor stages (stage I-II) and median volume of used plasma was even lower than 5 mL, thus impeding variant detection in plasma. Secondly, variants in cfDNA occur mainly at very low allele frequencies and often  $< 1\%$ <sup>5,11,47</sup>. Thus, the sensitivity of NGS approach might have been too low for detecting those less frequent mutations. The possibility of detecting variants below 1% is mostly only provided if sequencing of the tissue DNA was performed first to search for known variants in cfDNA. Nevertheless, such variants must fulfill certain quality parameters (e.g. coverage) to be considered as true-positive variants. The achievable sensitivity of cfDNA assays is dependent on several factors, including sampling volume, cfDNA input amount, and the fraction of tumor-derived ctDNA. Concerning the present study, targeted input amount of 5 ng DNA corresponds to 1500 copies and therefore to 750 diploid genome equivalents (DNA amount corresponding to a diploid genome of a cell) which in theory leads to a sensitivity of  $\sim 0.07\%$ <sup>16,48</sup>. However, evaluation of the detection limits of the applied molecular workflows demonstrated that such calculated sensitivities cannot easily be translated into practice as presented by the determined lower detection limits of NGS (1%) and dPCR (0.1%). As discussed, this is evoked by the biological and physiological variabilities and thus poses a considerable challenge for accurate and reproducible variant detection in plasma. Furthermore, it can also be assumed that observed discordance of tissue and plasma analysis is affected by technical assay variation, as we have demonstrated by the correlation of cfDNA input amount and the quality metrics coverage and on-target reads, and therefore the impact on confidence of plasma mutation detection.

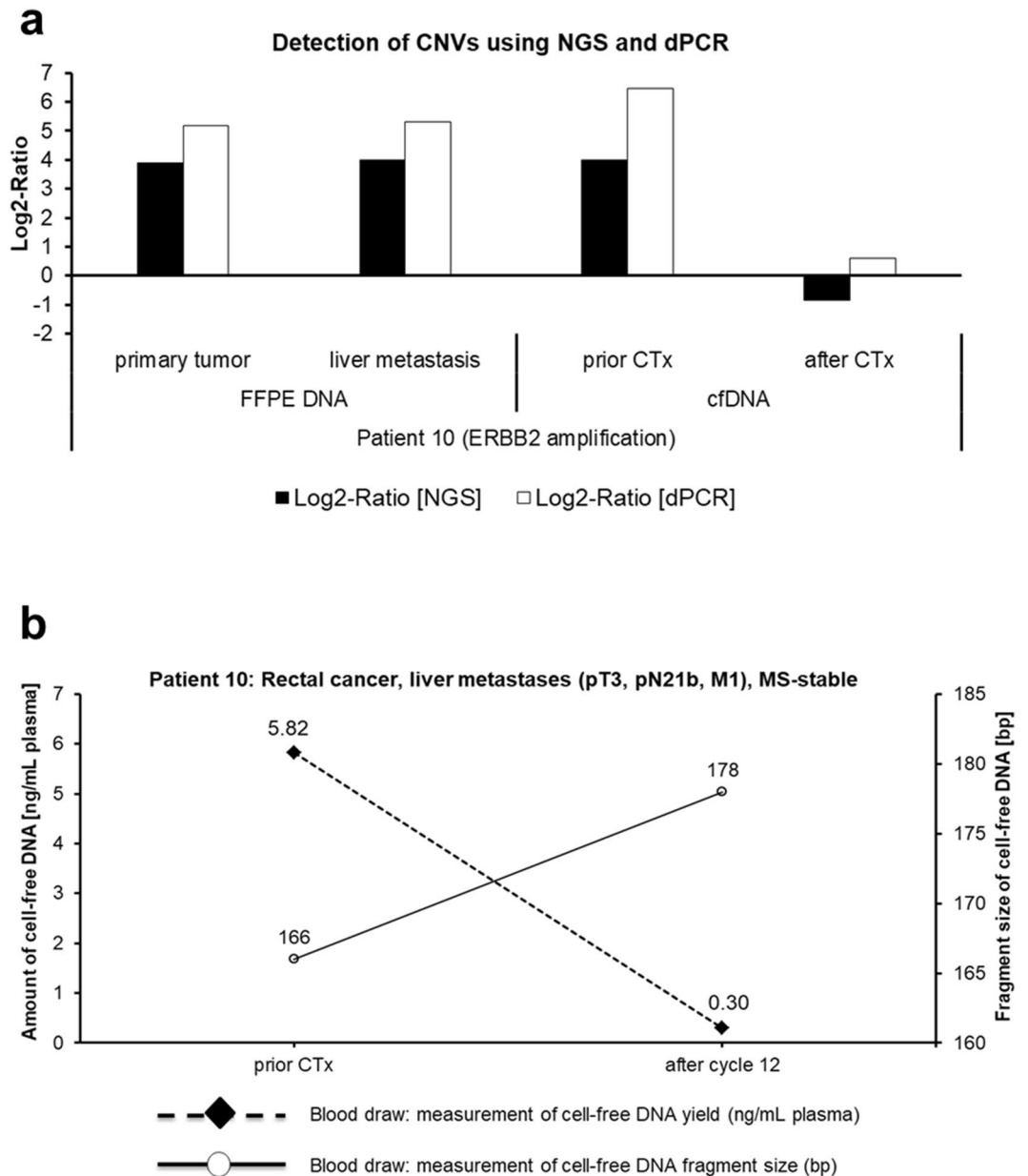
Another finding of the present study was that plasma NGS analysis revealed additional variants in cfDNA of three cases (20%), indicating that liquid biopsies enable in principle insights into tumor heterogeneity. Such

Patient	Gene	HGVS c.	HGVS p.	VAF FFPE DNA (%)	VAF cfDNA (%)
1	TP53	NM_000546.5:c.936del	NP_000537.3: p.Ser313Alafs*32	16.7	59.9
2	TP53	NM_000546.5:c.844C>T	NP_000537.3: p.Arg282Trp	59.2	5.7
5	TP53	NM_000546.5:c.994-1G>T	\	29.4	X
6	TP53	NM_000546.5:c.637C>T	NP_000537.3: p.Arg213*	64.2	3.2
	BRAF	NM_004333.4:c.1406G>C	NP_004324.2: p.Gly469Ala	58.0	3.1
	PIK3CA	NM_006218.3:c.1633G>A	NP_006209.2: p.Glu545Lys	56.5	3.4
7	TP53	NM_000546.5:c.672+1G>T	-	20.6	0.7
8	TP53	NM_000546.5:c.844C>T	NP_000537.3: p.Arg282Trp	27.1	X
		NM_000546.5:c.1146del	NP_000537.3: p.Lys382Asnfs*40	6.8	0.9
	BRAF	NM_004333.4:c.1799 T>A	NP_004324.2: p.Val600Glu	26.9	X
10	TP53	NM_000546.5:c.438G>A	NP_000537.3: p.Trp146*	32.8	35.4
		NM_000546.5:c.375+1G>T	-	28.9	31.8
	ERBB2	NM_004448.3:c.2317G>A	NP_004439.2: p.Val773Met	X	0.2
11	TP53	NM_000546.5:c.761 T>G	NP_000537.3: p.Ile254Ser	57.0	0.6
	BRAF	NM_004333.4:c.1780G>A	NP_004324.2: p.Asp594Asn	18.2	X
	KIT	NM_000222.2:c.1708del	NP_000213.1: p.Tyr570Thrfs*2	X	1.4
12	TP53	NM_000546.5:c.723del	NP_000537.3: p.Cys242Alafs*5	46.4	1.1
13	KRAS	NM_004985.4:c.35G>A	NP_004976.2: p.Gly12Asp	49.1	X
14	TP53	NM_000546.5:c.797G>A	NP_000537.3: p.Gly266Glu	63.6	X
	KRAS	NM_004985.4:c.38G>A	NP_004976.2: p.Gly13Asp	42.8	X
16	TP53	NM_000546.5:c.817C>T	NP_000537.3: p.Arg273Cys	49.8	X
		NM_000546.5:c.1146del	NP_000537.3: p.Lys382Asnfs*40	X	3.3
	NRAS	NM_002524.4:c.35G>T	NP_004976.2: p.Gly12Asp	47.4	X
17	TP53	NM_000546.5:c.325 T>A	NP_000537.3: p.Phe109Ile	39.2	X
18	TP53	NM_000546.5:c.844C>T	NP_000537.3: p.Arg282Trp	4.4	X
	KRAS	NM_004985.4:c.38G>A	NP_004976.2: p.Gly13Asp	2.0	X
19	TP53	NM_000546.5:c.637C>T	NP_000537.3: p.Arg213*	60.1	X

**Table 2.** Variant allele frequencies (VAF) of the comparative NGS analysis from nuclear DNA. X indicate variants which were not detected in the respective DNA from tissue or plasma. c.: coding DNA reference sequence, p.: protein reference sequence.

mutations could arise due to intratumoral heterogeneity<sup>49</sup> or metastasis-private events such as chromothripsis which has been shown to be prevalent in CRC<sup>50</sup>. Chromothripsis is defined as a single catastrophic phenomenon in which one or a few chromosomes in a cancer cell harbor clustered rearrangements and therefore oncogene amplifications. Indeed, NGS and dPCR analysis of metastatic patient 10 revealed an amplification of the oncogene *ERBB2* (*HER2*). The examination of *ERBB2* copy numbers is of emerging diagnostic value, because amplifications of *ERBB2* are a potential therapeutic target for metastatic CRC and have shown to be involved in acquired resistance to anti-*EGFR* antibodies<sup>51,52</sup>. Therefore, one interesting clinically relevant findings of the present study is that we were able to reliably detect this amplification in ctDNA of patient 10. Moreover, in a recent study with known *ERBB2*-positive metastatic CRC patients the use of ctDNA also showed accurate determination of *ERBB2* copy numbers and therefore was as reliable as tissue-based genotyping<sup>53</sup>. These results emphasize that liquid biopsies might be performed as a minimally invasive alternative to tissue-based detection of CNVs for assessment of therapy response or application of another therapy regimen with a better response.

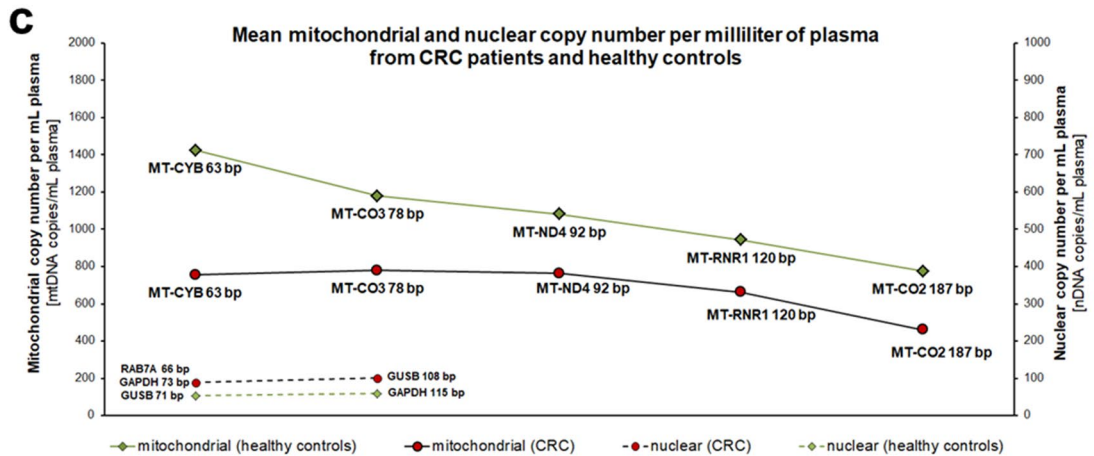
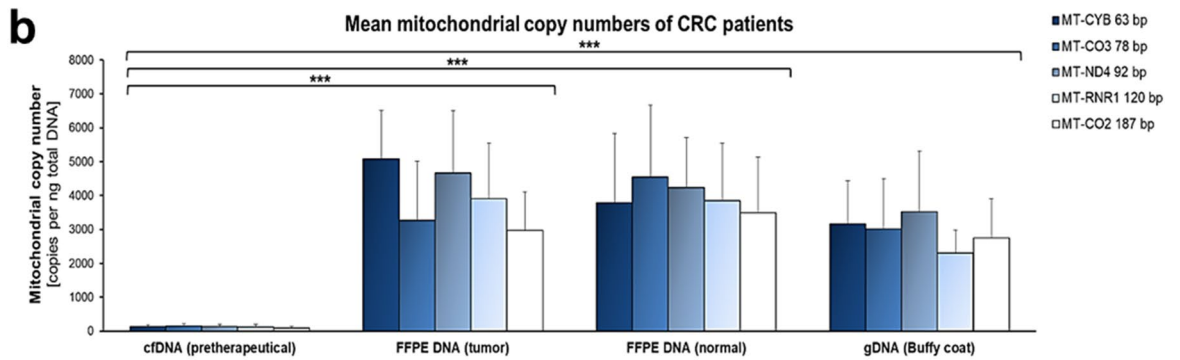
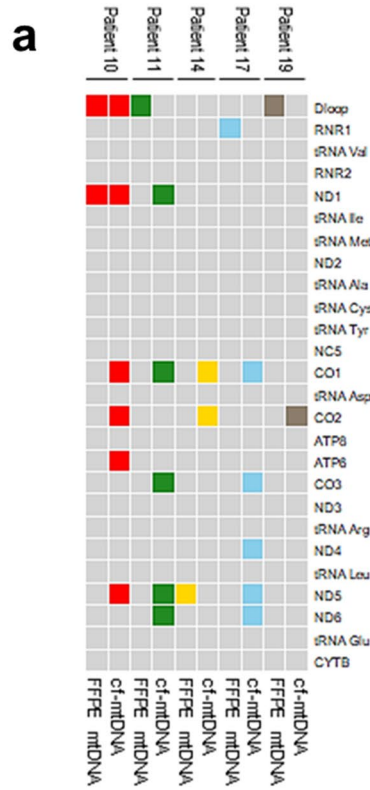
Next, we hypothesized that variants in cf-mtDNA would be more easily detectable as those in cf-nDNA due to higher copy number per cell and thus would require less sensitive applications to detect them. Overall, our results of comparative NGS analysis of mtDNA are consistent with a recently published study, also reporting limited ability in tracing tumor-specific cf-mtDNA variants by sequencing<sup>54</sup>. Most of the variants they identified from tumor tissue were undetectable in matched plasma mtDNA, resulting in a variant-level concordance (17%) even lower than ours (25%). Although in the present study NGS analysis of mitochondrial cfDNA revealed many additional variants, it can be assumed that these variants did not occur due to tumor heterogeneity and thus are not tumor-specific for CRC, because detection of the respective tumor tissue variants in matched cf-mtDNA was failing. The substantial discordance we observed between tissue and plasma mtDNA probably emerged from heteroplasmy, representing coexistence of mutated and wildtype mtDNA molecules in the same cell, tissue or organ<sup>55</sup>. The heteroplasmic variant pattern can differ within the tissues of an individual, often



**Figure 4.** Validation of copy number variation analysis using NGS and dPCR. **(a)** Comparison of Log<sub>2</sub>-Ratios from NGS using Archer Reveal ctDNA 28 Concordance Kit and dPCR TaqMan analysis showed good concordance. Gain of ERBB2 was detectable in tissue sample of primary tumor and liver metastasis, and also in cfDNA before chemotherapy (FOLFIRI + cetuximab), but not in cfDNA after it. **(b)** Course of amount and fragment size of cfDNA prior and after chemotherapeutic treatment of patient 10 was in agreement with variant detection. CTx: chemotherapy.

constituting heteroplasmic patient-specific variants from normal tissue but not tumor-specific heteroplasmic mtDNA variants<sup>56</sup>. However, cf-mtDNA, which is released from different cell types into circulation, represents the entirety of variants and therefore arising variant pattern is dependent on tissue origin of cf-mtDNA<sup>56,57</sup>.

Currently, it is still unclear why variant detection from plasma mtDNA is that challenging. There are several factors that should be considered: first, the variable mtDNA copy number of each cell can evoke that cells of non-tumor tissues have a higher mtDNA abundance than tumor cells. Thus, it might be that primarily non-tumor tissues are shedding mtDNA into blood circulation opposed to less frequent release of tumor-derived mtDNA. This leads to an accumulation of non-tumor variants and therefore an impeded detection of tumor-specific plasma mtDNA mutations due to a high wildtype background. Furthermore, a short fragment size of cf-mtDNA has been noticed by us and others which can be partially ascribed to the lack of nucleosome-associated histone proteins, resulting in a more fragmented pattern than its nuclear counterpart<sup>8,58,59</sup>. Another possible explanation



◀ **Figure 5.** Mitochondrial variants and copy numbers in tissue and plasma. (a) Heatmap showing results of variant detection in five CRC patients using QIAseq Targeted Panel DHS-105Z covering whole mitochondrial genome. Depicted variants are non-coding or non-synonymous and were categorized according to their occurrence in the different loci of mitochondrial genome. Colors represent variants from matched tissue and plasma samples originating from one patient. Overall, we found eight tumor-specific variants in tissue FFPE DNA of which two (25%) were also detectable in cf-mtDNA of patient 10 (patient-level concordance: 20%). (b) Column graph showing average mitochondrial copy numbers (mtDNA copies per 1 ng of applied DNA) of pretherapeutical cf-mtDNA, mtDNA from tumor and adjacent normal tissue, and from buffy coat of in total six CRC patients that had no distant metastases. Copy numbers were determined with dPCR TaqMan assays of varying amplicon length for mitochondrial DNA (*MT-CYB* 63 bp, *MT-CO3* 78 bp, *MT-ND4* 92 bp, *MT-RNR1* 120 bp, *MT-CO2* 187 bp) expressed as the mean  $\pm$  SD from all patients and the respective assay. Comparison revealed that cf-mtDNA ( $125 \pm 66$ ) had a significantly lower copy number than mtDNA from tumor ( $3979 \pm 1675$ ) and normal ( $3982 \pm 1723$ ) tissue and from buffy coat ( $2949 \pm 1301$ ). Welch's t-Test  $***P < 0.001$  (c) Comparison of average mitochondrial and nuclear plasma concentration (copy number per mL plasma) of CRC patients ( $n = 6$ ) and healthy controls ( $n = 7$ ) presented as the mean from all patients and the respective assay. Nuclear assays are combined to their fragment sizes, from short (*RAB7A* 66 bp, *GUSB* 71 bp, *GAPDH* 73 bp), to long (*GUSB* 108 bp, *GAPDH* 115 bp). Average mitochondrial plasma concentration was significantly higher than average nuclear plasma concentration in CRC patients (cf-mtDNA:  $684 \pm 357$ ; cf-nDNA:  $93 \pm 48$ ; Welch's t-test:  $P < 0.001$ ) and healthy controls (cf-mtDNA:  $1081 \pm 780$ ; cf-nDNA:  $55 \pm 19$ ; Mann-Whitney U test:  $P < 0.01$ ). Surprisingly, mitochondrial plasma concentration was higher in healthy controls than in CRC patients (healthy:  $1081 \pm 780$ ; CRC:  $684 \pm 357$ ; student's t-test:  $P < 0.05$ ). Overall, a decreasing copy number and therefore fragment size was noticeably in all individuals.

is that mtDNA might not be present in a cell-free form. A recent study described the occurrence of intact cell-free mitochondria with extracellular full-length mtDNA in blood<sup>60</sup>. Hence, it is possible that the mitochondrial fraction we have examined with the applied molecular approaches was not as expected primarily compound of short fragments but rather of long mtDNA which was not caught with the used techniques. Therefore, optimized protocols, such as size selection of cf-mtDNA fragments or an alternative method for library preparation like single-strand DNA libraries, are possible solutions for a promising outcome concerning analysis of cf-mtDNA. So far, there have been only a few studies investigating plasma mtDNA in CRC patients. In consistency with our results, those studies reported that mtDNA copy number is lower in plasma samples from CRC patients than in healthy subjects, also hampering detection of tumor-specific variants in plasma<sup>59,61</sup>. In many cancers, a reduction of mtDNA copy number is a common feature<sup>62</sup>. In addition, it has been reported that patients with low mtDNA content have increased benefit from chemotherapeutic treatment due to higher susceptibility to mitochondrial damage<sup>63</sup>. Therefore, in clinical settings it is conceivable that determination of mtDNA copy numbers may be helpful for prospective assessment of therapy response. However, this study included only a small sample size. For a comprehensive use of liquid biopsies in clinical practice further studies of paired tumor and plasma samples are necessary to evaluate the diagnostic accuracy of the analytical platforms, especially for variants with allele frequencies close to the limit of detection which should be interpreted with caution. Besides, defining standards for liquid biopsy approaches is still an ongoing process. Furthermore, the discordance of mutations detected from tumor tissue and plasma cfDNA reported in many studies is still one of the major challenges for application of liquid biopsies.

## Materials and methods

**Study design and patient cohort.** Between July 2017 and August 2020, we enrolled 22 colorectal cancer (CRC) patients that were treated surgically and/or with chemotherapy at the University Medical Center Mainz (Germany). All patients had a pathological diagnosis of primary colorectal adenocarcinoma. Of the enrolled 22 CRC patients, four patients (patient 4, 9, 15, 20) were subsequently excluded for one of the following reasons: lack of pretherapeutical tissue and/or plasma, inflammatory disease, or no detectable cfDNA. Thus, the final study cohort consisted of 18 patients with non-metastatic ( $n = 9$ ), locally advanced ( $n = 3$ ) or metastatic ( $n = 6$ ) CRC. Biological material, including DNA from biopsied or resected tissue specimen and cfDNA from plasma, was predominantly (in 17/18 patients) collected before any treatment for comparison of tissue and plasma mutation status and copy numbers. Levels of the protein tumor marker carcinoembryonic antigen (CEA) were also measured before treatment by the local diagnostic laboratory with CEA concentrations of  $< 5$  ng/mL considered normal. We also included a control group of 10 apparently healthy individuals with no current cancer diagnosis. Of these healthy controls, 80% were female and median age was 31 years (range 23–56 years).

**Sample collection and processing.** Formalin-fixed paraffin-embedded (FFPE) DNA from diagnostic biopsied or resected tumor and adjacent normal tissues was extracted from  $\sim 5$   $\mu$ m thick tissue sections using QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany). FFPE DNA quantification was performed using Qubit 1.0 Fluorometer (Thermo Fisher Scientific, Dreieich, Germany). Venous blood from healthy donors and CRC patients was taken with a blood collection set with preattached holder (Becton Dickinson, Heidelberg, Germany) and PAXgene Blood ccfDNA Tubes (Qiagen). From all individuals, we collected two tubes with a maximum volume of 10 mL each. After blood drawing, tubes were immediately inverted and directly processed. Whole-blood samples were centrifuged at  $2.000 \times g$  for 10 min at  $4^\circ\text{C}$  and plasma was transferred to a fresh tube and was centrifuged a second time at  $16.000 \times g$  for 10 min at  $4^\circ\text{C}$ . Plasma was then used for extraction of cfDNA with QIAamp MinElute ccfDNA Midi Kit (Qiagen) and buffy coat layer containing the white blood cells was

used for extraction of genomic DNA (gDNA) using QIAmp DNA Blood Mini Kit (Qiagen). Concentration of gDNA was measured with Qubit 1.0 Fluorometer (Thermo Fisher Scientific). Quality and quantity of extracted cfDNA was assessed via electropherograms by using Agilent High Sensitivity DNA Kit for 2100 Bioanalyzer instrument (Agilent Technologies, Waldbronn, Germany). Data were analyzed with 2100 Expert Software version B.02.08 (Agilent Technologies). cfDNA fragment size was defined as the average size of the main peak in the electropherogram and cfDNA level was calculated as the area under the main peak.

**Molecular analysis of nuclear DNA.** For comparison of tissue and plasma mutation status and copy numbers, FFPE DNA and cfDNA from CRC patients with sufficient material ( $n = 15$ ) were analyzed for somatic mutations using the targeted NGS panel Reveal ctDNA 28 Concordance Kit (ArcherDx, Boulder, USA) including 28 genes (*ALK, AKT1, AR, BRAF, CTNNB1, DDR2, EGFR, ERBB2, ESRI, FGFR1, HRAS, IDH1, IDH2, KIT, KRAS, MAP2K1, MAP2K2, MET, NRAS, NTRK1, NTRK3, PIK3CA, PDGFRA, RET, ROS1, SMAD4, MTOR, TP53*). This panel is based on anchored multiplex PCR (AMP), a target enrichment method that uses unidirectional gene-specific primers and molecular barcoded (MBC) adapters for amplification<sup>20</sup>. Molecular barcoding identifies reads originating from the same molecule, correcting for PCR or sequencing errors<sup>48,64–66</sup>. The panel enables detection of sequence variants and calling of copy number variations (CNV) by counting the uniquely tagged molecules to deduce the copy number from each target region in comparison to the baseline of normal samples. For workflow validation and evaluation of detection limit, we used 5 ng of the commercially available Multiplex I cfDNA Reference Standard Set HD780 (Horizon Discovery, Cambridge, UK). For generating NGS libraries, average input amount was 4.5 ng (range 1.0–13.6 ng) for cfDNA and 184.5 ng (range 12.9–250 ng) for FFPE DNA. Libraries were quantified using Qubit 1.0 Fluorometer (Thermo Fisher Scientific) and diluted to a concentration of 4 nM. Quality control of generated libraries was performed by using 2100 Bioanalyzer Instrument (Agilent Technologies). All libraries were denatured by adding 0.2 nM NaOH and diluted to 13 pM for FFPE DNA and 10 pM for cfDNA with hybridization buffer from Illumina (San Diego, USA). The number of pooled samples was adapted to type of library to ensure FFPE DNA libraries obtain 1 million reads and cfDNA libraries 5 million reads per sample. Sequencing was performed on MiSeq instrument using MiSeq Reagent Kit v3 from Illumina and paired-end sequencing with  $2 \times 151$  bp reads. FASTQ files were processed using Archer Analysis Software Version 6.2.3 (ArcherDx) with the appropriate SNV and CNV analysis workflow. For analysis of cfDNA, advanced settings of a read depth normalization of 10,000,000 reads, specifying the number of reads for randomly subsampling, and a variant downstream region of interest size of 150 bases, which defines the number of bases downstream of the gene-specific primers that are considered when searching for variants, were used. The software provides all secondary analysis (read trimming, de-duplication, error correction, alignment, mutation calling). Results are presented relative to hg19 (GRCh37). Molecular alterations were evaluated on entries in the cancer-specific variant databases COSMIC<sup>67</sup> and CIViC<sup>68</sup> as well as the variant interpretation database ClinVar<sup>69</sup>. Sequence variants were classified according to the recommendations of the American College of Medical Genetics and Genomics (ACMG). Coverage was defined as the reads from unique start sites in each target region with a quality control threshold of 50. All sequencing data of nuclear DNA are listed in Supplementary Table S1.

Evaluation of detection limit of dPCR (QuantStudio 3D Digital PCR System, Thermo Fisher Scientific) was performed by using 5 ng of the Multiplex I cfDNA Reference Standard Set HD780 (Horizon Discovery) and TaqMan dPCR assays from Thermo Fisher Scientific (*NRAS* p.Gln61Lys, assay ID: Hs000000079\_rm; *KRAS* p.Gly12Asp, assay ID: Hs000000051\_rm; *EGFR* p.Thr790Met, assay ID: Hs000000029\_rm). For validation of copy number status detected with NGS workflow, we used a TaqMan copy number assay for *ERBB2* (Hs00450668\_cn, Thermo Fisher Scientific). dPCR analysis enables an absolute copy number quantitation of gene of interest which is then normalized to a reference gene known to be present in two copies in a diploid genome. For each dPCR reaction, copy number assay of the target gene and the reference gene *RNase P* (assay ID: 4,403,326, Thermo Fisher Scientific) were run simultaneously with 1 ng FFPE DNA from tumor and matched normal tissue and with cfDNA from liquid biopsies of *ERBB2*-positive patient 10. dPCR was performed using QuantStudio 3D Digital PCR Master Mix and 3D PCR 20 K Chip Kit version 2 on a flat block thermocycler (ProFlex) from Thermo Fisher Scientific. Setup of dPCR reactions were as follows: hot start at 96 °C for 10 min, denaturation at 98 °C for 30 s, annealing/extension at 60 °C for 2 min for a total of 39 cycles, followed by a final extension step at 60 °C for 2 min. Data were analyzed with QuantStudio 3D AnalysisSuite Cloud Software version 3.2 (Thermo Fisher Scientific).

**Molecular analysis of mitochondrial DNA.** Mitochondrial DNA (mtDNA) of pretherapeutic tissue and plasma was analyzed by NGS for variant detection and by dPCR for determination of copy numbers. For excluding variants arising from clonal hematopoiesis as well as constitutional polymorphisms and as a control for copy numbers, we analyzed FFPE DNA from tumor and tumor-adjacent normal tissue (intestinal mucosa) as well as cfDNA from plasma and the corresponding leukocyte DNA from buffy coat. NGS libraries for variant detection were constructed with the QIAseq Targeted DNA Panel-Human Mitochondria Panel (DHS-105Z; Qiagen) covering the whole mitochondrial genome. This NGS panel is based on AMP as well and uses unique molecular identifiers (UMIs) to better differentiate NGS artifacts from real mutations at very low allele fractions. For generating cfDNA libraries, the enzymatic fragmentation was inhibited. DNA input amount of tissue samples ( $n = 10$ ) and buffy coat samples ( $n = 4$ ) was 100 ng, and average input amount of cfDNA samples ( $n = 5$ ) was 4.7 ng (range 0.5–10.0 ng). Quantification, quality control, dilution, and sequencing of the libraries were performed as described above for NGS analysis of nuclear DNA, but final concentration of all mtDNA libraries was 10 pM. Sequencing was carried out to ensure that buffy coat DNA libraries had 500,000 reads, FFPE DNA libraries 1 million reads, and cfDNA libraries 5 million reads. FASTQ files were processed using CLC Genomics Workbench version 12.0 (Qiagen) with a customized data analysis workflow (Supplementary

Table S2). Reads were aligned to Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA (NC\_012920). Mitochondrial variants were evaluated on entries of the database MITOMAP<sup>70</sup>. Quality threshold of coverage was 100. Supplementary Table S3 gives an overview of all mitochondrial sequencing data.

Determination of copy numbers was performed using dPCR (QuantStudio 3D Digital PCR System, Thermo Fisher Scientific) with customized TaqMan assays of varying amplicon length for mitochondrial DNA (*MT-CYB* 63 bp, *MT-CO3* 78 bp, *MT-ND4* 92 bp, *MT-RNR1* 120 bp, *MT-CO2* 187 bp) and nuclear DNA (*GAPDH* 73 bp, *GAPDH* 115 bp, *GUSB* 71 bp, *GUSB* 108 bp, *RAB7A* 66 bp). Panel of amplicons with various sizes was designed using Primer 3 (<http://primer3.ut.ee>) and assays were manufactured by Integrated DNA Technologies (Interleuvenlaan, Belgium). mtDNA assays contained a FAM dye labeled minor groove binding probe and nuclear DNA assays contained a HEX dye-labeled TAMRA probe. TaqMan assays were run simultaneously in a duplex dPCR reaction. Sequences of primers and probes and appropriate ratios are listed in Supplementary Table S4. TaqMan assays (0.75 µl each) were mixed with 7.5 µl QuantStudio 3D Digital PCR Master Mix and variable volumes of DNA and H<sub>2</sub>O to a final volume of 15 µl, of which 14.5 µl were put on a 3D PCR 20 K Chip version 2 and run on a flat block thermocycler (ProFlex) from Thermo Fisher Scientific with following PCR conditions: hot start at 96 °C for 10 min, denaturation at 98 °C for 30 s, annealing/extension at 58 °C for 2 min for a total of 39 cycles, followed by a final extension step at 58 °C for 2 min. Data were analyzed with QuantStudio 3D AnalysisSuite Cloud Software version 3.2 (Thermo Fisher Scientific). Copy number analysis was performed with in total 12 tissue (tumor and normal), and six buffy coat and plasma samples, with an input amount of 1 ng for FFPE DNA and gDNA, 0.6 ng for patients' cfDNA (n=6) and 0.5 ng for healthy individuals (n=7) (Supplementary Table S5).

**Statistical analysis.** All statistical analyses were performed with RStudio version 3.5.3 (R Foundation for Statistical Computing, Vienna, Austria) which is the graphical user interface of R programming language (R Core Team, 2020). Differences between experimental groups were considered statistically significant when *P* values were < 0.05. Shapiro-Wilk normality test was used to determine whether values follow a normal distribution. F-test was used to assess the equality of variances. Two-sided Mann-Whitney U test was used to compare nuclear cfDNA fragment sizes and levels of control and patient samples. T-Test (student's t-test or Welch's t-test) or Mann-Whitney U test was used for comparison of mitochondrial and nuclear copy numbers as indicated. Correlations were analyzed using the parametric Pearson correlation coefficient or the non-parametric rank-based Spearman correlation coefficient as indicated.

**Ethics approval.** All procedures performed in this study involving human participants were approved by the ethical standards of the local ethics committee (Ethics Committee of the Rhineland-Palatinate State Medical Association, Mainz, Germany, 11,194) and were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from all patients.

Received: 20 April 2021; Accepted: 20 July 2021

Published online: 18 August 2021

## References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–386. <https://doi.org/10.1002/ijc.29210> (2015).
2. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
3. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94. <https://doi.org/10.1038/nrclinonc.2017.166> (2018).
4. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat. Med.* **21**, 795–801. <https://doi.org/10.1038/nm.3870> (2015).
5. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990. <https://doi.org/10.1038/nm.1789> (2008).
6. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra224. <https://doi.org/10.1126/scitranslmed.3007094> (2014).
7. Parikh, A. R. *et al.* Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat. Med.* **25**, 1415–1421. <https://doi.org/10.1038/s41591-019-0561-9> (2019).
8. Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91. <https://doi.org/10.1126/scitranslmed.3001720> (2010).
9. Thierry, A. R. *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res.* **38**, 6159–6175. <https://doi.org/10.1093/nar/gkq421> (2010).
10. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548. <https://doi.org/10.1038/nrclinonc.2017.14> (2017).
11. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16368–16373. <https://doi.org/10.1073/pnas.0507904102> (2005).
12. Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin. Chim. Acta Int. J. Clin. Chem.* **313**, 139–142. [https://doi.org/10.1016/s0009-8981\(01\)00665-9](https://doi.org/10.1016/s0009-8981(01)00665-9) (2001).
13. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Can. Res.* **61**, 1659–1665 (2001).
14. Wan, J. C. M. *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238. <https://doi.org/10.1038/nrc.2017.7> (2017).
15. Kim, K. *et al.* Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: Diagnostic validity and significant reduction of cfDNA after surgical resection. *Ann. Surg. Treat. Res.* **86**, 136–142. <https://doi.org/10.4174/ast.2014.86.3.136> (2014).

16. Dasari, A. *et al.* ctDNA applications and integration in colorectal cancer: An NCI Colon and rectal-anal task forces whitepaper. *Nat. Rev. Clin. Oncol.* <https://doi.org/10.1038/s41571-020-0392-0> (2020).
17. Thierry, A. R. *et al.* Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat. Med.* **20**, 430–435. <https://doi.org/10.1038/nm.3511> (2014).
18. Crowley, E., Di Nicolantonio, F., Loupakis, F. & Bardelli, A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **10**, 472–484. <https://doi.org/10.1038/nrclinonc.2013.110> (2013).
19. Baker, M. Digital PCR hits its stride. *Nat. Methods* **9**, 541–544. <https://doi.org/10.1038/nmeth.2027> (2012).
20. Zheng, Z. *et al.* Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484. <https://doi.org/10.1038/nm.3729> (2014).
21. Jennings, L. J. *et al.* Guidelines for validation of next-generation sequencing-based oncology panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagn. JMD* **19**, 341–365. <https://doi.org/10.1016/j.jmoldx.2017.01.011> (2017).
22. Chiu, R. W. *et al.* Quantitative analysis of circulating mitochondrial DNA in plasma. *Clin. Chem.* **49**, 719–726. <https://doi.org/10.1373/49.5.719> (2003).
23. Copeland, W. C., Wachsman, J. T., Johnson, F. M. & Penta, J. S. Mitochondrial DNA alterations in cancer. *Cancer Investig.* **20**, 557–569. <https://doi.org/10.1081/cnv-120002155> (2002).
24. Hertweck, K. L. & Dasgupta, S. The landscape of mtDNA modifications in cancer: A tale of two cities. *Front. Oncol.* **7**, 262. <https://doi.org/10.3389/fonc.2017.00262> (2017).
25. van Osch, F. H. *et al.* Mitochondrial DNA copy number in colorectal cancer: Between tissue comparisons, clinicopathological characteristics and survival. *Carcinogenesis* **36**, 1502–1510. <https://doi.org/10.1093/carcin/bgv151> (2015).
26. An, Q. *et al.* The size of cell-free mitochondrial DNA in blood is inversely correlated with tumor burden in cancer patients. *Precis. Clin. Med.* **2**, 131–139. <https://doi.org/10.1093/pcmedi/pbz014> (2019).
27. Moulriere, F. & Rosenfeld, N. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3178–3179. <https://doi.org/10.1073/pnas.1501321112> (2015).
28. Heinemann, V. *et al.* FOLFIRI plus cetuximab or bevacizumab for advanced colorectal cancer: Final survival and per-protocol analysis of FIRE-3, a randomised clinical trial. *Br. J. Cancer* **124**, 587–594. <https://doi.org/10.1038/s41416-020-01140-9> (2021).
29. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216. <https://doi.org/10.1038/ng.3214> (2015).
30. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356. <https://doi.org/10.1038/nm.3967> (2015).
31. Vogelstein, B. *et al.* Cancer genome landscapes. *Science (New York, N.Y.)* **339**, 1546–1558. <https://doi.org/10.1126/science.1235122> (2013).
32. Lang, H. *et al.* Extended molecular profiling improves stratification and prediction of survival after resection of colorectal liver metastases. *Ann. Surg.* **270**, 799–805. <https://doi.org/10.1097/sla.0000000000003527> (2019).
33. Stahler, A. *et al.* Single-nucleotide variants, tumour mutational burden and microsatellite instability in patients with metastatic colorectal cancer: Next-generation sequencing results of the FIRE-3 trial. *Eur. J. Cancer (Oxford, England: 1990)* **137**, 250–259. <https://doi.org/10.1016/j.ejca.2020.07.003> (2020).
34. Fleischhacker, M. & Schmidt, B. Circulating nucleic acids (CNAs) and cancer—A survey. *Biochem. Biophys. Acta* **1775**, 181–232. <https://doi.org/10.1016/j.bbcan.2006.10.001> (2007).
35. Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437. <https://doi.org/10.1038/nrc3066> (2011).
36. Bronkhorst, A. J. *et al.* Characterization of the cell-free DNA released by cultured cancer cells. *Biochem. Biophys. Acta* **157–165**, 2016. <https://doi.org/10.1016/j.bbamcr.2015.10.022> (1863).
37. Pisetsky, D. S. & Fairhurst, A. M. The origin of extracellular DNA during the clearance of dead and dying cells. *Autoimmunity* **40**, 281–284. <https://doi.org/10.1080/08916930701358826> (2007).
38. Diaz, L. A. Jr. & Bardelli, A. Liquid biopsies: Genotyping circulating tumor DNA. *J. Clin. Oncol.* **32**, 579–586. <https://doi.org/10.1200/jco.2012.45.2011> (2014).
39. Moulriere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941. <https://doi.org/10.1016/j.molonc.2014.02.005> (2014).
40. Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389. <https://doi.org/10.1038/s41586-019-1272-6> (2019).
41. Tie, J. *et al.* Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26**, 1715–1722. <https://doi.org/10.1093/annonc/mdv177> (2015).
42. Frattini, M. *et al.* Quantitative and qualitative characterization of plasma DNA identifies primary and recurrent colorectal cancer. *Cancer Lett.* **263**, 170–181. <https://doi.org/10.1016/j.canlet.2008.03.021> (2008).
43. Beije, N. *et al.* Somatic mutation detection using various targeted detection assays in paired samples of circulating tumor DNA, primary tumor and metastases from patients undergoing resection of colorectal liver metastases. *Mol. Oncol.* **10**, 1575–1584. <https://doi.org/10.1016/j.molonc.2016.10.001> (2016).
44. Kato, S. *et al.* Genomic assessment of blood-derived circulating tumor DNA in patients with colorectal cancers: Correlation with tissue sequencing, therapeutic response, and survival. *JCO Precis. Oncol.* **3**, 1–16. <https://doi.org/10.1200/PO.18.00158> (2019).
45. Kim, S. T. *et al.* Prospective blinded study of somatic mutation detection in cell-free DNA utilizing a targeted 54-gene next generation sequencing panel in metastatic solid tumor patients. *Oncotarget* **6**, 40360–40369. <https://doi.org/10.18632/oncotarget.5465> (2015).
46. Strickler, J. H. *et al.* Genomic landscape of cell-free DNA in patients with colorectal cancer. *Cancer Discov.* **8**, 164–173. <https://doi.org/10.1158/2159-8290.Cd-17-1009> (2018).
47. Aravanis, A. M., Lee, M. & Klausner, R. D. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* **168**, 571–574. <https://doi.org/10.1016/j.cell.2017.01.030> (2017).
48. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet* **20**, 71–88. <https://doi.org/10.1038/s41576-018-0071-5> (2019).
49. Kim, T. M. *et al.* Subclonal genomic architectures of primary and metastatic colorectal cancer based on intratumoral genetic heterogeneity. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **21**, 4461–4472. <https://doi.org/10.1158/1078-0432.CCR-14-2413> (2015).
50. Voronina, N. *et al.* The landscape of chromothripsis across adult cancer types. *Nat. Commun.* **11**, 2320. <https://doi.org/10.1038/s41467-020-16134-7> (2020).
51. Mohan, S. *et al.* Changes in colorectal carcinoma genomes under anti-EGFR therapy identified by whole-genome plasma DNA sequencing. *PLoS Genet.* **10**, e1004271. <https://doi.org/10.1371/journal.pgen.1004271> (2014).
52. Meric-Bernstam, F. *et al.* Pertuzumab plus trastuzumab for HER2-amplified metastatic colorectal cancer (MyPathway): An updated report from a multicentre, open-label, phase 2a, multiple basket study. *Lancet Oncol.* **20**, 518–530. [https://doi.org/10.1016/s1470-2045\(18\)30904-5](https://doi.org/10.1016/s1470-2045(18)30904-5) (2019).
53. Siravegna, G. *et al.* Plasma HER2 (ERBB2) copy number predicts response to HER2-targeted therapy in metastatic colorectal cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **25**, 3046–3053. <https://doi.org/10.1158/1078-0432.Ccr-18-3389> (2019).



54. Weerts, M. J. A. *et al.* Tumor-specific mitochondrial DNA variants are rarely detected in cell-free DNA. *Neoplasia* **20**, 687–696. <https://doi.org/10.1016/j.neo.2018.05.003> (2018).
55. Wallace, D. C. Diseases of the mitochondrial DNA. *Annu. Rev. Biochem.* **61**, 1175–1212. <https://doi.org/10.1146/annurev.bi.61.070192.005523> (1992).
56. He, Y. *et al.* Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* **464**, 610–614. <https://doi.org/10.1038/nature08802> (2010).
57. Li, M., Schroder, R., Ni, S., Madea, B. & Stoneking, M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2491–2496. <https://doi.org/10.1073/pnas.1419651112> (2015).
58. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **17**, 100087. <https://doi.org/10.1016/j.bdq.2019.100087> (2019).
59. Liu, Y. *et al.* NGS-based accurate and efficient detection of circulating cell-free mitochondrial DNA in cancer patients. *Mol. Ther. Nucleic Acids* **23**, 657–666. <https://doi.org/10.1016/j.omtn.2020.12.017> (2021).
60. Al Amir Dache, Z. *et al.* Blood contains circulating cell-free respiratory competent mitochondria. *Faseb J.* **34**, 3616–3630. <https://doi.org/10.1096/fj.201901917RR> (2020).
61. Meddeb, R. *et al.* Quantifying circulating cell-free DNA in humans. *Sci. Rep.* **9**, 1–16. <https://doi.org/10.1038/s41598-019-41593-4> (2019).
62. Wallace, D. C. Mitochondria and cancer. *Nat. Rev. Cancer* **12**, 685–698. <https://doi.org/10.1038/nrc3365> (2012).
63. Weerts, M. J. A. *et al.* Low tumor mitochondrial DNA content is associated with better outcome in breast cancer patients receiving anthracycline-based chemotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **23**, 4735–4743. <https://doi.org/10.1158/1078-0432.Ccr-17-0032> (2017).
64. Haupts, A., Roth, W. & Hartmann, N. Liquid biopsy in colorectal cancer: An overview of ctDNA analysis in tumour diagnostics. *Pathologie* **40**, 244–251. <https://doi.org/10.1007/s00292-019-00698-3> (2019).
65. Keller, L., Belloum, Y., Wikman, H. & Pantel, K. Clinical relevance of blood-based ctDNA analysis: Mutation detection and beyond. *Br. J. Cancer* <https://doi.org/10.1038/s41416-020-01047-5> (2020).
66. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14508–14513. <https://doi.org/10.1073/pnas.1208715109> (2012).
67. Forbes, S. A. *et al.* COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–811. <https://doi.org/10.1093/nar/gku1075> (2015).
68. Griffith, M. *et al.* CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **49**, 170–174. <https://doi.org/10.1038/ng.3774> (2017).
69. Landrum, M. J. *et al.* ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–985. <https://doi.org/10.1093/nar/gkt1113> (2014).
70. Kogelnik, A. M., Lott, M. T., Brown, M. D., Navathe, S. B. & Wallace, D. C. MITOMAP: a human mitochondrial genome database. *Nucleic Acids Res.* **24**, 177–179. <https://doi.org/10.1093/nar/24.1.177> (1996).

## Acknowledgements

We would like to thank Klaus Rehfeldt and Andre Koper from Qiagen for supporting the development of a customized data analysis workflow for CLC Genomics Workbench. We thank Yasmin Isenbruck-Wenk and Ulrike Süßdorf for organization of blood sampling. We thank all healthy donors.

## Author contributions

Designed the experiments: A.H., N.H., W.R.; Performed the experiments: A.H., N.H., S.F., T.H., N.W., A.V.; Analyzed the data: A.H., N.H., W.R., M.M., H.L., W.K., T.H., M.H., A.M.; Wrote the paper: A.H., N.H.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95006-6>.

**Correspondence** and requests for materials should be addressed to A.H. or N.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021