Detection of circulating tumor DNA by tumor-informed whole-genome sequencing enables prediction of recurrence in stage III colorectal cancer patients

Frydendahl, Amanda; Nors, Jesper; Rasmussen, Mads H.; Henriksen, Tenna V.; Nesic, Marijana; Reinert, Thomas; Afterman, Danielle; Lauterman, Tomer; Kuzman, Maja; Gonzalez, Santiago; Glavas, Dunja; Smadback, James; Maloney, Dillon; Levativ, Jurica; Yahalom, Michael; Ptashkin, Ryan; Tavassoly, Iman; Donenhirsh, Zohar; White, Eric; Kandasamy, Ravi; Alon, Ury; Nordentoft, Iver; Lindskrog, Sia V.; Dyrskjøt, Lars; Jaensch, Claudia; Løve, Uffe S.; Andersen, Per V.; Thorlacius-Ussing, Ole; Iversen, Lene H.; Gotschalck, Kåre A.; Zviran, Asaf; Oklander, Boris; Andersen, Claus L.

Published in:
European Journal of Cancer

DOI:
10.1016/j.ejca.2024.114314

Publication date:
2024

Document version:
Final published version

European Journal of Cancer 211 (2024) 114314

Contents lists available at ScienceDirect

European Journal of Cancer

journal homepage: www.ejcancer.com

Original research

Detection of circulating tumor DNA by tumor-informed whole-genome sequencing enables prediction of recurrence in stage III colorectal cancer patients

Amanda Frydendahl [a,b], Jesper Nors [a,b], Mads H. Rasmussen [a,b], Tenna V. Henriksen [a,b], Marijana Nesic [a,b], Thomas Reinert [a,b], Danielle Afterman [c], Tomer Lauterman [c], Maja Kuzman [d], Santiago Gonzalez [d], Dunja Glavas [d], James Smadback [d], Dillon Maloney [d], Jurica Levativ [d], Michael Yahalom [d], Ryan Ptashkin [d], Iman Tavassoly [d], Zohar Donenhirsh [c], Eric White [d], Ravi Kandasamy [d], Ury Alon [c], Iver Nordentoft [a,b], Sia V. Lindskrog [a,b], Lars Dyrskjøt [a,b], Claudia Jaensch [e], Uffe S. Løve [f], Per V. Andersen [g], Ole Thorlacius-Ussing [h], Lene H. Iversen [b,i], Kåre A. Gotschalck [b,j], Asaf Zviran [d], Boris Oklander [c], Claus L. Andersen [a,b,*]

[a] Department of Molecular Medicine, Aarhus University Hospital, Denmark
[b] Department of Clinical Medicine, Aarhus University, Denmark
[c] C2i Genomics, Ltd., Haifa, Israel
[d] C2i Genomics Inc., New York, NY 10014, USA
[e] Department of Surgery, Herning Regional Hospital, Denmark
[f] Department of Surgery, Viborg Regional Hospital, Denmark
[g] Department of Surgery, Odense University Hospital, Denmark
[h] Department of Surgical Gastroenterology, Aalborg University Hospital, Denmark
[i] Department of Surgery, Aarhus University Hospital, Denmark
[j] Department of Surgery, Randers Regional Hospital, Denmark

ARTICLE INFO

Keywords:
Circulating tumor DNA
Cell-free DNA
Colorectal cancer
Whole-genome sequencing
CtDNA
CfDNA
Liquid biopsy
Surveillance
Recurrence detection
Biomarker

ABSTRACT

Introduction: Circulating tumor (ctDNA) can be used to detect residual disease after cancer treatment. Detecting low-level ctDNA is challenging, due to the limited number of recoverable ctDNA fragments at any target loci. In response, we applied tumor-informed whole-genome sequencing (WGS), leveraging thousands of mutations for ctDNA detection.

Methods: Performance was evaluated in serial plasma samples (n = 1283) from 144 stage III colorectal cancer patients. Tumor/normal WGS was used to establish a patient-specific mutational fingerprint, which was searched for in 20x WGS plasma profiles. For reproducibility, paired aliquots of 172 plasma samples were analyzed in two independent laboratories. *De novo* variant calling was performed for serial plasma samples with a ctDNA level > 10 % ($n = 17$) to explore genomic evolution.

Results: WGS-based ctDNA detection was prognostic of recurrence: post-operation (Hazard ratio [HR] 6.75, 95 % CI 3.18–14.3, $p < 0.001$), post-adjuvant chemotherapy (HR 28.9, 95 %CI 10.1–82.8; $p < 0.001$), and during surveillance (HR 22.8, 95 %CI 13.7–37.9, $p < 0.0001$). The 3-year cumulative incidence of ctDNA detection in recurrence patients was 95 %. ctDNA was detected a median of 8.7 months before radiological recurrence. The independently analyzed plasma aliquots showed excellent agreement (Cohens Kappa=0.9, r = 0.99). Genomic characterization of serial plasma revealed significant evolution in mutations and copy number alterations, and the timing of mutational processes, such as 5-fluorouracil-induced mutations.

Conclusion: Our study supports the use of WGS for sensitive ctDNA detection and demonstrates that post-treatment ctDNA detection is highly prognostic of recurrence. Furthermore, plasma WGS can identify genomic differences distinguishing the primary tumor and relapsing metastasis, and monitor treatment-induced genomic changes.

* Correspondence to: Department of Molecular Medicine, Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, Aarhus N DK-8200, Denmark.
E-mail address: cla@clin.au.dk (C.L. Andersen).

https://doi.org/10.1016/j.ejca.2024.114314
Received 22 March 2024; Received in revised form 22 August 2024; Accepted 2 September 2024
Available online 11 September 2024

1. Introduction

Circulating tumor DNA (ctDNA) has emerged as a promising biomarker for the detection of minimal residual disease (MRD) and recurrence surveillance following cancer treatment [1]. The perspective of ctDNA-based MRD detection and recurrence surveillance is to identify molecular recurrence as early as possible. Clinically, this has the potential to translate into a better selection of patients for adjuvant therapy and earlier recurrence intervention at a time when the efficacy of treatment intervention is expected to be high. Several studies have employed deep-targeted sequencing of circulating cell-free DNA (cfDNA) for MRD detection [2–7]. While these studies have yielded promising results, strategies relying on deep sequencing of a single or a few markers for ctDNA detection will be constrained by the number of ctDNA fragments recovered at the target loci. Consequently, the number of cfDNA genome equivalents (GEs) available will be a limiting factor, dictating the lowest ctDNA fraction detectable [8,9]. Increasing the number of targets can mitigate the limitation. Recently, proof-of-principle studies have demonstrated that employing whole-genome sequencing (WGS) for ctDNA detection has great potential [10,11]. This approach leverages the power of tumor and normal WGS to identify a comprehensive set of tumor-specific mutations and copy number alterations (CNAs), which are subsequently searched for within WGS cfDNA data. By aggregating the genome-wide signal from thousands of mutations and CNAs, rather than relying on the detection of a single or few markers, the number of potential ctDNA markers within a single copy of the genome is substantially increased. This means, that the required plasma volume for ctDNA analysis is only ~1 mL of plasma, whereas targeted strategies require up to 8 mL of plasma. Here, we assess the clinical performance of a WGS-based ctDNA strategy, in the context of Union for International Cancer Control (UICC) stage III colorectal cancer (CRC) at landmark time points including post-operation (post-OP), post adjuvant chemotherapy (post-ACT), and during surveillance after the end of treatment (EoT). Furthermore, we performed tumor-agnostic *de novo* genomic analysis of postoperative plasma samples with a high ctDNA level to explore the potential for characterizing the genomic evolution of the metastatic relapses.

2. Material and methods

2.1. Study population

This study included 146 patients with UICC stage III CRC treated with curative intent at six Danish hospitals between 2012 and 2019. Two patients were diagnosed with synchronous tumors, and all subsequent analyses included both tumors. All patients were treated and monitored according to National Guidelines, encompassing adjuvant chemotherapy (as standard combination of 5-fluorouracil and oxaliplatin) at the clinician's discretion and a recommended CT scan at 12 and 36 months after surgery. The Committees on Biomedical Research Ethics in the Central Region of Denmark approved the study (1–16-02–453-14 and 1–10-72–3-18). The study was performed in accordance with the Declaration of Helsinki, and all participants provided written informed consent. Plasma from 45 healthy individuals, aged 40 or older (53 % male; median age 62, interquartile range [IQR] 58–70), was anonymously collected through the blood bank at Aarhus University Hospital ($n = 15$) or available through the Colorectal Cancer Research Biobank at Aarhus University Hospital ($n = 30$).

2.2. Sample collection and preparation

2.2.1. Tumor and matched normal samples

Fresh frozen (FrFr) or formalin-fixed paraffin-embedded (FFPE) tumor tissue biopsies with a minimum tumor fraction of 20 % (histological assessment) were collected from all patients. Snap-frozen FrFr tissue was stored at −80 °C and FFPE at room temperature until DNA extraction. Peripheral mononuclear blood cells (PBMCs) were isolated from blood samples collected in K2-EDTA 10 mL tubes (Becton Dickinson) from all patients after centrifugation for 10 min at 4000 rpm. They were stored in TPP® cryotubes at −80 °C until DNA extraction. DNA was extracted from FFPE tissue using QiAamp DNA FFPE tumor tissue kit (Qiagen), from FrFr tumor tissue using Puregene DNA Purification Kit (Gentra Systems), and from PBMC (normal DNA) using the QIAamp DNA Blood Kit (Qiagen). Tumor and normal DNA were quantified using the Qubit™ dsDNA BR Assay Kit (ThermoFisher). Sequencing libraries were generated using xGen UDI-UMI Adapters (Integrated DNA Technologies Inc.) and the Twist Library Preparation Enzymatic Fragmentation Kit 1.0 (TWIST Bioscience). Libraries were prepared according to the manufacturer's protocol, with the following modifications: 1) 50 ng input and 10 min fragmentation for normal and FrFr DNA, 2) 200 ng input and 6 min fragmentation for FFPE DNA, and 3) seven PCR cycles of library amplification for all samples. Libraries were quantified using Qubit™ dsDNA BR Assay Kit (ThermoFisher) and library fragment size was estimated using TapeStation D1000 (Agilent).

2.2.2. Plasma samples

From patients, blood samples were collected pre-operatively (pre-OP), post-operatively (post-OP), and every 3 months for up to 3 years. Blood samples were collected in K2-EDTA 10 mL tubes (Becton Dickinson) from healthy controls and patients with CRC. Plasma was isolated within 2 h of blood collection by double centrifugation at 4000 rpm for 10 min and stored at −80 °C in TTP® cryotubes until DNA extraction. cfDNA was extracted using the QiaAMP Circulating DNA kit (Qiagen) and quantified using digital PCR, as previously described [12]. Before cfDNA extraction, all plasma samples were spiked with a fixed number of soybean CPP1 DNA fragments, which was used to estimate purification efficiency, as previously described [13]. To enable removal of samples with unintended lysis of blood cells during blood processing, all cfDNA samples were assessed for leukocyte contamination using a digital PCR assay targeting the VDJ-rearrangement of B-lymphocytes [13]. As another control for leukocyte DNA contamination, cfDNA ScreenTape analysis (Agilent) was performed on all cfDNAs to assess if the cfDNA fragment size distribution showed signs of contamination with high molecular weight DNA fragments. Sequencing libraries were prepared using cfDNA from 2 mL of plasma. cfDNA libraries were generated using xGen UDI-UMI Adapters (IDT) and KAPA HyperPrep kit (Roche). Post-ligation clean-up was performed with AMPURE beads in a 1.4x (beads/DNA) ratio to retain short fragments, while post-PCR clean-up was done using a 1.0x ratio. The libraries were amplified with seven cycles of PCR. Libraries were quantified using Qubit™ dsDNA BR Assay Kit (ThermoFisher) and library fragment size was estimated using TapeStation D1000 (Agilent). Libraries that did not show the usual bi-modal fragment size distribution of cfDNA were excluded before sequencing.

2.2.3. Whole genome sequencing

All samples underwent WGS on the NovaSeq platform (Illumina), employing paired-end sequencing with a read length of 150 base pairs (2 x 150 bp). Normal DNA and cfDNA samples were sequenced to a target coverage of 20x. Tumor DNA purified from FFPE tissue was sequenced to a target coverage of 60x. FrFr tumor samples with histologically estimated tumor purities above 30 % were sequenced to a target coverage of 30x, while samples with purities at or below 30 % were sequenced to 60x. Each flow cell was demultiplexed according to predetermined sample indices. Raw reads were translated into FASTQ files using the Illumina tool "bcl2fastq".

2.3. Data processing and variant calling

2.3.1. Preprocessing of tumor, normal, and plasma WGS data

The genetic concordance of FASTQ files from the same patient was confirmed using NGSCheckMate [14]. FASTQ files from all three sample

types (tumor, normal, and cfDNA) were trimmed with Skewer v0.2.2 [15] to remove adapter sequences. The trimmed FASTQ files were run through fastqc v0.11.9 [16] to determine median coverage, insert size, and duplication rate. The trimmed FASTQ files were aligned to the reference genome (GRCh38) with Burrows-Wheeler Aligner MEM v 0.7.17 [17]. The aligned BAM files were sorted and indexed using Samtools v1.14[18]. Each BAM file was marked for duplicate reads using GATKMarkDuplicatesSpark v.4.1.8.0 [19], resulting in a duplicate-marked BAM file that was passed for calculation and recalibration of the per-read base quality score using GATK BQSRPipelineSpark [20]. Each recalibrated BAM file was indexed and re-sorted by read name using Samtools v1.11 [18]. In situations where the same library was sequenced more than once (*e.g.* to reach the target coverages), GATK MarkDuplicatesSpark [19] was used to merge all BAM files, thereby generating the final coordinate-sorted BAM file for these samples. Alignment quality control metrics were computed on the final BAM file using Picard (QualityScoreDistribution, MeanQualityByCycle, CollectBaseDistributionByCycle, CollectAlignmentSummaryMetrics, CollectInsertSizeMetrics, CollectGcBiasMetrics, CollectOxoGMetrics) [21] and GATK (average coverage, percentage of mapped and duplicate reads) [22]. These metrics were used to identify potential problems in sequencing or preprocessing. For tumor and normal samples, a minimum median coverage of 20x and 10x, respectively, was required. For plasma samples, a minimum median coverage of 18x was required. A subset of samples did not reach the minimum coverage requirement in the first sequencing run and was resequenced using the same sequencing library. Samples not passing the listed coverage requirements were excluded (Fig. 1A).

2.3.2. Tumor/normal somatic mutation calling

Each tumor and matched normal BAM file were analyzed using GATK Mutect2 v4.2.4.1 [23] and Strelka2 v2.9.10 [24] to identify putative somatic single nucleotide variants (SNVs). These SNVs were filtered using GATK FilterMutectCalls [25] to retain PASS variants and remove variants corresponding to known single nucleotide polymorphism sites (SNPs) (retrieved from dbSNP v138 [26]) as annotated by GATK VariantAnnotator [27]. Only SNVs detected by both Mutect2 and Stelka2 were retained. Furthermore, the list was filtered for SNVs observed in plasma cfDNA from healthy individuals ($n = 45$). Somatic insertions and deletions (INDELs) were called by SVaba (v1.1.3) [28] and Mutect2 v4.2.4.1 [23] and the final list of INDELs comprised those detected by both methods. In cases where INDELs overlapped with a non-exact match, the largest INDEL was selected. The total mutational burden (TMB) was defined as the sum of SNVs and INDELs. CNAs in solid tumors were called using FACETS v0.6.2 [29] using base coverage and variant allele fraction (VAF) of heterozygous SNP positions of the tumor sample and paired normal sample as input. A list of heterozygous SNP positions in the normal samples was calculated using BCFtools mpileup [18] and a reference set of SNPs from the HapMap Project [30]. The SNP list was processed to revert all subclonal CNA calls to the nearest clonal CNA call.

2.3.3. Biological characterization of tumor samples

For the characterization of primary tumors, an in-house proprietary filtering method was applied to remove low-quality mutations with features characteristic of FFPE-induced artifacts. VEP v107.0 [31] was applied to the final list of SNVs and INDELs to determine the affected gene and functional impact. All non-synonymous variants were queried against the Intogen[32] and cancer biomarkers CGI [33] databases of cancer driver genes to identify the somatic mutations in genes of biological interest (Table S1).

Similarly, genes affected by CNAs were classified as biologically relevant if present in a list of candidate cancer genes as defined by previous studies [34,35]. Whole-genome doubling (WGD) status was determined by the stratification of each sample based on its ploidy and heterozygosity as previously described [34,35]. This resulted in two separate clusters of samples, which corresponded to near-diploid samples and WGD samples.

SigProfiler v1.1.20 [36] was used to extract SNV and INDEL signatures in a three-step workflow: (I) De-novo extraction of signatures. (II) Fitting of selected signatures from COSMIC v3.3[37,38] and artifact signatures used for the detection of spurious deviations. (III) A final fitting using cancertype exclusive signatures and manual inclusion of select signatures determined from steps (I) and (II), that correspond to the final set of signatures presented in the results (Fig. S2A).

2.3.4. Estimation of tumor fraction in solid tumor and cfDNA

Estimated tumor fraction (TF) of both solid tumor samples and plasma samples were estimated using an in-house proprietary bioinformatic data processing pipeline, based on a previously described method [10]. Tumor samples with an estimated TF of less than 10 % were excluded, as these samples contain an inadequate tumor signal to confidently call mutations. The approach used the complete set of somatic mutations, including SNVs, INDELs, and CNAs, found in the patient's tumor samples to generate a patient-specific tumor signature. Combined with an inhouse proprietary error-suppression model based on WGS of cfDNA from 45 healthy individuals, the patient-specific signature was then used to detect the tumor presence in cfDNA and estimate the ctDNA TF. The same threshold for calling a sample ctDNA-positive was applied to all samples in the study.

2.3.5. De novo detection of somatic alterations in high tumor fraction plasma samples

Plasma samples with a tumor fraction above 10 % were selected for de-novo calling of SNVs and CNAs. *De-novo* calling of mutations in plasma was performed using the same approach as described above in the section "*Tumor/normal somatic mutation calling*". The genomic changes identified in the plasma samples were compared to results from the patient's tumor sample to determine shared, tumor-unique, and plasma-unique mutations.

2.4. Microsatellite instability assessment

Microsatellite instability (MSI) was assessed by three different approaches. First, MSI status was assessed using COSMIC SBS mutational signatures v3 [38]. Samples were classified as MSI if they exhibited one or more of the following signatures: SBS6, SBS15, SBS21, SBS26, or SBS44. Conversely, samples lacking the presence of these signatures were classified as microsatellite stable (MSS). Secondly, we employed the bioinformatics tool MSIsensor [39] to get an MSIsensor score for each sample, as previously decribed[40]. The resulting MSIsensor score (ranging from 0 to 100) corresponds to the fraction of somatically mutated microsatellite loci. Thirdly, MSI status was assessed by immunohistochemistry (IHC) against the mismatch repair proteins MLH1, MSH2, MSH6, and PMS2, as part of the routine diagnostic workup. The IHC results were extracted from patient hospital records. Tumors were classified as MSS if they showed nuclear expression of all four proteins, and as MSI if any of the four proteins lacked expression.

2.5. Reproducibility test

We analyzed 15 complete sets of patient samples (tumor DNA, normal DNA, and plasma cfDNA from serially collected blood samples) at independent laboratories in Denmark and in the USA. Tumor and normal DNA was extracted and aliquoted in Denmark and distributed for sequencing both in Denmark and the USA. As part of the blood processing procedure, the plasma was aliquoted, and paired aliquots from the same samples were distributed for sequencing in Denmark and the USA. The entire sample preparation protocol and sequencing were carried out in parallel in both laboratories, and the sequencing data was analyzed with identical pipelines at the same location.

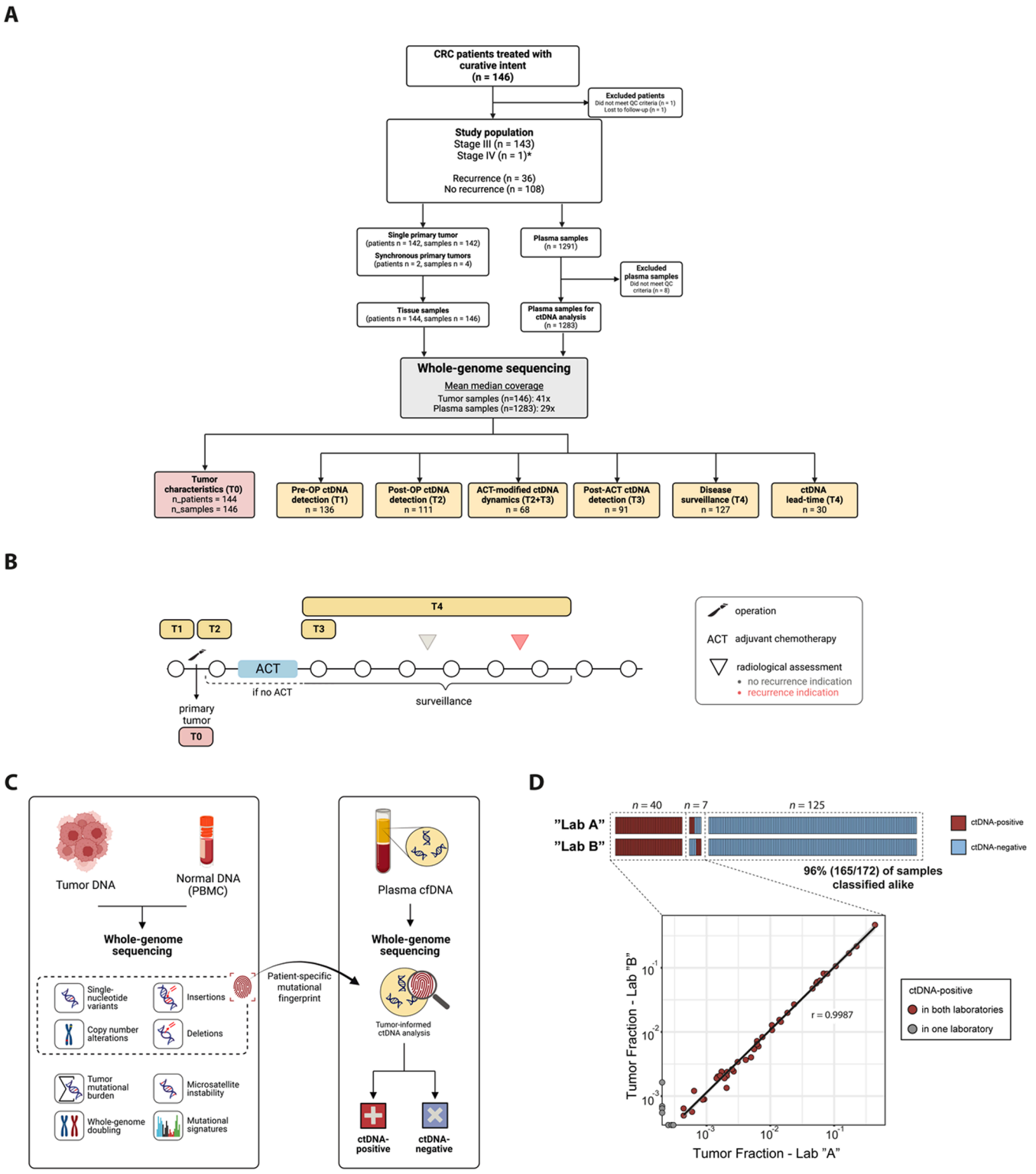

Fig. 1. Study design and cohort overview. A) Cohort overview with indication of the number of patients included in each subanalysis. n refers to the number of patients unless stated otherwise. Pre-OP: pre-operation, post-OP: post-operation, ACT: adjuvant chemotherapy, QC: quality criteria, CRC: colorectal cancer, ctDNA: circulating tumor DNA. After initial pathological evaluation, one patient (CRC-039, indicated by asterisk) was classified as a stage III patient and received treatment accordingly, however after clinical re-examination, the patient was re-classified as stage IV. B) Overview of the sample time point (T0-T4) included in each sub-analysis. Same subanalysis numbering as in (A). Samples in time points T1-T4 are defined as follows: T0) tumor from primary operation, T1) plasma collected before the operation, T2) first plasma collected after the operation (within 8 weeks) and before ACT, T3) plasma collected maximum 3 months after the end of ACT, T4) plasma collected after the end of treatment (surgery alone or surgery + ACT). **C)** WGS of tumor and normal DNA were used to identify tumor-specific changes. Single-nucleotide variants, copy number alterations, insertions, and deletions compiled a patient-specific mutational fingerprint and used for tumor-informed ctDNA analysis of WGS cfDNA. **D)** Top panel: ctDNA status of paired plasma samples processed in "Lab A" or "Lab B". Bottom panel: estimated cfDNA tumor fractions of plasma samples called ctDNA positive in both laboratories (n = 40) or just one laboratory (n = 7).
Figures A-C created with Biorender.com.

2.6. Statistics

The results presented in this study represent single measurements of distinct samples. Two-sided Wilcoxon rank-sum test was used to compare unmatched groups. Two-sided McNemar test and Cohen's Kappa were used to compare paired samples. Pearson's correlation was used to estimate the strength of a linear relationship between paired data. Logistic regression was used to relate clinicopathological risk factors to post-OP ctDNA detection. In recurrence-free survival (RFS) analysis, radiological recurrence (local or distant) or death was counted as an event. Patients were censored at the end of radiological follow-up. Survival analyses were conducted using the Cox proportional hazards regression analysis with Firth's penalized maximum likelihood bias reduction method. In serial ctDNA analysis, ctDNA status (positive/negative) was used as a time-dependent covariate. Survival was plotted using the Kaplan-Meier method. In the EoT analysis, we included serially collected plasma samples obtained after the EoT. For patients treated by operation alone, EoT was defined as after the operation, while for patients treated by both operation and ACT, EOT was defined as after the completion of adjuvant chemotherapy. The cumulative incidence of ctDNA detection was estimated in serial samples by censoring patients at the end of plasma collection. Statistical calculations were done using R (v.4.2.0) [41].

3. Results

3.1. Patient characteristics and WGS data generation

A total of 146 UICC stage III CRC patients (59 % male; median age 66 years, IQR 59–72 years) treated with curative intent were enrolled in this study. From each patient, tumor DNA and normal DNA, as well as cfDNA from serially collected blood samples were subjected to WGS. Two patients were subsequently excluded: one was lost to follow-up, and the other failed the required minimal WGS estimated tumor fraction (10 %). Out of 1291 plasma samples, eight samples were excluded because they failed to reach target coverage (Fig. 1A). This resulted in a final evaluable cohort of 144 patients, with a total of 1283 plasma samples available for analysis. The median number of plasma samples collected per patient was 10 (IQR: 6–13). The majority of patients received adjuvant chemotherapy (ACT) (88 %, 126/144). The recurrence rate was 25 % (36/144). The median follow-up for non-recurrence patients was 36 months (IQR: 35.6–36.5) (Fig. S1). The study overview, including patients, samples, and mean WGS genome coverage information, is presented in Fig. 1A-C.

3.2. ctDNA detection by tumor-informed WGS

For ctDNA detection in plasma, we applied a tumor-informed strategy that integrated the entire genome-wide compendium of somatic mutations and CNAs to generate a patient-specific mutational fingerprint. Each patient-specific mutational fingerprint was then applied to WGS data from that patient's cfDNA to determine the ctDNA status and ctDNA level. For initial validation of the robustness and reproducibility of the approach, we analyzed paired aliquots of the same plasma samples ($n_{samples} = 172$ plasma samples, $n_{patients} = 15$ patients) processed and sequenced in two independent laboratories in the USA and Denmark. We observed an excellent reproducibility, with 96 % (165/172) agreement between ctDNA classification (Cohen's kappa = 0.89, McNemar test p = 1) (Fig. 1D). Furthermore, a near-perfect correlation (Pearson's r = 0.9987, p < 2.2e-16) was observed between the estimated TFs of the ctDNA-positive samples (Fig. 1D).

3.3. Genomic tumor characteristics

WGS of the resected tumors allowed us to explore their genomic characteristics. This revealed 12 % (18/146) of the tumors to carry mutational signatures associated with MSI. Consistent with the suggested mismatch repair deficient phenotype, these 18 tumors had the highest TMBs (SNVs and INDELs), the highest MSIsensor scores, and frequently showed loss of mismatch repair protein expression by IHC (Fig. S2A-C). One tumor showed a MUTYH mutational signature (Fig. S2A), and in agreement, the hospital record revealed the patient to be a homozygous carrier of a pathogen MUTYH variant (c536A>G; p.TyrY179Cys). While the synchronous tumors were molecularly distinct, they nevertheless showed the same MSI status (one tumor pair was MSI, the other MSS) and showed similar mutational burdens (Fig. S2A). WGD was observed solely among the MSS tumors, of which 53 % (68/130) showed WGD (Fig. S2D). The most frequently mutated cancer driver genes were *APC*, *TP53*, *KRAS*, and *PIK3CA* for MSS tumors and *ARID1A*, *LRP1B*, *RNF43*, and *BRAF* for MSI tumors (Fig. S2E-F).

3.4. Pre-operative ctDNA detection

Pre-operative (pre-OP) plasma samples were available from 136 patients. Tumor-informed cfDNA WGS analysis detected ctDNA in 84 % (114/136) of patients (Fig. 2A). The detected tumor fractions (TFs) of ctDNA-positive samples ranged from 2×10^{-4} to 3×10^{-1} (median 1.7×10^{-3}, IQR 6×10^{-4} - 6.2×10^{-3}) (Fig. 2B). The ctDNA-positive patients had higher TMB (median 16900 mutations, IQR 11,578–22,447) than the ctDNA-negative patients (median 13,416, IQR 11,057–16,481) (p = 0.041) (Fig. 2C). Pre-OP ctDNA was detected in 93 % (13/14) and 83 % (101/122) of patients with MSI and MSS cancers, respectively (p = 0.47) (Fig. 2C).

3.5. Post-operative ctDNA detection and association to recurrence risk

At the post-OP landmark, defined as plasma samples collected within eight weeks after the operation but prior to ACT, plasma samples were available from 111 patients. Hereof, 17 % (19/111) were ctDNA-positive. Being post-OP ctDNA positive was associated with higher pN category and venous invasion in the primary tumor (Table S2). The rate of recurrence was 68 % (13/19) among the ctDNA-positive and 16 % (15/92) among the ctDNA-negative patients (Fig. 2D), which translates to a sensitivity of 46 % (13/28) and a specificity of 93 % (77/83) for the plasma WGS analysis at detecting the patients with incipient recurrence. The 3-year RFS was significantly lower for the ctDNA-positive (46 %, 95 % CI: 26 %−83 %) compared to the ctDNA-negative patients (90 %, 95 % CI: 83 %−99 %) (Hazard ratio [HR] 6.75, 95 % CI: 3.18–14.3; p < 0.001) (Fig. 2E). In multivariable analysis, post-OP ctDNA status was the only factor significantly associated with recurrence with a HR of 6.49 (95 % CI 2.75–15.3) (p < 0.001) (Table 1). In a previous study, we showed that surgical trauma leads to a temporary surge in wild-type cfDNA, which may challenge ctDNA detection in samples collected within 14 days after the operation [42]. To investigate the potential impact of this issue on plasma WGS analysis, we compared the false negative rate in samples obtained early (≤14 days after the operation) to those obtained later (≥15 days after the operation). The false negative rate was 60 % (9/15) in the early and 46 % (6/13) in the late samples (Fig. 2F; Fig. S3). Next, we compared the sensitivity and specificity of the plasma WGS approach in the early and late samples. In the early samples, sensitivity was 40 % (6/15) and specificity was 100 % (28/28), and in the late samples, sensitivity was 54 % (7/13) and specificity 89 % (49/55). Four of the six false positive late samples were from patients who received ACT.

3.6. ACT-modified ctDNA dynamics

Among the 127 patients who received ACT, 68 had both a post-OP and a post-ACT plasma sample available for ctDNA assessment. A total of 58 patients were post-OP ctDNA-negative, and 57 of these remained ctDNA-negative post-ACT (neg→neg), while one converted from negative to positive (neg→pos). Ten patients were post-OP ctDNA-positive.

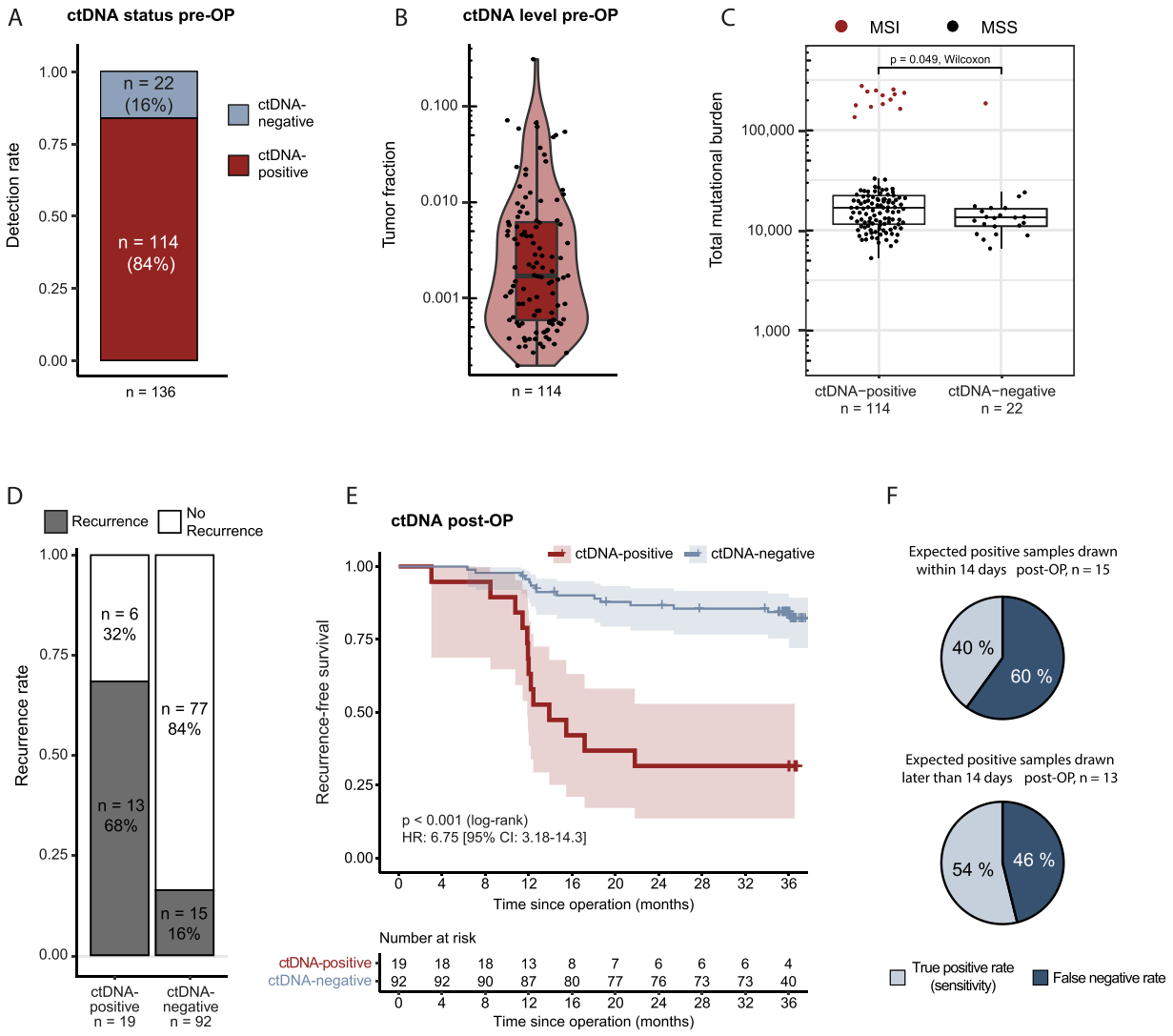

Fig. 2). Detection of ctDNA in pre- and post-operative plasma samples. A) ctDNA detection rate assessed in pre-operative plasma samples (n = 136). B) cfDNA tumor fractions of the ctDNA-positive pre-operative plasma samples (n = 114). Box limits represents the lower and upper quartile, center line indicates median TMB, whiskers represent the minimum and maximum values when excluding outliers. C) Total tumor mutational burden (TMB) of the primary tumors, stratified by pre-operative ctDNA status. The mutational burden is the sum of the SNVs and INDELs. For synchronous tumors, the highest TMB is plotted. Microsatellite status (MSI/MSS) is identified through the mutational signatures of the tumors. D) Recurrence rate in post-operative (post-OP) ctDNA-positive and ctDNA-negative patients. E) Kaplan-Meier plot showing recurrence-free survival stratified for ctDNA status assessed in plasma samples drawn within 8 weeks after surgery and before initiation of adjuvant chemotherapy. F) False negative rate of post-surgery samples drawn within the first 14 days after surgery (top) or later than 14 days after surgery (bottom). Samples are labeled as false/true negative according to recurrence status regardless of adjuvant chemotherapy.

Of these, six remained ctDNA-positive (pos→pos) and four converted from ctDNA-positive to ctDNA-negative (pos→neg) (Fig. 3A, left panel). The post-ACT ctDNA-positive patients ($n = 7$) all experienced disease recurrence, regardless of post-OP status. None of the patients who converted from ctDNA-positive to ctDNA-negative ($n = 4$) experienced recurrence, suggesting that ACT eliminated any MRD present after the operation (Fig. 3A, right panel). Of the 57 persistently ctDNA-negative patients, 9 % (5/57) experienced recurrence within the follow-up period. Hereof, ctDNA was detected in later plasma samples during surveillance for three patients (Fig. 3B).

3.7. Post-ACT ctDNA detection and association with recurrence risk

At the post-ACT landmark, defined as samples collected a maximum of 3 months after the end of ACT, plasma was available from 91 patients. ctDNA was detected in 11 % (10/91) of patients. Among these, the recurrence rate was 90 % (9/10), while it was 9 % (7/81) among the ctDNA-negative patients (Fig. 3C). This translates to a sensitivity of 56 % (9/16) and a specificity of 99 % (74/75) for detecting patients with incipient recurrence. Next, we assessed the association between post-ACT ctDNA status and RFS, which was measured from the end of ACT until recurrence or censoring at the end of follow-up. The post-ACT ctDNA-positive patients had significantly lower RFS than the ctDNA-negative patients (3-year RFS ctDNA-pos: 10 % [95 % CI:

1.5 %–64 %]; 3-year RFS ctDNA-neg: 92 % [95 % CI: 87 %–98 %]; HR 28.9, 95 % CI: 10.1–82.8; $p < 0.001$) (Fig. 3D). In multivariable analysis, post-ACT ctDNA status was the only factor significantly associated with recurrence with an HR of 22.61 (95 % CI 7.19–71.1; $p < 0.001$) (Table 1).

3.8. Serial recurrence monitoring using ctDNA

To examine the value of serial ctDNA monitoring, we assessed the cumulative incidence of ctDNA detection from the EoT until recurrence (Fig. 3E). Among recurrence patients ($n = 30$) the cumulative ctDNA detection was 95 % at the end of follow-up at 32 months after EoT. Most

Table 1
Patient characteristics, clinicopathological parameters, and ctDNA status stratified for recurrence.

		Patient characteristics		Univariable				Multivariable, Post-OP[b] N = 108[c]				Multivariable, Post-ACT[b] N = 88[c]			
Variable	**N**	Non-recurrence, N = 108[a]	Recurrence, N = 36[a]	**N**	**HR[b]**	**95% CI[b]**	**p-value**	**N**	**HR[b]**	**95% CI[b]**	**p-value**	**N**	**HR[b]**	**95% CI[b]**	**p-value**
Age	144														
<70		67 (62%)	21 (58%)	88	—	—									
>70		41 (38%)	15 (42%)	56	1.18	0.61, 2.29	0.622								
Sex	144														
F		42 (39%)	17 (47%)	59	—	—									
M		66 (61%)	19 (53%)	85	0.75	0.39, 1.44	0.377								
Tumor location	144														
Colon		96 (89%)	27 (75%)	123	—	—		93	—	—		73	—	—	
Rectum		12 (11%)	9 (25%)	21	2.41	1.14, 5.1	**0.030**	15	2.73	1.02, 7.34	0.059	15	1.58	0.49, 5.04	0.427
Resection	140														
R0		92 (88%)	25 (71%)	117	—	—		95	—	—		78	—	—	
R1-R2		13 (12%)	10 (29%)	23	2.37	1.14, 4.91	**0.029**	13	1.53	0.56, 4.19	0.418	10	1.29	0.25, 6.7	0.753
MMR status	143														
dMMR		15 (14%)	1 (3%)	16	—	—		12	—	—		8	—	—	
pMMR		92 (86%)	35 (97%)	127	3.24	0.62, 17	0.089	96	4.29	0.66, 27.7	0.068	80	1.47	0.06, 34.1	0.783
pT stage	144														
pT1-2		9 (8%)	3 (8%)	12	—	—		11	—	—		10	—	—	
pT3-4		99 (92%)	33 (92%)	132	0.86	0.28, 2.63	0.796	97	1.31	0.28, 6.02	0.711	78	0.66	0.15, 2.96	0.575
pN stage	144														
pN1		70 (65%)	18 (50%)	88	—	—		70	—	—		59	—	—	
pN2		38 (35%)	18 (50%)	56	1.70	0.88, 3.26	0.110	38	0.73	0.31, 1.74	0.463	29	1.45	0.48, 4.41	0.500
Histology	144														
adenocarcinoma		101 (94%)	31 (86%)	132	—	—									
mucinous/medullar		7 (6%)	5 (14%)	12	1.86	0.74, 4.66	0.212								
Differentiation	124														
Poorly differentiated		8 (9%)	6 (19%)	14	—	—									
Well/moderately differentiated		85 (91%)	25 (81%)	0.42	0.47	0.17, 1.0	0.070								
Venous invasion	143														
No venous invasion		67 (63%)	17 (47%)	84	—	—									
Venous invasion		40 (37%)	19 (53%)	59	1.77	0.92, 3.4	0.084								
ACT	144														
No ACT		10 (9%)	8 (22%)	18	—	—		16	—	—					
Received ACT		98 (91%)	28 (78%)	126	0.42	0.19, 0.92	**0.043**	92	0.43	0.17, 1.06	0.076				
Post-OP ctDNA	111														
Negative		77 (93%)	15 (54%)	92	—	—		90	—	—					
Positive		6 (7%)	13 (46%)	19	6.75	3.18, 14.3	**<0.001**	18	6.49	2.75, 15.3	**<0.001**				
Post-ACT ctDNA	91														
Negative		74 (99%)	7 (44%)	81	—	—						79	—	—	
Positive		1 (1%)	9 (56%)	10	28.9	10.1, 82.8	**<0.001**					9	22.61	7.19, 71.1	**<0.001**

[a] n (%)
[b] HR = Hazard ratio, CI = Confidence interval, OP = Operative, ACT = Adjuvant chemotherapy
[c] Patients with missing values for one or more of the variables in the multivariate models were excluded. In the Post-OP analysis n = 3 with missing resection status were excluded. In the Post-ACT analysis n = 3 were excluded: two with missing resection status and one with missing MMR status.

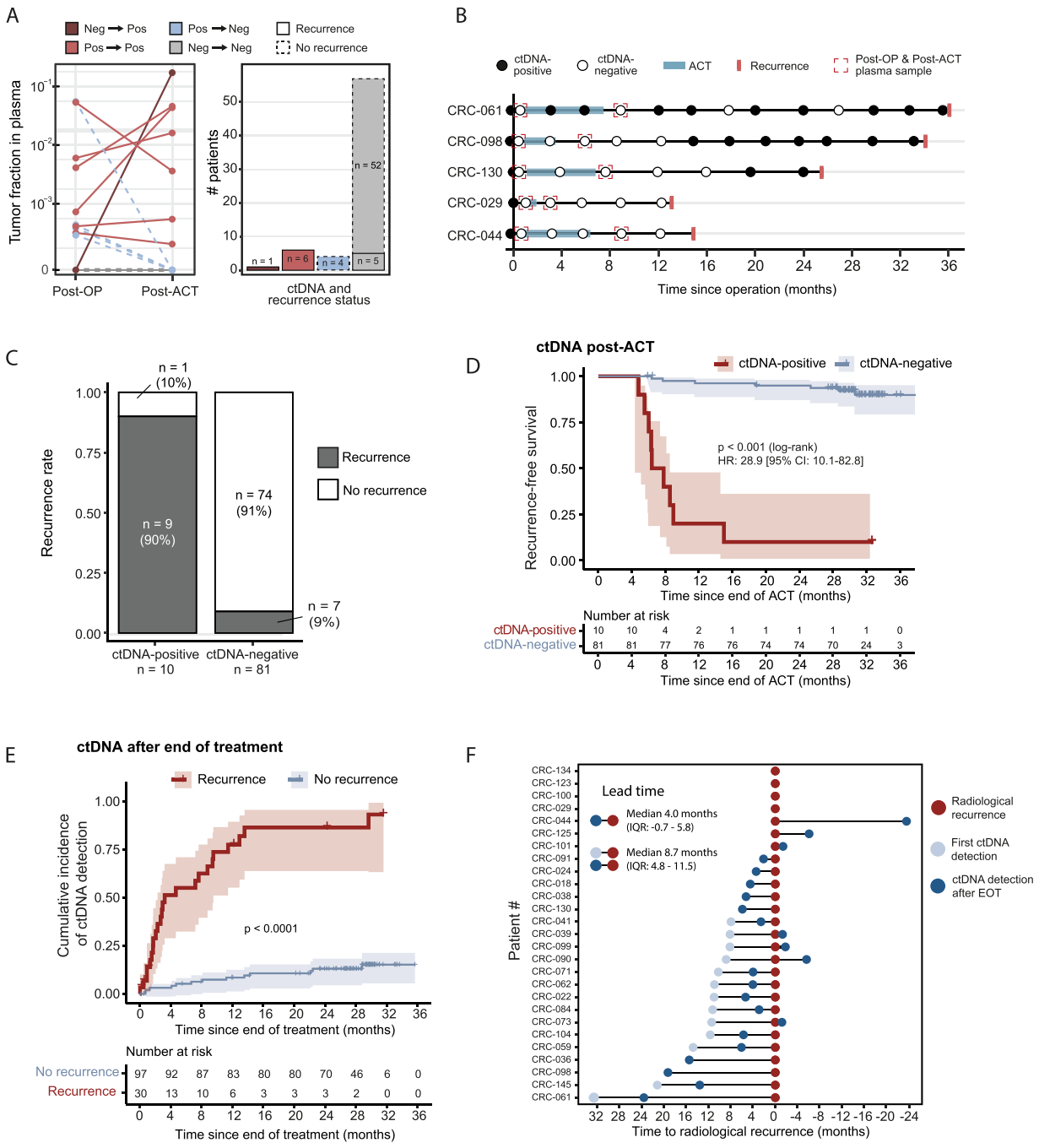

Fig. 3). Assessment of ctDNA post adjuvant chemotherapy and during surveillance. A) ctDNA status in plasma collected post-OP (before ACT) and post-ACT (after ACT) (T2 and T3 in Fig. 1B). Color indicates the change in ctDNA status. Dashed lines indicate no disease recurrence. B) Swimmer plot of the recurrence patients who were ctDNA negative both before and after ACT. The post-OP and post-ACT samples are marked by a red square. ctDNA was detected at a later time point for three out of the five patients. C) Recurrence rates of patients who were ctDNA positive and ctDNA negative after ACT (T3 in Fig. 1B). D) Kaplan-Meier plot showing recurrence-free survival stratified for ctDNA status of the post-ACT sample (sample time T3 in Fig. 1B). E) Cumulative incidence of ctDNA detection after end of definitive treatment (EoT) (T4 in Fig. 1B). EoT was defined as surgery alone or surgery + ACT. Each step represents ctDNA detection. Crosses mark censoring and indicate the last plasma sample. **F)** Lead time analysis, showing time to radiological recurrence (red) and time to first ctDNA detection (light blue: before EoT, dark blue: after EoT. ctDNA lead-time was calculated for the first ctDNA detection before EoT (light blue to red) and the first ctDNA detection after EoT (dark blue to red).

recurrence cases (77 %) became ctDNA-positive within 12 months after EoT. Comparatively, for non-recurrence patients ($n = 97$) the cumulative incidence of ctDNA detection was only 10 % at 12 months and 15 % at the end of follow-up at 36 months after EoT (Fig. 3E). A swimmer plot of the non-recurrence patients revealed that the few false positive samples were mainly sporadic (Fig. S4A). The finding was corroborated by a low overall false-positive rate of 3.5 % (24/694) among surveillance samples collected within the follow-up time from non-recurrence patients. Further analysis of the 24 false-positive samples revealed that their estimated TFs were significantly lower than true-positive samples, defined as surveillance samples from recurrence patients ($n = 44$) ($p = 0.0001$) and pre-OP samples ($n = 114$) ($p = 2.4e\text{-}7$) (Fig. S4B). Cox regression analysis using ctDNA as a time-dependent variable revealed an association between serial ctDNA detection and lower RFS (HR 22.8, 95 % CI 13.7–37.9).

3.9. Early detection of recurrence

In 74 % of recurrence patients (20/27), ctDNA was detected before diagnosis of recurrence by standard-of-care radiological imaging (Fig. 3F). A significant lead time was observed, from the first detection of ctDNA to the diagnosis of recurrence (median 8.7 months, IQR 4.8 – 11.5 months, p = 4.4e-4). When restricting the analysis to plasma samples collected after EoT, we observed a median lead time of 4.0 months (IQR: −0.7 - 5.8 months, p = 0.08) (Fig. 3F).

3.10. Tracing genomic evolution in recurrence patients through plasma WGS

In recurrence patients, genomic changes seen in postoperative plasma samples reflect the metastatic lesion and not the primary tumor (Fig. 4A). Using the plasma WGS data, we explored the genomic changes occurring *de novo* in the ctDNA of recurrence patients as a proxy of the metastatic lesion, *i.e.* SNVs, INDELs, and CNA not observed in the primary tumor (Fig. 4A). To enable robust mutation assessment, and hence a fair comparison to the tumor biopsy, the analysis was restricted to postoperative plasma samples with estimated TFs above 10 % ($n = 17$) as estimated from the tumor-informed ctDNA analysis. At least one such sample was available for 22 % (8/36) of the recurrence patients. All eight patients had at least one postoperative plasma sample, for which a large fraction of the called mutations (~20 % or more, up to 53 %) were not detected in the primary tumor (Fig. 4B). Similarly, the plasma samples were more affected by CNA when compared to the primary tumor (Fig. 4B, Fig. S5). This indicates substantial genetic differences between the primary tumor and metastatic lesions at recurrence, suggesting that considerable time has passed since the metastatic lesion was seeded from the primary tumor until recurrence detection. Driver gene analysis of the mutations unique to the postoperative plasma samples revealed non-silent driver gene mutations in *FAT4*, *JAK1*, and *TP53* in the plasma of 37.5 % (3/8) of patients. Mutational signature analysis revealed enrichment of mutations associated with 5-fluorouracil (5-FU) treatment activity among the plasma-unique mutations in the post-operative plasma of 63 % (5/8) of patients (Fig. 4C). In agreement, these five patients all received 5-FU containing adjuvant chemotherapy, while the 66 % (2/3) of patients without the 5-FU signature, did not (Fig. 4C). Next, we aimed to investigate the origin of the 5-FU signal, specifically assessing whether it derived from the metastatic lesion or the hematopoietic cells. For this, we performed *de novo* analysis of plasma samples collected at least one year after the end of ACT from the non-recurrence patients who received the longest duration of ACT (>168 days of ACT, $n_{\text{patients}} = 16$; $n_{\text{samples}} = 16$). None of the analyzed samples showed an association with a 5-FU signature (Fig. S6).

In two patients, we further explored the observed differences and used mutational profile, driver mutation, CNA, and mutational signature to reconstruct the mutational events that likely occurred from tumor initiation until postoperative plasma collection, including treatment-related mutational events.

For patient CRC-071, only 47 % of the *de novo* identified ctDNA SNVs were present in the primary tumor biopsy (Fig. 4D) while the two plasma samples (collected at day 264 and 372 after the operation) showed high similarity with ~91 % and 92 % of the observed mutations in the two samples being shared. Among all observed mutations in the patient (primary tumor and plasma samples) six mutations affected driver genes. Of those, only *KRAS* and *TERT*-promoter mutations were observed in both the tumor and the plasma samples (Fig. 4D). Two independent *TP53* mutations were observed. Notably, one in the primary tumor and the other in the plasma samples (primary tumor: chr17:7673803 G>A; plasma samples: chr17:7670716 splice variant) indicating convergent evolution after the metastasis seeding clone left the primary tumor. The two remaining driver mutations (*PIK3CA* and *PCBP1*) were exclusively present in the primary tumor (Fig. 4D). Assessment of mutational signatures showed common activity of SBS1 (aging), SBS5, and ROS signatures in both the tumor-unique and the plasma-unique mutations (Fig. 4D). The 372-day plasma sample, but not the 264-day sample, presented a signature associated with 5-FU treatment (Fig. 4D), which agrees with the patient receiving 5-FU treatment (ACT treated from day 28–106). Copy number analysis revealed the acquisition of WGD exclusively to the plasma samples (ploidy: tumor = 2.2, 264-day plasma = 3.4, 372-day plasma = 3.4). Further analysis of the interaction between WGD and the previously described driver mutations revealed that the WGD event occurred after the *TP53* mutation and prior to the 5-FU treatment (Fig. S7).

Patient CRC-091 had three postoperative plasma samples with TF above 10 %. On average, 68 % of the mutations detected in the three plasma samples were shared with the primary tumor (Fig. 4E). While the mutational profiles indicated substantial genetic differences between the primary tumor and plasma samples, this was not mimicked at the copy number level (Fig. 4B, Fig. S5). Mutational signature analysis revealed that the primary tumor and the relapsing metastasis had been exposed to the similar mutational processes. Further analysis revealed that the SBS93 signature was identified among the tumor-unique and plasma-unique mutations, but not in the mutations shared between the two (Fig. 4E).

4. Discussion

The analysis of ctDNA holds great promise as a minimally invasive tool for detecting MRD and identifying recurrence. In this study, we demonstrated the clinical potential of employing a WGS-based strategy for ctDNA analysis, specifically in the post-operative management of stage III CRC patients.

Consistent with other studies [3,5,7,43], we showed that WGS-based ctDNA detection at any time point was associated with a high risk of relapse. At the post-OP and post-ACT landmark time points, the plasma WGS ctDNA detection approach, here applied to a stage III CRC cohort, showed sensitivities of 46 % and 56 % and specificities of 93 % and 99 %, respectively. The performance is comparable to other studies using digital PCR and deep-targeted sequencing approaches [3,5,7,44]. Surgical trauma is often associated with a temporary surge in wild-type cfDNA. If post-OP samples are collected early in the window of the surge, then the extraordinarily high wild-type cfDNA levels may negatively impact ctDNA assessment [42]. Consistent with this, the sensitivity of our plasma WGS ctDNA detection approach was higher in the post-OP samples collected after day 14, compared to those collected before day 14. Consequently, it is recommended to collect post-OP samples after day 14. Additional steps to increase sensitivity could be to reduce background noise, thereby allowing detection of even lower-frequency variants, and increasing sequencing depth to increase the robustness of the results. Alternatively, repeat measurements could be a way to confirm initially ctDNA negative results. However, current trials investigating the utility of ctDNA-guided treatment are using methods with similar sensitivities as demonstrated in our study [45,46]. The results from these trials will more definitively answer the limitations of current ctDNA sensitivities.

Compared to the post-ACT landmark, we observed a lower specificity at the post-OP landmark. This is a general observation for studies of stage III CRC and is explained by most stage III patients receiving ACT after the collection of the post-OP sample [3,5,7]. The fraction of the post-OP ctDNA-positive patients who are cured by ACT will inevitably appear as false positives when the outcome is recurrence. Consistent with this notion, among patients where both a post-OP and post-ACT sample were available, we observed 40 % (4/10) of the post-OP ctDNA-positive patients converting to ctDNA-negative after ACT. None of them recurred, while the 60 % (6/10) who remained ctDNA positive all recurred. Thus, while the current ACT regimen is beneficial for some patients, it is insufficient for others. Several, ongoing clinical studies are exploring new therapeutic strategies in this setting (The PEGASUS trial [NCT04259944]; the GALAXY study [UMIN000039205])[6,47] and are

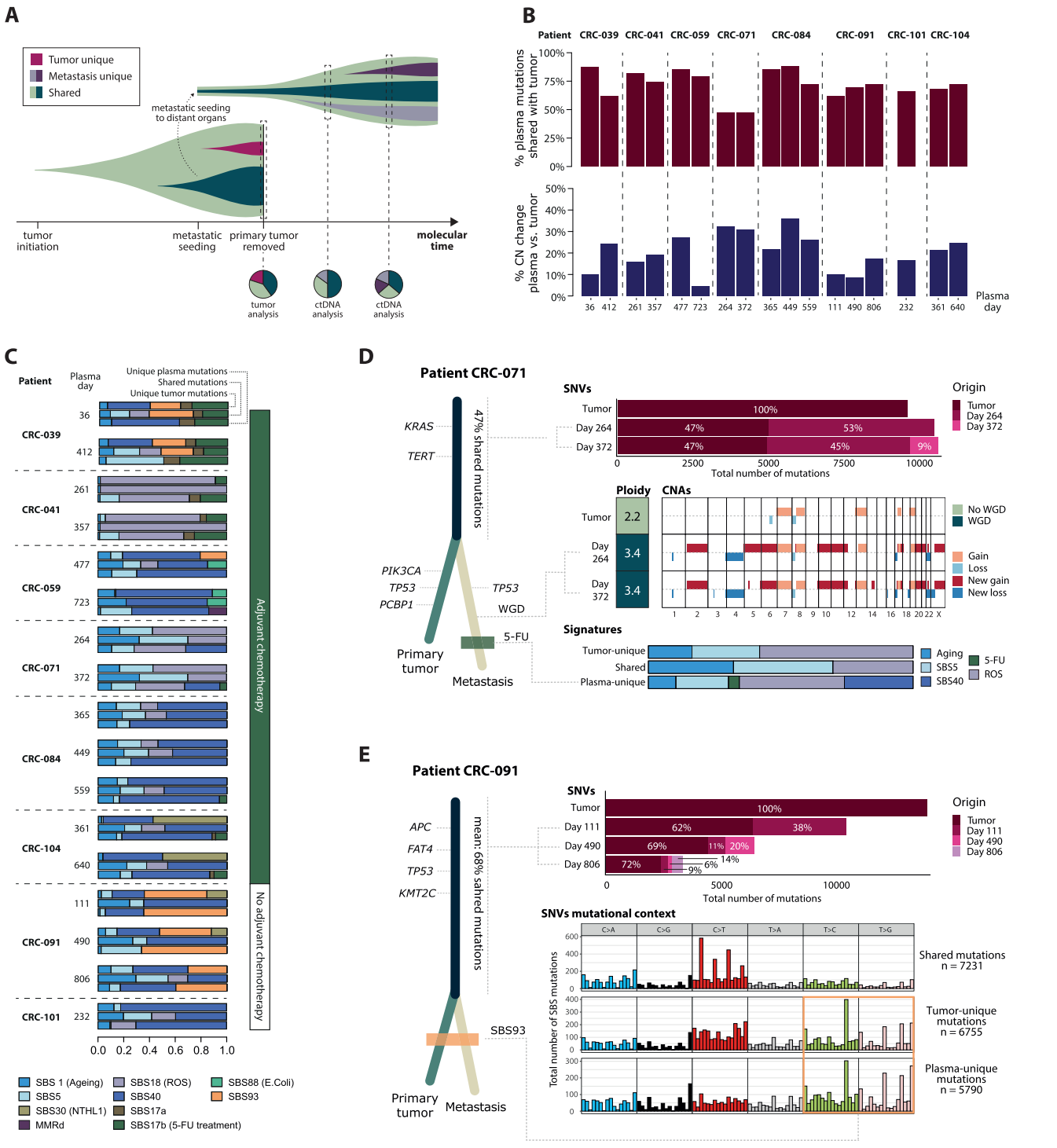

Fig. 4). De novo analysis of postoperative plasma samples with a high tumor fraction. A) Schematic illustration representing mutational evolution from tumor initiation through metastatic seeding in a recurrence patient. Mutations acquired prior to metastatic seeding are shared between the primary tumor and metastatic lesion. Mutations occurring in the primary tumor after metastatic seeding are tumor-unique, while any mutations acquired in the metastatic lesions after seeding are unique to the metastasis. Analysis of the tumor tissue reflects the mutational composition of the tumor at the time of removal. Analysis of postoperative plasma samples from recurrence patients reflects the composition of the metastatic lesion. B) Upper panel: Percentage of de novo identified mutations in plasma samples with TF above 10 % that are shared with mutations observed in the primary tumor. Lower panel: Percentage of the genome that has changed its copy number (CN) composition in the plasma sample compared with the primary tumor. C) Relative contribution of mutational signatures in primary tumor and plasma samples. Each patient is represented by three bars, which are the signature activity corresponding to (upper) mutations present only in the primary tumor, (middle) mutations present both in the primary tumor and ctDNA, and (lower) mutations present exclusively in the ctDNA. D) Explorative analysis of the mutational differences between high TF plasma samples and the primary tumor from patient CRC-071. The left panel represents a schematic reconstruction of the timeline from tumor initiation through metastatic seeding based on the comparison of plasma samples and primary tumor. E) Explorative analysis of the mutational differences between high TF plasma samples and the primary tumor from patient CRC-091. Left panel represents a schematic reconstruction of the timeline from tumor initiation through metastatic seeding based on the comparison of plasma samples and primary tumor.

using post-ACT ctDNA analysis as an effective approach to identify the high-risk patient population from which patients can be recruited. Preliminary results from the PEGASUS trial indeed indicate that additional treatment may salvage patients initially testing ctDNA-positive after three months of ACT. Here, 11/23 (48 %) of post-ACT ctDNA-positive patients reverted to ctDNA negative and avoided relapse after additional treatment with FOLFIRI [48].

For ctDNA-positive patients, curative intended surgical or local intervention is a potential alternative to systemic therapy. However, this requires that early ctDNA detection can create a window of opportunity for radiological imaging to pinpoint the location of the residual disease when it is still eligible for curative intended intervention. Here, we showed that serial plasma WGS-based ctDNA analysis detected ctDNA with a substantial lead time of several months compared to standard-of-care radiological recurrence detection. This agrees with other studies using other ctDNA detection approaches [3,5,43]. Whether the lead time of ctDNA-guided surveillance is sufficient to change treatment and improve the patient outcome is currently being explored, *e.g.*, in the randomized trial IMPROVE-IT2 (NCT04084249) [49], which compares ctDNA-guided and standard-of-care surveillance in ACT-treated CRC stage III and high-risk stage II patients. For ctDNA-guided surveillance, it might still be warranted to conduct a 36-month CT scan in persistently ctDNA negative patients to catch any missed recurrences. However, our cumulative ctDNA-detection rate was 95 % in recurrence patients, indicating that consistently ctDNA negative patients will be at low risk of recurrence, likely making the cost of radiological examination too great for the potential benefits.

WGS-based ctDNA analysis is an attractive option for clinical use with a low requirement for input (~1 mL of plasma) compared to targeted approaches often requiring 4–10 times as much input material [46,50,51]. The simple workflow, which does not require custom assays yet still allows patient customization, would greatly aid in clinical implementation, as resources can be spent on data acquisition rather than individualized sample handling. This simplified workflow may also allow for shorter turnaround times, which are crucial in the ACT-guiding setting, considering sample timing should be late enough to facilitate higher sensitivity. Additionally, the rapid decline in sequencing cost facilitates the use of WGS-based approaches, even in large studies [52]. While this approach requires an up-front investment in powerful sequencing machinery, the wide-spread adoption of sequencing equipment at hospital departments, for general genetic analyses, ushers a future where WGS of plasma for ctDNA detection could be a feasible strategy. The cost-effectiveness of using WGS for ctDNA detection in a clinical workflow needs further examination. However, compared to other approaches for ctDNA detection, the simplicity of the streamlined WGS workflow and decreasing sequencing costs makes WGS a more financially reasonable option every day.

To facilitate the clinical implementation of ctDNA analysis, inter-lab concordance is essential to ensure uniform results. Here, we report a remarkable reproducibility of samples processed in two independent sites with a near-perfect correlation of estimated TFs. The discrepancy between ctDNA status was observed in samples with low TFs. Thus, samples close to the detection threshold should potentially be called with caution and could be re-analyzed to confirm the initial observation. In a previous study, we observed an exponential increase in ctDNA levels leading up to recurrence [3]. Thus, if there are minute amounts of ctDNA present, analysis of another plasma sample collected a few weeks later could push the ctDNA level across the detection threshold and confirm ctDNA presence.

Analysis of pre-OP samples may offer some insight into the baseline performance of a ctDNA detection methodology. Using tumor-informed WGS, we detected ctDNA in 84 % of samples before surgery, which is within the 80–90 % range reported using other methods [4,5,53,54]. Notably, we observed the patients who were pre-OP ctDNA-positive to have a higher tumor mutational burden than those who were ctDNA-negative, likely reflecting that these tumors offered more targets for ctDNA detection using WGS. Possibly, higher sequencing coverage could compensate for the smaller amount of targets in tumors with lighter mutational burden, in turn increasing sensitivity across the board.

Our analysis of high TF plasma samples illustrates how plasma WGS approaches provide unique opportunities to interrogate the plasma for genomic changes occurring *de novo* in the plasma to explore the genetic differences between the primary tumor and metastatic lesions, both in terms of mutational processes acting solely in the primary tumor or metastases or acting both in the primary tumor and metastases. Our data suggest that while the tumor and metastatic lesion develop independently, they may be affected by similar mutational processes. This was observed for both patient CRC-071 and CRC-091, where the same mutational signatures (CRC-071: SBS1, SBS5, and ROS; CRC-091: SBS93) were observed among the tumor-unique and plasma-unique mutations. Furthermore, the *de novo* analysis revealed 5-FU related mutations in the plasma of recurrence patients who received ACT. Notably, in patient CRC-071, the 5-FU signature was solely observed in the late plasma samples, collected at day 372, and not in the sample collected at day 264. This observation indicates that clonal expansion of selected cells with 5-FU-induced mutations occurred between day 264 and 372. Furthermore, absence of 5-FU related signature in non-recurrence patients who received ACT suggests that the signal arises from the metastatic lesion and not the hematopoietic cells.

While the reported results are promising, our study also has some limitations. The modest number of patients in each sub-analysis may limit the generalizability of our findings. Especially the absolute number of ctDNA-positive post-OP (n = 19) and post-ACT (n = 10) patients is limited, making estimates of sensitivity vulnerable. Further studies of larger cohorts are necessary to validate and expand upon our results. However, the consistency of the results compared to similar studies documents the robustness. This was an observational study, preventing assessment of the clinical impact of ctDNA-guided postoperative management of these patients. Further, the Danish radiological recurrence-surveillance program does not include investigations more than 36 months after surgery. Therefore, we have only included 36 months of follow up, which does not enable assessment of the ctDNA sensitivity for late recurrences. Additionally, specificity measures may be skewed if a "false-positive" ctDNA surveillance call was truly an indication of a recurrence outside of our follow up. Lastly, while the presented WGS approach in principle fits all cancer types, our study focused specifically on CRC patients. Therefore, to document the generalizability of the approach, studies of other cancer types are needed.

In conclusion, our study highlights the significant clinical potential of WGS-based ctDNA analysis in detecting MRD and monitoring recurrence in stage III CRC patients. Facilitated by the ease of performing WGS, rather than designing bespoke ctDNA assays for each patient, WGS-based ctDNA analysis is an attractive approach for ctDNA analysis. Furthermore, we illustrate how WGS-based ctDNA analysis provides a unique opportunity to explore genomic changes occurring *de novo* in the plasma, *i.e.,* not present in the resected tumor. Lastly, the robust inter-laboratory reproducibility is very promising as it supports the feasibility for clinical implementation.

Code availability

The fully documented computer code, software versions, and the R statistical computing environment for the analyses related to this article are deposited on GitHub (https://github.com/lindbjerg-group/CRC_ctDNA_WGS_Analysis).

CRediT authorship contribution statement

Danielle Afterman: Writing – review & editing, Software, Project administration, Methodology, Investigation, Formal analysis. **Lene Hjerrild Iversen:** Writing – review & editing, Data curation. **Thomas**

Reinert: Writing – review & editing, Project administration, Data curation, Conceptualization. **Ole Thorlacius-Ussing:** Writing – review & editing, Data curation. **Maja Kuzman:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Asaf Zviran:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Tomer Lauterman:** Writing – review & editing, Investigation, Data curation. **Kåre Andersson Gotschalck:** Writing – review & editing, Data curation. **Dunja Glavas:** Writing – review & editing, Investigation, Formal analysis. **Santiago Gonzales:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Boris Oklander:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **James Smadback:** Writing – review & editing, Project administration, Investigation. **Claus Lindbjerg Andersen:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Formal analysis, Data curation, Conceptualization, Funding acquisition. **Jurica Levativ:** Writing – review & editing, Investigation. **Dillon Maloney:** Writing – review & editing, Investigation. **Ryan Ptashkin:** Writing – review & editing, Data curation. **Michael Yahalom:** Writing – review & editing, Data curation. **Zohar Donenhirsh:** Writing – review & editing, Data curation. **Iman Tavassoly:** Writing – review & editing, Data curation. **Ravi Kandasamy:** Writing – review & editing, Project administration, Data curation. **Eric White:** Writing – review & editing, Data curation. **Ury Alon:** Writing – review & editing, Project administration, Data curation. **Amanda Frydendahl:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sia Viborg Lindskrog:** Writing – review & editing, Data curation. **Iver Nordentoft:** Writing – review & editing, Data curation. **Claudia Jaensch:** Writing – review & editing, Data curation. **Mads Heilskov Rasmussen:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Lars Dyrskjøt:** Writing – review & editing, Supervision, Investigation. **Jesper Nors:** Writing – review & editing, Project administration, Investigation, Formal analysis, Data curation. **Per Vadgaard Andersen:** Writing – review & editing, Data curation. **Marijana Nesic:** Writing – review & editing, Investigation, Data curation. **Uffe Schou Løve:** Writing – review & editing, Data curation. **Tenna Vesterman Henriksen:** Writing – review & editing, Resources, Investigation, Data curation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. CLA reports collaborations with C2i Genomics and Natera. AZ, BO, RV, TL, and DM report stock options at C2i Genomics. AZ is the co-founder and a member of the board of directors of C2i Genomics. BO is the co-founder and CTO of C2i Genomics. SG, MK, JL, DG, RP, JS, DA, TL, YC, ZD, IT, and UA are employees of C2i Genomics. The opinions, results, and conclusions reported in this article are those of the authors and are independent of any competing interests. LD has sponsored research agreements with C2i Genomics, Natera, AstraZeneca, Photocure, and Ferring and has an advisory/consulting role at Ferring, MSD and UroGen. LD has received speaker honoraria from AstraZeneca, Pfizer and Roche and received travel support from MSD. LD is a board member at BioXpedia.

Data availability

To protect the privacy and confidentiality of patients in this study, personal data including clinical and sequence data are not made publicly available in a repository or the supplementary material of the article. The data can be requested at any time from the corresponding author. Any requests will be reviewed within a time frame of 2 to 3 weeks by the data assessment committee to verify whether the request is subject to any intellectual property or confidentiality obligations. All data shared will be de-identified. Access to clinical data and processed sequencing data output files (Mutect2 v4.2.4.1, Strelka2 v2.9.10, and FACETS v0.6.2) used in the article requires that the data requestor (legal entity) enter into Collaboration and Data Processing Agreements, with the Central Denmark Region (the legal entity controlling and responsible for the data). Request for access to raw sequencing data furthermore requires that the purpose of the data re-analysis is approved by The Danish National Committee on Health Research Ethics. Upon a reasonable request, the authors, on behalf of the Central Denmark Region, will enter into a collaboration with the data requestor to apply for approval. Additional info can be found at https://genome.au.dk/library/GDK000005/.

Acknowledgments

We extend our thanks to the patients and their families. We acknowledge the Danish Cancer Biobank and Colorectal Cancer Research Biobank at Aarhus University Hospital for providing access to blood and tissue materials. This study was supported by the NEYE Foundation (AF), the Danish Cancer Society (AF), the Novo Nordisk Foundation [grant numbers NNF17OC0025052 and NNF22OC0074415 (CLA)] and the Danish Cancer Society [grant numbers R133-A8520–00S41 (CLA), R146-A9466–16-S2 (CLA), R231-A13845 (CLA), and R257-A14700 (CLA)]. C2i Genomics covered the cost of sequencing. The opinions, results, and conclusions reported in this article are those of the authors and are independent of funding.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejca.2024.114314.

References

[1] Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017;17:223–38.
[2] Gale D, Heider K, Ruiz-Valdepenas A, Hackinger S, Perry M, Marsico G, et al. Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer. Ann Oncol 2022;33:500–10.
[3] Henriksen TV, Tarazona N, Frydendahl A, Reinert T, Gimeno-Valiente F, Carbonell-Asins JA, et al. Circulating tumor DNA in Stage III colorectal cancer, beyond minimal residual disease detection, toward assessment of adjuvant therapy efficacy and clinical behavior of recurrences. Clin Cancer Res 2022;28:507–17.
[4] Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017;9.
[5] Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, et al. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. JAMA Oncol 2019;5:1124–31.
[6] Taniguchi H, Nakamura Y, Kotani D, Yukami H, Mishima S, Sawada K, et al. CIRCULATE-Japan: Circulating tumor DNA-guided adaptive platform trials to refine adjuvant therapy for colorectal cancer. Cancer Sci 2021;112:2915–20.
[7] Tie J, Cohen JD, Wang Y, Christie M, Simons K, Lee M, et al. Circulating Tumor DNA Analyses as Markers of Recurrence Risk and Benefit of Adjuvant Therapy for Stage III Colon Cancer. JAMA Oncol 2019;5:1710–7.
[8] Kurtz DM, Soo J, Co Ting Keh L, Alig S, Chabon JJ, Sworder BJ, et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. Nat Biotechnol 2021;39:1537–47.
[9] Haque IS, Elemento O. Challenges in using ctDNA to achieve early detection of cancer. bioRxiv 2017:237578.
[10] Zviran A, Schulman RC, Shah M, Hill STK, Deochand S, Khamnei CC, et al. Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. Nat Med 2020;26:1114–24.
[11] Widman AJ, Shah M, A. Frydendahl, Halmos D, Khamnei CC, Øgaard N, et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. Nat Med 2024;30(6):1655–66. PMID: 38877116.
[12] Reinert T, Schøler LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. Gut 2016;65:625–34.
[13] Pallisgaard N, Spindler KL, Andersen RF, Brandslund I, Jakobsen A. Controls to validate plasma samples for cell free DNA quantification. Clin Chim Acta 2015;446:141–6.

[14] Lee S, Lee S, Ouellette S, Park WY, Lee EA, Park PJ. NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. Nucleic Acids Res 2017;45:e103.
[15] Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinforma 2014;15:182.
[16] Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010.
[17] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.
[18] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience 2021;10.
[19] Broad Institute GT. MarkDuplicatesSpark. 2021.
[20] Broad Institute GT. BQSRPipelineSpark (BETA). 2020.
[21] Broad Institute GT. Picard. 2022.
[22] Broad Institute GT. Genome Analysis Toolkit (GATK). 2023.
[23] Broad Institute GT. Mutect2. 2019.
[24] Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods 2018;15:591–4.
[25] Broad Institute GT. FilterMutectCalls. 2021.
[26] Sirotkin Smigielski EM, Ward K, Sherry M. ST. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 2000;28:352–5.
[27] Broad Institute GT. VariantAnnotator (BETA). 2019.
[28] Wala JA, Bandopadhayay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 2018;28:581–91.
[29] Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 2016;44:e131.
[30] Consortium TIH. The International HapMap Project. Nature 2003;426:789–96.
[31] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122.
[32] Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer 2020;20:555–72.
[33] Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med 2018;10:25.
[34] Wagner AH, Walsh B, Mayfield G, Tamborero D, Sonkin D, Krysiak K, et al. A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. Nat Genet 2020;52:448–57.
[35] Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 2019;575:210–6.
[36] AlexandrovLab. SigProfilerExtractor. 2021.
[37] Sondka, Dhir Z, Carvalho-Silva NB, Jupe D, Madhumita S, McLaren K, et al. COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Res 2024;52:D1210–7.
[38] Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.
[39] Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics 2014;30:1015–6.
[40] Johansen AFB, Kassentoft CG, Knudsen M, Laursen MB, Madsen AH, Iversen LH, et al. Validation of computational determination of microsatellite status using whole exome sequencing data from colorectal cancer patients. BMC Cancer 2019;19:971.
[41] Team R.C.R.: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022.
[42] Henriksen TV, Reinert T, Christensen E, Sethi H, Birkenkamp-Demtröder K, Gögenur M, et al. The effect of surgical trauma on circulating free DNA levels in cancer patients-implications for studies of circulating tumor DNA. Mol Oncol 2020;14:1670–9.
[43] Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med 2016;8. 346ra92.
[44] Tarazona N, Gimeno-Valiente F, Gambardella V, Zuñiga S, Rentero-Garrido P, Huerta M, et al. Targeted next-generation sequencing of circulating-tumor DNA for tracking minimal residual disease in localized colon cancer. Ann Oncol 2019;30:1804–12.
[45] Tie J, Cohen JD, Lahouel K, Lo SN, Wang Y, Kosmider S, et al. Circulating Tumor DNA Analysis Guiding Adjuvant Therapy in Stage II Colon Cancer. N Engl J Med 2022;386:2261–72.
[46] Kotani D, Oki E, Nakamura Y, Yukami H, Mishima S, Bando H, et al. Molecular residual disease and efficacy of adjuvant chemotherapy in patients with colorectal cancer. Nat Med 2023;29:127–34.
[47] Lonardi SM, C, Pietrantonio F, Elez E, Sartore-Bianchi A, Tarazona N, Sciallero S, et al. The PEGASUS trial: Post-surgical liquid biopsy-guided treatment of stage III and high-risk stage II colon cancer patients. J Clin Oncol 2020;38. TPS4124-TPS.
[48] Lonardi S, Pietrantonio F, Llavero TN, Viladot CM, Bianchi AS, Zampino MG, et al. The PEGASUS trial: Post-surgical liquid biopsy-guided treatment of stage III and high-risk stage II colon cancer patients. ESMO Congress 2023. Ann Oncol 2023:S1254–335.
[49] Nors J, Henriksen TV, Gotschalck KA, Juul T, Søgaard J, Iversen LH, et al. IMPROVE-IT2: implementing noninvasive circulating tumor DNA analysis to optimize the operative and postoperative treatment for patients with colorectal cancer - intervention trial 2. Study protocol. Acta Oncol 2020;59:336–41.
[50] Frydendahl A, Rasmussen MH, Jensen S, Henriksen TV, Demuth C, Diekema M, et al. Error-Corrected Deep Targeted Sequencing of Circulating Cell-Free DNA from Colorectal Cancer Patients for Sensitive Detection of Circulating Tumor DNA. Int J Mol Sci 2024;25.
[51] Henriksen TV, Demuth C, Frydendahl A, Nors J, Nesic M, Rasmussen MH, et al. Unraveling the potential clinical utility of circulating tumor DNA detection in colorectal cancer-evaluation in a nationwide Danish cohort. Ann Oncol 2024;35:229–39.
[52] Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). 2023.
[53] Jensen SO, Ogaard N, Orntoft MW, Rasmussen MH, Bramsen JB, Kristensen H, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer-a clinical biomarker discovery and validation study. Clin Epigenetics 2019;11:158.
[54] Andersen L, Kisistok J, Henriksen TV, Bramsen JB, Reinert T, Ogaard N, et al. Exploring the biology of ctDNA release in colorectal cancer. Eur J Cancer 2024;207:114186.