Single tube liquid biopsy for advanced non-small cell lung cancer

de Wit, Sanne; Rossi, Elisabetta; Weber, Sabrina; Tamminga, Menno; Manicone, Mariangela; Swennenhuis, Joost F.; Groothuis-Oudshoorn, Catharina G. M.; Vidotto, Riccardo; Facchinetti, Antonella; Zeune, Leonie L.

Published in:
International Journal of Cancer

DOI:
10.1002/ijc.32056

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

Single tube liquid biopsy for advanced non-small cell lung cancer

Corresponding author

Prof. Harry. J.M. Groen, MD, PhD
Faculty of Medical Sciences, Lung diseases
Hanzeplein 1, 9713 GZ Groningen, The Netherlands
Phone: +31 50 361 1546
Email: h.j.m.groen@umcg.nl

Key words

Liquid biopsy, Circulating tumour cells (CTC), Circulating tumour DNA (ctDNA), Extracellular vesicles (EV), Survival, Non-small cell lung cancer (NSCLC), EpCAM, Biomarkers

Abbreviations

CK = cytokeratins

CTC = circulating tumour cells

ctDNA = circulating tumour DNA

EpCAM = epithelial cell adhesion molecule

$EpCAM^{high}$ CTC = EpCAM high expressing CTC

$EpCAM^{low}$ CTC = EpCAM low expressing CTC

NSCLC = Non-small cell lung cancer

tdEV = tumour-derived Extracellular Vesicles

Research article, category Tumor Markers and Signatures

Novelty and Impact

Using a single blood draw, multiple cancer biomarkers as CTC, extracellular tumour vesicles and circulating tumour DNA, were analysed in 97 advanced NSCLC patients. The presence of EpCAM high expressing CTC and elevated levels of tumour vesicles and tumour DNA were associated with a poor clinical outcome, whereas the presence of EpCAM low expressing CTC was not. This single tube approach enables simultaneous analysis of several biomarkers to explore their potential as a liquid biopsy.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.32056

Single tube liquid biopsy for advanced non-small cell lung cancer

Sanne de Wit[1], Elisabetta Rossi[2,3], Sabrina Weber[4], Menno Tamminga[5], Mariangela Manicone[3], Joost F. Swennenhuis[1], Catharina G.M. Groothuis-Oudshoorn[6], Riccardo Vidotto[3], Antonella Facchinetti[2,3], Leonie L. Zeune[1,7], Ed Schuuring[5], Rita Zamarchi[3], T. Jeroen N. Hiltermann[3], Michael R. Speicher[4], Ellen Heitzer[4,8], Leon W.M.M. Terstappen[1], Harry J.M. Groen[5]

[1] Department of Medical Cell BioPhysics, University of Twente, Enschede, The Netherlands

[2] DISCOG, University of Padua, Padua, Italy

[3] Veneto Institute of Oncology IOV – IRCCS, Padua, Italy

[4] Institute of Human Genetics, Diagnostic and Research Centre for Molecular BioMedicine, Medical University of Graz, Graz, Austria

[5] Department of Pulmonary Diseases, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

[6] Department of Health Technology & Services Research, University of Twente, Enschede, the Netherlands

[7] Department of Applied Mathematics, University of Twente, Enschede, the Netherlands

[8] Christian Doppler Laboratory for Liquid Biopsies for Early Detection of Cancer, Medical University of Graz, Graz, Austria

Running Title: Liquid biopsy biomarkers in NSCLC

Corresponding author: Harry Groen (h.j.m.groen@umcg.nl)

Abstract

The need for a liquid biopsy in non-small cell lung cancer (NSCLC) patients is rapidly increasing. We studied the relation between overall survival (OS) and the presence of four cancer biomarkers from a single blood draw in advanced NSCLC patients: EpCAM^{high} circulating tumour cells (CTC), EpCAM^{low} CTC, tumour derived extracellular vesicles (tdEV) and cell-free circulating tumour DNA (ctDNA). EpCAM^{high} CTC were detected with CellSearch, tdEV in the CellSearch images and EpCAM^{low} CTC with filtration after CellSearch. ctDNA was isolated from plasma and mutations present in the primary tumour were tracked with deep sequencing methods. In 97 patients, 21% had ≥2 EpCAM^{high} CTC, 15% had ≥2 EpCAM^{low} CTC, 27% had ≥18 tdEV and 19% had ctDNA with ≥10% mutant allele frequency. Either one of these four biomarkers could be detected in 45% of the patients and all biomarkers were present in 2%. In 11 out of 16 patients (69%) mutations were detected in the ctDNA. Two or more unfavourable biomarkers were associated with poor OS. The presence of EpCAM^{high} CTC and elevated levels of tdEV and ctDNA was associated with a poor OS; however, the presence of EpCAM^{low} CTC was not. This single tube approach enables simultaneous analysis of multiple biomarkers to explore their potential as a liquid biopsy.

Introduction

Advanced non-small cell lung cancer (NSCLC) patients are characterized by gradual growing metastases in different organs, increasing tumour load and comorbidities that grimly determines their fate. Invasive diagnostics are often difficult by inability to perform invasive procedures or due to inaccessible metastases. Liquid biopsies may provide a convenient and patient-friendly approach to obtain information on prognosis and prediction of the best treatment management [1]. Liquid biopsy approaches include the sampling and analysis of circulating components from blood and other body fluids [2]. While the clinical utility of circulating tumour cells (CTC) and cell-free circulating tumour DNA (ctDNA) has been extensively investigated in recent years, other components such as tumour derived extracellular vesicles (tdEV) have only recently been put to the focus of research [3–7].

CTC are epithelial cells disseminated into the blood from primary or metastatic sites. The presence of CTC is predictive of relatively short survival in several types of cancer, including breast, prostate, colon, small and non-small cell lung carcinoma [8–14]. CTC are rare events; they are surrounded by ~$5{\cdot}10^6$ white blood cells and ~$5{\cdot}10^9$ red blood cells per mL [15,16]. For this reason the appropriate marker selection for enrichment is a crucial factor. In most cases, CTC detection is based on the expression of the cell surface epithelial cell adhesion molecule (EpCAM), as it is expressed by the majority of epithelial derived cancers, while hematopoietic cells show no or only very little expression [17,18]. However, the sole use of

EpCAM for CTC isolation might lead to an underestimation of CTC numbers because tumour cells expressing low amounts of EpCAM might be missed by the system. While EpCAM expressing CTC have shown to be highly clinically relevant, recently, the relevance of the presence of CTC expressing no or low EpCAM in cancer patients, is the subject of debate. While many subpopulations can be described, the clinical utility of these cells is barely addressed [19–23].

The use of ctDNA as a clinical response marker for NSCLC patients has already moved into the clinical routine and EGFR T790M testing from plasma has been proven to complement tissue-based testing [5]. However, this can be applied to only a subset of patients harbouring an activating EFGR mutation, while an untargeted approach, which does not require prior knowledge of the mutation status the tumour, would facilitate a more widespread application.

Tumour derived Extracellular Vesicles (tdEV) comprise of a variety of vesicles secreted or budded of from cancer cells and are known to play an important role in many tumour biological processes [7,11,24,25]. In a previous work we have demonstrated that a subset of tdEV, which expresses both EpCAM and cytokeratin but not CD45 or DNA, can be enriched and enumerated using the CellSearch and their presence was strongly associated with poor overall survival [26].

While all of these biomarkers are promising for predicting survival, the predictive ability of the combined biomarkers may yield complementary information and thereby improve diagnostic sensitivity. We hypothesized that a comprehensive, multi-parameter approach with different highly specific tumour shedding products will predict those patients with a relative good prognosis from those with a poor prognosis. Therefore, we determined the presence of four biomarkers in one tube of blood in advanced NSCLC patients. Two CTC subpopulations were discriminated: EpCAM expressing CTC (referred to as $EpCAM^{high}$ CTC) detected using the CellSearch® system and no or low EpCAM expressing CTC (referred to as $EpCAM^{low}$ CTC) detected on microsieves after filtration [21]. The tdEV were identified in the images from the CellSearch using the open source imaging program ACCEPT, whereas plasma from the same tube was used for cell-free ctDNA extraction followed by an untargeted tumour allele fraction analysis [27–29]. Moreover, for a subset of patients tumour-specific mutations were tracked in plasma DNA using deep sequencing or Safe-SeqS [30]. In total 97 NSCLC patients were included and the presence of these different biomarkers in an all-in-one liquid biopsy was explored.

Methods

Patients and healthy donors

Patients with stage IIIB and IV NSCLC were staged according to IASLC staging system (7th Edition) and diagnosed using FDG-PET/CT imaging and different techniques to procure tumour tissue. In total 97 patients were processed: 60 patients were enrolled at University Medical Centre Groningen (The Netherlands) and 37 patients at Veneto Institute of Oncology IOV – IRCCS, Padua (Italy). All patients provided written informed consent and the study protocol was approved by the medical ethical committees. In total 12 different healthy donors donated 35 blood samples, which were used as controls and provided informed consent prior to blood donation, in accordance to the study protocol approved by the METC Twente ethics committee.

Blood and plasma collection

Peripheral blood samples were drawn by vena puncture into 10 mL CellSave blood collection tubes (Menarini Silicon Biosystems, Huntingdon Valley PA, USA) and in an additional EDTA blood collection tube. EDTA blood collection was performed routinely as part of the diagnostic process that included tumour tissue procurement. For CellSearch analysis, the blood from patients was processed within 96 hours, whereas blood samples from healthy donors were processed within 24 hours. Blood from the CellSave tube was transferred to a CellSearch conical tube and centrifuged for 10 minutes at 800g without using the brake. Thereafter, plasma was aspirated without disturbing the buffy coat into a sterile 2 mL Eppendorf tube and stored at -80°C. For CellSearch CTC enumeration the same volume of plasma was replaced with CellSearch Dilution buffer and again centrifuged at 800g for 10 minutes without using the brake. Finally, the sample was placed on the CellTracks Autoprep for CTC analysis. Plasma from EDTA blood was removed immediately after sampling. Blood from EDTA collection tubes was transferred to 15 mL Falcon polypropylene tubes. Samples were centrifuged for 10 minutes at 200g at room temperature with both brake settings set to slow and followed by a second centrifugation step at 1,600g for 10 minutes. The upper plasma layer was transferred to a Falcon polypropylene tube, avoiding contact with the buffy coat layer and again centrifuged at 1,600g for 10 minutes. Afterwards, the supernatant was transferred to Eppendorf tubes without disturbing the cell pellet and stored at -80°C. Circulating DNA from EDTA tubes was extracted within 96 hours.

Plasma DNA extraction

A total of 97 patients were included in the study. Plasma from 31 patients was available from two different tubes: CellSave tubes and EDTA tubes. These paired samples were used to evaluate differences in ctDNA recovery. For the remaining 66 patients plasma was only available from either CellSave (n=23) or EDTA tubes (n=43), respectively. Taken together, plasma DNA was extracted from 128 plasma

samples using the QIAamp Circulating Nucleic Acid Kit (Qiagen) including EDTA plasma (n=74) (mean 1.0 mL, range 0.5-2.0 mL) or CellSave plasma (n=54) (mean 1.7 mL, range 0.8-2.0 mL) and eluted in 60 µL to 90 µL nuclease-free H2O, depending on the input volume of plasma. Plasma DNA was quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Waltham MA, USA).

Stratification of plasma DNA samples based on tumour fraction using mFAST-SeqS

Tumour fractions were assessed using the mFAST-SeqS assay, which is based on the selective amplification of uniquely mappable *LINE1* (*L1*) sequences and can be used as an overall measure of aneuploidy and therefore corresponds to the plasma tumour fraction. *L1* amplicon libraries were prepared as previously described [29]. Briefly, using target-specific *L1* primers, 5 µL plasma DNA was amplified with Phusion Hot Start II Polymerase for 8 PCR cycles. PCR products were purified with AMPure Beads (Beckman Coulter, Brea CA, USA), and 10 µL was directly used for a second PCR with 18 cycles to add Illumina specific adaptors and indices. *L1* amplicon libraries were pooled equimolarly and sequenced on an Illumina MiSeq generating 150 bp single reads aiming for at least 100,000 reads. Aligned sequence reads were counted and normalized using an in-house script. In order to assess over- and under-representation of read counts of each chromosome arm, a z-score statistic was applied by comparing read counts to a set of healthy individuals. In order to get a general overview of aneuploidy, a genome-wide z-score was calculated by normalizing read counts per chromosome arms and squaring and summing them up. Based on previous comparisons with genome-wide z-scores and mutant allele frequencies of somatic mutations, a z-score of 5 correlated with a tumour allele frequency of approximately 10% [29]. Plasma DNA-samples were stratified based on genome-wide z-scores with high tumour allele frequency (z-score ≥5) and low tumour allele frequency (z-score <5).

Tracking primary tumour mutations in of plasma DNA

In 16 patient samples, in which the mutational status of the primary tumour was available, mutations identified in *BRAF*, *EGFR*, *KRAS* and *NRAS* were tracked in plasma DNA using two deep sequencing approaches: conventional deep-Seq and Safe-SeqS. Validation with respect to analytical sensitivity was done with reference material harbouring pre-defined variant allele frequencies. To assess specificity we additionally sequenced a set of healthy control samples. Since the average error rates of the respective *EGFR* sequence regions were much lower compared to the other hotspots (e.g. 0.02 for *EGFR* mutations in codons T790 and L858 versus 1.36 for *KRAS* codons 12 and 13), we achieved a maximal sensitivity of 0.1% for conventional deep sequencing of *EGFR* without molecular barcodes for error correction. However, due to the error prone sequence context of *BRAF* and *KRAS,* a sensitivity of 0.1% was only

achieved when using Safe-SeqS, which employs molecular barcoding of individual DNA template strands to track all sequencing reads back to a single original templates and correct for PCR errors during library preparation after error correction. For Safe-SeqS, on average 11.7 ng plasma DNA (range 7.9-20.0 ng) was amplified by Phusion polymerase (Thermo Fisher) using amplicon specific primers whereby the sense primer contains a 12-base unique molecular identifier (UMI). After 12 cycles of PCR, products were purified using Ampure XP beads (Beckman Coulter) and eluted in nuclease-free H_2O. In a second PCR with 35 cycles, Illumina specific adapters and indices were added. Products were again purified and subjected to quality control and quantification on an Agilent Bioanalyzer DNA 7500 chip (Agilent Technologies). All samples were pooled equimolarly and sequenced on an Illumina MiSeq in a 2x 150 bp paired-end run. Generated reads were then grouped to read families according to the UMI. A consensus sequence of each read family and a FastQ-file from this sequence was generated and aligned to the human reference genome (hg19) using Burrows-Wheeler transformation, SAMtool and alignments visualized in the "Integrative Genomes Viewer" to detect variants. For Deep-Seq, on average 5.2 ng (range 3.3-9.6 ng) was used for a target-specific PCR and amplified in 25 cycles using FastStart HiFi Polymerase (5 U/µL) followed by a Ampure XB beads (Beckman Coulter) purification. Illumina specific adapters and indices were added in a second PCR for 25 cycles. Analysis was performed as described above but without collapsing the read to a consensus sequence.

EpCAM^{high} CTC detection by CellSearch

CTC were enumerated in aliquots of 7.5 mL of blood with CellSearch® Circulating Tumour Cell Kit (Menarini Silicon Biosystems). Blood samples were enriched for EpCAM^{high} cells and stained with DAPI, Cytokeratin-PE and CD45-APC on the CellTracks Autoprep. Image acquisition of the stained cartridges was performed on the CellTracks Analyzer II and all images were stored for review by an independent trained operator.

EpCAM^{low} CTC detection by filtration after CellSearch

After immunomagnetic selection of EpCAM^{high} cells, the CellTracks Autoprep transports the remaining blood sample to a waste container. These samples can be used for identification of residual tumour cells, as described previously [21]. In short, microsieves (VyCAP, Deventer, The Netherlands) were used to filter tumour cells from these samples, containing mostly leukocytes and EpCAM^{low} CTC. The microsieves contain 111,800 pores of 5 µm in diameter and are spaced 14 µm apart on a total surface area of 8 by 8 mm. After filtration, the microsieve was washed once with a permeabilization buffer containing PBS, 1% bovine serum albumin (Sigma-Aldrich, St. Louis MO, USA) and 0.15% saponin (Sigma-Aldrich) and was

incubated in this buffer for 15 min at room temperature. Subsequently, a cocktail of fluorescently labelled antibodies was used to stain the cells on the sieve for 15 min at 37°C. The staining solution consisted of the following monoclonal antibodies: three CK antibody clones targeting CK 4, 5, 6, 8, 10, 13, 18 (clone C11) conjugated to PE (not commercialized), CK 1-8 (clone AE3) and CK 10, 14, 15, 16 and 19 (clone AE1), both conjugated to eFluor570 (ThermoFisher Scientific, Waltham, MA, USA), and one antibody targeting CD45 (clone HI30) labelled with PerCP (Themo Fisher Scientific). After removal of the staining cocktail, the microsieve was washed once and then incubated for 5 min at room temperature with PBS/1%BSA and fixed using PBS with 1% formaldehyde (Sigma-Aldrich) for 10 min at room temperature. Removal of the fluid during each of the staining and washing steps was performed by bringing the bottom of the microsieve in contact with an absorbing material using a staining holder (VyCAP). The microsieve was subsequently covered with ProLong® Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific). A custom cut glass coverslip of 0.85 by 0.85 cm^2 (Menzel-Gläser, Saarbrükener, Germany) was placed on both sides of the microsieve for immediate analysis or stored at -30°C awaiting further analysis.

Images covering the entire 0.64 cm^2 surface of the microsieves were acquired on a Nikon fluorescence microscope equipped with computer controlled X, Y, Z stage, a 20X microscope objective with a NA of 0.45 and a LED as a light source. The following filters were used: DAPI (DAPI-50LP-A-NQF) with excitation 377/50 nm, dichroic 409 nm LP, emission 409 nm LP; PE (TRITC-B-NQF) with excitation 543/22 nm, dichroic 562 nm LP, emission 593/40 nm and PerCP (FF02-435/40, FF510-Di02 and FF01-676/29 (customized filter cube)) with excitation 435/40 nm, dichroic 510 nm LP, emission 676/29 nm. All cubes were acquired via Nikon (Semrock, Rochester, NY, USA).

Scoring of CTC by CellSearch and on microsieves

Analysis of the fluorescent images generated from the CellSearch cartridges was performed according the instructions of the manufacturer. Images of $EpCAM^{high}$ CTC candidates were identified by the CellTracks Analyzer II and presented to an operator for CTC classification. Cell candidates were assigned as “CTC” when the objects were larger than 4 µm in diameter, stained with DAPI and CK, lacked CD45 staining and had morphological features consistent with that of a cell [18]. The fluorescent images from the microsieves were analysed for identification of $EpCAM^{low}$ CTC using a plugin for the open-source software ICY [31]. Operators were asked to annotate every DAPI+/CK+/CD45– event and classify the event as a CTC when morphological features were consistent with that of a cell.

Analysis of tdEV with ACCEPT

The CellTracks images from every cartridge were analysed with the open source image analysis program ACCEPT (www.github.com/LeonieZ/ACCEPT) [27,28,32]. The ACCEPT toolbox detects all events present in the images by an advanced multi-scale segmentation approach and extracts several fluorescence intensities and shape measurements for every event it has found. The tdEV identified here are relatively large as they have been pelleted with the blood cell fraction after centrifugation at 800g. The selection criteria used for tdEV were: CK mean intensity ≥60, CK maximum intensity ≥90, CK standard deviation of intensity ≥0.15, CK size <150 μm^2, CK perimeter ≥3.2 μm (≥5 pixels), CK roundness <0.80 (where 0 is perfectly round and 1 is a perfect line), CK perimeter to area <1.1, DNA mean intensity <5, CD45 mean intensity <5. Objects that fit the selected definition are depicted in blue, whereas all other objects present in the cartridge are depicted in grey. Scatter plots of all parameters for tdEV are presented in Supplementary Figure 1.

Statistical analysis

Patient variables and $EpCAM^{high}$ CTC, $EpCAM^{low}$ CTC, tdEV and ctDNA data were gathered in an independent way and blindly merged into one data set. For $EpCAM^{high}$ and $EpCAM^{low}$ CTC a cut-off of 2 CTC was used as threshold [18,33]. For tdEV the cut-off threshold was set at 18 (see results). A previously established cut-off of 5 was used for the genome-wide mFAST-SeqS z-score to estimate high versus low tumour allele frequency [29]. To determine associations between biomarkers the non-parametric Spearman's Rho correlation coefficient was used. Kaplan-Meier curves for overall survival (OS) were constructed and differences between groups were tested by the log-rank test. OS was defined from the first diagnosis to death or loss of follow-up. Subsequently, a multivariable Cox regression analysis was used to evaluate the discriminative power of favourable (above cut-off threshold) versus unfavourable (below cut-off threshold) biomarkers. To analyse the added predictive value of the biomarkers a multivariable analysis of changes in concordance index (C-index) was used [34]. Statistical analysis was performed in SPSS (version 24, SPSS Inc., Chicago IL, USA) and R (R Foundation, Vienna, Austria). A nominal p-value <0.05 was considered to be significant.

Results

Patients and healthy donors

In this study, 97 patients with advanced NSCLC, median age of 65 years, with 91% ECOG performance score 0-1 and 20% non-smokers were included (Table 1). No difference in survival was found between the two locations; the median survival of IOV was 10.2 months and the median survival of UMCG was

9.9 months (p=0.693). Healthy donors (n=35) aged 20–55 years and without prior history of cancer or blood transmittable disease were used as controls.

Classification of EpCAM^{high} CTC, EpCAM^{low} CTC and tdEV

EpCAM^{high} CTC were identified in the thumbnail images presented to the operator by the CellTracks Analyzer II. The EpCAM^{low} CTC were manually identified in the images scanned from the microsieves with the open source imaging ICY software. Three typical images of EpCAM^{high} CTC (panel A), EpCAM^{low} CTC (panel B) and tdEV (panel C) are displayed using the ACCEPT software in Figure 1. In this patient sample a total of 40,094 events were detected by ACCEPT. After application of the criteria for an event to be assigned as a tdEV, in total 113 objects were identified as a tdEV (panel D).

Presence of CK positive EpCAM^{high} and EpCAM^{low} cells and tdEV in healthy donors

In order to assess the specificity of our classification system, blood of 35 healthy donor samples were processed for the detection of CK positive EpCAM^{high} and EpCAM^{low} cells and tdEV. While only one EpCAM^{high} cell was found in a single control sample (2.9%), EpCAM^{low} cells were detected in five of the 35 samples (14.3%) (mean 0.2, ±0.5 SD, range 0-2), and of these samples, two samples (5.7%) were on the ≥2 CTC threshold. The mean tdEV count in these samples was 5.1 (median 3, range 0-36) with a standard deviation (SD) of 6.7. For tdEV, the cut-off threshold was established at 18; based on the mean tdEV plus two standard deviations (5.1 + 2*(6.7) = 18.4). One sample was above the ≥18 tdEV threshold.

Presence of EpCAM^{high} CTC, EpCAM^{low} CTC, tdEV and ctDNA in NSCLC patients

In 20 patients (21%) ≥2 EpCAM^{high} CTC were detected, in 15 patients (15%) ≥2 EpCAM^{low} CTC, in 29 patients (30%) ≥18 tdEV and 18 patients (19%) showed high (>10%) tumour allele frequency with genome-wide mFAST-SeqS. EpCAM^{high} CTC, tdEV and ctDNA were significantly correlated with each other, but not with EpCAM^{low} cells. The frequency distribution is illustrated in Figure 2 and in more detail in Supplementary Table S1. ctDNA fraction was determined in the plasma from either the CellSave tube (n=23) used for CTC enumeration or from an additional EDTA tube (n=74). Concentration of plasma total DNA ranged from 11.9 to 407 ng per mL plasma (mean 69.87 ng/mL) for CellSave tubes and 4.6 to 780.3 ng per mL plasma (mean 98.7 ng/mL) for EDTA tubes. To determine if CellSave plasma yields different ctDNA levels as conventional EDTA plasma, we evaluated the concordance of ctDNA fractions from 31 patients, of which both tubes were available. Due to the fact that z-scores below 3 cannot be used as quantitative measures, only a moderate but significant correlation (r=0.493, p=0.005) was observed when all samples were considered. However, after assigning samples into various z-score categories

(low: 0-5, elevated: 5-10, and high: 10-50), all samples fell into the same category indicating a high consistency between the two tubes.

High resolution analysis of tumour mutation in plasma

Molecular analysis of the primary tumour was performed in 46 of the 97 (47%) of the patients; however, only in 16 patients a specific mutation was detected. Due to the limited sensitivity of mFAST-SeqS and to test whether high-resolution sequencing methods can also be applied to CellSave plasma we searched for these mutation in plasma. We were able to tracked tumour-specific mutations with ultra-deep sequencing in 11 of these 16 patients (69%). Consistent with the low mFASt-SeqS z-scores for these patients (range 0.47-3.99) the detected variant allele frequencies (VAF) detected were also low ranging from 0.1-5.3%. Detailed information on the mutations and their VAFs, as wells as CTC and tdEV counts, for these patients are shown in Table 2.

Single blood biomarkers and overall survival of NSCLC patients

To study the discriminative value of the biomarkers NSCLC patients were stratified in those with favourable and unfavourable biomarker status according to the threshold cut-off values (Figure 3). $EpCAM^{high}$ CTC was associated with prolonged overall survival (HR 2.1, 95% CI 1.2-3.7; p=0.014) with a median OS of 4.2 months (range 1-21) for the unfavourable group (≥2 CTC) and 12.2 months (range 1-30) for the favourable group (<2 CTC) (panel A). Secondly, tdEV was associated with overall survival (HR 2.0, 95% CI 1.2-3.5; p=0.014) with a median OS of 4.2 months (range 1-19) for the unfavourable group (≥18) versus 12.2 months (range 1-30) for the favourable group (<18) (panel C). Thirdly, ctDNA was associated with overall survival (HR 1.9, 95% CI 1.1-3.4; p=0.032) with a median OS of 5.2 months (range 1-26) for the unfavourable group with high tumour allele frequency (≥10%) versus 11.5 months (range 1-30) for the favourable group (tumour allele frequency <10%) (panel D). However, the presence of $EpCAM^{low}$ CTC did not associate with overall survival (HR 1.2, 95% CI 0.6-2.3, p=0.579) with a median OS of 6.8 months (range 1-30) for the unfavourable group (≥2 CTC) versus 11.0 months (range 1-29) for the favourable group (<2 CTC) (panel B). To study the predictive ability of the four biomarkers, the concordance index was calculated. For $EpCAM^{high}$ CTC the C-index was 0.561, for $EpCAM^{low}$ CTC 0.512, for tdEV 0.565 and for ctDNA 0.551 (Table 3).

Comprehensive multi-parameter blood biomarker

The significant biomarkers from the univariate analysis were $EpCAM^{high}$ CTC, tdEV and ctDNA, but not $EpCAM^{low}$ CTC (Supplementary Table S2). The values of $EpCAM^{high}$ CTC were correlated with those of

tdEV (0.66), ctDNA (0.35) but not with EpCAM^{low} CTC (0.08). The correlation between tdEV and ctDNA was low (0.25). EpCAM^{low} CTC was not correlated with tdEV (0.05) or ctDNA (-0.02). To study the discriminative power of the three significant biomarkers from the univariate regressions, all were simultaneously included as categorical values (favourable and unfavourable) in a multivariable Cox proportional regression model. In the model none versus one unfavourable biomarker was not significantly different from each other (HR 1.0, 95% CI 0.4-2.1, p=0.909), whereas two (HR 2.3, 95% CI 1.0-5.0, p=0.038) or all three (HR 2.9, 95% CI 1.4-6.0, p=0.005) unfavourable biomarkers were significantly different compared to none unfavourable biomarkers (panel A in Figure 4). Therefore, we stratified the patients based on the presence of none and one unfavourable biomarker versus two and three unfavourable biomarkers and determined the OS of these two groups (panel B). The patients with none and one unfavourable biomarker had a median OS of 12 months (range 1-30) versus 6 months (range 1-19) for the patients with two and three unfavourable biomarkers (HR 2.6, 95% CI 1.5-4.6, p=0.001). The predictive ability of all three biomarkers in the multivariable C-index model provides a significant contribution of 0.575 (p=0.047), but the biomarkers themselves become non-significant. When each of the biomarkers were taken of the model, the drop in C-index was extremely small (Table 3). Moreover, the effect size of each biomarker in the combined model became smaller and non-significant (EpCAM^{high} CTC HR 1.4; tdEV HR 1.5 and ctDNA HR 1.5).

Discussion

Blood may contain different tumour derived cells, vesicles and DNA molecules that offer a simple, patient friendly approach, to study clone diversity that may be most relevant to determine treatment options for patients with advanced NSCLC. In the CANCER-ID consortium (www.cancer-id.eu) CTC and tumour related nucleic acids in blood are being extensively explored for their potential to serve as a predictive or prognostic factor in advanced NSCLC. In this study, members of this consortium explored CTC, tumour derived extra cellular vesicles (tdEV) and plasma nucleic analysis of 97 NSCLC patients whether a combined analysis of multiple liquid biopsy components is feasible on the same tube of blood and what information could be obtained from such analysis. From the blood collected in 10 mL CellSave tubes for subsequent CTC analysis, on average 1.7 mL (0.8-2.0 mL) of plasma could be harvested and stored for ctDNA analysis before processing the samples on the CellSearch system. Comparison of the CellSave plasma for ctDNA analysis with plasma from EDTA blood showed a strong correlation, indicating that all tests can be reliably obtained from CellSave tubes, thereby facilitating a single tube liquid biopsy approach.

For CTC analysis, the FDA cleared CellSearch system was used and in 21% of the advanced NSCLC patients CTC were detected. These CTC are referred here as EpCAM^{high} CTC and their presence has been reported to be associated with poor survival which was confirmed in this study [14,21]. We also confirmed previous findings that the presence of EpCAM^{low} CTC, present in 15% of the patients, after filtration of the EpCAM depleted blood was not significantly associated with survival [21]. This may question the cancerous origin of the EpCAM^{low} CTC. For single cell analysis of these EpCAM^{low} CTC, technology will need to be developed that can determine the molecular composition of these CTC on the microsieves.

In metastatic prostate cancer, objects smaller than cells expressing cytokeratin and lacking CD45 in the EpCAM enriched cell suspensions were associated with poor survival, similarly with CTC in these patients [26]. Using the recently introduced open source imaging program ACCEPT, the identification of these objects in CellSearch image sets was automated and in our NSCLC cohort these objects were identified with a relatively high density, with elevated levels in 27% of the patients. Moreover, patients with elevated tdEV numbers showed significantly worse survival, confirming the earlier observations of a strong relation between poor outcome and the presence of tdEV [35].

For ctDNA analysis we determined the tumour allele frequency in plasma DNA using mFAST-SeqS assay, which measures the aneuploidy fraction of circulating DNA [29]. It is of note that mFAST-SeqS has a limited analytical sensitivity and a correlation of z-scores with tumour allele frequency can only be provided in patients with high tumour allele frequency (≥10%). mFAST-SeqS z-scores in the lower range are not informative, but indicate a low tumour DNA content. Nevertheless, the intent of this study was not an absolute quantification of tumour-derived fragments in plasma but rather the assessment of a fast and cost-effective method to stratify patients into groups of high and low tumour allele frequencies and to combine these data with other liquid biopsy components. Patients with high tumour allele frequencies (19%) had significantly poor survival, similar to elevated EpCAM^{high} CTC and tdEV.

Blood derived tumour markers EpCAM^{high} CTC, tdEV and ctDNA in advanced NSCLC were each associated with poor survival. Two or three unfavourable biomarkers – all shedding from the tumour – discriminates poor prognosis better than one biomarker, but the predictive contribution of each biomarker is small, as was shown by the drop in C-index after removing each biomarker from the model. We questioned whether there is still room for their own contribution to survival since serious collinearity arises with higher correlations between biomarkers. In other words, these biomarkers come from the same underlying biological processes but still may have their own dynamics that may influence

survival. However, the lack of power – that is the low number of patients where all three biomarkers were present – prohibited significance for the predictive accuracy.

Although EpCAM^{high} CTC, tdEV and ctDNA may be useful to identify a subset of NSCLC patients with a relatively poor prognosis, it does not address the question whether information can be retrieved to predict whether patients are eligible for targeted therapy. Expression of mutated proteins, such as EGFR, can be assessed on CTC and tdEV, but only in those harbouring such mutations (i.e. ~20% of patients with advanced NSCLC). In this cohort of patients, mutations in the primary tumour were identified in 16 of the 46 NSCLC patients (34%), of whom molecular profiling of the tumor was performed. Based on mFAST-SeqS the ctDNA fractions of these patients were very low. The z-scores ranged from 0.47-3.99 and were therefore below the dynamic range of this method. However, specific mutations can easily be tracked in ctDNA with a much higher analytical sensitivity [36–38]. To test whether such ultra-deep sequencing methods can also be applied to CellSave plasma, we tracked tumour-specific mutations in these patients. In 11 of the 16 patients (69%) mutations could be identified in plasma with variant allele frequencies ranging from 0.1 to 5.3%, which is consistent with the respective low z-scores. Yet, our concordance rate was slightly lower than those reported for metastatic NSCLC patients that range between 74-85% [39]. However, given our small samples size and the low amount of DNA input (which might lead to sampling errors at low variant allele frequencies), our data might not be representative. Nevertheless, these data show that the high-resolution assessment of mutations might yield in increased ctDNA detection rates, and therefore improve patients stratification based on tumour fraction.

The true potential of a liquid biopsy lies in determining genetic alterations associated with therapy resistance or new mutations occurring during the course of the disease. A variety of different liquid biopsy approaches have been evaluated for their clinical potential in NSCLC. Due to the low efficiency to retrieve high CTC numbers (30% of the patients have 1 or more CTC; 8% with >5 CTC per 7.5 mL blood) and elevated plasma DNA tumour fractions all fall short when it comes to a broad patient coverage [33,40–42]. In this study we detected one of the biomarkers in 45% of the patients, whereas each individual biomarker was detected in 15-27% of the patients. A potential solution to increase these percentages even further is to increase the blood volume that can be analysed which can be obtained through a diagnostic leukapheresis [43]. Studies in advanced NSCLC are currently being conducted in the CANCER-ID consortium to evaluate whether this approach can yield sufficient number of CTC or ctDNA to yield a liquid biopsy for the majority of NSCLC patients [44].

Taken together, here we report for the first time a single tube approach enabling a simultaneous analysis of EpCAM^{high} CTC, EpCAM^{low} CTC, tdEV and ctDNA. Except for EpCAM^{low} CTC, the presence of each component was associated with a poor clinical outcome in advanced NSCLC patients. Two or more biomarkers discriminated an unfavourable subgroup of advanced NSCLC.

Additional information

Conflict of interest: The authors declare no conflict of interest.

Funding: This study was supported by the EU FP7 # 305341 "CTC-Trap", the EU IMI # 115749-1 "CANCER-ID" and the Italian Ministry of Health, Proposal No: # GR-2010-2303193A, "Individualized treatments of patients with advanced NSCLC: potential application for CTCs molecular and phenotypical profiling" (PI: E.R.).

Authors' contributions: Study design: L.W.M.M.T. and H.J.M.G.; Preparation, data collection and writing of the manuscript: S.dW., E.H., L.W.M.M.T. and H.J.M.G.; Patient recruitment and clinical data: E.R., M.T., E.S., R.Z., T.J.N.H. and H.J.M.G.; Performing experiments CTC: S.dW., E.R., M.M., J.F.S., R.V., and A.F.; Image analysis ACCEPT: S.dW. and L.L.Z.; Study design and performing experiments ctDNA: S.W., E.H. and M.R.S.; Statistical analysis; S.dW. and C.G.M.G-O; Visualization: S.dW. All authors reviewed and approved the manuscript.

Ethics approval and consent to participate: The study protocol was approved by the ethics committee at each centre, and written informed consent was obtained for all patients. Healthy donors - used as controls - provided informed consent prior to blood donation, in accordance to the study protocol approved by the METC Twente ethics committee.

Availability of data and material: Data is available upon request at the corresponding author.

References

1. Tang Y, Qiao G, Xu E, Xuan Y, Liao M, Yin G. Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer. *Onco Targets Ther* 2017;Volume 10:4527–34.

2. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of

cancer. *Nat Rev Clin Oncol* 2017;14:531–48.

3. Heitzer E, Auer M, Ulz P, Geigl JB, Speicher MR. Circulating tumor cells and DNA as liquid biopsies. *Genome Med* 2013;5:73.

4. Gallo M, De Luca A, Maiello MR, D'Alessio A, Esposito C, Chicchinelli N, Forgione L, Piccirillo MC, Rocco G, Morabito A, Botti G, Normanno N. Clinical utility of circulating tumor cells in patients with non-small-cell lung cancer. *Transl lung cancer Res* 2017;6:486–98.

5. Singh AP, Cheng H, Guo X, Levy B, Halmos B. Circulating Tumor DNA in Non–Small-Cell Lung Cancer: A Primer for the Clinician. *JCO Precis Oncolocy* 2017;

6. Reclusa P, Taverna S, Pucci M, Durendez E, Calabuig S, Manca P, Serrano MJ, Sober L, Pauwels P, Russo A, Rolfo C. Exosomes as diagnostic and predictive biomarkers in lung cancer. *J Thorac Dis* 2017;9:S1373–82.

7. Cui S, Cheng Z, Qin W, Jiang L. Exosomes as a liquid biopsy for lung cancer. *Lung Cancer* 2018;116:46–54.

8. Cristofanilli M, Budd G, Ellis M, Stopeck A, Matera J, Miller M, Reuben J, Doyle G, Allard W, Terstappen L, Hayes D. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *Engl J* 2004;351:781–91.

9. Janni WJ, Rack B, Terstappen LWMM, Pierga J-Y, Taran F-A, Fehm T, Hall C, de Groot MR, Bidard F-C, Friedl TWP, Fasching PA, Brucker SY, et al. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clin Cancer Res* 2016;22:2583–93.

10. de Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle G V, Terstappen LWWM, Pienta KJ, Raghavan D. Circulating Tumor Cells Predict Survival Benefit from Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clin Cancer Res* 2008;14:6302–9.

11. Cohen SJ, Punt CJA, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, Picus J, Morse M, Mitchell E, Miller MC, Doyle G V, Tissing H, et al. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients With Metastatic Colorectal Cancer. *J Clin Oncol* 2008;26:3213–21.

12. van Dalum G, Stam G-J, Scholten LFA, Mastboom WJB, Vermes I, Tibbe AGJ, De Groot MR, Terstappen LWMM. Importance of circulating tumor cells in newly diagnosed colorectal cancer. *Int J Oncol* 2015;46:1361–8.

13. Hiltermann TJN, Pore MM, van den Berg A, Timens W, Boezen HM, Liesker JJW, Schouwink JH, Wijnands WJ a, Kerner GSM a, Kruyt F a E, Tissing H, Tibbe a GJ, et al. Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor. *Ann Oncol* 2012;23:2937–42.

14. Krebs MG, Sloane R, Priest L, Lancashire L, Hou J-M, Greystoke A, Ward TH, Ferraldeschi R, Hughes A, Clack G, Ranson M, Dive C, et al. Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J Clin Oncol* 2011;29:1556–63.

15. de Wit S, van Dalum G, Terstappen LWMM. Detection of Circulating Tumor Cells. *Scientifica (Cairo)* 2014;2014:1–11.

16. Andree KC, van Dalum G, Terstappen LWMM. Challenges in circulating tumor cell detection by the CellSearch system. *Mol Oncol* 2016;10:395–407.

17. Rao CG, Chianese D, Doyle G V, Miller MC, Russell T, Sanders RA, Terstappen LWMM. Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. *Int J Oncol* 2005;27:49–57.

18. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AGJ, Uhr JW, Terstappen LWMM. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients With Nonmalignant Diseases. *Clin Cancer Res* 2004;10:6897–904.

19. Lustberg MB, Balasubramanian P, Miller B, Garcia-Villa A, Deighan C, Wu Y, Carothers S, Berger M, Ramaswamy B, Macrae ER, Wesolowski R, Layman RM, et al. Heterogeneous atypical cell populations are present in blood of metastatic breast cancer patients. *Breast Cancer Res* 2014;16:R23.

20. Armstrong AJ, Marengo MS, Oltean S, Kemeny G, Bitting RL, Turnbull JD, Herold CI, Marcom PK, George DJ, Garcia-Blanco MA. Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers. *Mol Cancer Res* 2011;9:997–1007.

21. de Wit S, van Dalum G, Lenferink ATM, Tibbe AGJ, Hiltermann TJN, Groen HJM, van Rijn CJM, Terstappen LWMM. The detection of EpCAM+ and EpCAM– circulating tumor cells. *Sci Rep* 2015;5:12270.

22. Francart M-E, Lambert J, Vanwynsberghe AM, Thompson EW, Bourcy M, Polette M, Gilles C. Epithelial-mesenchymal plasticity and circulating tumor cells: Travel companions to metastases.

Dev Dyn 2017;

23. Alix-Panabières C, Mader S, Pantel K. Epithelial-mesenchymal plasticity in circulating tumor cells. *J Mol Med* 2017;95:133–42.

24. van der Pol E, Boing AN, Harrison P, Sturk A, Nieuwland R. Classification, Functions, and Clinical Relevance of Extracellular Vesicles. *Pharmacol Rev* 2012;64:676–705.

25. Azmi AS, Bao B, Sarkar FH. Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review. *Cancer Metastasis Rev* 2013;32:623–42.

26. Coumans FAW, Doggen CJM, Attard G, de Bono JS, Terstappen LWMM. All circulating EpCAM+CK+CD45- objects predict overall survival in castration-resistant prostate cancer. *Ann Oncol* 2010;21:1851–7.

27. de Wit S, Zeune L, Hiltermann T, Groen H, Dalum G, Terstappen L, de Wit S, Zeune LL, Hiltermann TJN, Groen HJM, Dalum G van, Terstappen LWMM. Classification of Cells in CTC-Enriched Samples by Advanced Image Analysis. *Cancers (Basel)* 2018;10:377.

28. Zeune LL, van Dalum G, Terstappen LWMM, van Gils SA, Brune C. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM J Imaging Sci* 2017;10:111–46.

29. Belic J, Koch M, Ulz P, Auer M, Gerhalter T, Mohan S, Fischereder K, Petru E, Bauernhofer T, Geigl JB, Speicher MR, Heitzer E. Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clin Chem* 2015;61:838–49.

30. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci* 2011;108:9530–5.

31. ICY [Internet]. Available from: http://icy.bioimageanalysis.org/

32. Zeune LL, van Dalum G, Decraene C, Proudhon C, Fehm T, Neubauer H, Rack B, Alunni-Fabbroni M, Terstappen LWMM, van Gils SA, Brune C. Quantifying HER-2 expression on circulating tumor cells by ACCEPT. *PLoS One* 2017;12:e0186562.

33. Krebs MG, Hou J-M, Sloane R, Lancashire L, Priest L, Nonaka D, Ward TH, Backen A, Clack G, Hughes A, Ranson M, Blackhall FH, et al. Analysis of circulating tumor cells in patients with non-small cell lung cancer using epithelial marker-dependent and -independent approaches. *J Thorac Oncol* 2012;7:306–15.

34. Kattan MW. Evaluating a new marker's predictive contribution. *Clin Cancer Res* 2004;10:822–4.

35. Nanou A, Coumans FAW, van Dalum G, Zeune LL, Dolling D, Onstenk W, Crespo M, Fontes MS, Rescigno P, Fowler G, Flohr P, Brune C, et al. Circulating tumor cells, tumor-derived extracellular vesicles and plasma cytokeratins in castration-resistant prostate cancer patients. *Oncotarget* 2018;9:19283–93.

36. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, Dawson S-J, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, et al. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Sci Transl Med* 2012;4:136ra68-136ra68.

37. Dietz S, Schirmer U, Mercé C, von Bubnoff N, Dahl E, Meister M, Muley T, Thomas M, Sültmann H. 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.

38. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, Stehr H, Liu CL, Bratman S V, Say C, Zhou L, Carter JN, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* 2016;34:547–55.

39. Yang N, Li Y, Liu Z, Qin H, Du D, Cao X, Cao X, Li J, Li D, Jiang B, Duan L, Yang H, et al. The characteristics of ctDNA reveal the high complexity in matching the corresponding tumor tissues. *BMC Cancer* 2018;18:319.

40. Hanssen A, Wagner J, Gorges TM, Taenzer A, Uzunoglu FG, Driemel C, Stoecklein NH, Knoefel WT, Angenendt S, Hauch S, Atanackovic D, Loges S, et al. Characterization of different CTC subpopulations in non-small cell lung cancer. *Sci Rep* 2016;6:28010.

41. Tong B, Xu Y, Zhao J, Chen M, Xing J, Zhong W, Wang M. Prognostic significance of circulating tumor cells in non-small cell lung cancer patients undergoing chemotherapy. *Oncotarget* 2017;8:86615–24.

42. Pailler E, Oulhen M, Borget I, Remon J, Ross K, Auger N, Billiot F, Ngo Camus M, Commo F, Lindsay CR, Planchard D, Soria J-C, et al. Circulating Tumor Cells with Aberrant *ALK* Copy Number Predict Progression-Free Survival during Crizotinib Treatment in *ALK* -Rearranged Non–Small Cell Lung Cancer Patients. *Cancer Res* 2017;77:2222–30.

43. Stoecklein NH, Fischer JC, Niederacher D, Terstappen LWMM. Challenges for CTC-based liquid

biopsies: low CTC frequency and diagnostic leukapheresis as a potential solution. *Expert Rev Mol Diagn* 2016;16:147–64.

44. Andree KC, Tamminga M, Mentink A, et al. Circulating tumor cells in the peripheral blood and leukapheresis product of non-small cell lung cancer patients. In: American Association for Cancer Research. 2018.

Figures and tables

Figure 1 – Gallery of CTC and tdEV

Thumbnail gallery of EpCAM^{high} CTC from CellSearch (A), EpCAM^{low} CTC from microsieves (B) and EpCAM^{high} tdEV from CellSearch (C), showing fluorescent signal for DAPI and/or CK (red circle drawn by the ACCEPT software). The scale bar in the overlay thumbnails is 6.4 µm. Panel D shows a scatter plot of every object present in the cartridge for characteristics in size (y-axis) and fluorescent mean intensity (x-axis). The tdEV are identified with a multi parameter gate and are visualized as blue dots. The remaining objects that do not fit the multi parameter gate are visualized as grey dots.

Figure 2 – Frequency distribution of all biomarkers in NSCLC patients

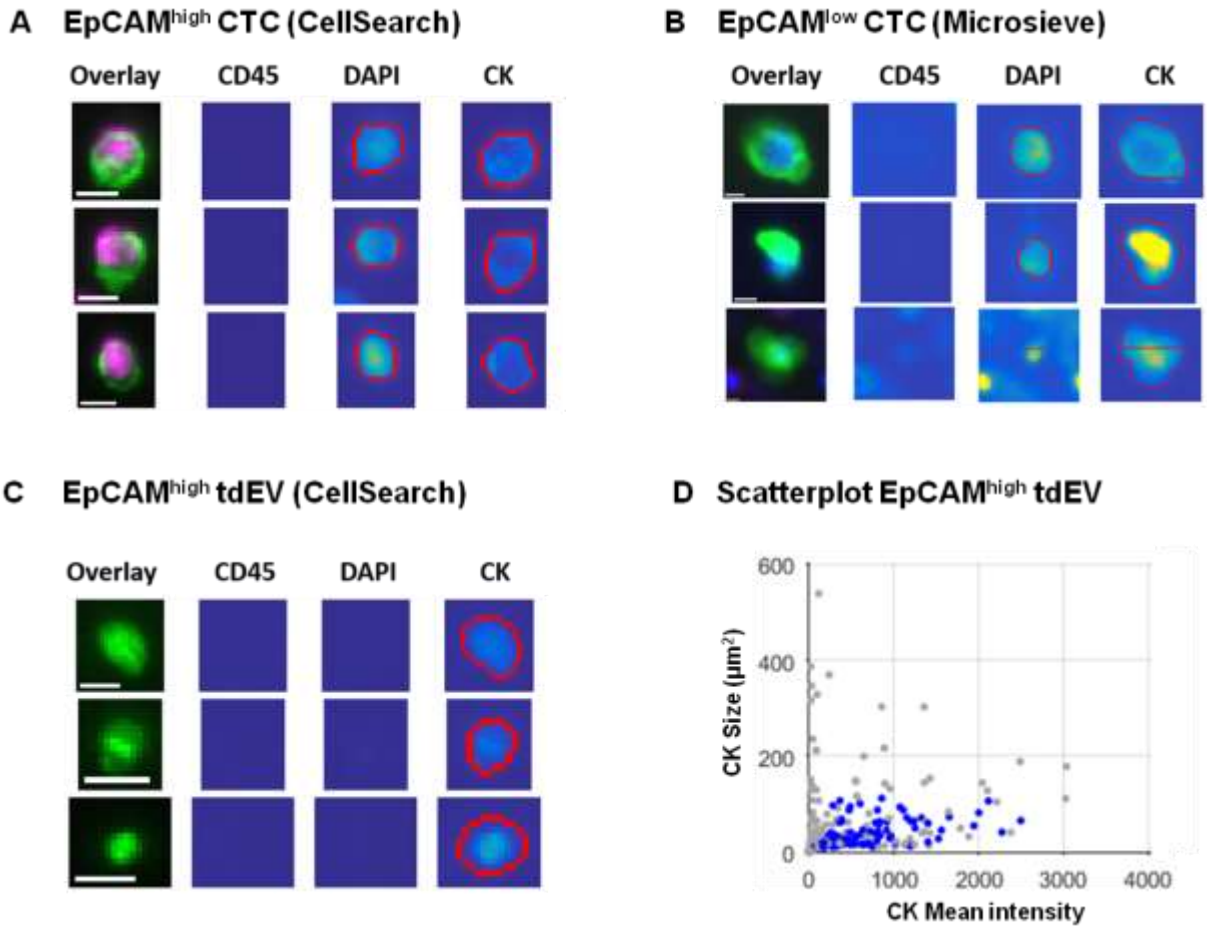
Frequency distribution of EpCAM^{high} CTC, EpCAM^{low} CTC, tdEV and z-score for ctDNA of 97 NSCLC patients. Percentages displayed above the black bar represent the patients that score above the threshold for that biomarker. Thresholds are: two CTC for EpCAM^{high} and EpCAM^{low}, 18 vesicles for tdEV and a z-score of 5 for ctDNA, representing approximately 10% mutant DNA alleles. z-score was determined in 74 samples from EDTA tubes (open circle) and in 23 samples from CellSave tubes (filled circle).

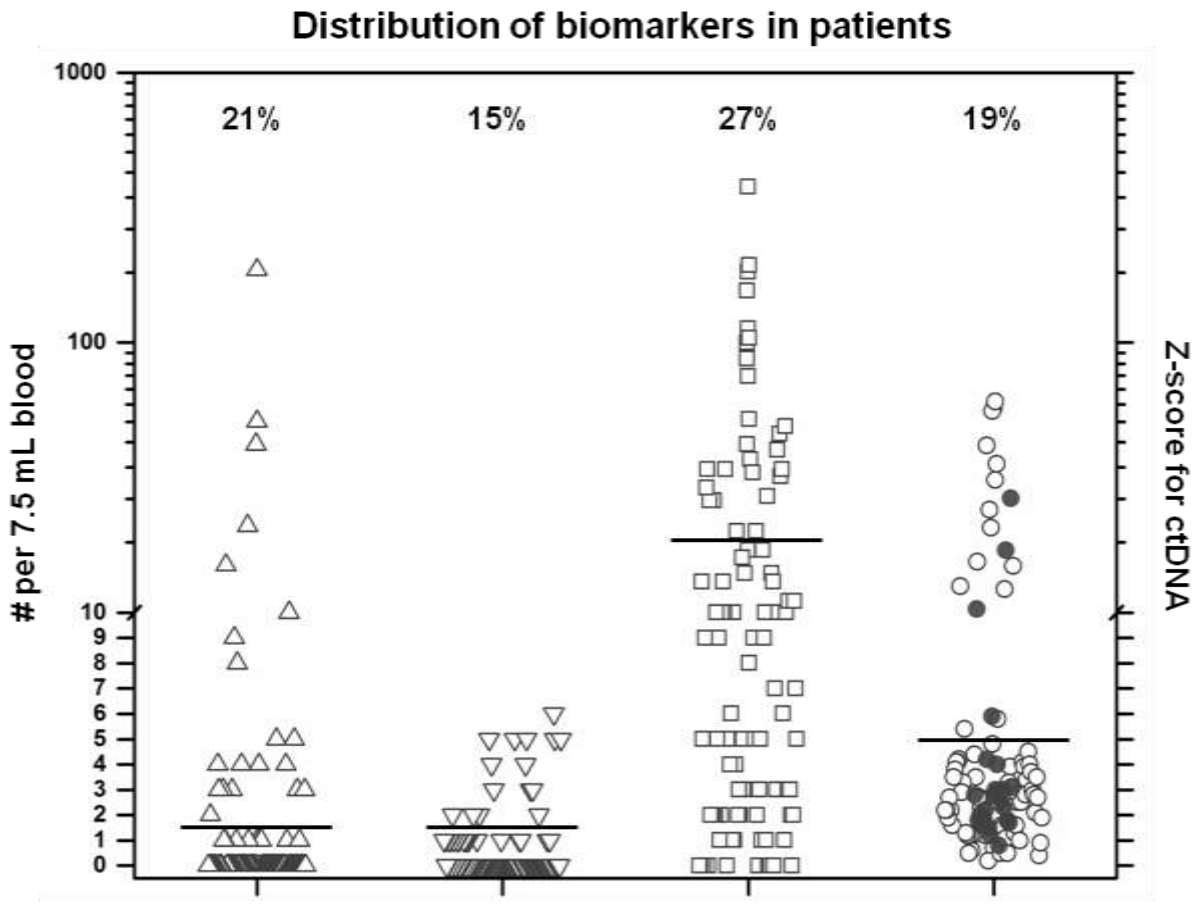
Figure 3 – Survival plots for each biomarker in NSCLC patients

Kaplan-Meier plots of probabilities of overall survival of 97 advanced NSCLC patients with favourable or unfavourable EpCAM^{high} CTC (A), EpCAM^{low} CTC (B), tdEV (C) and ctDNA (D). To separate between favourable and unfavourable groups, the threshold for CTC was 2, for tdEV 18, and for ctDNA 10% mutant alleles (z-score of 5).

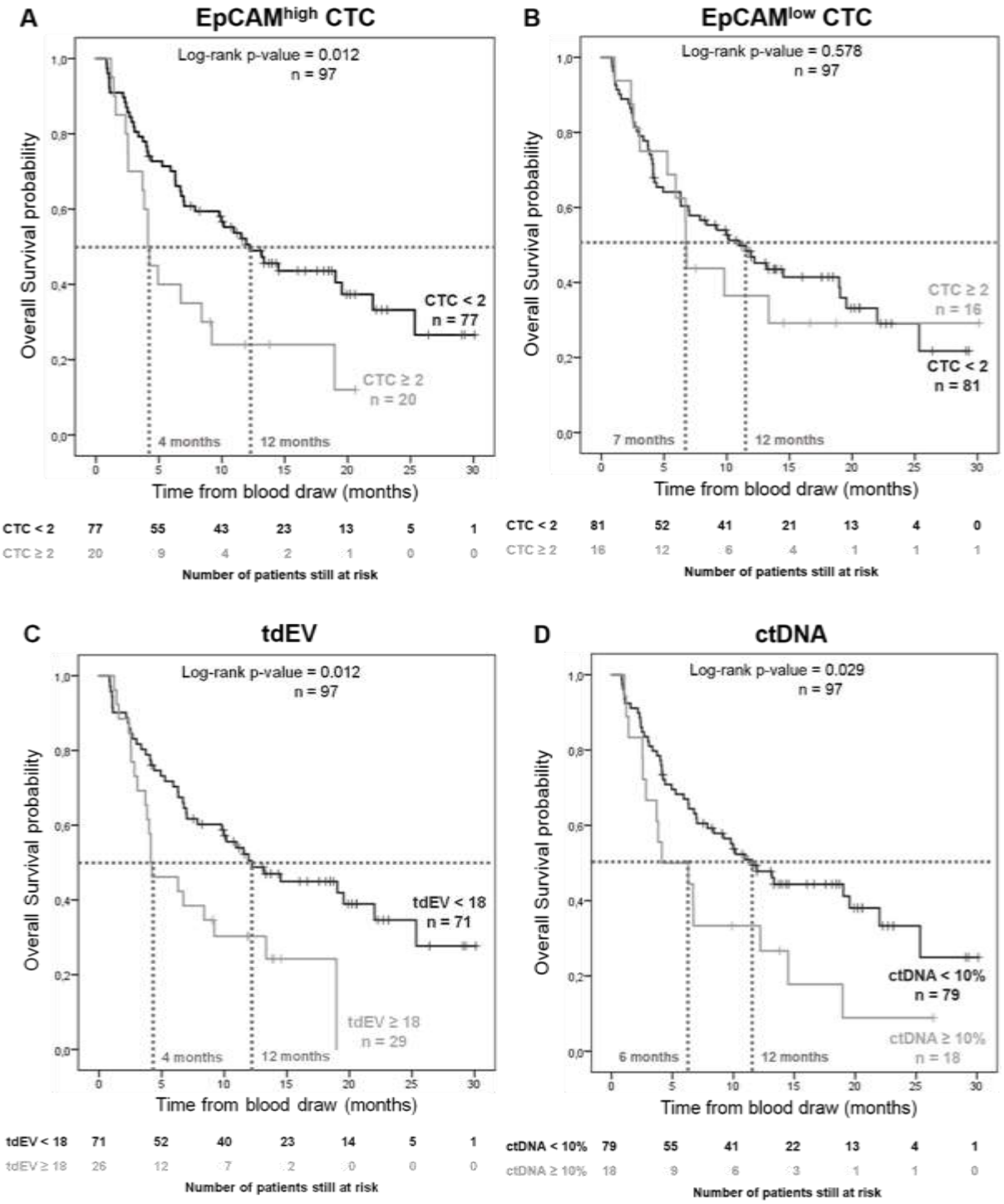
Figure 4 – Survival plots for grouped biomarkers in NSCLC patients

Kaplan Meier plot of probabilities of overall survival of 97 advanced NSCLC patients stratified for the amount of unfavourable biomarkers (EpCAM^{high} CTC ≥2, tdEV ≥18 and ctDNA ≥10%). Group 0 includes patients with no unfavourable biomarkers, group 1 with one unfavourable biomarker, group 2 with two unfavourable biomarkers, and group 3 with all unfavourable biomarkers (A). Two groups stratified for the presence of zero/one unfavourable biomarker and two/three unfavourable biomarkers (B).





	EpCAM^{high} CTC	EpCAM^{low} CTC	tdEV	ctDNA		
				EDTA	CellSave	all
Median	0	0	7	2.8	2.6	2.8
Mean	4	1	24	7.2	4.5	6.6
Min	0	0	0	0.2	0.8	0.2
Max	186	6	378	60.4	26.5	60.4



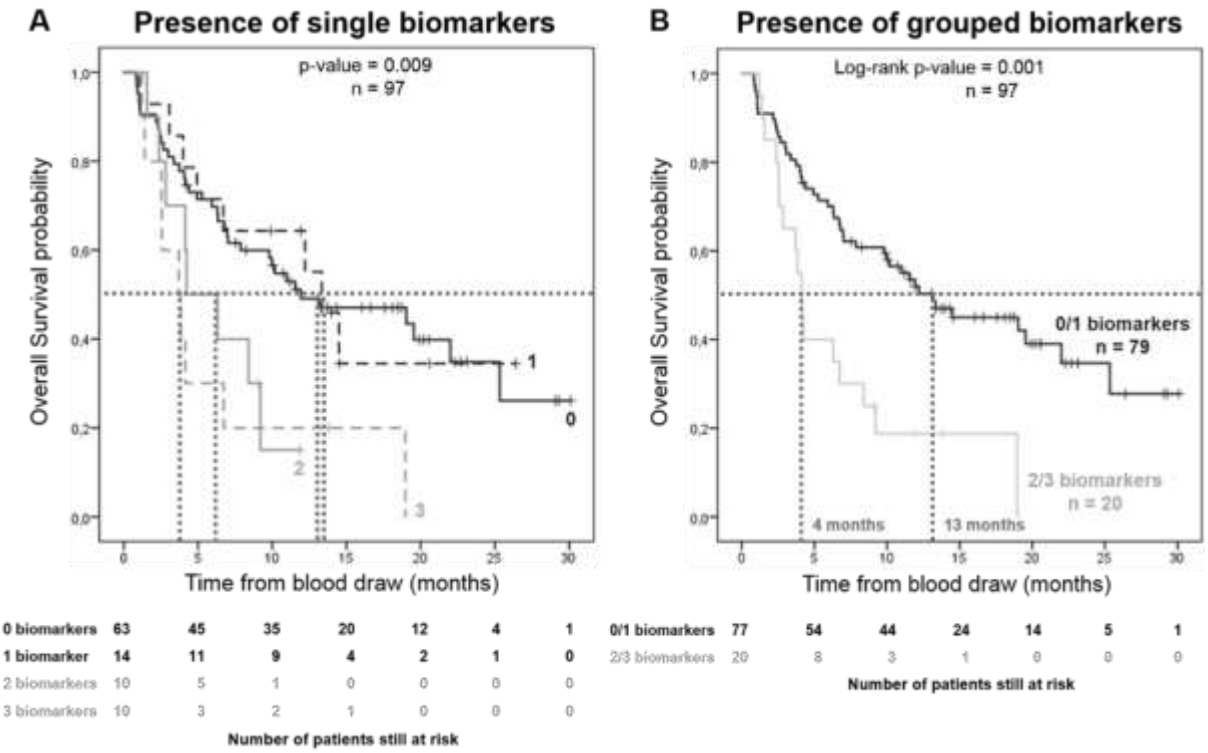


Table 1 – Patient demographics

Characteristics of the patients enrolled in this study.

Stage IIIB and IV non-small lung cancer patients	***n = 97***
Site location	
UMCG, the Netherlands	60 (62%)
Veneto Institute of Oncology IOV, Italy	37 (38%)
Age (years)	
Median (range)	65 (40-82)
Gender	
Male	47 (48%)
Female	50 (52%)
Smoking	
Never	19 (20%)
Smoker	59 (60%)
Unknown	19 (20%)
ECOG Performance Status	
0	54 (56%)
1	34 (35%)
2	6 (6%)

3	2 (2%)
4	1 (1%)
Therapy type	
Chemotherapy	41 (42%)
Targeted therapy	23 (23%)
Immunotherapy	24 (25%)
Unknown	10 (10%)
Cancer type	
Adenocarcinoma	59 (61%)
Squamous cell carcinoma	38 (39%)
Mean follow-up time in months *(min-max)*	
Alive	16 (4-30)
Dead	7 (1-25)
Status at last follow-up	
Alive	36 (37%)
Dead	61 (63%)

Table 2 – Tumour mutations located in 16 NSCLC patients

Primary tumour mutations and the shedding of CTC, tdEV and ctDNA in sixteen patients with adenocarcinoma of the lung.

Pt	Mutation primary tumor	EpCAM^{high} CTC	EpCAM^{low} CTC	tdEV	Z-score	Wild type reads	Mutated reads	VAF %*	Method
1	KRAS: c.37 G>T; p.G13C	3	0	52	0.91	16,577	476	2.79	SS
2	EGFR: p.L747_P753delinsS EGFR: c.2369 C>T; p.T790M	1	6	9	2.06	179,591	0	0.00	DS
3	KRAS: c.38 G>A; p.G13D	1	3	33	3.99	60,274	2,911	4.61	SS
4	KRAS: c.35 G>C; p.G12A**	0	5	5	1.95	48,546	151	0.31	SS
5	ALK: c.3616 T>G;p.S1206A	0	3	2	2.23	410,108	1,179	0.29	DS
6	EGFR: c.2315_2316insGTT; p.P772_H773insF	0	3	0	1.15	430,091	611	0.14	DS
7	BRAF: c.1406 G>T; p.G469V	0	1	17	0.49	45,137	459	1.01	SS
8	NRAS: c.182 A>G; p.Q61R	0	1	1	1.23	36,371	52	0.14	SS
9	KRAS: c.34 G>T; p.G12C EGFR: c.2305 G>T; p.V769L	0	1	0	2.18	203,64	177	0.09	SS
10	KRAS; c.34 G>T; p.G12C	0	0	14	2.96	166	0	0.00	SS
11	EGFR: c.2573 T>G; p.L858R EGFR: c.2369 C>T; p.T790M	0	0	14	0.57	37,493	0	0.00	SS
12	EGFR: c.2126 A>C; p.E709A EGFR: c.2156 G>C; p.G719A	0	0	10	2.00	619,082 620,329	1,489 474	0.24 0.10	DS
13	KRAS: c.183 A>C; p.Q61H	0	0	5	2.63	10,98	0	0.00	SS
14	EGFR: c.2236_2250del15; p.E746_A750del	0	0	5	0.54	514,839	29,039	5.34	DS
15	EGFR: c.2236_2250del15; p.E746_A750del	0	0	3	0.96	527,145	3,715	0.70	DS
16	EGFR: c.2240_2254del15; p.L747_T751delLREAT	0	0	2	0.47	503,499	5	0.00	DS

* VAF (%) indicates the percentage of variant allele frequency found among the wild type alleles in ctDNA; ** Additional KRAS mutation: c.35G>T; p.G12A was found with 2.50 VAF%; SS = Safe-SeqS; DS = Deep Sequencing.

Table 3 – C-index for all biomarkers

Concordance index for biomarkers calculated in a univariate and multivariate setting.

Biomarker (univariate)	C-index	Biomarkers (multivariate)	C-index
$EpCAM^{high}$ CTC	0.561	$EpCAM^{high}$ CTC & tdEV & ctDNA	0.575
$EpCAM^{low}$ CTC	0.512	$EpCAM^{high}$ CTC & tdEV	0.570
tdEV	0.565	$EpCAM^{high}$ CTC & ctDNA	0.575
ctDNA	0.551	tdEV & ctDNA	0.573