OPINION

Early stage NSCLC — challenges to implementing ctDNA-based screening and MRD detection

Christopher Abbosh, Nicolai J. Birkbak and Charles Swanton

Abstract | Circulating tumour DNA (ctDNA) refers to the fraction of cell-free DNA in a patient's blood that originates from a tumour. Advances in DNA sequencing technologies and our understanding of the molecular biology of tumours have resulted in increased interest in exploiting ctDNA as a tool to facilitate earlier detection of cancer and thereby improve therapeutic outcomes by enabling early intervention. ctDNA analysis might also have utility in the adjuvant therapeutic setting by enabling the identification of patients at a high risk of disease recurrence on the basis of the detection of post-surgical minimal (or molecular) residual disease (MRD). This approach could provide the capability to adapt clinical trials in the adjuvant setting in order to optimize risk stratification, and we argue that this objective is achievable with current technologies. Herein, we evaluate contemporary next-generation sequencing (NGS) approaches to ctDNA detection with a focus on non-small-cell lung cancer. We explain the technical and analytical challenges to low-frequency mutation detection using NGS-based ctDNA profiling and evaluate the feasibility of ctDNA profiling in both screening and MRD assessment contexts.

Strategies to improve cure rates in patients with lung cancer are an unmet medical need. Detection of non-small-cell lung cancer (NSCLC) at an early stage results in favourable survival outcomes. For example, on the basis of the eighth edition of the tumour, node, metastasis (TNM) classification of lung cancer by the International Association for the Study of Lung Cancer (IASLC) Lung Cancer Staging Project[1], clinical stage IA1 NSCLC (T1aN0M0) is associated with a 5-year overall survival of 92% compared with 0% for clinical stage IVB NSCLC (TxNxM1c). In the adjuvant setting (following surgical excision of NSCLC), platinum-based doublet chemotherapy modestly improves absolute 5-year overall survival of patients with NSCLC by ~5%[2], whereas in patients with metastatic NSCLC, chemotherapy results in a median survival benefit of only a small number of months[3]. This contrast possibly reflects an increased susceptibility of low-volume postoperative residual disease, rather than higher-burden metastatic disease, to systemic chemotherapy. These observations suggest that, in addition to preventive efforts, focus should be maintained on developing strategies for early cancer detection and enhancing adjuvant therapy to improve the outcomes of patients with NSCLC. Circulating tumour DNA (ctDNA) has demonstrated potential as a biomarker for the detection and localization of early stage cancers[4,5] and for the assessment of post-surgical minimal residual disease (MRD; also known as molecular residual disease)[6–11]. In patients with solid tumours, MRD describes a disease state after the initiation of radical therapy that is associated with clinically occult residual disease and inferior recurrence-free survival compared with that of patients without detectable residual disease[12,13]. Thus, MRD could be a useful parameter to stratify patients into high-risk cohorts, driving forward innovative adjuvant clinical trial designs predicated on small but more relevant populations of patients deemed to be at a high risk of tumour recurrence.

In this Perspectives article, we discuss contemporary next-generation sequencing (NGS) (BOX 1) approaches to ctDNA evaluation in the context of early stage NSCLC and post-surgical MRD. We provide an overview of the effects of artefactual and biological noise on the detection of low-frequency mutations in plasma cell-free DNA and summarize the technical and analytical approaches used to circumvent these issues. We highlight how low quantities of cell-free DNA in the plasma of patients with NSCLC can limit ctDNA detection and summarize data on the correlation between tumour burden and ctDNA mutant allele frequency (MAF) (BOX 1). Finally, we discuss how these challenges might affect the use of this technology in an early detection or post-surgical MRD setting and discuss the steps needed to move MRD profiling into the clinic.

ctDNA in early stage NSCLC and MRD

We compiled published data from studies of ctDNA in patients with early stage NSCLC[4,5,7,8] (Supplementary data set) encompassing four different NGS platforms (TABLE 1) and examined the MAFs of the most abundant mutations (maximum detectable MAFs; FIG. 1a) and the least abundant mutations (minimum detectable MAFs; FIG. 1b) reported as confidently detected in plasma of individual patients before definitive treatment. We also determined the proportion of patients who were analysed using each platform that had detectable ctDNA while their tumours remained in situ (FIG. 1c). The ctDNA analysis platforms used in these studies were cancer personalized profiling by deep sequencing (CAPP-Seq)[7], targeted error correction sequencing (TEC-Seq)[4], and the Tracking Cancer Evolution Through Therapy (TRACERx, Signatera)[8] and CancerSEEK[5] multiplex PCR (mPCR) assays (BOX 1; TABLE 1). These platforms are not directly comparable in terms of sensitivity and limit of detection parameters owing to the different contexts in which they were designed to operate. For example, TEC-Seq and CancerSEEK were designed as early

Box 1 | **Glossary**

Cancer personalized profiling by deep sequencing
(CAPP-Seq). CAPP-Seq describes a capture-hybridization next-generation sequencing (NGS) platform designed to target regions of the lung cancer genome that are recurrently mutated across different patients to maximize lung cancer mutation detection within minimal genomic space. Multiple CAPP-Seq panel iterations have been described, including a 125 kbp panel by Newman et al.[6] in 2014, a 203 kbp cell-free DNA panel and a 300 kbp tumour-tissue panel by Newman et al.[14] in 2016, and a 188 kbp panel used by Chaudhuri et al.[7] in 2017. In 2016, Newman et al.[14] described integrated digital error suppression (iDES) — an enhanced version of CAPP-Seq that incorporates strategies to reduce artefactual errors associated with the platform.

CancerSEEK
A non-invasive early cancer detection platform that incorporates NGS of cell-free DNA plus protein biomarkers and has been evaluated for early stage cancer detection across patients with cancers of the ovary, liver, stomach, pancreas, oesophagus, colorectum, lung, or breast.

Multiplex PCR
(mPCR). A common technique whereby a combination of PCR primers amplifies multiple genomic regions in a single reaction mixture. mPCR can be used as an alternative to capture hybridization to enrich DNA samples for targeted NGS. This enrichment technique was used in the CancerSEEK platform and the Tracking Cancer Evolution Through Therapy (TRACERx) platform (Signatera).

Mutant allele frequency and ctDNA fraction
Mutant allele frequency (MAF) describes the frequency of DNA molecules containing a mutation compared with that of their wild-type counterparts. Within the context of circulating tumour DNA (ctDNA), the combined MAFs of the mutations in cell-free DNA that are known to be present in all cells within a tumour (that is, clonal mutations) broadly equate to the fraction of total cell-free DNA that is tumour-derived, which is referred to as the ctDNA fraction.

Next-generation sequencing
High-throughput DNA-sequencing technologies whereby millions to billions of DNA strands are sequenced in parallel.

Omega score
The omega (Ω) score is a confidence score used in classifying the ctDNA status of patient samples in the CancerSEEK manuscript by Cohen and colleagues[5]. This score was derived by comparing the MAFs of variants identified in a sample to control-sample and cancer-sample reference sets. An omega score of >3 was considered by the authors as a high-confidence indicator that a mutation was detected.

Targeted error correction sequencing
(TEC-Seq). A NGS approach for non-invasively identifying early stage tumours. This platform involves the use of an 81 kbp capture-hybridization enrichment technology and error-control strategies to detect low-frequency mutations across a range of patients with early stage colorectal, breast, lung, or ovarian cancers.

The safe-sequencing system
(Safe-Seq). An NGS approach first reported in 2011 (REF.[20]) that incorporates the use of unique identifiers (UIDs) to identify duplicate sequencing reads arising from the same original DNA molecule. The use of UIDs facilitates in silico error-control strategies that enable the detection and quantification of rare mutations, and variations of this approach are now in widespread use with various cell-free DNA NGS platforms.

cancer detection tools, TEC-Seq solely as a ctDNA detection tool and CancerSEEK as an instrument capable of integrating ctDNA and protein biomarker information (only mutations detected with a high degree of confidence in ctDNA using the CancerSEEK mPCR assay are displayed in FIG. 1). By contrast, CAPP-Seq and the TRACERx mPCR platform were designed to track mutations in cell-free DNA in an MRD setting, in which advanced knowledge of a patient's tumour mutation profile exists. Targeted mutation detection can increase the lower limit of ctDNA detection compared with mutation discovery without prior knowledge of the particular mutations present in the patient's tumour[14,15] (see the 'Calling of mutations in cell-free DNA' section). Thus, rather than a comparison between platforms, the data presented (FIG. 1a–c) should be interpreted as an overview of ctDNA detection in patients with early stage NSCLC. Notably, several different iterations of CAPP-Seq capture panels have been reported that target between 125 kbp[6] and ~300 kbp (REF.[14]) of the genome (BOX 1); in FIG. 1a–c, we present only data derived from the 188 kbp CAPP-Seq panel that was predominantly used by Chaudhuri et al.[7] (TABLE 1).

When ctDNA was confidently detected, maximum MAFs per cell-free DNA sample ranged from 0.01% to 9.3% (median 0.31%) in patients with stage I NSCLC, from 0.03% to 13.9% (median 0.48%) in patients with stage II NSCLC, and from 0.1% to 26.5% (median 1.48%) in those with stage III NSCLC (FIG. 1a); the minimum MAFs per cell-free DNA sample were 0.01–3.64% (median 0.07%), 0.01–9.13% (median 0.16%), and 0.02–26.1% (median 0.5%), respectively (FIG. 1b). The largest proportion of patients with undetectable ctDNA using the TRACERx mPCR, CancerSEEK mPCR, or TEC-Seq platforms had stage I NSCLC (FIG. 1c). These data suggest that the cohort of patients with stage I disease is enriched for tumours that either do not release ctDNA or release ctDNA at frequencies below the limit of detection of current technologies. Contrary to this observation, CAPP-Seq enabled detection of pretreatment ctDNA in 5 of 5 patients with stage I NSCLC compared with 4 of 6 patients with stage II NSCLC (FIG. 1c). Notably, tissue from tumour biopsy samples was analysed using CAPP-Seq in 3 of 5 patients with stage I NSCLC to generate advanced knowledge of tumour-associated mutations, whereas tumour-tissue analysis was performed in only 1 of 6 patients with stage II NSCLC[7]. Prior knowledge of the mutations present in the tumour can improve the analytical sensitivity of CAPP-Seq to detect mutant alleles compared with biopsy-free plasma genotyping[14]; therefore, this variation in tumour-tissue availability might explain why a higher proportion of stage I tumours than stage II tumours was associated with detectable pretreatment ctDNA in this study[7]. This pattern was not observed using the TRACERx platform[8] because advanced knowledge of tumour-associated mutations existed for all patients.

Data regarding ctDNA evaluation in a post-surgical NSCLC MRD context are currently limited[7,8]. In TRACERx[8], we evaluated the presence of ctDNA in plasma samples that were taken following surgical excision of early stage NSCLC. In 13 of 14 patients who had post-surgical relapse, ctDNA was detected before or at the point of disease recurrence diagnosed through clinical and chest radiography follow-up assessments; ctDNA was not detected in the other patient who had post-surgical relapse[8]. Investigation of the relationship between ctDNA mutation frequency and lead time (defined as time from ctDNA detection to radiologically confirmed relapse) revealed that the 5 patients with protracted lead

Table 1 | **Studies on NGS analysis of ctDNA for early detection or MRD assessment of NSCLC**

Study	Panel size (total amount of DNA analysed)	Enrichment technology	NSCLCs profiled (stage)					Application
			I	II	III	IV	Total	
CAPP-Seq[7]	128 genes (188 kbp)	Hybridization (NimbleGen SeqCap)	5 (15.6%)	6 (18.8%)	21 (65.6%)	0 (0%)	32	Biopsy-free genotyping and MRD detection[a]
TEC-Seq[4]	58 genes (80.9 kbp)	Hybridization (Agilent SureSelect)	29 (41.4%)	31 (44.3%)	5 (7.1%)	5 (7.1%)	70	Non-invasive detection
CancerSEEK[5]	61 ROI, including 16 genes (4.6 kbp)	mPCR (Safe-Seq)	46 (44.7%)	26 (25.2%)	31 (30.1%)	0 (0%)	103	Non-invasive detection (ctDNA in combination with protein biomarkers)
TRACERx[8]	Median of 18 patient-specific SNVs (~1.5 kbp)	mPCR (Signatera)	59 (61.5%)	23 (24.0%)	14 (14.6%)	0 (0%)	96	MRD detection[a] and MRD phylogenetic characterization[a]

CAPP-Seq, cancer personalized profiling by deep sequencing; ctDNA, circulating tumour DNA; mPCR, multiplex PCR; MRD, minimal residual disease; NGS, next-generation sequencing; NSCLC, non-small-cell lung cancer; ROI, regions of interest; Safe-Seq, safe-sequencing system; SNV, single-nucleotide variant; TEC-Seq, targeted error correction sequencing; TRACERx, Tracking Cancer Evolution Through Therapy. [a]Requires prior knowledge of primary tumour-associated variants determined through tumour-tissue or preoperative plasma analyses.

times of >100 days had a median ctDNA fraction at MRD detection of <0.1% (FIG. 1d). This finding suggests that NGS platforms applied in a NSCLC MRD context should enable calling of mutations present at a level below this threshold. In support of this suggestion, Chaudhuri et al.[7] used CAPP-Seq of plasma ctDNA to evaluate MRD in a cohort of 40 patients who had received various definitive treatments for localized lung cancer — 5 of 40 patients in this cohort underwent surgical intervention, and the remaining patients received chemotherapy and/or radiotherapy. Of the five patients who underwent surgery and chemotherapy, one had disease relapse at the time of reporting. In this patient, ctDNA-based MRD was detected at a MAF of 0.04% within 4 months of treatment, which was >20 months before progression was defined radiologically according to the Response Evaluation Criteria In Solid Tumours 1.1 (RECIST 1.1)[7].

Detection of low-frequency mutations

The available data demonstrate the importance of detecting mutations present at a MAF of <0.5% in plasma cell-free DNA for ctDNA-based screening for stage I or stage II NSCLC (FIG. 1a) and at a MAF of <0.1% for identifying post-surgical MRD (FIG. 1d). This contrasts with the later-stage disease setting, in which ctDNA MAFs can reach >10%[16–19] (FIG. 1a). Technically, detection of low-frequency mutations using NGS technologies is limited by artefactual errors that accumulate during library preparation and sequencing, which can result in false-positive mutation discovery and mask the detection of true biological variants[14,20,21]. Through profiling of cell-free DNA from individuals without cancer, Newman et al.[14] observed increasingly evident background errors in sequencing data at allele fractions of <0.2%; below an allele frequency of 0.02%, artefactual errors were observed for >50% of genomic positions. These errors could be attributed to the capture-hybridization process, DNA polymerase errors, and cytosine deamination events[14,22–24]. Further artefactual errors can be introduced by Illumina instruments during sequencing; the Illumina-Phred-like quality score (Q score) is used to measure error, and bases with a Q score of ≥30 (equating to an error in base calling with probability of 1 in 1,000 or less) are considered a benchmark for quality in NGS[25,26].

In 2011, Kinde et al.[20] proposed a method to suppress artefactual NGS errors and thereby improve the calling of low-frequency mutations, which was termed the safe-sequencing system (Safe-Seq) (BOX 1). The method is centred on assigning unique identifiers (UIDs) to each DNA molecule being analysed. These UIDs consist of both endogenous UIDs, referring to the start and stop coordinates of a fragment of DNA, and exogenous UIDs, which are unique oligonucleotide sequences used to label DNA molecules during library preparation[20]. Because each starting DNA molecule is uniquely labelled, PCR duplicate reads arising from the same starting DNA template can be identified during data analysis; instead of standard removal of PCR duplicates, duplicates are grouped into molecular families originating from a unique DNA template. If most duplicate reads (≥95%) within a family contain a mutation, that mutation can be considered as a true positive, whereas mutations that are present in only a minority of duplicates are regarded as erroneous. The CancerSEEK mPCR[5], CAPP-Seq (integrated digital error suppression (iDES)-enhanced versions[7,14]; see below), and TEC-Seq[4] platforms all include adaptations of this UID strategy to control for sequencing artefacts in cell-free DNA analyses. By contrast, the mPCR approach used in TRACERx did not include UIDs, instead relying on statistical modelling of individual variant error profiles through analyses of samples from donors without cancer[8]. Newman et al.[14] also described the development of the iDES 'polishing' technique to suppress the pattern of artefactual mutations associated with the CAPP-Seq approach to ctDNA profiling, the majority of which are G>T transversions and, to a lesser extent, C>T and G>A transitions. Implementation of iDES yielded similar error rates to those achieved with error suppression through use of UIDs (~1×10^{-4} errors per base), and iDES synergized with UID use to reduce background error rates to 1.5×10^{-5} errors per base[14].

Calling of mutations in cell-free DNA

Mutation calling algorithms must consider the background error rates that exist in cell-free DNA sequencing data and include steps to limit the incorrect reporting of false-positive mutations. The required stringency of these algorithms largely depends on the number of bases queried because the false discovery rate will increase as a function of the number of bases analysed[27]. In the context of early cancer detection, in which hundreds to thousands of possible mutation combinations exist (depending on panel size), rigorous calling parameters are required to maintain platform specificity — potentially at the expense of analytical sensitivity in terms of detecting low-frequency mutations. For example, the TEC-Seq developers designed calling parameters expected to identify >99%

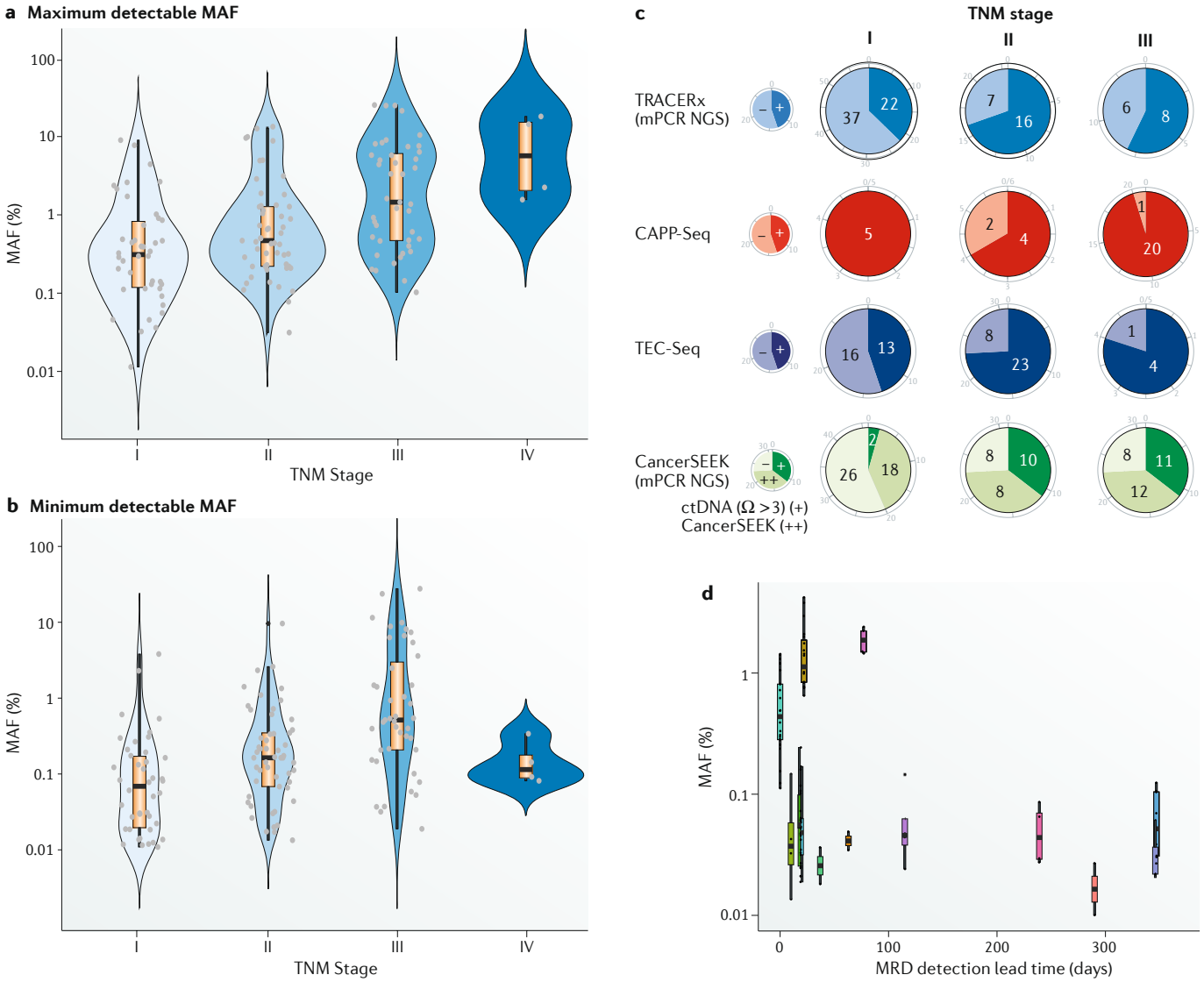


Fig. 1 | **Detection of ctDNA in patients with NSCLC.** This figure provides an overview of the circulating tumour DNA (ctDNA) fractions observed in next-generation sequencing (NGS) analyses on the basis of published supplementary data from four independent cohorts of patients with non-small-cell lung cancer (NSCLC)[4,5,7,8]. **a** | The maximum and **b** | minimum detectable mutant allele frequencies (MAFs) (BOX 1) for 142 patients with detectable ctDNA, from a total of 301 patients analysed. The maximum MAFs and minimum MAFs refer to the frequency of the most abundant and least abundant confidently detected mutations reported as present in a patient's plasma sample. The values plotted in the graphs in parts **a** and **b** of this figure differ for patients with multiple detected alterations reported in their plasma sample (96 of 142 patients with detectable ctDNA); for patients with only a single detected alteration reported in their plasma sample (46 of 142 patients with detectable ctDNA), the corresponding MAF was plotted in both part **a** and part **b**. For the CancerSEEK platform, only high-confidence detected mutants — defined as an omega (Ω) score >3 (BOX 1), as per the supplementary methods of the paper by Cohen et al.[5] — are included in this figure. The *y* axis is on a $\log_{10}$ scale. The lower and upper hinges of each box plot correspond to the 25th and 75th percentiles of the data distribution for each stage category. The upper and lower whiskers (black lines extending from the box plot) extend to the largest and smallest values no more than 1.5 times the interquartile range from the hinge, and the blue shaded areas display the continuous distribution in each category. The graph in part **a** highlights the limit of detection parameters required to detect stage I–IV NSCLCs and demonstrates that to detect stage I NSCLCs, NGS platforms capable of detecting mutations present at MAFs of <0.5% in plasma are desirable. Part **b** demonstrates that current NGS approaches can detect mutations present in plasma at a MAF as low as 0.01%. **c** | Pie charts demonstrating the proportion of patients with ctDNA detected using different platforms, stratified by clinical tumour, node, metastasis (TNM) stage. The number of patients with a positive (+) or a negative (–) test result is indicated within the respective segment of the pie chart. The pie charts for the CancerSEEK data indicate the proportion of patients with high-confidence ctDNA detection according to the aforementioned criteria (dark green; +) and the additional proportion of patients with a positive test result after incorporation of the CancerSEEK ctDNA and protein biomarker parameters (light green; ++). Notably, in all patients with stage I disease apart from those analysed using cancer personalized profiling by deep sequencing (CAPP-Seq), the capacity to detect ctDNA seems to be limited (see the main text for a possible explanation). **d** | Data from the Tracking Cancer Evolution Through Therapy (TRACERx) study[8] demonstrating the median clonal MAFs (ctDNA fraction) (BOX 1) at the time point of minimal residual disease (MRD) detection using personalized NGS panels (*y* axis; $\log_{10}$ scale) versus the lead time achieved through MRD detection (that is, the time between ctDNA detection and clinical diagnosis of relapsed disease; *x* axis). The lower and upper hinges of each box plot correspond to the 25th and 75th percentiles of the data distribution for each patient analysed, and the horizontal line in the centre of the box plot corresponds to the median clonal MAF value. This plot demonstrates the importance of detecting mutations with MAFs <0.1% in order to detect relapsing NSCLC >100 days in advance of clinical detection. mPCR, multiplex PCR; TEC-Seq, targeted error correction sequencing.

of alterations with a MAF of 0.5% at the anticipated sequencing depth of the platform; in analytical validation studies, these thresholds had a specificity of >99.9999% for mutation calls[4]. To achieve this level of accuracy, the developers of this platform used calling parameters based on criteria including MAF and the hotspot status of an alteration. Hotspot alterations with MAFs >0.1% were called if an alteration was present in at least three unique sequencing reads, with each read supported by at least three duplicate reads containing the alteration; for MAFs between 0.05% and 0.1%, four distinct sequencing reads containing the alteration supported by at least four duplicate reads containing the hotspot mutation were required[4]. The increased stringency for confident calling of a hotspot mutation with a MAF of <0.1% reflects the higher risk of sequencing artefacts at frequencies below this threshold. For non-hotspot alterations, the lower limit of detection of the platform was a MAF of 0.1%, and at MAFs between 0.1% and 0.2%, seven unique sequencing reads containing a mutation were required to call a mutation as present[4]. The developers used more stringent parameters to call non-hotspot versus hotspot alterations because the probability of hotspot alterations being tumour-derived is higher than that of non-hotspot alterations, translating into a lower risk of false-positive events at hotspot loci. Furthermore, candidate alterations were excluded if they were present in >1% of the reads from a panel of unmatched cell-free DNA samples from individuals without cancer or if they corresponded to either germline mutations present in the general population or known haematopoietic expansion-related variants[4]. Only 12 of 313 (3.8%) true somatic mutations called in cell-free DNA across the TEC-Seq patient cohort (independent of cancer type) were present at a MAF of <0.1%[4], possibly reflecting these rigorous calling parameters.

By contrast, in the MRD setting, one need only query the presence or absence of a specific set of mutations known in advance to be associated with a specific tumour. Thus, specificity can be maintained with less rigorous calling parameters, potentially increasing the analytical sensitivity of a platform compared with that of early cancer detection approaches. For example, with prior knowledge of the tumour-associated mutations, iDES-enhanced CAPP-Seq enabled the confident detection of ctDNA at MAFs as low as 0.004% in plasma from a patient with NSCLC[14]. In TRACERx[8], we controlled for artefactual mutations by generating an error-propagation model for each single-nucleotide variant targeted by a particular mPCR panel on the basis of analyses of cell-free DNA from individuals without cancer. The false-positive rate associated with this approach in terms of errors per base queried was 3.7×10^{-3} (REF.[8]), which is higher than that observed with IDES-enhanced CAPP-Seq (1.5×10^{-5} errors per base[14]) and TEC-Seq ($<3.3 \times 10^{-7}$ errors per base[4]). However, because the TRACERx platform queried only a limited number of known mutations per patient (a median of 18 mutations in preoperative plasma samples), the platform maintained a specificity of >99% for ctDNA detection using a two-variant threshold, meaning at least two known variants had to be detected to confidently call the presence of ctDNA[8]. The TRACERx mPCR platform had an analytical sensitivity of 84% for mutations present at a frequency of 0.05–0.1% in plasma cell-free DNA. Reflecting this sensitivity, 30% (162 of 534) of mutations confidently detected in preoperative plasma cell-free DNA from 96 patients with early stage NSCLC were present at MAFs of <0.1%[8].

Input DNA and physical detection limit

The physical limit of detection of a ctDNA analysis platform will depend on whether a mutant DNA molecule circulating at low frequencies in plasma is captured in a blood sample and whether that mutant molecule is incorporated into the final library and sequenced. The concentration of cell-free DNA molecules in plasma directly influences these parameters. To further investigate the physical limit of detection, we analysed the available data on cell-free DNA concentrations from the CancerSEEK[5], CAPP-Seq[7], and TEC-Seq[4] cohorts and plotted the distribution of cell-free DNA concentrations measured in preoperative plasma samples from patients with stage I–III NSCLC ($n = 200$) (FIG. 2a). Median cell-free DNA concentrations varied by cohort, with median concentrations of 3.31 ng/ml in the CancerSEEK cohort ($n = 103$), 9.12 ng/ml in the CAPP-Seq cohort ($n = 32$), and 16.84 ng/ml in the TEC-Seq cohort ($n = 65$). This variation between cohorts could be related to differences in plasma sample handling, varying cell-free DNA extraction techniques, or different cell-free DNA measurement techniques. The median cell-free DNA concentration across the three cohorts was 7.69 ng/ml, which is similar to the median concentration of 5.87 ng/ml reported in a previous analysis of preoperative plasma samples from a cohort of 50 patients undergoing surgery for early stage NSCLC[28].

On the basis of the observed distribution of cell-free DNA concentrations (FIG. 2a), we estimated the probabilities of sampling mutant DNA molecules present at a MAF of 0.1% or 0.01% from 10 ml of plasma (equating to 15–20 ml of blood), a volume that is feasible to obtain in routine clinical practice. In our calculations, we assumed a library preparation efficiency of 40% (describing the number of sequenced unique cell-free DNA molecules as a percentage of total input cell-free DNA molecules), which reflects the fact that library preparation and sequencing are subject to loss of cell-free DNA fragments at multiple steps[4,6,14]. For patients with plasma cell-free DNA concentrations in the lowest quartile of observed values, a lack of input cell-free DNA from 10 ml of plasma means that sampling of a ctDNA molecule present at a MAF of 0.1% is not certain and that sampling of a mutant DNA molecule circulating at a MAF of 0.01% carries a probability of <50% (FIG. 2b, solid lines). For early detection efforts using ctDNA screening, the TEC-Seq platform requires the detection of more than four unique DNA molecules containing the same alteration to confidently call a hotspot mutation present at a MAF <0.1% in plasma cell-free DNA[4]; we estimate that the sampling of four or more DNA molecules containing the same alteration at 0.01% MAF is likely to occur in only patients with cell-free DNA concentrations in the top quartile of the cell-free DNA distribution (FIG. 2b, hashed lines). These data highlight that the sampling of low-abundance DNA molecules in 10 ml of plasma might be a physical limitation to ctDNA analyses, particularly in the context of early cancer detection.

In an MRD setting, one can track multiple mutations known in advance to be present within a patient's tumour. Focusing analyses on tracking multiple mutations — rather than a single mutation — as an indicator of MRD increases the physical limit of detection that can be achieved with a limited quantity of cell-free DNA[7,14]. To illustrate this point, we estimate that tracking five mutations improves the probability of sequencing a single low-frequency ctDNA-associated mutation at plasma cell-free DNA concentrations in the lower range of the distribution observed in patients with early stage NSCLC when compared with tracking a single tumour-associated mutation (FIG. 2c). This hypothesis probably explains the observation that tracking multiple mutations

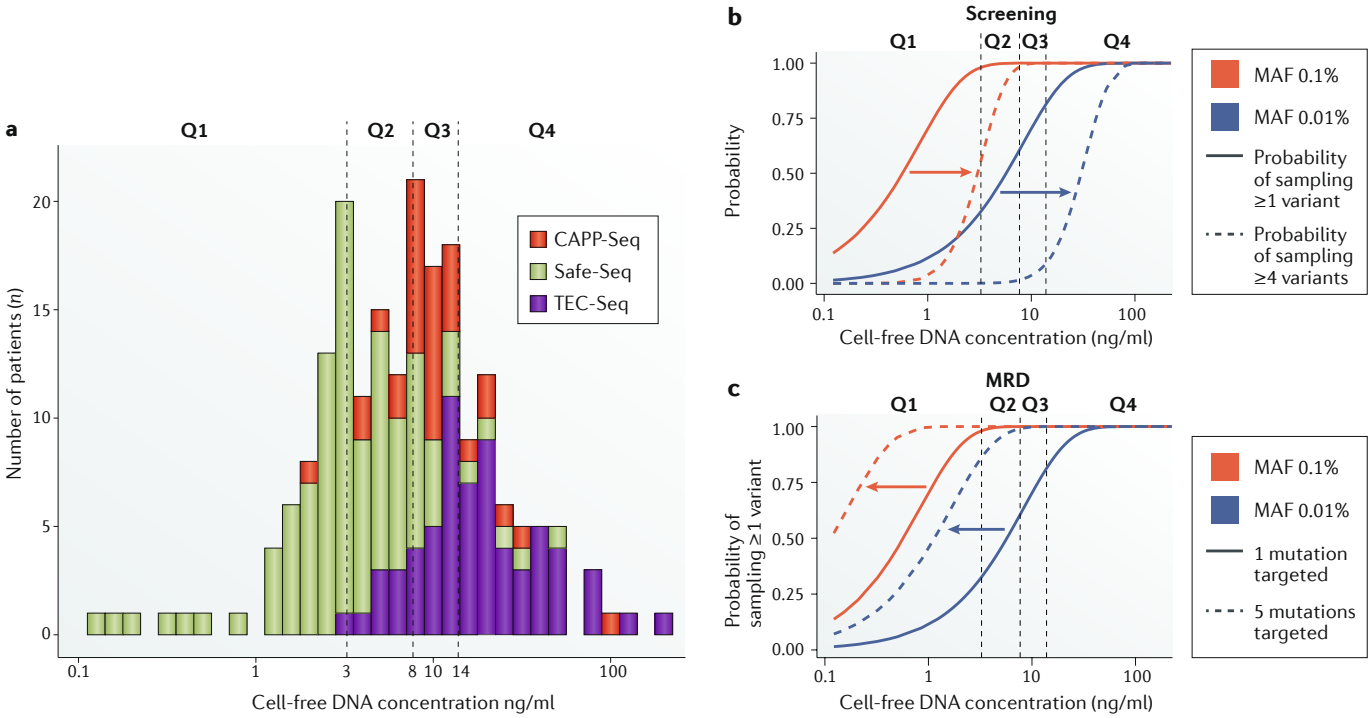


Fig. 2 | **Circulating cell-free DNA concentrations detected in patients with early stage (stage I–III) NSCLC and associated probabilities of detecting mutations in ctDNA. a** | Distribution of plasma cell-free DNA concentrations reported by Phallen et al.[4] using targeted error correction sequencing (TEC-Seq; purple bars), Chaudhuri et al.[7] using cancer personalized profiling by deep sequencing (CAPP-Seq; red bars), and Cohen et al.[5] using multiplex PCR (mPCR and the safe-sequencing system (Safe-Seq); green bars) in a total of 200 patients across all studies with early stage non-small-cell lung cancer (NSCLC). The *x* axis is on a $\log_{10}$ scale and dashed vertical lines indicate the quartile boundaries. **b** | Estimation of probability of sampling and sequencing ≥1 mutant circulating tumour DNA (ctDNA) molecules containing the same alteration (non-dashed lines) or ≥4 mutant ctDNA molecules containing the same alteration (dashed lines) across cell-free DNA concentrations visualized in part **a** (*x* axis is on a $\log_{10}$ scale). Red and blue lines illustrate estimations based on sampling of a mutant ctDNA molecule circulating at a mutant allele frequency (MAF) of 0.1% and of 0.01%, respectively. This scenario equates to an early detection setting in which multiple unique individual DNA molecules containing the same alteration must be sequenced to facilitate differentiation of true-positive low-frequency mutations from sequencing-associated artefacts. **c** | Estimation of sampling and sequencing ≥1 mutant ctDNA molecules at a MAF of 0.1% (red lines) or 0.01% (blue lines) when targeting one single mutation (solid lines) or when sampling a five-mutation tumour signature (dashed lines) across cell-free DNA concentrations indicated in part **a** (*x* axis is on a $\log_{10}$ scale). This scenario equates to a minimal residual disease (MRD) setting in which prior knowledge of tumour-associated mutations exists, either from non-invasive genotyping of pretreatment plasma or from sequencing of primary tumour tissues. All estimations are based on a Poisson distribution and calculated using the R function ppois() with lower.tail set to FALSE. Estimations assume that cell-free DNA is extracted from 10 ml of plasma and that the library preparation efficiency — that is, the number of sequenced unique cell-free DNA molecules as a percentage of total input cell-free DNA molecules — is 40%.

in cell-free DNA improves the sensitivity of MRD detection in patients with NSCLC[7,8].

Biology and ctDNA detection

Several biological challenges are associated with ctDNA detection. A true biological mutation detected in plasma (rather than an artefactual error) might not be specific for a cancer cell population. The majority of cell-free DNA is thought to be released from cells of the haematopoietic lineage[29,30], and haematopoietic progenitor cells accumulate mutations during ageing[31,32]; therefore, somatic driver mutations detected in plasma can arise from haematopoietic progenitors. Somatic driver variants known to be associated with haematological malignancies are carried by 5–6% of individuals aged >70 years[31]. The term clonal haematopoiesis of undetermined potential (CHIP) is used to describe the situation in which these mutations lead to clonal expansion but do not result in haematological neoplasia[33]. In an abstract presented in 2017 (REF.[34]), deep sequencing of cell-free DNA across a broad gene panel (508 genes covering ~2 Mb) revealed the presence of many somatic nonsynonymous mutations (a total of 1,072) that were also detected in matched genomic DNA from white blood cells (as non-germline mutations) in both patients with cancer and individuals without cancer. The most commonly mutated genes identified were *DNMT3A*, *TET2*, *PPM1D*, and *TP53* — genes consistent with clonal haematopoiesis[34]. In 2018, the Oxnard group[35] reported the detection of clonal haematopoiesis associated with *KRAS*, *TP53*, and *JAK2* mutations in patients with advanced-stage NSCLC. Mutations in cell-free DNA that have a non-tumour origin — but are also commonly associated with cancer — present a challenge for both MRD profiling (when a patient's tumour mutation profile has not been established through analyses of tumour tissue) and early cancer detection efforts. Thus, deep sequencing of both white blood cell DNA and cell-free DNA to identify and filter out CHIP-related mutations might be required in these settings in order to reduce false positive ctDNA detection rates.

The biology of early stage NSCLCs might not be uniformly conducive to ctDNA release, which would present an additional challenge to early detection efforts. In the prospective TRACERx cohort of 96 patients with early stage NSCLC[8], we confidently detected 30 of 31 lung squamous cell carcinomas (97%) but only 11 of 58 lung adenocarcinomas (19%); however, this finding was not reproduced in the CAPP-Seq and TEC-Seq NSCLC data sets[4,7],

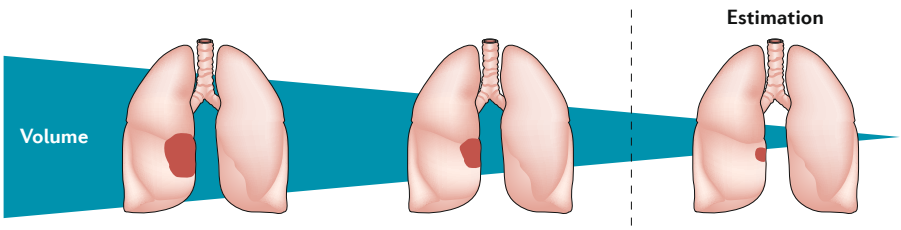
probably owing to differences in TNM stage distributions between the cohorts (TABLE 1). In particular, a higher proportion of adenocarcinomas analysed in TRACERx were TNM stage I (39 of 58; 67%)[8] than the proportion of adenocarcinomas analysed using either the TEC-Seq platform (15 of 38; 40%)[4] or the 188 kbp CAPP-Seq panel (2 of 16; 12.5%)[7]. Apart from histological subtype, ctDNA detection in the TRACERx cohort was positively correlated with 2-[^{18}F] fluoro-2-deoxyglucose (FDG) avidity on PET imaging, Ki67 proliferation indices, the degree of necrosis in tumour specimens, and the presence of lymphovascular invasion (LVI) noted on histological examination of the tumours[8]. In NSCLCs, FDG avidity and Ki67 expression are associated with rapid tumour doubling times and are indicators of a poor prognosis, as are LVI and, potentially, tumour necrosis[36–40]. These data suggest that ctDNA detection in patients with early stage NSCLC is associated with a more aggressive tumour biology and faster tumour growth. Validation of these associations between tumour phenotype, prognosis, and ctDNA detection will be required in future prospective studies.

The relationship between disease burden and the ctDNA fraction of plasma cell-free DNA is another aspect of ctDNA biology that is relevant to both cancer screening and MRD assessment. In 2014, Diaz et al.[41] analysed patients with chemotherapy-refractory metastatic colorectal cancer and estimated that a tumour burden equivalent to 44 million *KRAS*-mutant tumour cells was sufficient to lead to one mutant *KRAS* molecule per millilitre of serum. Correlations between the quantity of ctDNA in plasma and tumour burden have also been demonstrated in patients with early stage and advanced-stage NSCLC[6,7,42,43]. In TRACERx[8], we correlated radiological volumetric measurements obtained through analyses of preoperative CT scans with the mean plasma MAF of clonal mutations detected using a patient-specific mPCR panel and subsequently predicted the MAFs corresponding to a 10 cm^3 NSCLC tumour (0.1%) and a 100 cm^3 NSCLC tumour (1.4%). On the basis of the relationship observed, we made a preliminary estimation that a tumour of 1 cm^3 equates to a mean plasma MAF of clonal alterations of 0.008% (95% CI 0.002–0.03%)[8]. Clinicians use unidimensional tumour measurements to establish the T stage of NSCLCs (according to the eighth edition of the TNM staging criteria[1]). Assuming a spherical tumour nodule, a 10 cm^3 tumour would equate to a T1c NSCLC and a 1 cm^3 tumour would equate to a T1b NSCLC (FIG. 3). Given the difficulties discussed above regarding sampling and calling mutations present at MAFs of <0.1% (particularly in a screening context), these data suggest that the low tumour cell burden will limit our capacity to detect the smallest stage I tumours using published ctDNA-profiling approaches.

These predictions relate to primary early stage NSCLC, and we do not fully understand how this relationship between tumour volume and the ctDNA fraction of plasma cell-free DNA holds up in the metastatic or post-surgical MRD setting. The Brenton group[19] demonstrated a higher ctDNA fraction to tumour volume ratio in patients with relapsed ovarian cancer than in those with newly diagnosed ovarian cancer. In TRACERx[8], we detected ctDNA in a postoperative MRD context when CT scans showed no evidence of disease. This finding was surprising given the correlation between tumour volume and ctDNA fraction observed when evaluating patients with in situ primary tumours[8]. Explanations for this observation in the MRD setting could include sampling of ctDNA originating from multiple micrometastatic tumours each too small to be apparent on CT imaging but with a combined burden sufficient to result in detectable levels of ctDNA and/or a shift in tumour cell death dynamics towards a biology that is more conducive to ctDNA release in the metastatic setting.

ctDNA screening in NSCLC

Increasing cure rates in patients with lung cancer requires diagnosis of the disease at its earliest stages, when the tumour burden is limited: according to the eighth edition of the TNM classification[1], the 5-year overall survival of patients with clinical stage IA1 (maximum tumour diameter of ≤1 cm; T1aN0M0), stage IA2 (1–2 cm; T1bN0M0), and stage IA3 (2–3 cm; T1cN0M0) NSCLC is 92%, 83%, and 77%, respectively. In the National Lung Screening Trial[44], the effectiveness of lung cancer screening with annual low-dose CT (LDCT) scans for 3 years versus a single chest radiograph was compared in 53,454 asymptomatic participants at high risk of the disease; a 20% relative reduction in mortality from lung cancer with LDCT screening was reported. Of the 1,060 lung cancers diagnosed in the LDCT group, 416 (40%) were stage IA, compared with 196 of 941 (21.1%) lung cancers identified on chest radiographs — similar numbers of stage IB, stage II, and stage III cancers were detected using each modality[44]. Fewer stage IV lung cancers were diagnosed in the LDCT group, suggesting that LDCT generally enabled the identification of lung cancers at an earlier stage than was possible with chest radiography (resulting in stage shifts), increasing the feasibility of cure[44]. This finding suggests that, in order to demonstrate superior sensitivity to chest



MAF	1.4% (0.62–3.1%)	0.1% (0.06–0.18%)	0.008% (0.002–0.03%)
Nodule diameter	5.8 cm	2.6 cm	1.2 cm
Nodule volume	100 cm^3	10 cm^3	1 cm^3
T stage	T3	T1c	T1b

Fig. 3 | **The correlation between the abundance of ctDNA, tumour volume, tumour diameter, and T stage.** This figure demonstrates the correlation between estimated mean clonal mutant allele frequency (MAF) in circulating tumour DNA (ctDNA) isolated from plasma, tumour volume, and predicted tumour diameter and how this measurement equates to the clinical T stage according to the eighth version of the tumour, node, metastasis (TNM) staging system[1]. The estimates were calculated on the basis of the Tracking Cancer Evolution Through Therapy (TRACERx) data[8] and suggest that early detection of small non-small-cell lung cancers (<2 cm; T1a–T1b) using ctDNA will be limited by the technical and physical constraints of detecting mutations present at a low MAF (<0.1%), which are discussed in this Perspectives article. Figure adapted from REF.[8], Macmillan Publishers Limited.

radiography-based NSCLC screening, ctDNA profiling must also have utility in diagnosing stage IA disease.

Herein, we have highlighted several challenges in terms of detecting ctDNA in patients with stage I NSCLC. This group contains the largest proportion of patients with undetectable ctDNA using current NGS approaches (FIG. 1c). In those patients with detectable ctDNA, maximum ctDNA MAFs in plasma observed in this group ranged from 0.01% to 9.3% (median 0.31%) (FIG. 1a), and the correlation between tumour burden and mean clonal mutation MAF in plasma suggests that ctDNA fractions are even lower in patients with low-volume stage I tumours capable of releasing ctDNA (for example, T1a–T1b tumours <2 cm in diameter) (FIG. 3). When attempting to discover ctDNA molecules present at such low frequencies using current NGS technologies, multiple unique molecules might need to be sequenced to differentiate between sequencing artefacts and true biological variants[4]. Considering the distribution of plasma cell-free DNA concentrations observed in patients with early stage NSCLC, detection of multiple ctDNA molecules will be possible in only a small proportion of patients without using substantially greater input blood volumes than those routinely used in clinical practice (FIG. 2). This conclusion is shared by Haque and Elemento[45], who have conducted an independent analysis of the TEC-Seq data reported by Phallen and colleagues[4]. In their currently non-peer-reviewed pre-print paper, Haque and Elemento[45] estimate that 150–300 ml of input blood would be required to detect a ctDNA molecule present at a MAF of 0.01% with 95% confidence. These challenges are not technically insurmountable. For example, a shift towards acquisition of cell-free DNA from larger blood volumes is achievable when we consider that up to 470 ml of blood is collected during blood donation. Furthermore, screening blood tests are unlikely to solely rely on mutation detection in ctDNA but perhaps will also incorporate analyses of cell-free DNA methylation, circulating tumour cells, protein biomarkers, exosomal RNAs, and/or imaging parameters to increase the sensitivity and specificity at which low-frequency mutations can be called. For example, with the CancerSEEK early cancer detection tool[5], ctDNA and protein biomarkers are evaluated together, and incorporation of the protein biomarkers increases the sensitivity of this platform to identify patients with NSCLC compared with that achieved through ctDNA profiling alone (FIG. 1c). Similarly, combining exosomal RNA and ctDNA analyses in patients with advanced-stage NSCLC improved the sensitivity of *EGFR* mutation detection compared with the use of ctDNA alone[46]. Finally, we reported that ctDNA detection in patients with NSCLC was associated with a malignant adenocarcinoma phenotype characterized by high rates of tumour cell proliferation, death, and LVI[8]; thus, ctDNA-based screening could potentially overcome one of the limitations of LDCT-based early detection of NSCLC — the detection of indolent tumours and potential for overdiagnosis[47].

NSCLC MRD detection using ctDNA

Established metastatic NSCLC cannot be cured with chemotherapy, and yet in the adjuvant setting, clinical studies have shown a 5% absolute benefit of platinum-based chemotherapy on 5-year overall survival[2]; therefore, chemotherapy can be curative within the context of clinically indiscernible postoperative MRD. Drug development efforts in the adjuvant setting are under-resourced, reflecting the challenges associated with conventional adjuvant clinical trials that make them unattractive pursuits. The greatest challenge in this setting is the low event rates after resection of early stage NSCLC: post-surgical 5-year relapse rates range from ~20% in patients with stage I disease to ~50% in those with stage III disease[48–52]. Thus, demonstrating the effect of an adjuvant therapeutic intervention on NSCLC cure rates often requires the recruitment of thousands of patients to a study with a long follow-up duration. For example, the current practice of adjuvant platinum-based chemotherapy is based on the results of multiple trials and a meta-analysis involving 4,584 patients[2,53]. This situation highlights the need for novel diagnostic approaches to enrich adjuvant clinical trial cohorts for patients at a high risk of metastatic recurrence in order to enable these studies to be conducted in smaller populations of patients in which relapse rates are higher. Because ctDNA-based MRD detection in the postoperative setting is highly specific for impending disease relapse[8–10], postoperative ctDNA profiling could potentially lead to a new generation of adjuvant clinical studies wherein randomization of patients to treatments, including targeted therapy directed at *EGFR* or *ALK* alterations, is based on MRD positivity.

We argue that MRD detection in patients with early stage NSCLC is within the capabilities of current technologies. In a similar fashion to early detection of NSCLC, detection of MRD requires calling of low-frequency mutant molecules in plasma (FIG. 1d). The challenges associated with early detection, however, are lessened in an MRD setting because advanced knowledge of tumour-associated mutations can be obtained through analysis of tumour tissue. This increases the confidence by which low-frequency mutations can be called and controls for biological nonspecificity related to clonal haematopoiesis. Furthermore, advanced knowledge of tumour-associated mutations can also facilitate the tracking of tens to hundreds of tumour-associated variants in MRD settings using personalized NGS panels covering only kilobases of the genome[8,54]. By contrast, non-customized 'off-the-shelf' NGS assays, such as CAPP-Seq, are limited in terms of the number of mutations tracked per patient, particularly in non-smokers, who often have a low tumour mutation burden[55]. For example, by overlaying the genomic regions captured by the original 125 kb CAPP-Seq panel onto whole-exome sequencing data acquired for 183 lung adenocarcinomas, Newman et al.[6] demonstrated that the CAPP-Seq panel captured mutations present in 88% of patients, with a median of four single-nucleotide variants per patient. The larger 203 kbp CAPP-Seq cell-free DNA capture panel used in a subsequent study by Newman et al.[14] captured mutations present in 95.3% of lung adenocarcinomas from an in silico validation cohort of 148 patients and was able to detect at least 8 mutations per tumour in 48.6% of these patients. Theoretically, tracking a greater number of mutations using personalized NGS panels could improve the sensitivity of these platforms in detecting low-volume MRD by facilitating broader sampling of the tumour genome in the context of limited plasma concentrations of cell-free DNA (FIG. 2c). Further studies are required to determine the clinical utility of personalized versus off-the-shelf panels in an MRD setting, in particular, by defining the number of mutations that need to be analysed in these contexts. Using these technologies, we envisage a therapeutic setting in which patients without detectable MRD would be spared from the toxicity of unnecessary adjuvant treatment, whereas patients with MRD detected would receive aggressive multi-modal adjuvant therapy while their disease burden is low, subclonal heterogeneity is limited, and tumour microenvironmental remodelling is incomplete in order to further reduce the disease burden, delay or prevent disease recurrence, and prolong postoperative overall survival.

Taking MRD detection to the clinic

Mounting evidence indicates that ctDNA can be used to detect MRD in advance of clinical disease relapse[7–10], although the value of this approach in terms of improving patient outcomes will require investigation in prospective interventional trials. In these trials, ctDNA profiling at pre-specified postoperative time points should be used to identify patients with MRD. If a patient has detectable MRD, they will be randomly assigned to receive either experimental systemic adjuvant therapy or the routine standard of care; the primary end point should be relapse-free survival. An example of such a trial is the phase II c-TRAK-TN study in patients with resectable triple-negative breast cancer (NCT03145961), which has been designed to assess whether ctDNA screening (over a 1-year period following completion of primary treatment) enables the detection of postoperative MRD; patients with detectable MRD will be randomly assigned in a 2:1 ratio to receive either the anti-programmed cell death protein 1 (PD-1) antibody pembrolizumab or observation.

MRD studies will present novel logistical challenges, particularly if an NGS-based MRD assay personalized to a patient's tumour mutation profile is used. For example, sequencing of excised tumour tissue will need to be performed prospectively for all patients being screened for MRD in order to identify tumour-associated mutations. Mutation calls from these analyses will then be used to generate amplicon or capture-hybridization panels tailored to an individual patient's tumour mutation profile, which will subsequently be applied to cell-free DNA extracted from postoperative plasma samples for classification of the patient's MRD status. This process will be time sensitive and must be rapid enough to usefully inform clinical decision-making; ideally, profiling of tumour tissue and generation of sequencing panels should be performed within a 2-month postoperative window. This consideration is based on an analysis of the National Cancer Database[56], which revealed that the median time to commencing adjuvant chemotherapy for NSCLC is 48 days (range 18–127 days). Such endeavours will require a consortium-led approach with clinical, academic, and industry input. The cost of generating and applying a personalized ctDNA panel for MRD detection will vary with sample throughput, number of mutations targeted, and tumour sequencing parameters. In TRACERx[8], generation and application of personalized mPCR NGS panels to five plasma samples was estimated to cost US$1,750 per patient.

Conclusions

We have discussed technical and biological challenges associated with ctDNA detection in patients with early stage NSCLC and in the postoperative setting. NGS platforms operating in these clinical contexts must enable confident detection of mutations in plasma at a MAF of <0.1% and should therefore incorporate strategies to control for sequencing artefacts. Platforms adopting these approaches differ from the standard NGS platforms that are currently approved by the FDA, such as the Memorial Sloan Kettering–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) panel[57]. This panel is designed to identify clinically actionable mutations in tumour tissue and has a MAF detection limit of ~2% for hotspot mutations[58], which as discussed within this Perspectives article, is insufficient for analyses of low-frequency variants in cell-free DNA.

Further research is required to determine the optimal NGS approaches for low-frequency mutation detection in plasma cell-free DNA, necessitating cross-platform validation studies comparing different library construction and enrichment strategies[59]. In contrast to ctDNA detection approaches that typically interrogate a single locus and have low multiplexing capabilities, such as digital droplet PCR[60], NGS methodologies can be used to interrogate larger portions of the tumour genome and track multiple tumour-associated mutations. In an MRD setting, tracking multiple mutations per patient has been shown to improve platform sensitivity for ctDNA detection[7,8]. We have yet to establish the necessary number of mutations to track in an MRD context while also considering the balance between financial and logistical requirements to implement personalized versus off-the-shelf ctDNA assays as well as the analytical sensitivity parameters that are associated with these assays.

On the basis of the data discussed herein, we conclude that further technical improvements and perhaps sampling of greater quantities of blood will be required to demonstrate the validity of ctDNA profiling as a screening tool in the detection of early stage NSCLC. Additionally, the use of multimodality approaches to blood evaluation could result in improved sensitivity compared with ctDNA profiling alone. In the adjuvant setting, an urgent need exists to implement approaches such as MRD assessment in order to improve risk stratification and drive forward drug development. The optimal technology for MRD detection remains to be determined; however, we argue that broad NGS-based assays capable of interrogating multiple tumour-associated mutations are already technically feasible and should be assessed in consortia-led interventional studies designed to evaluate MRD detection as a patient stratification tool in the adjuvant NSCLC setting.

*Christopher Abbosh[1], Nicolai J. Birkbak[1,2] and Charles Swanton[1,2]**

[1]Cancer Research UK Lung Cancer Centre of Excellence London and Manchester, University College London Cancer Institute, London, UK.

[2]Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, London, UK.

**e-mail: charles.swanton@crick.ac.uk*

https://doi.org/10.1038/s41571-018-0058-3

Published online 3 July 2018

1. Goldstraw, P. et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).
2. Pignon, J. P. et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J. Clin. Oncol.* **26**, 3552–3559 (2008).
3. Non-Small Cell Lung Cancer Collaborative Group. Chemotherapy for non-small cell lung cancer. *Cochrane Database Syst. Rev.* **2**, CD002139 (2000).
4. Phallen, J. et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl Med.* **9**, eaan2415 (2017).
5. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
6. Newman, A. M. et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
7. Chaudhuri, A. A. et al. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov.* **7**, 1394–1403 (2017).
8. Abbosh, C. et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
9. Tie, J. et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl Med.* **8**, 346ra392 (2016).
10. Garcia-Murillas, I. et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl Med.* **7**, 302ra133 (2015).
11. Beaver, J. A. et al. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin. Cancer Res.* **20**, 2643–2650 (2014).
12. Pantel, K. & Alix-Panabières, C. Tumour microenvironment: informing on minimal residual disease in solid tumours. *Nat. Rev. Clin. Oncol.* **14**, 325–326 (2017).
13. Burgener, J. M., Rostami, A., De Carvalho, D. D. & Bratman, S. V. Cell-free DNA as a post-treatment surveillance strategy: current status. *Semin. Oncol.* **44**, 330–346 (2017).
14. Newman, A. M. et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
15. Forshew, T. et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl Med.* **4**, 136ra168 (2012).
16. Dawson, S.-J. et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
17. Adalsteinsson, V. A. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).

18. Guibert, N. et al. Amplicon-based next-generation sequencing of plasma cell-free DNA for detection of driver and resistance mutations in advanced non-small cell lung cancer. *Ann. Oncol.* **29**, 1049–1055 (2018).
19. Parkinson, C. A. et al. Exploratory analysis of *TP53* mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. *PLOS Med.* **13**, e1002198 (2016).
20. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108**, 9530–9535 (2011).
21. Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).
22. Park, G. et al. Characterization of background noise in capture-based targeted sequencing data. *Genome Biol.* **18**, 136 (2017).
23. Hestand, M. S., Houdt, J. V., Cristofoli, F. & Vermeesch, J. R. Polymerase specific error rates and profiles identified by single molecule sequencing. *Mutat. Res.* **784–785**, 39–45 (2016).
24. Chen, G., Mosier, S., Gocke, C. D., Lin, M. T. & Eshleman, J. R. Cytosine deamination is a major cause of baseline noise in next-generation sequencing. *Mol. Diagn. Ther.* **18**, 587–593 (2014).
25. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
26. Manley, L. J., Ma, D. & Levine, S. S. Monitoring error rates in Illumina sequencing. *J. Biomol. Tech.* **27**, 125–128 (2016).
27. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
28. Szpechcinski, A. et al. Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease. *Br. J. Cancer* **113**, 476–483 (2015).
29. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
30. Lui, Y. Y. et al. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin. Chem.* **48**, 421–427 (2002).
31. Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
32. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
33. Steensma, D. P. et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).
34. Razavi, P. et al. Cell-free DNA (cfDNA) mutations from clonal hematopoiesis: implications for interpretation of liquid biopsy tests. *J. Clin. Oncol.* **35**, 11526 (2017).
35. Hu, Y. et al. False positive plasma genotyping due to clonal hematopoiesis. *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-18-0143 (2018).
36. Higgins, K. A. et al. Lymphovascular invasion in non-small-cell lung cancer: implications for staging and adjuvant therapy. *J. Thorac. Oncol.* **7**, 1141–1147 (2012).
37. Warth, A. et al. Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role. *Br. J. Cancer* **111**, 1222 (2014).
38. Duhaylongsod, F. G. et al. Lung tumor growth correlates with glucose metabolism measured by fluoride-18 fluorodeoxyglucose positron emission tomography. *Ann. Thorac. Surg.* **60**, 1348–1352 (1995).
39. Higashi, K. et al. ^{18}F-FDG uptake as a biologic prognostic factor for recurrence in patients with surgically resected non-small cell lung cancer. *J. Nucl. Med.* **43**, 39–45 (2002).
40. Gkogkou, C., Frangia, K., Saif, M. W., Trigidou, R. & Syrigos, K. Necrosis and apoptotic index as prognostic factors in non-small cell lung carcinoma: a review. *SpringerPlus* **3**, 120 (2014).
41. Diaz, L. A. Jr et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
42. Winther-Larsen, A. et al. Correlation between circulating mutant DNA and metabolic tumour burden in advanced non-small cell lung cancer patients. *Br. J. Cancer* **117**, 704–709 (2017).
43. Pécuchet, N. et al. Base-position error rate analysis of next-generation sequencing applied to circulating tumor DNA in non-small cell lung cancer: a prospective study. *PLOS Med.* **13**, e1002199 (2016).
44. Aberle, D. R. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365**, 395–409 (2011).
45. Haque, I. S. & Elemento, O. Challenges in using ctDNA to achieve early detection of cancer. Preprint at *bioRxiv* https://doi.org/10.1101/237578 (2017).
46. Krug, A. K. et al. Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma. *Ann. Oncol.* **29**, 700–706 (2017).
47. Patz, E. F. Jr et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern. Med.* **174**, 269–274 (2014).
48. Winton, T. et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *N. Engl. J. Med.* **352**, 2589–2597 (2005).
49. Lou, F., Sima, C. S., Rusch, V. W., Jones, D. R. & Huang, J. Differences in patterns of recurrence in early-stage versus locally advanced non-small cell lung cancer. *Ann. Thorac. Surg.* **98**, 1755–1761 (2014).
50. Kelsey, C. R. et al. Local recurrence after surgery for early stage lung cancer: an 11-year experience with 975 patients. *Cancer* **115**, 5218–5227 (2009).
51. Sugimura, H. et al. Survival after recurrent nonsmall-cell lung cancer after complete pulmonary resection. *Ann. Thorac. Surg.* **83**, 409–417 (2007).
52. Dziedzic, D. A., Rudzinski, P., Langfort, R. & Orlowski, T. Risk factors for local and distant recurrence after surgical treatment in patients with non-small-cell lung cancer. *Clin. Lung Cancer* **17**, e157–e167 (2016).
53. Artal Cortés, Á., Calera Urquizu, L. & Hernando Cubero, J. Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res.* **4**, 191–197 (2015).
54. Ng, S. B. et al. Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery. *Sci. Rep.* **7**, 40737 (2017).
55. Govindan, R. et al. Genomic landscape of non-small cell lung cancer in smokers and never smokers. *Cell* **150**, 1121–1134 (2012).
56. Salazar, M. C. et al. Association of delayed adjuvant chemotherapy with survival after lung cancer surgery. *JAMA Oncol.* **3**, 610–619 (2017).
57. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
58. Cheng, D. T. et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).
59. Merker, J. D. 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.* **36**, 1631–1641 (2018).
60. Denis, J. A., Guillerm, E., Coulet, F., Larsen, A. K. & Lacorte, J.-M. The role of BEAMing and digital PCR for multiplexed analysis in molecular oncology in the era of next-generation sequencing. *Mol. Diagn. Ther.* **21**, 587–600 (2017).

Acknowledgements
The work of C.A. is supported by a Stand Up To Cancer–LUNGevity–American Lung Association Lung Cancer Interception Dream Team translational research grant and the Rosetrees Trust. The work of C.S. has been supported by the Breast Cancer Research Foundation, Cancer Research UK (CRUK; Tracking Cancer Evolution Through Therapy (TRACERx) and CRUK Cancer Immunotherapy Catalyst Network), the CRUK Lung Cancer Centre of Excellence, the CRUK University College London Experimental Cancer Medicine Centre, a Stand Up To Cancer–LUNGevity–American Lung Association Lung Cancer Interception Dream Team Translational Research Grant, the European Research Council (THESEUS project), the Medical Research Council, the UK National Institute for Health Research, the Novo Nordisk Foundation (ID 16584), the Prostate Cancer Foundation, the Rosetrees Trust, the University College London Hospitals Biomedical Research Centre, and the Wellcome Trust.

Author contributions
C.A. and N.J.B. researched the data for the article, and C.A. and C.S. wrote the manuscript. All authors contributed to discussions of content and reviewed and/or edited the manuscript.

Competing interests
C.A. and C.S. submitted a patent with University College London (UCL) business PLC (provisional patent number 1618485.5) based on a phylogenetic approach to analysis of circulating tumour DNA. C.S. has received grant support from AstraZeneca; personal fees from Boehringer Ingelheim, Celgene, Eli Lilly, GlaxoSmithKline, Novartis, Pfizer, and Roche; has stock options in Achilles Therapeutics, ApoGen Biotechnologies, EPIC Bioscience, and GRAIL; and is a co-founder of Achilles Therapeutics. N.J.B. declares no competing interests.

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

Reviewer information
Nature Reviews Clinical Oncology thanks B. Li, C. Rolfo, and Y.-L. Wu for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41571-018-0058-3.

RELATED LINKS
ClinicalTrials.gov database: https://clinicaltrials.gov/