Article

Identifying key circulating tumor DNA parameters for predicting clinical outcomes in metastatic non-squamous non-small cell lung cancer after first-line chemoimmunotherapy

Haolun Ding[1], Min Yuan [2,4] ✉, Yaning Yang[1] & Xu Steven Xu [3,4] ✉

Circulating tumor DNA (ctDNA) provides valuable tumor-related information without invasive biopsies, yet consensus is lacking on optimal parameters for predicting clinical outcomes. Utilizing longitudinal ctDNA data from the large phase 3 IMpower150 study (NCT02366143) of atezolizumab in combination with chemotherapy with or without bevacizumab in patients with stage IV non-squamous Non-Small Cell Lung Cancer (NSCLC), here we report that post-treatment ctDNA response correlates significantly with radiographic response. However, only modest concordance is identified, revealing that ctDNA response is likely not a surrogate for radiographic response; both provide distinct information. Various ctDNA metrics, especially early ctDNA nadirs, emerge as primary predictors for progression-free survival and overall survival, potentially better assessing long-term benefits for chemoimmunotherapy in NSCLC. Integrating radiographic and ctDNA assessments enhances prediction of survival outcomes. We also identify optimal cutoff values for risk stratification and key assessment timepoints, notably Weeks 6–9, for insights into clinical outcomes. Overall, our identified optimal ctDNA parameters can enhance the prediction of clinical outcomes, refine trial designs, and inform therapeutic decisions for first-line NSCLC patients.

Circulating tumor DNA (ctDNA) plays a pivotal role as a valuable biomarker for efficient clinical trials and drug development regulation. Due to its minimally invasive nature[1], ctDNA holds promise for cancer diagnosis, prognosis assessment, and guiding treatment decisions[2,3]. In clinical trials, ctDNA can be measured at multiple time points during treatment, facilitating treatment monitoring and assessment of clinical efficacy (e.g., responses)[4,5].

In the context of precision immune-oncology (IO) for NSCLC, studies suggest that ctDNA response may serve as an early endpoint indicating clinical benefit[6–9]. In addition, machine learning models of ctDNA data have demonstrated success in predicting overall survival (OS)[10]. Furthermore, initial investigations using small clinical cohorts from Canadian Cancer Trials Group (CCTG) BR.36 trial[11] have been conducted to assess the potential of utilizing ctDNA for evaluating

[1]Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui, China. [2]Department of Health Data Science, Anhui Medical University, Hefei, Anhui, China. [3]Clinical Pharmacology and Quantitative Science, Genmab Inc., Princeton, NJ, USA. [4]These authors jointly supervised this work: Min Yuan, Xu Steven Xu. ✉e-mail: myuan@ustc.edu.cn; sxu@genmab.com

radiographic response and radiographic progression-free survival (PFS) and determining the optimal definition of ctDNA molecular response along with the appropriate timing of assessment.

Despite the progress, the current focus has primarily centered around restricted, notable ctDNA features at snapshot time points using data with limited sample size. Owing to the intricate nature of longitudinal data, it remains uncertain which specific summary ctDNA metrics (concentrations as tumor molecules per mL plasma (TMPMP) or variant allelic frequency (VAF), or mutation numbers, etc), longitudinal parameters/features (molecular response/reduction, progression, or tumor load), optimal assessment times, and risk-stratification cutoffs are of most effective in predicting radiographic responses, PFS and OS. We hypothesized that different clinical outcomes (radiographic response, PFS, and OS) may be associated with different longitudinal ctDNA features and identification of optimal ctDNA features, corresponding timing of assessments, as well as their cutoff points can potentially optimize treatment decisions and design of clinical trials.

In this work, we leverage the extensive dataset derived from the Phase 3 Impower150 trial in response to the limitations in the existing literature and to address pertinent research questions. Various ctDNA features predict outcomes in 1 L NSCLC post chemoimmunotherapy, with nadir of ctDNA tumor load as a key predictor of PFS and OS, while the maximum percent reduction from baseline showing modest concordance with radiographic responses. Combining ctDNA with radiographic assessments enhances long-term survival predictions and may serve as an accelerated endpoint for drug development. We also establish ctDNA cutoffs for risk stratification and optimal timing of assessments, aiding trial design and therapy optimization.

Results
Study population
Summary statistics of clinical and demographic characteristics of patients in this study are presented in Supplementary Table 2. The study cohort comprised individuals with a median age was 63 years (range, 56–68), and 61% of patients were male. Eighty-one percent of patients had a history of smoking. High PD-L1 expression was detected in 22% of the patients. In the ABCP arm, median PFS and OS were 9.8 (95% CI: 8.3–11.9) and 22.3 (95% CI: 18.9–26.1) months, respectively.

Similarly, in the ACP arm, median PFS and OS were 6.9 (95% CI: 5.8–8.3) and 21.3 (95% CI: 16.4–24.4) months, respectively. The patient characteristics and clinical outcomes in the ctDNA analysis were comparable to those in the intention-to-treat population of the IMpower150 trial[12,13].

Genomic alterations identified through ctDNA analysis closely mirrored those previously reported in non-squamous cell NSCLC[14,15]. TP53 (45%), KRAS (23%), SPTA1 (21%), LRP1B (19%), FAT3 (19%), STK11 (15%), and KEAP1 (14%) emerged as the most commonly identified genomic alterations (Fig. 1). There are 48 patients with undetectable ctDNA (zero mutations detected) at baseline. Compared to those with detectable ctDNA at baseline, patients with undetectable ctDNA at baseline tend to have slightly lower overall response rate, similar PFS, but seemingly better OS (Supplementary Fig. 1).

Concordance between ctDNA and radiographic response
Responders (PR/CR) and non-responders (SD/PD) exhibited distinct trajectories in ctDNA over time (Supplementary Fig. 2). After FDR correction for multiple comparisons, three ctDNA longitudinal features exhibit statistically significant correlation with the best overall radiographic response (BOR) (Fig. 2A). Based on the proposed composite ranking, the optimal predictors of BOR among all ctDNA summary metrics were identified as the maximum % reduction in total mutations and known/likely pathogenic mutations (Fig. 2A). These metrics demonstrated the strong association with BOR with large effect size on both training and test datasets.

Overall, there was only modest concordance observed between different ctDNA molecular responses (i.e., ctDNA reduction) and radiographic response (Fig. 2A). The area under the curve (AUC) values, reaching 0.71 and 0.75, were achieved for the maximum % change in total mutations in the training and test dataset, respectively. At the optimal cutoff of 20% reduction, the sensitivity and specificity were determined to be 0.84 and 0.62, respectively on the test data. Consistent with existing literature[6,11,16], substantial clearance of ctDNA concentrations, such as a maximum >95% reduction according to median AF or TMPMP (Fig. 2A), was also indicative of radiographic response (Fig. 1A), with an AUC of 0.76 on the test set. However, these results were not statistically significant after FDR correction.
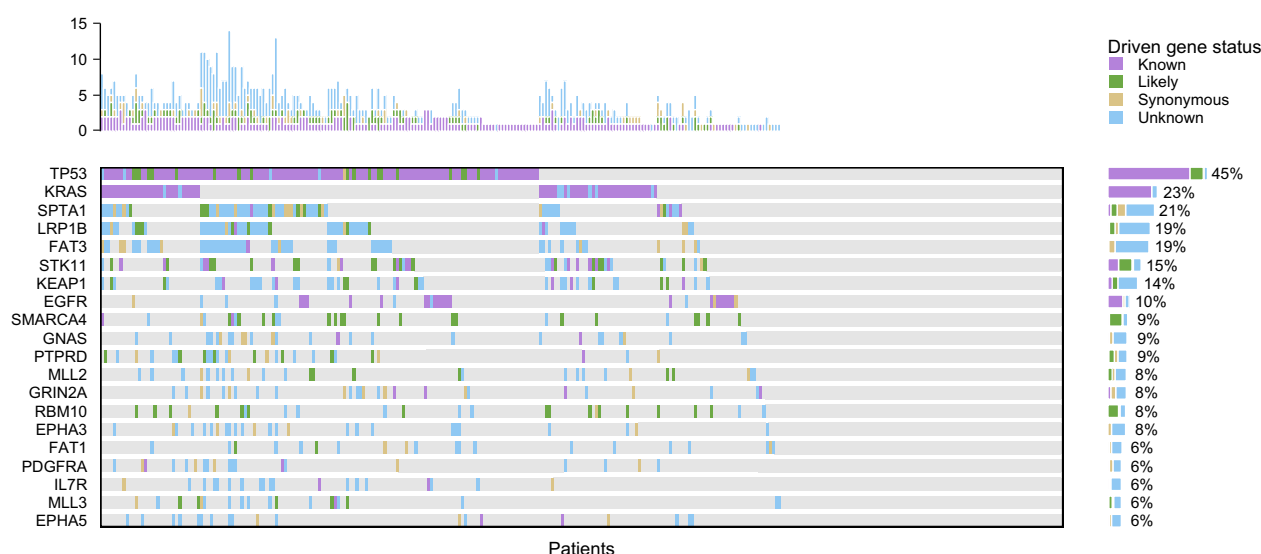

Fig. 1 | Summary of plasma variants detected pretreatment in patients with metastatic non-squamous 1 L NSCLC. The mutation count per sample is shown at the top, while alteration prevalence for each gene is listed on the right. Different colors represent the status of driver genes. Among a total of 262 patients, 48 patients exhibited undetectable tumor-derived mutations pretreatment. Source data are provided as a Source Data file.
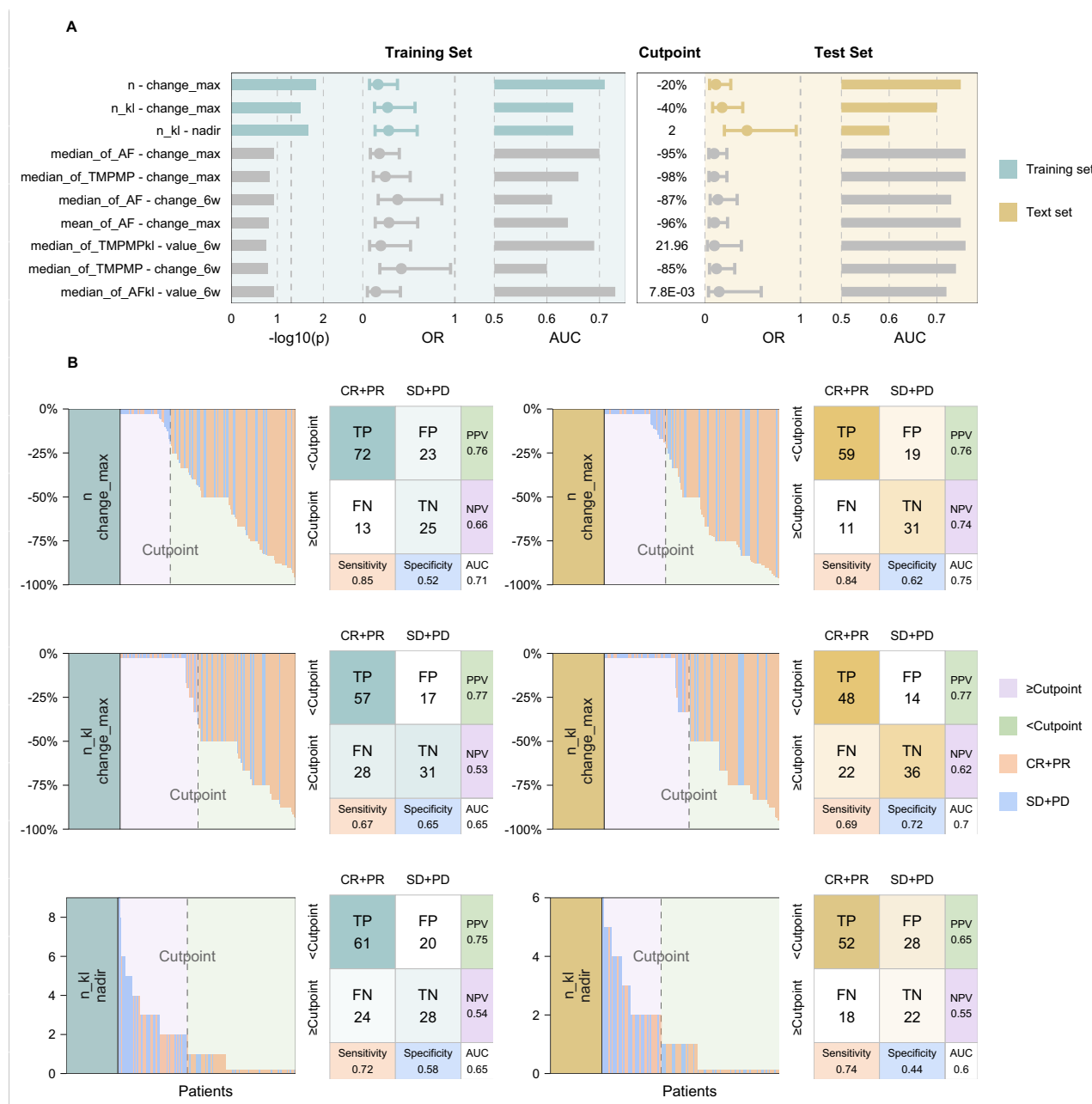
Fig. 2 | Composite ranking of concordance between ctDNA features and radiographic response. A barplot illustrating −log10 (*P*-values after FDR correction) of the association between Best Objective Response and the top 10 ctDNA longitudinal features based on logistic regression models on the training set (gray bars represents *P*-values ≥ 0.05 after FDR correction); *P*-values are calculated by a two-sided *Z* test using logistic regression models; forest plots depicting estimated odds ratio and barplots showing AUC on both training and test datasets based on the optimal cutoff point identified from the training set; error bars represent the exponential of estimated coefficients ± 1.96*SEM (*N* = 253). **B** Waterfall/distribution plots and confusion matrices for training and test sets utilizing optimal cutoff points identified from the training set for the top 3 ctDNA features (max % change of overall mutation numbers (*n*), max % change of n_kl (known or likely pathogenic mutation numbers), nadir of n_kl). Cut-points were determined by maximizing the Youden index (Youden = sensitivity + specificity − 1). Source data are provided as a Source Data file.

We further analyzed ctDNA profiles in the subset of patients with RECIST SD and with unconfirmed PD (Supplementary Fig. 2). It is apparent that in patients with RECIST SD and those with unconfirmed PD, ctDNA tended to decrease over time. Even in patients with confirmed PD, initial reduction in ctDNA concentrations was observed. This may further explain the moderate concordance between ctDNA response and radiographic response since ctDNA decrease occurred in both responders and non-responders (SD and PD patients).

At specific time points, none of the ctDNA features showed significant associations with BOR after FDR correction (Supplementary Fig. 4). However, among the snapshot ctDNA reductions, those at Week 6 (C3D1) seem to offer more consistent predictions of BOR (Fig. 2A and Supplementary Fig. 4). It is noteworthy that the maximum overall percentage reduction surpassed all fixed-time snapshot ctDNA reduction measures (Fig. 2A).

Prognostic value of ctDNA for radiographic PFS

Distinct trajectories of ctDNA over time were discerned between censored patients and those who experienced death or disease progression (Supplementary Fig. 3). Various ctDNA features emerged as
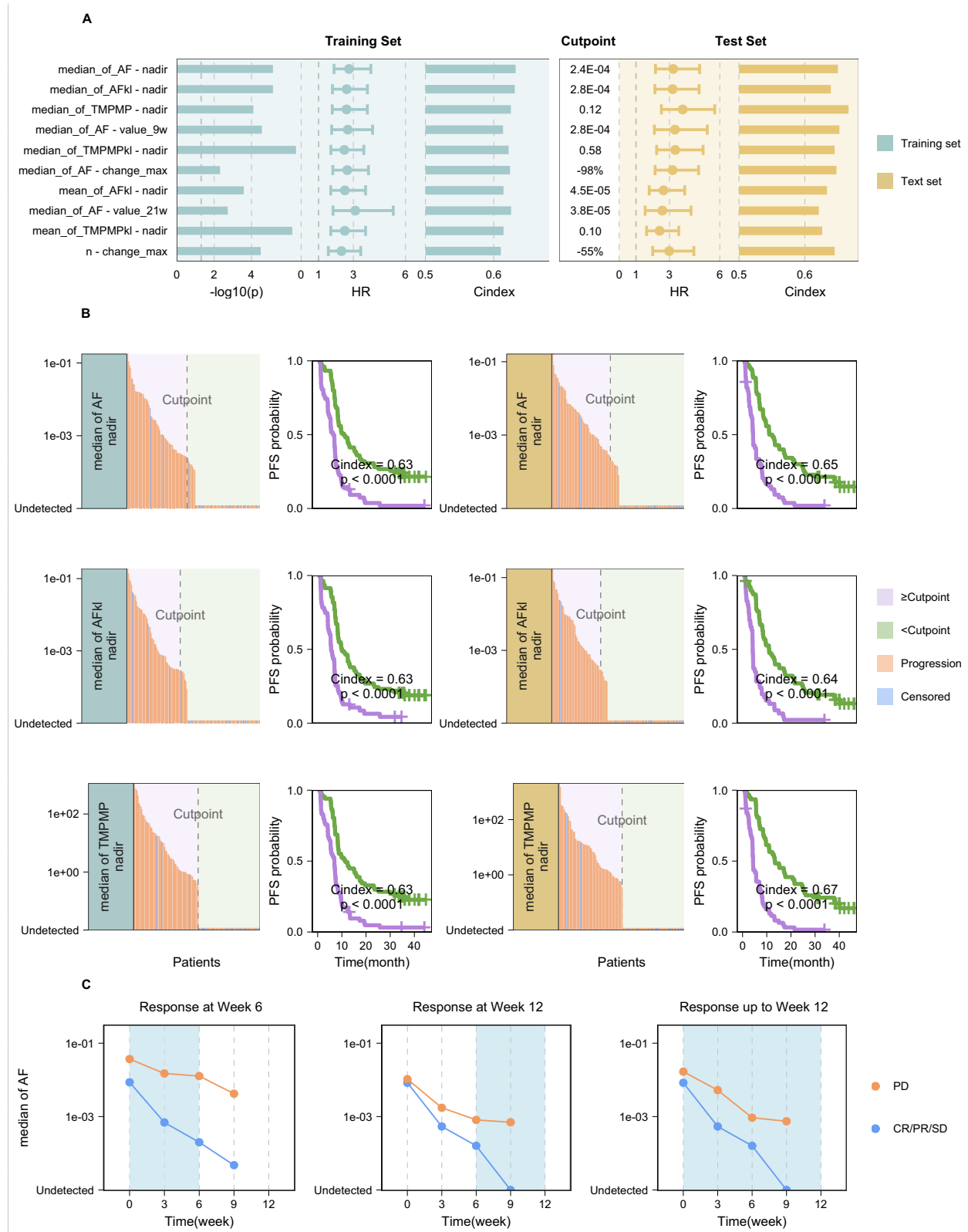
effective predictors for PFS (Fig. 3A). Notably, the ctDNA nadir concentration based on median allelic fraction (AF) appears to be the most predictive ctDNA summary measure for PFS based on the composite ranking criteria (Fig. 3A).

Optimal cutoff was determined for the studied ctDNA features using the training set and validated on the test set (Fig. 3A/B). Utilizing the identified optimal cutoff for each ctDNA feature, the NSCLC

patients were stratified into a molecular high-risk group (ctDNA feature ≥ cutoff) and a molecular low-risk group (ctDNA feature <cutoff) (Fig. 3B). For instance, based on nadir of median AF, the molecular high-risk group exhibited a 2.74 times higher risk of death or disease progression compared to the molecular low-risk group (95% CI: 1.75–4.30; $P < 0.0001$) on the test set. The median PFS was 11.7 months (95% CI: 9.0–17.5 months) in the molecular low-risk group, contrasting

Fig. 3 | Composite ranking of predictive ctDNA features for progression-free survival (PFS). A barplot illustrating −log10 (*P*-values after FDR correction) on the training set; *P*-values are calculated by a two-sided *Z* test using Cox proportional hazards models; forest plots depicting estimated hazard ratio and barplots showing c-index on both training and test datasets based on the optimal cutoff point identified from the training set (Top 10 ctDNA longitudinal features); error bars represent the exponential of estimated coefficients ± 1.96*SEM (*N* = 262).
B distributions and Kaplan–Meier plots for training and test sets utilizing optimal cutoff points identified from the training set for the top 3 ctDNA features (nadir of median AF, median AFkl, and median TMPMP); *P*-values are determined using a two-sided log-rank test to compare Kaplan–Meier curves. The exact *P*-values, from left to right by row, are 4.17e-8 (Training set: 63 patients ≥ cutpoint; 74 patients <cutpoint), 1.44e-9 (Test set: 55 patients ≥ cutpoint; 70 patients <cutpoint), 1.92e-7 (Training set: 56 patients ≥ cutpoint; 81 patients <cutpoint), 3.51e-9 (Test set: 47

patients ≥ cutpoint; 78 patients <cutpoint), 2.91e-7 (Training set: 70 patients ≥ cutpoint; 67 patients <cutpoint), and 1.57e-11 (Test set: 63 patients ≥ cutpoint; 62 patients <cutpoint). **C** average ctDNA kinetics (median AF) for early progressors (progressive disease at Week 6, Week 12, and up to Week 12) and non-early progressors (CR/PR/SD). In the 'Response at Week 6' plot, 18, 18, 6, and 4 early progressors, and 235, 226, 213, and 205 non-early progressors were ctDNA evaluable at Weeks 0, 3, 6, and 9, respectively. In the "Response at Week 12" plot, 11, 11, 11, and 10 early progressors, and 208, 199, 193, and 191 non-early progressors were ctDNA evaluable at Week 0, 3, 6, and 9, respectively. In the "Response up to Week 12" plot, 25, 25, 13, and 11 early progressors, and 208, 199, 193, and 191 non-early progressors were ctDNA evaluable at Week 0, 3, 6, and 9, respectively. Blue shaded areas represent response evaluation periods. Source data are provided as a Source Data file.

with 4.4 months (95% CI: 4.1–7.2 months) in the molecular high-risk group. The c-index was determined to be 0.63 and 0.65 on the training set and test set, respectively, at the identified optimal cutoff point for nadir of median AF (Fig. 3). Also, it is worth noting that the determined optimal cutoffs for the ctDNA nadirs are very close to undetectable level (Fig. 3B).

Interestingly, unlike the observations for radiographic responses, where molecular response in terms of percent reduction in ctDNA best predicted the radiographic response (Fig. 2), percent reduction in ctDNA appears less effective in predicting PFS when compared to the absolute molecular ctDNA tumor burden such as nadir of median AF. Among all ctDNA features based on percent change from baseline, only the maximum percent changes in median AF and total mutation numbers were identified in the top 10 most predictive features for PFS, according to the composite ranking.

Within the context of snapshot data at fixed time points, the Week 9 (C4D1) median AF concentrations exhibited the strongest association with PFS (Fig. 3A and Supplementary Fig. 4). Additional analyses were performed to evaluate if the median of allele fraction at Week 9 was the most indicative of early progression. The longitudinal ctDNA trajectories in early progressors (EPD; patients who progressed during the first 2 radiographic assessments (up to Week 12)) and non-early progressor (NEPD; patients who achieved and maintained a response of SD or better (SD, PR, or CR) during the first 2 radiographic assessments) were compared (Fig. 3C). The divergence in ctDNA trajectories between EPD and NEPD became evident as early as 3 weeks into treatment (C2D1), with NEPD showing deeper suppression of ctDNA tumor load. By Week 6 (C3D1) and Week 9 (C4D1), the separation between ctDNA curves widened. The most substantial separation in ctDNA reduction between EPD and NEPD was observed at Week 9, supporting its importance as the identified optimal time-point for predicting progression.

Prognostic value of ctDNA for OS
Clear distinctions in ctDNA kinetics were also observed between deaths and patients who were still alive (Supplementary Fig. 3). A range of ctDNA features appear to be effective in predicting OS (Fig. 4A). Similar to the findings for PFS, the ctDNA nadirs exhibit excellent predictive performance for OS. Particularly, the nadir of median TMPMP emerged as the primary predictor for OS according to the composite ranking (Fig. 4A).

Like what was observed for PFS, the optimal cutoffs identified for ctDNA nadirs predicting OS are in close proximity to the undetectable level (Fig. 4B). These findings underscore the significance of complete elimination of ctDNA in predicting OS. For nadir of median TMPMP, among the NSCLC patients classified as molecular high-risk (nadir of median TMPMP ≥ 0.51) in the test set, a 3.63 times higher mortality risk was observed (95% CI: 2.28–5.77; *P* < 0.0001) compared to molecular low-risk patients (nadir of median TMPMP < 0.51) (Fig. 4B). The median OS was 30.6 months (95% CI: 24.7 – NR months) in the molecular low-

risk group, in contrast to 10.2 months (95% CI: 7.4–15.7 months) in the molecular high-risk group. The c-index at the identified optimal cutoff point for nadir of median TMPMP was 0.65 on the training set and 0.68 on the test set (Fig. 4B).

Regarding snapshot data at specific time points, a number of Week 21 (C8D1) ctDNA features emerged as top 10 predictors based on the composite ranking (Fig. 4). However, Week 6 (C3D1) median TMPMP also exhibited strong association with OS (Supplementary Fig. 4), aligning with existing literature regarding the optimal time of ctDNA assessment where Week 6 (C3D1) ctDNA concentration according to TMPMP best predicted OS[10].

Integrating radiographic response and ctDNA for survival prediction
In an effort to leverage the distinct prognostic information carried by radiographic assessments and ctDNA measurements, we attempted to integrate both factors for predicting PFS and OS in patients evaluable for both response and ctDNA at the landmark of Week 6 (C3D1) where both radiographic response and ctDNA were assessed. For OS analysis, patients were categorized into molecular low- and high-risk groups based on nadir of median TMPMP up to the landmark of Week 6. In the case of PFS analysis, molecular low- and high-risk patients were defined by a nadir of median AF up to the landmark of Week 6.

As depicted in Fig. 5, the c-index for the OS model utilizing only radiographic response assessments (response vs. non-response) at Week 6 was 0.57 (Fig. 5A). Conversely, the OS model using molecular risk stratification (high vs. low risk) achieved a c-index of 0.63 (Fig. 5B). Notably, the combination of radiographic response and molecular risk stratification markedly enhanced the prediction of OS, yielding a c-index of 0.66 (Fig. 5C).

Similarly, the univariate PFS model relying solely on radiographic response at Week 6 yielded a c-index of 0.57 (Fig. 5D). In contrast, the PFS model incorporating molecular ctDNA stratification achieved a c-index of 0.62 (Fig. 5E). The integration of radiographic assessment and ctDNA measurements resulted in a notable improvement in the prediction of PFS, with a c-index of 0.65 (Fig. 5F).

These findings suggest that integration of radiographic response and ctDNA molecular response further improves the OS and PFS predictions when compared to predictions with radiographic response and ctDNA molecular response alone. Clear, further separation of PFS and OS curves were observed within not only the radiographic response categories (stratified by ctDNA risk groups) but also the ctDNA risk groups (stratified by radiographic response) (Fig. 5C/F).

Discussion
ctDNA is recognized for its potential as an early endpoint in clinical trials, contributing to increased efficiency in both clinical trials and drug development processes as a potential regulatory tool[4,5]. Despite the significant strides made in advocating for the role of ctDNA molecular response as an early endpoint of therapeutic response[6–9,17–21],
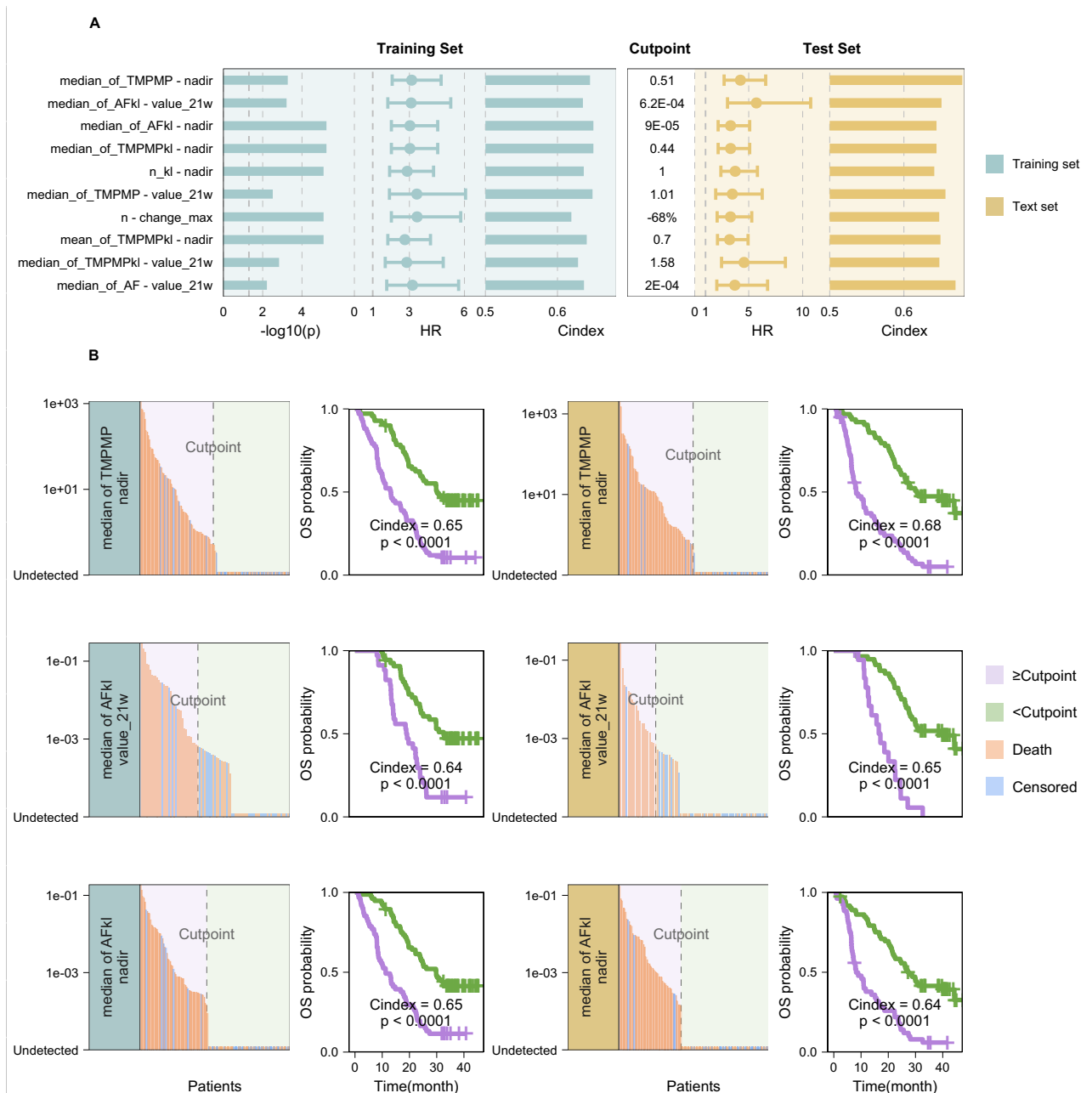
Fig. 4 | Composite ranking of predictive ctDNA features for overall survival (OS). A barplot illustrating −log10 (*P*-values after FDR correction) on the training set; *P*-values are calculated by a two-sided *Z* test using Cox proportional hazards models; forest plots depicting the estimated hazard ratio and barplots showing c-index on both training and test datasets based on the optimal cutoff point identified from the training set (Top 10 ctDNA longitudinal features); error bars represent the exponential of estimated coefficients ± 1.96*SEM (*N* = 262). **B** distributions and Kaplan–Meier plots for training and test sets utilizing optimal cutoff points identified from the training set for the top 3 ctDNA features (nadir of

median TMPMP, median AFkl at Week 21, and nadir of median AFkl); *P*-values are determined using a two-sided log-rank test to compare Kaplan–Meier curves. The exact *P*-values, from left to right by row, are 1.80e-8 (Training set: 68 patients ≥ cutpoint; 69 patients <cutpoint), 6.37e-12 (Test set: 62 patients ≥ cutpoint; 63 patients <cutpoint), 1.17e-5 (Training set: 35 patients ≥ cutpoint; 53 patients <cut-point), 1.93e-9 (Test set: 19 patients ≥ cutpoint; 58 patients <cutpoint), 2.22e-8 (Training set: 70 patients ≥ cutpoint; 67 patients <cutpoint) and 6.49e-8 (Test set: 63 patients ≥ cutpoint; 62 patients <cutpoint). Source data are provided as a Source Data file.

there is currently no consensus regarding the definition of ctDNA molecular response. The definitions of molecular response vary from any reduction[21], a 50% reduction in concentration[8,15,22], to the complete elimination of ctDNA[6,11,16]. Anagnostou et al.'s research suggested that ctDNA clearance according to maxMAF is the most predictive of radiographic responses[11].

Our comprehensive analysis, utilizing the extensive Impower150 dataset, reveals that ctDNA response (e.g., maximum percentage change from baseline based on total or known/likely pathogenic

mutation numbers) following ICI treatment exhibits stronger alignment with radiographic response compared to absolute ctDNA levels. Maximum overall ctDNA response generally proves to be a more effective predictor of radiographic response than snapshot ctDNA response at specific time points. Specifically, a maximum overall response of 20% reduction in total mutation numbers is optimal for reflecting the BOR. Despite this, only modest concordance (AUC = 0.75, sensitivity = 0.84, specificity = 0.62 on the test set) is observed, indicating that ctDNA response is not a surrogate for radiographic
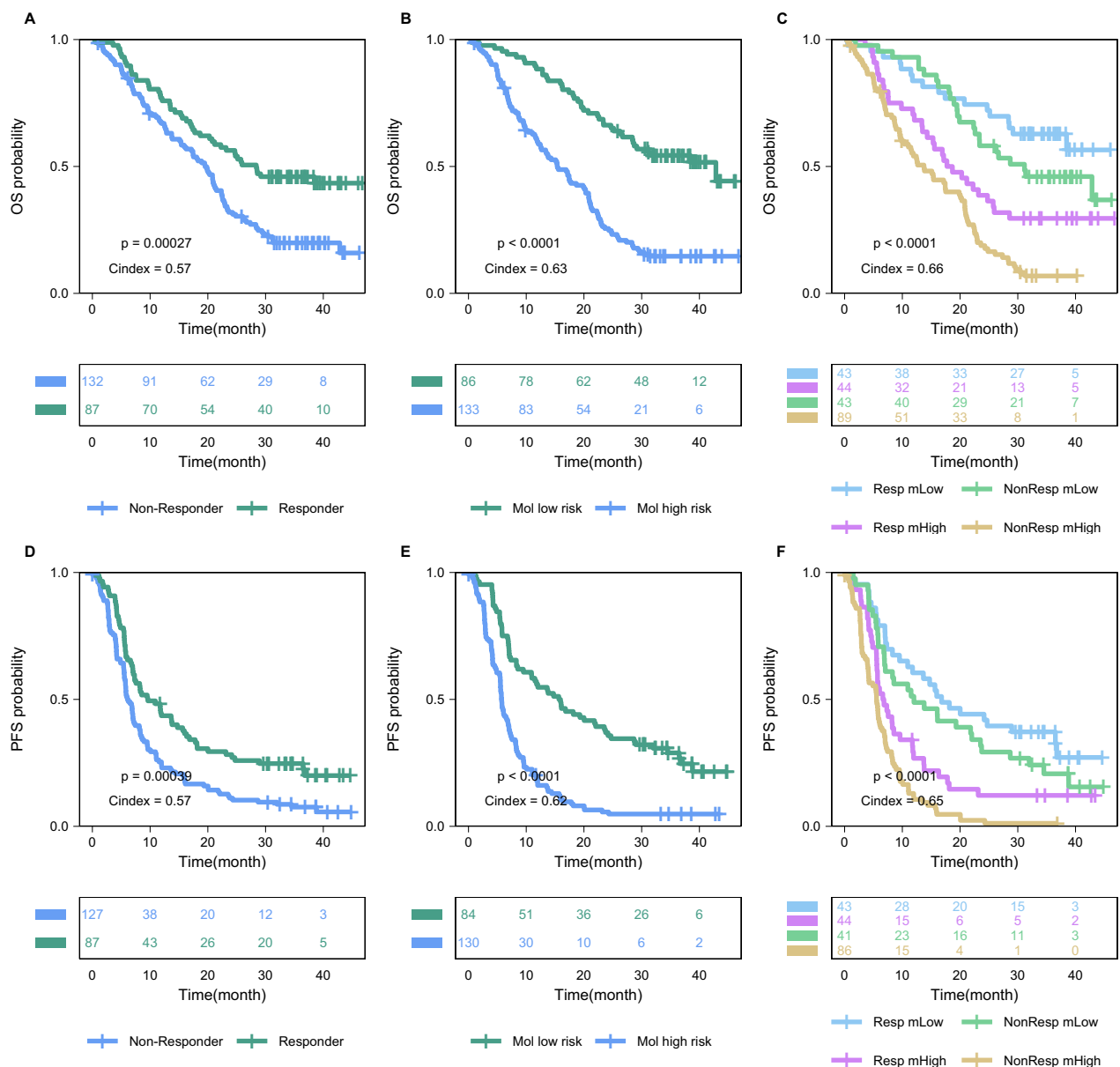
Fig. 5 | . Landmark analysis for comparing prediction of overall survival (OS; N = 219) and progression-free survival (PFS; N = 214) using radiographic response or/ and molecular risk stratification. **A** OS by radiographic response; **B** OS by molecular risk stratification; **C** OS by combining radiographic response and molecular risk stratification; **D** PFS by radiographic response; **E** PFS by molecular risk stratification; and **F** PFS by combining radiographic response and molecular risk stratification. For OS analysis, patients were categorized into molecular low- and high-risk groups based on nadir of median TMPMP up to the landmark of Week 6. In the case of PFS analysis, molecular low- and high-risk patients were defined by a nadir of median AF up to the landmark of Week 6. Radiographic response at Week 6 was used in the landmark analysis; P-values are calculated using a two-sided log-rank test to compare Kaplan–Meier curves, yielding P-values of 2.7e-4 (**A**), 1.24e-10 (**B**), 7.55e-12 (**C**), 3.9e-4 (**D**), 3.57e-10 (**E**), and 1.83e-11 (**F**) respectively. Source data are provided as a Source Data file.

response, and both assessments carry distinct information. The estimated concordance from our analysis is relatively lower than that reported in the smaller BR.36 cohort (AUC = 0.77, sensitivity = 0.82, specificity = 0.75)[11]. Substantial ctDNA decrease is observed in non-responders (SD and PD patients), which may explain the moderate concordance between ctDNA response and radiographic response.

When predicting long-term survival outcomes, a number of absolute ctDNA tumor load (ctDNA levels) surpasses ctDNA response (percent reduction) in performance. Absolute tumor load determined by the overall ctDNA nadir (i.e., nadir of median TMPMP for OS and nadir of median AF for PFS) also outperforms ctDNA tumor load at snapshot time points. The concept of achieving "ctDNA clearance" is similar to our identified nadirs as the depth of response according to

different ctDNA metrics. The identified optimal cutoff values for ctDNA nadirs are very close to undetectable level (Figs. 3B, 4B), supporting clearance of ctDNA tumor load may be clinically meaningful and strongly correlated with improved survival outcomes. Therefore, these data indicate that the nadir of early ctDNA tumor load (or ctDNA clearance) may be a suitable endpoint for evaluating long-term clinical benefits, such as PFS and OS benefits, particularly in the context of IO therapies for NSCLC.

Nevertheless, for some features at specific time points (Supplementary Fig. 4), while patients exhibiting ctDNA clearance at these specific time points were consistently categorized as low risk, a small subset of patients without clearance also showed favorable survival outcomes. This is consistent with a recent report where any ctDNA

decrease from baseline was associated with improved survival outcomes across multiple cutoffs for patients treated with chemoimmunotherapy[23]. In addition, only 20–68% reductions according to maximum change in total mutation numbers were necessary to predict BOR, PFS, and OS (Supplementary Fig. 5B), suggesting that eliminating non-driver mutations may not be required for predicting these outcomes.

Identifying the optimal timing for ctDNA analysis for immunotherapy is challenging. Recent studies have suggested that the optimal time for evaluating ctDNA response falls within the timeframe of 4 to 9 weeks following ICI treatment[6,11,18]. Insights from the BR.36 trial suggest that the most effective time for ctDNA assessment during immunotherapy is at C3D1 (Week 6) after two cycles of pembrolizumab[11]. Additionally, Pellini et al. have demonstrated that continuous monitoring throughout treatment up to C4D1 (Week 9), as well as evaluation at C4D1 alone, provide valuable insights for response assessment and risk stratification for PFS and OS after chemoimmunotherapy[23]. In our present study, ctDNA at Week 6 demonstrated correlation with RECIST-based radiographic response. Conversely, the median allele fraction at Week 9 proved to be the most effective predictors for PFS. Moreover, our analysis revealed that the most significant disparity in ctDNA kinetics between EPD and NEPD occurred at Week 9, further supporting its critical role as the identified optimal time-point for predicting disease progression in 1 L NSCLC. For OS, several Week 21 (C8D1) ctDNA features emerged as top 10 predictors based on the composite ranking, but Week 6 median TMPMP also demonstrate strong predictive performance for OS.

Across BOR, PFS, and OS, integrating the individual composite rankings for these endpoints reveals that maximum percent reduction in total mutation counts consistently and effectively predicts these outcomes (Supplementary Fig. 5). Similarly, across survival outcomes like PFS and OS, the nadirs of median of TMPMP, allele fraction for known/likely pathogenic mutations, and TMPMP for known/likely pathogenic mutations emerged as consistent predictive metrics.

While our analysis has furnished insights into the concordance between molecular response and radiographic response and has established the association of ctDNA with long-term survival outcomes at the individual patient level, future research at the trial level is essential. Evaluating the consistency between the effect size of ctDNA endpoints, and the effect size of survival endpoints across multiple randomized trials is necessary. This research is crucial to position ctDNA as a potential predictor for treatment benefit within a specific study population[5]. Also, due to the continuous nature of most ctDNA features, future research utilizing g-computation for the continuous ctDNA metrics may be more effectively to estimate their causal effects on time-to-event outcomes accounting for confounding factors and time-varying relationships[24]. Moreover, the importance of ctDNA clearance (or nadir close to undetectable) suggests that further refinement is necessary to develop more scalable and sensitive tissue-agnostic ctDNA assays for the detection of ctDNA minimal residual disease (MRD). Such assays should ensure reliable sensitivity while mitigating the risk of false-positive results due to signals from CHIP signals. Examination of mutation-level data indicates that few emergent ctDNA alterations were detected within the first 21 weeks following chemoimmunotherapy, suggesting that this data is more reflective of intrinsic resistance rather than acquired resistance. Sampling over a longer period of time may be required to evaluate acquired resistance to chemoimmunotherapy in 1 L NSCLC. Additionally, besides analyzing a single training/test split, similar results were obtained from a 4-fold cross validation, demonstrating the robustness of the analysis (Supplementary Figs. 6–8).

Finally, our findings underscore the enhanced predictive performance and risk stratification achieved through the integration of radiographic assessments and ctDNA measurements as opposed to relying on individual tools. We showed that combining radiographic assessment and ctDNA measurements provided a significant improvement in the prediction of PFS and OS compared to solely relying on individual assessment tools. These findings provide valuable insights into the optimal utilization of liquid biopsies in IO clinical trials and have the potential to enhance and refine clinical decision-making for patients undergoing IO interventions.

In summary, we identified various ctDNA features predicting outcomes in 1 L NSCLC post chemoimmunotherapy. Nadirs of ctDNA tumor load were primary predictors for PFS and OS, while metrics related to maximum percent reduction from baseline showed modest concordance with radiographic endpoints like BOR. These findings indicate the importance of developing more sensitive tissue-agnostic ctDNA assays for detecting ctDNA MRD in 1 L NSCLC. Combining ctDNA measures with radiographic assessments enhances long-term survival predictions, indicating its potential and usefulness in assessing antitumor response and possibly serving as an accelerated endpoint for drug approvals when paired with radiographic response and safety data. Additionally, our established optimal ctDNA cutoffs allow for risk stratification, aiding in clinical trial design to enrich high-risk patients and potentially reducing trial sizes and durations. Conversely, low-risk patients may avoid unnecessary toxicity. The risk stratification may also optimize therapy switching via ctDNA monitoring in real-world clinics (e.g., intensified therapy for high-risk patients, de-escalation for low-risk). Finally, ctDNA assessments from Week 6 to Week 9 can provide useful insights on different clinical outcomes (BOR, PFS, and OS) although Week 21 also appears to be important for projecting OS in 1 L NSCLC following chemoimmunotherapy.

Methods

Study and patient population

The current retrospective analysis utilized anonymized data from participants of the IMpower150 trial (NCT02366143), a pivotal phase III, open-label, randomized clinical study designed to evaluate the efficacy and safety of various treatment regimens in patients with stage IV or recurrent metastatic non-squamous non-small cell lung cancer (NSCLC) who had not previously undergone chemotherapy. Comprehensive details on the study's methodology, including eligibility criteria, have been previously published[12,13]. Participants were randomly allocated into one of three treatment groups: the ACP group (receiving atezolizumab, carboplatin, and paclitaxel), the ABCP group (administered atezolizumab, bevacizumab, carboplatin, and paclitaxel), or the BCP group (treated with bevacizumab, carboplatin, and paclitaxel). The treatment was administered until the occurrence of disease progression, intolerable adverse effects, or patient withdrawal. Sex was determined by self-reporting and both males and females were enrolled in this study. Details on trial periods and clinical sites are available at https://clinicaltrials.gov/study/NCT02366143?tab=table.

Clinical endpoints

The response to treatment via tumor image classified in the RECIST system[25] including complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Patients achieving a best response of CR or PR are classified as responders (CR + PR), while those with a best response of SD or PD are termed non-responders (SD + PD).

The measurement of survival benefit is estimated by OS which refers to the duration from randomization to death for any reason, and PFS, assessed by investigators according to RECIST v1.1 criterion.

ctDNA sample collection and sequencing

Longitudinal ctDNA samples for baseline and treatment assessment were collected during the screening phase and on the first day of the subsequent four treatment cycles (C2D1 [Week 3], C3D1 [Week 6], C4D1 [Week 9], C8D1 [Week 21]) of the study[10]. Details of ctDNA sample collection, processing, and development of the ctDNA assay were

described elsewhere[10,26,27]. Briefly, baseline and on-treatment plasma samples were analyzed separately. The baseline plasma samples underwent retrospective analysis using a previously established assay method[26]. Subsequent on-treatment samples were then assessed with a custom 330 kb assay targeting 311 genes[27]. To construct the assay's hybrid capture panel, alterations detected in baseline samples were combined and refined, excluding known germline variants, CHIP genes (*TET2, DNMT3A, CBL, PPM1D, CHEK2, JAK2, ASXL1, SF3B1*), noncoding variants, and repetitive regions with <100× coverage. Paired analysis of white blood cells was implemented to further filter germline and CHIP variants for all collected samples at different timepoints. The assay demonstrated >99% sensitivity for short variants with allele frequency >0.5%, >95% sensitivity for allele frequency ranging from 0.25% to 0.5%, and 70% sensitivity for allele frequency between 0.125% and 0.25%. Additionally, no false positives were identified in 62 samples from healthy volunteers[27]. Reportable genomic alterations were identified as known or likely functional driver alterations based on their presence in the Catalog of Somatic Mutations in Cancer (COSMIC)[28]. The limit of detection (LOD) of the assay were determined to be at an allele frequency of 0.1%.

Types of ctDNA longitudinal profiles

Six different ctDNA measurements were used to characterize ctDNA levels. These measurements include cell-free DNA concentration (cfDNA), allele frequency (AF), AF with known or likely pathogenic mutations (AF_kl), tumor molecules per ml plasma (TMPMP), TMPMP_kl, and the number of mutations. For AF, AF_kl, TMPMP, and TMPMP_kl, three summary statistic metrics (mean, median, and maximum values) were calculated for each type of ctDNA measurement. As a result, 16 types of ctDNA longitudinal profiles were derived to describe the longitudinal evolution of the ctDNA data (Supplementary Table 1)[10].

Longitudinal ctDNA features

We extracted 12 different features from each type of ctDNA longitudinal profile. These features are (1) overall nadir (the lowest level at any time point) for a patient, (2) maximum overall percent reduction/change from baseline (representing maximum ctDNA/molecular response), (3) time to nadir, (4) percent reduction/change from baseline at snapshot time points (C2D1, C3D1, C4D1, and C8D1), and (5) absolute level at snapshot time points (baseline, C2D1, C3D1, C4D1, and C8D1). Analysis of covariance revealed low to modest correlations among most of the studied ctDNA features, with 80% of correlations below 0.4 and 97% below 0.8 (data not shown).

Statistical analysis

The partition of the training/test sets of this dataset is detailed elsewhere[10], divided 1:1 at baseline across patients with ctDNA measurements in the ABCP, ACP, and BCP arms ($n = 466$). After excluding the patients from the BCP cohort and the patients without detectable ctDNA (zero mutations detected) at baseline, a total of 262 ctDNA-evaluable patients treated with chemoimmunotherapy (ABCP and ACP) were included in this study. The training dataset included 137, 130, 117, 114, and 88 patients while 125, 123, 103, 96, and 77 patients were included in the test data at baseline, C2D1, C3D1, C4D1, and C8D1, respectively.

The concordance between radiographic responses (including responder vs. non-responder) and individual ctDNA features was analyzed using logistic regression models. AUC, representing the area under the ROC curve, was used to measure classification performance. Optimal cut-points of ctDNA features were identified by maximizing the Youden index for the binary outcomes utilizing the "cut-pointr" function in the R "cutpointr" package[29]. The model's performance was assessed through a confusion matrix, which determined sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV). These metrics were derived from the model's predictions using optimal cut points and were compared against the actual outcomes in both the training and test datasets.

The relationship between OS/PFS and ctDNA features was evaluated using a Cox proportional hazards model implemented in the "survival" package in R 3.5.3[30]. The predictive performance of the Cox model was assessed using the c-index[31]. Optimal cut-points for the ctDNA features in the survival analysis were identified by selecting the maximal rank statistics utilizing the "surv_cutpoint" function in the R "survminer" package[32]. All ctDNA features are first modeled as continuous variables using the training set to calculate *p*-values, which are reported as two-sided after using Benjamini–Hochberg procedure to correct for multiple testing and false discovery rate (FDR).

To compare different ctDNA features, we implemented a composite ranking system to incorporate statistical significance (p-value), effect size (odds ratio for radiographic response and hazard ratio for survival endpoints), and concordance with observed clinical outcomes (AUC for radiographic response and c-index for survival endpoints). The ctDNA feature exhibiting the lowest p-value, largest effect size, or greatest concordance is assigned a rank of 1. The composite ranking is derived by summing the individual rankings of p-value, effect size, and concordance from the training set, along with the effect size and concordance from the test set.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The ctDNA/clinical data used in this study are available under restricted access in the European Genome-Phenome Archive (EGA) under accession code EGAS00001006703. Access was obtained upon request via EGA. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

References

1. Rossi, G. & Ignatiadis, M. Promises and pitfalls of using liquid biopsy for precision medicine. *Cancer Res.* **79**, 2798–2804 (2019).
2. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
3. Narayan, P. et al. State of the science and future directions for liquid biopsies in drug development. *Oncologist* **25**, 730–732 (2020).
4. FDA draft guidance for industry: use of circulating tumor DNA for early stage solid tumor drug development; Available from: https://www.fda.gov/media/158072/download (2022).
5. Vellanki, P. J. et al. Regulatory implications of ctDNA in immuno-oncology for solid tumors. *J. Immunother. Cancer* **11**, e005344 (2023).
6. Anagnostou, V. et al. Dynamics of tumor and immune responses during immune checkpoint blockade in non-small cell lung cancer. *Cancer Res.* **79**, 1214–1225 (2019).
7. Vega, D. M. et al. Changes in circulating tumor DNA reflect clinical benefit across multiple studies of patients with non-small-cell lung cancer treated with immune checkpoint inhibitors. *JCO Precis. Oncol.* **6**, e2100372 (2022).
8. Goldberg, S. B. et al. Early assessment of lung cancer immunotherapy response via circulating tumor DNA. *Clin. Cancer Res.* **24**, 1872–1880 (2018).
9. Guibert, N. et al. Targeted sequencing of plasma cell-free DNA to predict response to PD1 inhibitors in advanced non-small cell lung cancer. *Lung Cancer* **137**, 1–6 (2019).

10. Assaf, Z. J. F. et al. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer. *Nat. Med.* **29**, 859–868 (2023).

11. Anagnostou, V. et al. ctDNA response after pembrolizumab in non-small cell lung cancer: phase 2 adaptive trial results. *Nat. Med.* **29**, 2559–2569 (2023).

12. Socinski, M. A. et al. IMpower150 final overall survival analyses for atezolizumab plus bevacizumab and chemotherapy in first-line metastatic nonsquamous NSCLC. *J. Thorac. Oncol.* **16**, 1909–1924 (2021).

13. Socinski, M. A. et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N. Engl. J. Med.* **378**, 2288–2301 (2018).

14. Collisson, E. A. et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

15. Thompson, J. C. et al. Serial monitoring of circulating tumor DNA by next-generation gene sequencing as a biomarker of response and survival in patients with advanced NSCLC receiving pembrolizumab-based therapy. *JCO Precis. Oncol.* **5**, 510–524 (2021).

16. Sivapalan, L. et al. Dynamics of sequence and structural cell-free DNA landscapes in small-cell lung cancer. *Clin. Cancer Res.* **29**, 2310–2323 (2023).

17. Váraljai, R. et al. Application of circulating cell-free tumor DNA profiles for therapeutic monitoring and outcome prediction in genetically heterogeneous metastatic melanoma. *JCO Precis. Oncol.* **3**, PO.18.00229 (2020).

18. Seremet, T. et al. Undetectable circulating tumor DNA (ctDNA) levels correlate with favorable outcome in metastatic melanoma patients treated with anti-PD1 therapy. *J. Transl. Med.* **17**, 303 (2019).

19. Kim, S. T. et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat. Med.* **24**, 1449–1458 (2018).

20. Keller, L. et al. Early circulating tumour DNA variations predict tumour response in melanoma patients treated with immunotherapy. *Acta Derm. Venereol.* **99**, 206–210 (2019).

21. Bratman, S. V. et al. Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. *Nat. Cancer* **1**, 873–881 (2020).

22. Zhang, Q. et al. Prognostic and predictive impact of circulating tumor DNA in patients with advanced cancers treated with immune checkpoint blockade. *Cancer Discov.* **10**, 1842–1853 (2020).

23. Pellini, B. et al. Circulating tumor DNA monitoring on chemo-immunotherapy for risk stratification in advanced non-small cell lung cancer. *Clin. Cancer Res.* **29**, 4596–4605 (2023).

24. Seaman, S., Dukes, O., Keogh, R. & Vansteelandt, S. Adjusting for time-varying confounders in survival analysis using structural nested cumulative survival time models. *Biometrics* **76**, 472–483 (2020).

25. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).

26. Gandara, D. R. et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat. Med.* **24**, 1441–1448 (2018).

27. Clark, T. A. et al. Analytical validation of a hybrid capture-based next-generation sequencing clinical assay for genomic profiling of cell-free circulating tumor DNA. *J. Mol. Diagn.* **20**, 686–702 (2018).

28. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, 777–783 (2017).

29. Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks; Available from: https://cran.r-project.org/web//packages/cutpointr/cutpointr.pdf (2022).

30. Survival Analysis; Available from: https://cran.r-project.org/web/packages/survival/survival.pdf (2023).

31. Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B. & Wei, L. J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* **30**, 1105–1117 (2011).

32. Drawing Survival Curves using 'ggplot2'; Available from: https://cran.r-project.org/web//packages/survminer/survminer.pdf (2021).

Author contributions

X.S.X., Y.Y., and M.Y. contributed to design of the research. M.Y., X.S.X., and Y.Y. contributed to data acquisition. H.D., M.Y., and X.S.X. contributed to data analysis. X.S.X., H.D., Y.Y., and M.Y. contributed to data interpretation. X.S.X., H.D., M.Y., and Y.Y. wrote the manuscript; and all authors critically reviewed the manuscript and approved the final version.

Competing interests

X.S.X. is an employee of Genmab, Inc. Genmab did not provide funding for this study. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51316-7.

Correspondence and requests for materials should be addressed to Min Yuan or Xu Steven Xu.

Peer review information *Nature Communications* thanks Christian Rolfo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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