Systematic or Meta-analysis Studies

Whole exome sequencing of cell-free DNA – A systematic review and Bayesian individual patient data meta-analysis

Manouk K. Bos\textsuperscript{a,⁎,1}, Lindsay Angus\textsuperscript{a,1}, Kazem Nasseringjad\textsuperscript{b}, Agnes Jager\textsuperscript{a}, Maurice P.H.M. Jansen\textsuperscript{a}, John W.M. Martens\textsuperscript{a}, Stefan Sleijfer\textsuperscript{a}

\textsuperscript{a}Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands
\textsuperscript{b}HOVON Data Center, Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Cell-free DNA
Whole exome sequencing
Bayesian meta-analysis
Liquid biopsy
Cancer

ABSTRACT

Molecular profiling of tumor derived cell free DNA (cfDNA) is gaining ground as a prognostic and predictive biomarker. However to what extent cfDNA reflects the full metastatic landscape as currently determined by tumor tissue analysis remains controversial. Though technically challenging, whole exome sequencing (WES) of cfDNA enables thorough evaluation of somatic alterations. Here, we review the feasibility of WES of cfDNA and determine the sensitivity of WES-detected single nucleotide variants (SNVs) in cfDNA on individual patient data level using paired tumor tissue as reference (\(\frac{\text{shared SNVs}}{\text{All tissue SNVs}} \times 100\%\)). The pooled sensitivity was 50% (95% credible interval (CI): 29–72%). The tissue mutant allele frequency (MAF) of variants exclusively identified in tissue was significantly lower (12.5%, range: 0.5–18%) than the tissue MAF of variants identified in both tissue and cfDNA (23.9%, range: 17–38%), \(p = 0.004\). The overall agreement (\(\frac{\text{shared SNVs}}{\text{All SNVs}} \times 100\%\)) between SNVs in cfDNA and tumor tissue was 31% (95% CI: 15–49%). The number of detected SNVs was positively correlated with circulating tumor DNA (ctDNA) fraction (\(p = 0.016\)). A sub analysis of samples with ctDNA fractions ≥ 25% improved the sensitivity to 69% (95% CI: 46–89%) and agreement to 46% (95% CI: 36–59%), suggesting that WES is mainly feasible for patients with high ctDNA fractions. Pre- and post-analytical procedures were highly variable between studies rendering comparisons problematic. In conclusion, various aspects of WES of cfDNA are largely in its investigative phase, standardization of methodologies is highly needed to bring this promising technique to its clinical potential.

Introduction

Next generation sequencing of tumor tissue is increasingly being performed since more and more targeted treatments require presence of specific genomic alterations [1–3]. Although metastatic tissue can be obtained for this analysis, it is a cumbersome procedure for patients and repetitive sampling is frequently not feasible. Therefore, genomic profiling of plasma derived cell-free DNA (cfDNA) is considered as a minimally-invasive surrogate to predict outcome and predict or monitor treatment efficacy [4].

cfDNA consists of short fragments of DNA derived from normal- and tumor cells (ctDNA). Contrary to a single tumor tissue biopsy, ctDNA might give a more accurate representation of the entire mutational profile present across the different lesions within an individual cancer patient [5–7]. Although significant progress has been made for tracking previously detected tumor mutations using targeted gene panels or single gene assays [8], whole exome sequencing (WES) enables a more comprehensive analysis covering the complex landscape of somatic alterations. Hence, can be used as a tool to gain insight into tumor biology, for example by which genomic mechanisms tumor cells can confer resistance.

In addition, WES enables the identification of genomic signatures such as tumor mutational burden (TMB) and microsatellite instability (MSI), all being recognized as biomarkers for selected therapies such as immunotherapy [9,10]. So, compared to targeted panels comprising a relatively limited number of genes, WES analyses of ctDNA holds great promise to identify emerging genes that are of interest in treatment resistance and to capture DNA signatures important for treatment decision making. However, WES on cfDNA is technically challenging due to the often low tumor fractions in a high background of normal cfDNA.

https://doi.org/10.1016/j.ctrv.2019.101951

Received 17 October 2019; Received in revised form 9 December 2019; Accepted 11 December 2019

© 2019 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
The aims of this systematic review were to (1) describe to what extent WES of cfDNA in cancer patients is technically feasible and which approaches are being used, and to (2) analyze the sensitivity of WES-detected single nucleotide variants (SNVs) in cfDNA using tumor tissue as reference \( \times 100\% \) as well as the agreement between cfDNA and tumor tissue \( \times 100\% \).

**Methods**

**Literature search**

PubMed was searched from May 2013 to July 2019 to find full publications. Search terms included *cell free DNA* and *whole exome sequencing*. Also synonyms of the terms and MeSH terms were used (Table A1). For the technical feasibility analysis, studies were eligible if (1) they were written in English (2) WES was used for molecular profiling of cfDNA, and (3) patients had solid tumors. Exclusion criteria were: (1) solely focusing on bioinformatics pipeline not presenting unique data, (2) cfDNA derived from other liquids than blood, and (3) patients without cancer. Subsequently for sensitivity and agreement meta-analyses, studies that reported WES-detected SNVs in cfDNA and matched tumor tissue were included. Studies were excluded if: (1) time between collection of tumor tissue and cfDNA for individual cases exceeded 2 months, and if (2) SNVs in tumor tissue and cfDNA were not reported on individual patient level.

**Data extraction**

Two authors (M.B. and L.A.) independently performed the article selection and data extraction. For all studies the following data were extracted using a data-extraction form (Table A2): year of publication, sample size, cancer type, time between plasma and tissue collection, pre-analytical variables (amount of DNA input, cfDNA fraction), analytical conditions (sequencing methods and coverage), post-analytical conditions (variant calling and analysis), and the mutant allele frequency (MAF) of detected variants. An overview of used source files is available in Table A3. In case of discords the authors reached agreement during a consensus meeting.

**Pooled sensitivity and agreement analysis**

To calculate a pooled sensitivity and agreement rate of WES-detected SNVs in paired cfDNA and tumor tissue (irrespective of primary- or metastatic lesion) we extracted individual patient data from each study. Per sample we collected the number of “shared SNVs” (SNVs detected in both tumor tissue and cfDNA), SNVs only found in tissue and SNVs only present in cfDNA. Also cfDNA input and sequencing coverage were extracted on individual sample level. Using SNVs detectable in tissue as reference, sensitivity was calculated as follows:

\[
\frac{\text{Shared SNVs}}{\text{All tissue SNVs}} \times 100\%.
\]

The agreement rate between WES-detected SNVs in tumor tissue and cfDNA was calculated as follows:

\[
\frac{\text{Shared SNVs}}{\text{All SNVs}} \times 100\%,
\]

in which “all SNVs” was defined as: SNVs only detected in tissue + SNVs only detected in cfDNA + shared SNVs. Patients without detectable SNVs in tumor tissue were excluded from the sensitivity and also from the agreement analysis to keep both groups comparable. We did not calculate specificity, since we were unable to calculate the numbers of true negatives (wild type genes).

**Additional WES-detected SNVs in cfDNA**

For all studies included in the meta-analysis, we extracted the number of additionally detected SNVs in cfDNA for each sample pair and calculated the fraction of uniquely detected variants versus all variants in cfDNA:

\[
\frac{\text{Additional SNVs in cfDNA}}{\text{All cfDNA variants}} \times 100\%.
\]

Per study we displayed the median of the individual sample data. To score the clinical potential of exclusively detected SNVs in plasma, we used the clinical annotation database OncoKB [11] (September 1st, 2019). Additionally, per study we scored whether variants detected exclusively in cfDNA had been described previously in the corresponding tumor type using cBioPortal for Cancer Genomics (September 4th, 2019) [12]. We reported SNVs with a MAF \( \geq 2\% \).

**Statistical analysis**

An individual patient data (IPD) meta-analysis was used to estimate the overall sensitivity and agreement rates across all the studies. Taking into account the heterogeneity, the patient-specific effects and the study-specific effects were employed as random-effects in the (multi-level) model. For this purpose, Bayesian IPD meta-analyses were employed. Results of these analyses were shown using a forest plot, where the median and the 95% highest probability density (hpdp) of credible intervals (CI) were reported for each study separately and pooled in an overall sensitivity and agreement rate. A sub-analysis was performed to estimate the sensitivity and agreement on a subset of cfDNA samples which contained an estimated tumor fraction \( \geq 25\% \).

Computations and graphics were performed in R program language [13]. All Bayesian computations were performed using the Markov Chain Monte Carlo (MCMC) sampler through Jags [14] interface in R program language and relatively non-informative priors were used for the parameters in the model. The MCMC sampling was run for each analysis for 200 k iterations after discarding the first 200 k iterations (burn-in) to reach the convergence.

To assess the correlation between the total number of SNVs with ctDNA fraction, a Spearman’s \( \rho \) was used. To compare the MAF in tumor tissue versus cfDNA a Mann-Whitney U test was performed.

**Results**

**Data retrieval and study characteristics**

In total, 20 studies were included in this review of which the individual patient data of 12 studies were included for the meta-analyses of SNV sensitivity and agreement (Fig. 1) [5,15–33].

**Feasibility of WES of cfDNA**

To evaluate the technical feasibility of WES of cfDNA we summarized pre-analytical and analytical parameters of all studies performing WES of cfDNA (Table 1). In total, WES has been performed on 303 samples, with a median coverage of 137X (range: 43-500X) using a median of 15 ng cfDNA (range: 2–100 ng). Most studies \( (n = 7) \) extracted cfDNA from plasma collected in EDTA tubes or did not report the tube type used \( (n = 6) \). Four studies performed only WES on samples with a high tumor fraction, i.e. \( \geq 10\% \) [15,17] and \( \geq 25\% \) [33] or “high” was not further specified [26]. Overall, the median tumor fraction of all samples from which individual tumor fractions were available, was 37% based on estimation by different platforms such as ultra-low pass whole genome sequencing (ULP-WGS), Sequenza [34] or maximum MAF of variants.

To focus on tumor-specific SNVs, all studies except for Dietz et al. [20] sequenced germline DNA derived from leukocytes or normal tissue. In addition, most studies used a combination of databases such as dbSNP [35], 1000 Genomes Project [36] or Exome Sequencing Project [37] to filter out single nucleotide polymorphisms (SNPs) present in germline DNA. Final selection of variants based on MAF, coverage and sequencing quality was highly variable (Table A4). For example, some studies only called SNVs based on a minimum MAF ranging from 1 [31,33] to 5% [16,17]. Sequencing quality scores involved either in-house developed algorithms or Phred scores with different cut-offs. Finally, not all studies performed an exome-wide final analysis and only reported data on cancer-associated genes [16] or...
genes involved in MAPK-pathway analysis [17] limiting the number of detected SNVs per patient. We further studied the correlation between total numbers of detected SNVs in cfDNA and ctDNA fraction on IPD which showed a positive correlation, p = 0.016 (Fig. A1). Since only the minority of studies provided IPD on cfDNA input or coverage, these variables were not individually tested.

In addition to SNV detection, WES of cfDNA also allowed analysis of copy number variation [19,23,26], mutational signatures [15], fusion genes [33], rearrangements [33], predicted neoantigens [15] and TMB [25]. Mutational signature fractions (i.e. APOBEC and BRCA-like) [15] and TMB [25] were identified and positively correlated between cfDNA and tumor tissue. The Spearman’s correlation coefficient was 0.92 (no CI given) for mutational signature fractions and 0.85 (no CI given) for TMB. Sensitivity and agreement of the genomic signatures in cfDNA versus tumor tissue were not reported.

Out of the 303 cfDNA samples on which WES was performed, WES data of matched tumor tissue was available for only 71 unique sample pairs. To calculate a pooled sensitivity and agreement between WES-detected SNVs in cfDNA versus tumor tissue we performed a Bayesian random-effect meta-analysis on this subset (Table 2; Table A5 for IPD). Most studies compared SNVs between metastatic tumor tissue and cfDNA [15–17,21,26,27,32], whilst three studies analyzed primary tumors [20,29,38] and two studies analyzed both [19,23]. The pooled sensitivity of WES-detected SNVs in cfDNA using tumor tissue as reference was 50% (95% CI: 29–72%) (Fig. 2). The tissue MAF of variants exclusively identified in tissue was significantly lower (12.5%, range: 0.5–18%) than the tissue MAF of shared variants (23.9%, range: 17–38%), p = 0.004. For cfDNA, the median MAF of variants detected in both tumor tissue and cfDNA (12.2%, range: 2.1–26.9%) was higher than the MAF of variants detected in cfDNA only (4.6%, range: 0.4–9.0%), although not statistically significant p = 0.093. The pooled sensitivity and agreement rate of cell-free DNA versus tumor tissue

Fig. 1. Flowchart of the selection of publications included in the feasibility analysis and meta-analysis
| First author | Year of publication | Pre-selection for WES analysis? | Sample size for WES (N) | Tumor type | Library Prep | Sequencing method | Median input cfDNA (ng) | Median coverage (X) | Estimated median ctDNA fraction | Method for estimation of ctDNA fraction | Genetic alterations other than mutations identified in cfDNA |
|-------------|-------------------|-------------------------------|-----------------------|------------|--------------|------------------|------------------------|-----------------------|-----------------------------|----------------------------------|--------------------------------------------------|
| Adalsteinsson¹⁵ | 2017 | Yes, samples with > 10% ctDNA based on ULP-WGS | 41 | Breast and prostate | Kapa Hyper Prep kit with custom adapters (IDT and Broad Institute) | HiSeq2500/4000 | 20 | 208 | 0.34 | ULP-WGS | BRCA-like signature and APOBEC signature in tumor tissue and matched cfDNA were strongly correlated; Number of predicted neoantigens in tumor tissue and matched cfDNA were strongly correlated |
| Ahlborn¹⁹ | 2019 | no | 44 | Pan-cancer | NEBNext Ultra II Library Prep Kit (New England Biolabs) and MedExome (Roche)/SureSelect (Agilent) | NextSeq/HiSeq | 10 | 210 | 0.9 | Maximum MAF of identified SCAA per patient |
| Ahlborn¹⁸ | 2018 | Yes, progression samples with known high MAF (>10%) for BRAF by dPCR | 4 | Colorectal- and bile duct | NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) and SeqCap EZ MEDExome (Roche NimbleGen) | Illumina | 10 | > 50 | 0.17, 0.12 | Maximum MAF of identified SCAA per patient |
| Butler¹⁸ | 2015 | No | 2 | Sarcoma; Breast | SureSelectXT Human All Exon V4 + UTRs (Agilent) | HiSeq 2000 | 100 | 416.5 | 0.227 | Maximum MAF of identified SCAA per patient |
| Chicard¹⁹ | 2018 | no | 36 | Neuroblastoma | Kapa Library Preparation Kit (Kapa Biosystems) and SeqCap EZ Human Exome Kit v3.0 (Roche NimbleGen) | HiSeq 2500 | 100 | NR | 0.5 | Sequenza¹ | CNAs were concordant between plasma and tumor tissue in 93% of all cases with an additional amplification of 15q harboring KIF11 in cfDNA |
| Dietz²⁰ | 2016 | no | 6 | NSCLC | ThruPLEX-FD Prep Kit (Rubicon Genomics) | HiSeq2000 | 10 | 68.5 | NR | No information on ctDNA fraction available |
| Goodall²¹ | 2017 | no | 6 | Prostate | Kapa Library Prep Kit (Kapa Biosystems) and SeqCap EZ Human Exome Kit v3.0 (Roche NimbleGen) | NextSeq500 | 40 | NR | 0.21 | Maximum MAF identified by WES per patient |
| Gremel²² | 2016 | no | 1 | Musosal melanoma | Accel NGS 25 DNA Library Kits (Swift Biosciences) | NextSeq | 10-25 | NR | 0.36 | Maximum MAF identified by WES extracted from graph |
| Huang⁵ | 2017 | no | 5 | HCC | Agilent and in-house 50 Mb system V4 (BGI Xome) | HiSeq4000 | 50-100 | 226.2 | NR | No information on ctDNA fraction available |
| Jiménez²³ | 2019 | no | 18 | Renal tumors | Kapa Library Preparation Kit (Kapa Biosystems) and SeqCap EZ Human Exome Kit (Roche NimbleGen) | HiSeq2500 | 60.9 | NR | 0.295 | Sequenza¹ | Mutational signatures with the largest contribution in tissue were found in plasma, some additional signatures in plasma were identified |
| Klevebring²⁴ | 2014 | no | 9 | Breast; Prostate | ThruPLEX-FD (Rubicon Genomics) and SeqCap EZ Exome Library Version 1 (Roche NimbleGen) | HiSeq2500 | 5 | 43 | 0 | Maximum MAF identified by WES per patient |

(continued on next page)
Table 1 (continued)

| First author         | Year of publication | Pre-selection for WES analysis? | Sample size for WES (N) | Tumor type                  | Library Prep                                      | Sequencing method | Median input cfDNA (ng) | Median coverage (X) | Estimated median ctDNA fraction | Method for estimation of ctDNA fraction | Genetic alterations other than mutations identified in cfDNA |
|----------------------|---------------------|---------------------------------|-------------------------|------------------------------|--------------------------------------------------|-------------------|------------------------|-----------------------|-------------------------------------|----------------------------------------|----------------------------------------------------------|
| Koeppel              | 2017                | Samples with at least 6 ng cfDNA remaining for analysis | 32                      | Pan-cancer                  | NebNext Ultra II DNA Library Prep (New England Biolabs) and SureSelect XT (Agilent) | HiSeq4000         | 25                     | 140                   | 0.19                  | Maximum MAF identified by WES per patient | TML in plasma and corresponding tissue was correlated for ctDNA positive patients with significant ctDNA fraction |
| Murtaza              | 2013                | Longitudinal samples with high allelic fraction | 19                      | Breast- and ovarian         | ThruPLEX-FD (Rubicon Genomics) and TrueSeq Exome Enrichment Kit (Illumina) | HiSeq2500         | 10                     | 110                   | 0.52                  | Genomewide aggregated allelic loss (GAAL) | CNAs were concordant between plasma and tumor tissue and more prominent in samples with ctDNA fraction > 50% |
| Murtaza              | 2015                | Yes only sample T1, T2 and T9, reason not further specified | 3                       | Breast                      | ThruPLEX-FD (Rubicon Genomics) and TrueSeq Exome Enrichment Kit (Illumina) | HiSeq2500         | 2–18                   | 79.6                  | 0.215                 | Mean MAF of stem mutations | |
| Olmedillas-Lopez     | 2018                | no                              | 30                      | Colorectal                  | SeqCap EZ HGS C Vrome Kit (Roche NimbleGen) | NextSeq00         | NR                     | 40-80                 | NR                   | No information on ctDNA fraction available | |
| Song                 | 2018                | no                              | 1                       | Thyroid                     | SureSelect Human All Exon V5 (Agilent) | HiSeq2000         | NR                     | 133                   | NR                   | No information on ctDNA fraction available | |
| Sun                  | 2019                | no                              | 6                       | Glioma                      | Nebnext Ultra II DNA Library Prep Kit (New England Biolabs) and NERB multiplex Oligo (New England Biolabs) kit | HISEQ Xten       | 61                     | 500                  | NR                   | No information on ctDNA fraction available | |
| Tailor               | 2019                | no                              | 33                      | Lung cancer; benign lung lesion | SureSelect All Exon V5 + UTR (Agilent) | HiSeq2500         | > 10                    | 49                    | NR                   | No information on ctDNA fraction available | |
| Toledo               | 2018                | no                              | 1                       | Colorectal                  | ThruPLEX PlasmaSeq (Rubicon Genomics) SureSelectXT Target Enrichment System (Agilent) | HiSeq4000         | 15                     | 183                   | 0.47                 | Maximum MAF identified by WES per patient | |
| Vandekerkhove        | 2017                | Samples with ≥ 25% ctDNA fraction | 11                      | Bladder cancer              | SeqCap EZ MedExome kit (Roche NimbleGen) | MiSeqV3 600 or HiSeq2500 | 10-25                  | 176                   | 0.40                 | MAF of autosomal somatic mutations corrected for mutations with LOH, mutations with detectable amplification were excluded | Novel fusion genes (FGFR-ADD1) and rearrangements were detected and validated |

Abbreviations: ULP-WGS: Ultra low-pass whole genome sequencing; NR: not reported; SCAA: selected cancer-associated alteration; MAF: Mutant Allele fraction; LOH: Loss of heterozygosity.

C = plasma concentration of tumor-derived DNA, Nnondel = number of sequenced reads carrying the nondeleted alleles in the tumor tissues, Ndel = number of sequenced reads carrying the deleted alleles in the tumor tissues.

1 Sequenza: estimates ploidy in samples with a tumor content > 30%.

2 Analysis of the allelic counts for SNPs exhibiting loss of heterozygosity in the tumor by the following equation: C = (Nnondel − Ndel)/Nnondel.

3 Stem mutations: common to all tumor biopsies.

4 ctDNA fraction = 2/(1/(1/mutant allele fraction) + 1).

* the original article mentions 'an average sequence depth of > 500 bp and genome coverage > 98%' which we interpreted as 500× depth.
## Table 2
Overview of studies investigating WES-detected SNVs in cfDNA and tissue.

| First author | Year of publication | Sample size for concordance analysis | Tumor type | Location of tissue biopsy | Input (ng) | Coverage (X) | Tumor purity (%) | Analysis performed exome wide? |
|--------------|---------------------|--------------------------------------|------------|---------------------------|------------|--------------|-----------------|-------------------------------|
| Adalsteinsson15 | 2017 | 24 | Breast- and prostate metastatic lesions | cfDNA | 20 | NR | 208 | 173 | 58 | yes |
| Ahlborn16 | 2019 | 13 | Pan-cancer metastatic lesions | cfDNA | 10 | 10 | 210 | > 100 | NR | yes but only selected cancer associated alterations were reported |
| Ahlborn17 | 2018 | 2 | Bile duct metastatic lesions | cfDNA | 10 | 10 | > 50 | NR | 52 | no, focused on MAPK-pathway genes |
| Chicard19 | 2018 | 14 | Neuroblastoma primary tumor and metastatic lesions | cfDNA | 100 | 100 | NR | NR | 88 | yes |
| Dietz20 | 2016 | 6 | NSCLC primary tumor | cfDNA | 10 | 50 | 68.5 | 61 | NR | no, 58 known most frequently mutated genes |
| Goodall21 | 2017 | 1 | Prostate metastatic lesions | cfDNA | 40 | NR | NR | NR | 30 | yes |
| Huang22 | 2017 | 5 | HCC primary tumor, multiple regions | cfDNA | 50 | 500-2000 | 226 | 211 | 70 | yes |
| Jiménez23 | 2019 | 2 | Renal tumors primary tumor and metastatic lesions | cfDNA | 61 | NR | NR | NR | 35 | yes |
| Murtaza24 | 2013 | 1 | Ovarian metastatic lesions | cfDNA | 17.5 | NR | 110 | NR | NR | yes |
| Murtaza25 | 2015 | 1 | Breast metastatic lesions | cfDNA | 10 | NR | 77.3 | 16.6 | 70 | yes |
| Toledo26 | 2018 | 1 | Colorectal metastatic lesions | cfDNA | 15 | 70-110 | 183 | 88.5 | NR | yes |
| Song27 | 2018 | 1 | Follicular thyroid carcinoma primary tumor | cfDNA | NR | NR | 33 | 140 | NR | yes |

Abbreviations: NR: not reported; NSCLC: non-small cell lung cancer; HCC: hepatocellular carcinoma.

Shared SNVs: Single nucleotide variants in both plasma and tissue.
Tissue SNVs: All single nucleotide variants in tissue (independent of plasma).
All SNVs: All single nucleotide variants in plasma and/or tissue.

1 Values are calculated medians unless ranges are reported.
agreement was 31% (95% CI: 15-49%) (Fig. 3). Since we found a positive correlation between total numbers of SNVs and ctDNA fraction, we conducted a sub-analysis of 40 unique sample pairs with a ctDNA fraction ≥ 25%, which resulted in a sensitivity of 69% (95% CI: 46-89%) and agreement of 46% (95% CI: 36-59%).

**Additional value of WES of cfDNA**

To evaluate the additional value of WES of cfDNA to identify clinically useful SNVs that were not present in tissue, we calculated the number of additional SNVs unique to cfDNA and the ratio between ctDNA variants unique to plasma versus all variants detected in ctDNA (Table 3). Of all plasma-detected SNVs per sample a median of 43% (range: 0-96%) was exclusively detected in plasma. The median number of additionally detected SNVs per sample was 17 (range: 0-2840). Of these additionally detected SNVs, 36 variants detected in 20 out of 53 patients (38%) were detected in cancer associated genes as reported in eBioportal. Matching IPD with targetable genes according to OncoKB, we identified in 11 variants in 9 out of 53 patients (17%). The targetability of these variants ranged from level 1 (FDA-approved) to level 4 (biological evidence) (Table 3).

**Discussion**

The increasing interest to capture the complex genomic landscape of individual cancer patients real time and in a minimally-invasive way, has initiated efforts on technical developments in the field of WES of cfDNA. The main purpose of this systematic review was to evaluate the technical feasibility of WES of cfDNA and to analyze the sensitivity and agreement of WES-detected SNVs in cfDNA using tumor tissue as reference.

It has become clear that there was significant variability between studies in the pre- and post-analytical conditions used (Table 1; Table A4) which severely impacted comparability of results. Differences between studies were observed regarding technical aspects of sequencing including sequencing coverage and amount of cfDNA used. Especially
sequencing coverage was highly variable, ranging from 43 to 500X coverage, which theoretically results in lower limits of variant detection of 2.3% and 0.2% respectively, assuming that the variant is heterozygous and the genome is diploid. Although cfDNA input generally consisted of 10–20 ng, inputs ranged from 2 ng to 100 ng. In addition, the post-analytical part in which different bioinformatics pipelines were used also impacted the final variant calling since most studies used their own set of criteria to filter SNPs and to perform final variant calling (Table A4).

Clonal hematopoiesis has been identified as an important factor affecting accurate variant interpretation. During the process of aging different mutations accumulate in hematopoietic stem cells. This phenomenon occurs frequently in the elderly and its prevalence has been estimated at 31% [39]. The mutations resulting from clonal hematopoiesis are often detected during cfDNA sequencing analysis, since the majority of cfDNA is derived from leukocytes. Recently, it has been demonstrated that 53.2% of all mutations detected by cfDNA sequencing analysis result from clonal hematopoiesis, indicating the need for collection and sequencing of leukocytes as a reference [39].

Taken into account that for only 71 out of 303 cfDNA samples WES data of matching tumor tissue was available, the merit of our comparison is that we performed the meta-analysis on IPD level which allowed us to adjust for patient-specific effects in our model. Compared to large studies describing agreement between cfDNA and tumor tissue based on targeted sequencing approaches covering pre-specified gene sets [40], the number of sample pairs which was analyzed by WES is considerably limited. Most studies lacked IPD on cfDNA input and sequencing coverage hampering analysis of the impact of those variables on sensitivity and agreement.

Some studies only performed WES on samples with a minimum tumor fraction [15,17,26,27,33]. Overall, the samples selected for WES consisted of a median tumor fraction of 37%, much higher than generally occurs in cancer patients [41]. By using techniques as ultralow-pass whole-genome sequencing with 10% tumor fraction as a cutoff value to pre-select samples, Adalsteinsson et al. [15] showed that only 34% of cfDNA samples from metastatic breast- and prostate cancer patients were feasible for WES analysis, including samples from all treatment lines. This implies that the number of samples with a sufficient tumor fraction in earlier lines of treatment might be even lower. Notably, studies have not reported success rates of WES in correlation to tumor fraction.

Our meta-analysis of the sensitivity of WES-detected SNVs in cfDNA versus tumor tissue has shown that 50% of SNVs present in tumor tissue are also detected in cfDNA. The reason for the rather low sensitivity of WES of cfDNA is probably multifactorial including technical and biological aspects. A major technical issue is the generally low sequencing coverage used for WES resulting in false-negative results for cfDNA variants present below the limit of detection. This is supported by the comparison of WES (coverage 226X) with targeted deep sequencing (coverage 1806X) on the same sample [5], demonstrating that some variants with low MAPs (<5%) were detected by targeted deep sequencing only [5]. However, with the introduction of unique molecular identifiers and Elimination of Recurrent Artefacts and Stochastic Errors (ERASE-Seq) [42] discrimination of sequencing artefacts from true variants can be improved, enabling detection DNA variants at ultralow frequency.

Another technical aspect possibly affecting WES sensitivity is the type used for blood collection. Previous studies have demonstrated that cfDNA isolated from serum instead of plasma increased the background of normal DNA by the release of germline DNA (gDNA) due to lysis of leukocytes during coagulation [43,44]. This might partly explain the low sensitivity of Dietz et al. [20]. Size selection of short fragments before sequencing might positively influence the ratio between gDNA and cfDNA as well [45]. Biological challenges are the amount of cfDNA available for sequencing, generally low overall tumor fraction present in the sample and the subclonal presence of clones bearing alterations associated with treatment resistance [19]. The importance of the cfDNA fraction for the sensitivity of WES is supported by our finding that sensitivity improves when analyzing samples with a ctDNA fraction ≥ 25%.

Our findings demonstrate that in addition to tumor-tissue detected SNVs, WES of cfDNA also discovers SNVs exclusively detected in cfDNA. When calculating the ratio of these SNVs versus all cfDNA variants, we observed that the fraction of variants unique to cfDNA was highly variable amongst studies. The variability might partly be explained by factors as sequencing coverage and cfDNA input. This assumption, however, could not be substantiated by our data as availability of sequencing coverage and cfDNA input on individual patient level was insufficient. Nevertheless, SNVs which are exclusively detected in cfDNA potentially reflect intra- and inter tumor heterogeneity. Whether these additionally detected SNVs in plasma are derived from clonal or subclonal fractions in tumor tissue remains a topic of interest. Adalsteinsson et al. [15] estimated clonality and subclonality of SNVs detected in plasma and demonstrated that on average 88% of the clonal

### Table 3

| First author | Year of publication | Tumor type | Number of samples | Ratio of cfDNA variants unique to plasma/all cfDNA variants | Median number of additional SNVs in cfDNA per patient | Mutated genes additionally identified in cfDNA | OncoKB actionable genes (level of evidence for this tumor type) |
|--------------|---------------------|------------|-------------------|----------------------------------------------------------|-----------------------------------------------------|---------------------------------|-----------------------------------------------------|
| Adalsteinsson [15] | 2017 | breast | 23 | 0.39 | 27 | | ESR1(3A); PIK3CA(1); mTOR(4); NF1(4); ERBB2(3A) |
| Ahlborn [16] | 2019 | prostate | 13 | 0.38 | 0 | | |
| Ahlborn [17] | 2018 | bile duct | 2 | 0.96 | 154 | | mTOR(4); FGFR3(4) |
| Chicard [19] | 2018 | neuroblastoma | 14 | 0.49 | 16 | | ARID1A; POL; DOT1; ANKR3DC |
| Dietz [26] | 2016 | NSCLC | 6 | 0.69 | 999 | | |
| Goodall [21] | 2017 | prostate | 1 | 0.07 | 2 | | mTOR(4) |
| Huang [22] | 2017 | HCC | 5 | 0.73 | 135 | | PTEN (4)* |
| Jiménez [23] | 2019 | clear cell sarcoma | 2 | 0.42 | 3 | | |
| Murtaza [26] | 2013 | breast | 1 | 0 | 0 | | |
| Murtaza [27] | 2015 | breast | 1 | 0.21 | 13 | | |
| Song [29] | 2018 | Thyroid | 1 | 0.21 | 56 | | |
| Toledo [32] | 2018 | colorectal | 1 | 0.21 | 14 | | MDC1; PARP3; RTN4; CHD7 |

Abbreviations: NR: not reported; NSCLC: non-small cell lung cancer; HCC: hepatocellular carcinoma.
and 45% of the subclonal mutations were confirmed in the tumor. Assuming that all plasma detected SNVs are true variants, i.e. free of sequencing artefacts, these results imply that the majority of SNVs exclusively detected in cfDNA were subclonal in this study. Furthermore, SNVs indicated as subclonal in cfDNA might be of clonal origin in tumor tissue from other metastatic sites.

Studies comparing cfDNA to multiple tumor region sampling support this hypothesis [5,25]. Huang et al. found that nearly all exclusively detected SNVs in plasma by WES were also detected in tumor samples from different liver lesions using targeted deep sequencing (average 98.7%, range: 69.3–100%) [5]. Another study also showed that when two tissue biopsies were taken and compared to cfDNA, the number of overlapping alterations between cfDNA and tumor tissue increased [25]. Importantly, some of these exclusively identified mutations were previously associated with therapy resistance (ESR1, ERBB2 and NFI) [46] and treatment outcome (PIK3CA) [47] in breast cancer. The clinical relevance of these findings and to what extent the MAF and its dynamics will impact outcome on certain therapies have thus far not been elucidated in prospective clinical studies. Secondly, the number of additional identified targetable mutations in cfDNA is currently very limited, but might be improved by efforts unraveling new actionable targets or profiles.

The added value of WES currently thus mainly resides in the discovery of resistance mechanisms and genomic alterations for which a wide coverage of the genome is needed such as TMG and mutational signatures. Goodall et al. [21] demonstrated this discovery-capacity of WES by identifying frameshifts in germline and somatic DNA repair mutations as mechanisms of resistance to PARP inhibitors in prostate cancer. For estimation of TMB, large targeted sequencing panels can be used [48]. However when taking estimation of TMB by whole genome sequencing (WGS) as reference, 30% of patients were misclassified – either false negative or false positive – when targeted sequencing panels were used. Concordance improved by increasing number of megabases (Mb) covered by the targeted panel [49]. Furthermore, reported correlations between cfDNA and tissue TMB using targeted panels, Spearman’s correlation coefficient of 0.64 and 0.6 [50] are lower than correlations reported using WES, 0.85 [25].

To this end, further design and validation of highly needed targeted sequencing panels for TMB estimation are currently ongoing.

Altogether WES is an attractive tool for identification of genomic signatures and discovery of resistance mechanisms. In this IPD meta-analysis we show that the sensitivity of WES of cfDNA is 50% and that the overall agreement is 31%. Furthermore we describe large variability in pre- and post-analytical conditions of WES of cfDNA. Moreover, our results underline that the applicability of WES mainly resides in a selected group of patients with high tumor fractions. We recognize that WES is still in its developmental phase and that implementation of methods such as unique molecular barcoding and ERASE-seq will further improve sensitivity of WES. However, standardization of methodologies is highly needed to further define the clinical utility of this promising approach.

Contributors

MKB and LA wrote the manuscript, which was edited, reviewed and approved by AJ, MPHMJ, JWMM and SS. KN performed the statistical analyses.

Acknowledgements

The authors would like to express their sincere gratitude to Marcel Smid for critically reading the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ctrv.2019.101951.

References

[1] Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, et al. Improved survival with MEK inhibition in BRAF-mutated melanoma. N Engl J Med 2012;367:107–14.
[2] McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummett R, et al. Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. Lancet Oncol 2014;15:323–32.
[3] Souich MJ, Wise MD, Rowland A, Kiehnadassie G, McKinnon RA, Karapetis CS. Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. Ann Oncol 2015;26:13–21.
[4] Frenel JS, Carreira S, Goodall J, Roda D, Perez Lopero R, Tanurri N, et al. Serial next-generation sequencing of circulating cell-free DNA evaluating tumor clone response to molecularly targeted drug administration. Clin Cancer Res 2015;21:4586–96.
[5] Huang A, Zhao X, Yang XR, Li PQ, Zhou XL, Wu K, et al. Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma. J Hepatol 2017;67:293–301.
[6] Merker JD, Oxnard GR, Compton C, Diehm M, Hurley P, Lazard AJ, et al. Circulating tumor DNA analysis in patients with cancer: American society of clinical oncology and college of American pathologists joint review. J Clin Oncol 2018;36:1613–41.
[7] De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng KC, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free DNA: a proof-of-principle. Ann Oncol 2014;25:1729–35.
[8] cobas EGFR Mutation Test v2. 2016.
[9] Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aslakd LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357:409–13.
[10] Ruiz NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel J, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in small cell lung cancer. Science 2015;348:124-8.
[11] Chakravarty D, Gao J, Phillips SM, Kundan R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. J Precis Oncol 2017 2017.
[12] cBioPortal for cancer genomics. http://cbioportal.org. 2019 [accessed 4 September 2019].
[13] Gore Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2019.
[14] Plummer M. JAGS: Bayesian Graphical Models using MCMC. R package version 4-6. 2016.
[15] Adalsteinsson VA, Ha G, Freeman SS, Choudhary AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun 2017;8:1324.
[16] Ahlborn LB, Rohrberg KS, Gabrielie M, Tuxen IV, Yde CW, Spanggard I, et al. Application of cell-free DNA for genomic tumor profiling: a feasibility study. Oncotarget 2019;10:13888–98.
[17] Ahlborn LB, Tuxen IV, Moullier F, Kinalis S, Schmidt AY, Rohrberg KS, et al. Circulating tumor DNA as a marker of treatment response in BRAF V600E mutated non-melanoma solid tumors. Oncotarget 2018;9:32570–9.
[18] Butler TM, Johnson-Camacho K, Peto M, Wang NJ, Macey TA, Korkola JE, et al. Exome sequencing of cell-free DNA from metastatic cancer patients identifies clinically actionable mutations distinct from primary disease. PLoS ONE 2015;10:e0136407.
[19] Chicard M, Colmet-Daage L, Clement N, Danzon A, Bocch M, Bernard V, et al. Whole-exome sequencing of cell-free DNA reveals temporospatial heterogeneity and identifies treatment-resistant clones in neuroblastoma. Clin Cancer Res 2018;24:939–49.
[20] Dietz S, Schirmer U, Merc M, von Bubnow N, Dahl E, Meister M, et al. Low input whole-exome sequencing to determine the representation of the tumor exome in circulating DNA of non-small cell lung cancer patients. PLoS ONE 2016;11:e0161012.
[21] Goodall J, Mateo J, Yuan W, Moposs H, Porta N, Miranda S, et al. Circulating cell-free DNA to guide prostate cancer treatment with PARP inhibition. Cancer Discov 2017;7:1066–17.
[22] Gremel G, Lee RJ, Girotti MR, Mandal AK, Valpione S, Garner G, et al. Distinctive subclonal tumor responses to therapy revealed by circulating cell-free DNA. Ann Oncol 2016;27:1595–69.
[23] Jimenez I, Chicard M, Colmet-Daage L, Clement N, Danzon A, Lapouble E, et al. Circulating tumor DNA analysis enables molecular characterization of pediatric renal tumors at diagnosis. Int J Cancer 2019;144:68–79.
[24] Klevebring D, Neiman M, Sundling S, Eriksson L, Darzi Ramqvist E, Celebioglu F, et al. Evaluation of exome sequencing to estimate tumor burden in plasma. PLoS ONE 2014;9:e104417.
[25] Koeppel F, Blanchard S, Jovelet C, Genin B, Marcailou C, Martin E, et al. Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients. PLoS ONE 2017;12:e0181874.
[26] Kurtmaa M, Dawson SJ, Tzui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 2013;497:108–12.
[27] Kurtmaa M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumor DNA in a case of metastatic breast cancer. Nat Commun 2015;6:8760.
[28] Olmedillas-Lopez S, Garcia-Olmo DC, Garcia-Arranz M, Peiro-Pastor R, Aguado B, et al. Improved clinically actionable mutations distinct from primary disease. PLoS ONE 2019;14:1-17.

M.K. Bos, et al.
Cancer Treatment Reviews 83 (2020) 101951

9
Garcia-Olmo D. Liquid biopsy by NGS: differential presence of exons (DPE) in cell-free DNA reveals different patterns in metastatic and nonmetastatic colorectal cancer. Cancer Med 2018;7:1706–16.

Song J, Yang Z. Case report: Whole exome sequencing of circulating cell-free tumor DNA in a follicular thyroid carcinoma patient with lung and bone metastases. J Circ Biomark 2018;7:184945418763725.

Sun J, Zhou W, Mao K, He Y, Yao J, Tang Z, et al. Examination of plasma cell-free DNA of glioma patients by whole exome sequencing. World Neurosurg 2019.

Tailor TD, Rao X, Campa MJ, Wang J, Gregory SG, Pate Jr. EF. Whole exome sequencing of cell-free DNA for early lung cancer: a pilot study to differentiate benign from malignant CT-detected pulmonary lesions. Front Oncol 2019;9:317.

Toledo RA, Garralda E, Mitzl M, Pons T, Monsech J, Vega E, et al. Exome sequencing of plasma DNA portrays the mutation landscape of colorectal cancer and discovers mutated VEGFR2 receptors as modulators of antiangiogenic therapies. Clin Cancer Res 2018;24:3550–9.

Vandekerckhove G, Todenhofer T, Annala M, Strauss WJ, Hong A, Beja K, et al. Circulating tumor DNA reveals clinically actionable somatic genome of metastatic bladder cancer. Clin Cancer Res 2017;23:6487–97.

Favero F, Joshi T, Marquard AM, Birkbak NJ, Kryzhanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol 2015;26:64–70.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan I, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.

Exome Variant Server. NHLBI Exome Sequencing Project (ESP), Seattle, WA.

[Other references omitted for brevity]