Individualized tumor-informed circulating tumor DNA analysis for postoperative monitoring of non-small cell lung cancer

Kezhong Chen, MD[1,2,3,7,†,*], Fan Yang, MD[2,†], Haifeng Shen, MD[2,†], Chenyang Wang, PhD[4,†], Xi Li, PhD[4], Olga Chervova, PhD[5], Shuailai Wu, PhD[4], Fujun Qiu, MS[4], Di Peng, PhD[4], Xin Zhu, PhD[4], Shannon Chuai, PhD[4], Stephan Beck, PhD[5], Nnennaya Kanu, PhD[5], David Carbone, MD&PhD[6], Zhihong Zhang, PhD[4,*] , Jun Wang, MD[2,*]

Affiliations:

[1] Thoracic Oncology Institute, Peking University People's Hospital, Beijing, 100044, China

[2] Department of Thoracic Surgery, Peking University People's Hospital, Beijing, 100044, China

[3] Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, University College London, 72 Huntley St, London, WC1E 6DD, UK;

[4] Burning Rock Biotech, Guangzhou, 510300, China

[5] University College London Cancer Institute, University College London, 72 Huntley St, London, WC1E 6DD, UK

[6] James Thoracic Oncology Center, Ohio State University, Columbus, OH 43026, USA

[7] Lead contact

*To whom correspondence should be addressed: Email: chenkezhong@pkuph.edu.cn (K.C.); zhihong.zhang@brbiotech.com (Z.Z.); wangjun@pkuph.edu.cn (J.W.)

† These authors contributed equally to the work.

Summary：

We report a personalized tumor-informed technology, Patient-specific pROgnostic and Potential tHErapeutic marker Tracking (PROPHET) using deep sequencing of 50 patient-specific variants to detect molecular residual disease (MRD) with a limit of detection of 0.004%. PROPHET and state-of-the-art fixed-panel assays were applied to 760 plasma samples from 181 prospectively enrolled early-stage non-small cell lung cancer patients. PROPHET shows higher sensitivity of 45% at baseline with circulating tumor DNA (ctDNA). It outperforms fixed-panel assays in prognostic analysis and demonstrates a median lead-time of 299 days to radiologically confirmed recurrence. Personalized non-canonical variants account for 98.2% with prognostic effects similar to canonical variants. The proposed TNMB classification surpasses TNM staging for prognostic prediction at the decision point of adjuvant treatment. PROPHET shows potential to evaluate the effect of adjuvant therapy and serve as an arbiter of the equivocal radiological diagnosis. These findings highlight the potential advantages of personalized cancer techniques in MRD detection.

Introduction:

Patients with localized non-small cell lung cancer (NSCLC) are generally offered treatments with curative intent [1]. Early detection of disease recurrence is conducive to improving the cure rate and prolonging survival. Current routine clinical surveillance after surgery involves serial radiographic imaging. However, the utility is limited by its low sensitivity, radiation exposure, and the demand for equipment and facility, hence leaving unmet needs for dynamic postoperative monitoring [2]. Such surveillance tests often lag behind the actual disease recurrence and become even more challenging in patients with equivocal scans due to posttreatment normal tissue changes. Therefore, tests to detect molecular residual disease (MRD) biomarkers before recurrence with superior sensitivity and precision could potentially help improve the clinical decision-making paradigm.

Recent advances in the detection of ctDNA show promise in reflecting MRD and monitoring recurrence in patients with nonmetastatic cancer, including NSCLC [3-5]. However, MRD assays face many technical challenges. The variant allele fraction (VAF) of mutations detected in the plasma reflects the amount of ctDNA released, usually proportional to the tumor volume [6]. After curative-intent treatment, the ctDNA signal from the residual disease imputed by mean VAF drops to 0.1%−0.003% in NSCLC [6,7], while ctDNA was undetectable in the majority of patients. Detection of extremely low ctDNA concentration is also impeded by a limited volume of blood drawn and/or a low amount of cell-free DNA. Therefore, assay sensitivity is crucial for MRD monitoring in patients with low disease burden.

The technologies applied for ctDNA detection could be broadly classified as tumor-informed and tumor-agnostic. The sensitivity of MRD detection can theoretically be enhanced by prior knowledge of the genetic profiling information from the corresponding tumor tissue (tumor-informed), thus enabling monitoring the presence of *a priori* tumor cell-carried mutations in plasma samples [6,8], as opposed to variant-calling without prior information about tumor mutations (tumor-agnostic). In tumor-

informed MRD assays, the personalized assay and the fixed-panel assay are used. The personalized approach allows deeper sequencing exclusively targeting mutations known to exist in the tumor, hence exponentially increasing the opportunity of capturing and identifying ctDNA fragments carrying targeted mutations. Previously, the TRACERx study uses the personalized assay targeting up to 200 variants per patient, achieves a specificity of 99.3%, and reveals that ctDNA is detected at or before relapse in 82.2% (37/45) patients who experience a relapse of their primary tumor [9]. In its latest updates, the authors claim 119 days of overall median lead time for non-recurrent NSCLC patients after 4.6 years of median follow-up time, reaching a longitudinal specificity of 95% [10]. Recently, the LUCID study uses personalized assays tracking up to 48 variants per patient and reports ctDNA presence after treatment in 64.3% of patients who had clinical recurrence [11]. Zhang et al. utilize a tumor-informed fixed-panel MRD assay (338-gene panel) and identify 87.2% (41/47) relapsed patients within 19.7 months (median follow-up time) [12]. Direct comparison of the sensitivity and specificity of different MRD assays proposes great challenges because of differences among the study cohorts. Therefore, one of the main purposes of our study was to compare the detection and quantification of MRD through personalized, tumor-informed, and tumor-agnostic fixed-panel assays in the same patients using a head-to-head approach.

Recent studies have made consistent observations that MRD presence after surgery is strongly correlated with shorter disease-free survival (DFS) (landmark hazard ratio [HR]: 3.95–14.8; longitudinal HR: 8.55–50)[12-15]. However, several important questions remain to be elucidated. First, what is the optimal MRD testing timepoint at the landmark, taking into consideration both accuracy and convenience? Second, when a positive/negative MRD status is determined in either a landmark or surveillance scenario, how do the odds of having a relapse or remaining relapse-free change over time? This leads to another important question – how often should plasma samples be collected during long-term surveillance? Third, it is intuitive and well-established that tracking canonical mutations (tier I & II)[16] is important to MRD

monitoring[1]. However, how much additional sensitivity can be gained by adding the non-canonical mutations (non-tier I & II)[16] in MRD monitoring? Is MRD positivity detected from non-canonical mutations also predictive for poor prognosis as canonical mutations? Fourth, MRD has been consistently reported to prelude radiological detection in relapse monitoring with a lead time of 3 to 7 months [11,12,17]. Therefore, it would be intriguing to see whether such MRD tests could help make earlier and more accurate clinical decisions in the case of equivocal radiological findings. Finally, previous studies mainly focused on binary ctDNA status, while the utility of ctDNA quantitative dynamics during long-term surveillance has been rarely analyzed. Could the ctDNA quantitative change further help improve predictive accuracy and precision beyond the binary status?

Here, we report results from a prospective study in patients with early-stage NSCLC, aiming to investigate the clinical utility of a Patient-specific pROgnostic and Potential tHErapeutic marker Tracking (PROPHET) technology for ctDNA detection after surgery, for some patients, and after recurrence.

Results:

Patient Characteristics and Experiment Design

Using personalized PROPHET designed from WES data, we analyzed 760 plasma samples from 181 patients prospectively recruited for MRD analysis from the MEDAL cohort (NCT03634826, Figure 1). This cohort predominantly consisted of patients aged 62.7 years old with 58.0% males, including stage I lung cancer at diagnosis (n = 115, 63.5%) and adenocarcinoma (n = 145, 80.1%, Table S1). The overall median follow-up period was 1071 days (987−1137 days). In total, 55 patients had adjuvant therapies, out of which 48 received chemotherapy (87.3%). A total of 7589 single nucleotide variants (SNVs) and insertions/deletions (INDELs) were included in the panel design, each panel containing selected genomic sites from two patients (n = 91) or one patient (n = 4) (Table S2). Among these sites, most (7511, 99.0%) were private mutations to a specific patient, while only 0.04% were shared

among no less than 3 patients. At the same time, we applied a previously validated LungPlasma ctDNA assay [18] covering 168 lung cancer-related genes to the same cohort with two different methods to identify ctDNA variants (tumor-agnostic and tumor-informed), in order to compare the three MRD methods directly (Table S3, see STAR Methods).

Establishment of PROPHET

To determine the optimal conditions for PROPHET, the dynamic change for LOD was first simulated and plotted for different combinations of the number of monitored variants (m) and effective sequencing depth (n), with background noise rate $(q) = 0.0001\%$ (Figure 2A, see STAR Methods). For each designated LOD level, the effective sequencing depth tends to saturate after 50 variants. A pilot study was then performed using five human cancer cell lines (H2009, H2126, HCC1395, H1437, and HCC38) to mimic tumor cells and a transformed B lymphocyte cell line GM24385 as the normal. A panel integrating 353 tumor cell line-specific variants from the 5 cell lines was applied to a series of contrived samples mimicking different "ctDNA" fractions from 0.0008% to 0.5%, prepared using enzyme-fragmented DNA from these 5 cancer cell lines and diluted into GM24385 as the background (Table S4). Each contrived sample was prepared using Burning Rock UHS DNA library prep kit, enriched by the 353-variant panel, and sequenced to >100,000× raw coverage. The raw sequencing data were analyzed to corresponding cell line variants. Random sampling of different numbers of variants on each cell line mix was applied to assess the detection sensitivity. The site-specific background noise was in the range of 0.1% to 0.00001% for each targeted mutation (Figure S1A–D, see STAR Methods) with the 90th percentile of SNVs at 0.002%. For each iteration, the significance site was defined as a cutoff of p-value < 0.05, and the sample was deemed to be positive if two or more sites were significant and the sample-level p-value < 0.005 (see STAR Methods). As expected, the result showed that more variants lead to more sensitive MRD detection for contrived samples while the LOD improvement tends to become marginal after more than 50 variant sites selected (Figure 2B). To further evaluate the

false positive rate, the sample level p-value of those non-intended variants of each contrived sample was evaluated. For example, for HCC38 contrived sample, the 5–60 variants designed for H2009, H2126, HCC1395, and H1437 were randomly selected to perform PROPHET and evaluated, in which the procedure was repeated 200 times. With a threshold of sample-level p-value < 0.005 of the PROPHET algorithm, the false positive rates were all under 0.5% (Figure 2C). In this study, the significant site ratio was calculated for each iteration stated in Figure 2B. Even when samples were diluted to 0.0008%, cancer cell lines showed a significant difference compared with the negative cell line, GM24385 ($p < 0.001$, Figure 2D), indicating the potential low LOD capability of the PROPHET assay. In conclusion, the PROPHET assay was established with minimal 20 ng cfDNA input up to 50 personalized variants, 100,000× raw depth of sequencing.

LODs Study and Specificity Confirmation

The LODs and specificity studies were conducted using a newly designed 100-variant H2009-specific panel. Fragmented H2009 DNA was serial diluted into GM24385 background with a proportion of 0.0008%–0.1% using the gravimetric method. Mixed samples with different DNA inputs (20–100 ng) were tested by PROPHET. For DNA input between 20 ng and 60 ng, the impact on effective sequencing depth by target sequencing depth revealed a tendency to saturate at 100,000× (Figure 2E). To assess the LOD at 50 loci for the PROPHET design, similar to the above cell-line study, 200 iterations of 50 randomly selected loci were applied to evaluate sample level sensitivity and specificity. The sample level significance was determined by the threshold aforementioned. Using Probit analysis, the LODs at 20 ng and 60 ng input were 0.0036% and 0.0018%, respectively (Figure 2F). In order to assess the specificity, randomly selected 50 unaltered positions, which is proximate to the intended sites, were selected. Using these sites, the sample-level specificity was greater than 99% at all cfDNA input amounts and dilution ratios (Figure 2G). A near-perfect correlation (Pearson, $R = 0.99$, $p < 0.001$) was shown between estimated and expected ctDNA fractions (Figure 2H). Significant site ratios at different levels of

DNA input levels were evaluated (Figure 2I). Even at a 0.002% dilution level, the detected significant site ratio could be differentiated from the negative samples, and the higher input amounts showed higher significant site ratios.

In a separate validation study using Seracare ctDNA MRD Panel Mix, a commercially available reference material, the contrived samples containing 4 gradients of spiked ratios (0, 0.005%, 0.05%, and 0.5%) and 3 replicates for each gradient (Table S4) demonstrated the LOD and accuracy similar to the above experimental results (Figure S1E and F), further verifying the analytical performance of PROPHET.

Clinical validation with pre-operative plasmas

ctDNA was detected by PROPHET in 63 of 151 pre-operative plasma samples (41.7%), with a positive rate of 15.7% (11/70), 40.0% (12/30), 75.0% (21/28), and 82.6% (19/23) among stage IA, IB, II, and III patients, respectively (Figure 3A and S2A). In this study, the majority of personalized panels contain tracking variants designed for two patients. For each patient's blood sample tested, when the PROPHET analysis was performed using the opposite variant set designed for the other patient, a negative result would be expected. We used this method to assess the specificity. Among the 743 ctDNA samples produced on 91 panels, 18 samples were detected as positive, giving a specificity estimate of 97.6% (Table S4). Patients with lung squamous cell carcinomas (LUSC) showed higher ctDNA fraction levels (Figure S2B), and larger tumor volume was significantly positively correlated with higher levels of ctDNA fraction (Figure S2C). PROPHET showed a higher positive rate than the tumor-informed and tumor-agnostic fixed-panel assays (41.7% vs. 21.9% vs. 17.2%, $p < 0.001$, Figure 3A). The ctDNA fraction estimated by PROPHET reached as low as 0.001% (Table S3) and increased as clinical stage (Kruskal-Wallis, $p = 2e{-}07$, Figure 3B). It was significantly correlated with the maximum allelic fraction (maxAF) estimated by the tumor-informed fixed-panel assay (Pearson $R = 0.68$, $p = 1.4e{-}5$, Figure 3C).

We next explored whether patient-specific mutation tracking translated into superior sensitivity. The median ctDNA fraction of the 30 patients detected by PROPHET alone was significantly lower than the 25 patients detected by all three MRD assays (0.01% vs. 1.0%, p = 2.9e−12, Figure 3D). For these 30 patients, 96.9% (377/389) of detected mutations by PROPHET were not included in the fixed-panel assays (Figure 3E). Moreover, PROPHET demonstrated a higher predictive power for DFS and overall survival (OS) at baseline than the tumor-informed fixed panel and tumor-agnostic assays (DFS HR: 14.11 vs. 4.57 vs. 4.37, OS HR: 8.84 vs. 4.99 vs. 6.64, Figure S2D and E).

We further compared genetic alterations and clinicopathological factors between shedders (refers to ctDNA positive before operation) and non-shedders (refers to ctDNA negative before operation either due to tumor biological features or technical limitations). Regardless of subtypes, shedders carried more abundant somatic mutations with significantly higher *TP53* mutation frequency (Figure 3F) and tumor mutation burden (TMB) value ($p < 0.001$, Figure S2F) than non-shedders. Shedders were also correlated with several clinicopathologic variables (Figure S2G). After propensity score matching (PSM) adjustment for TMB and other shedding-related clinical variables, the pathways of P53 signaling and transcriptional mis-regulation in cancer were enriched among shedders (Figure S2H).

Prognosis prediction by landmark MRD status

For primary landmark analysis, plasma samples at timepoint B (3–7 days post-operation, n = 172) and C (1 month post-operation, n = 162) were collected and assessed. The landmark MRD status determined by PROPHET showed a significant association with DFS (timepoint B HR 5.31, $p < 0.001$; timepoint C HR 16.40, $p < 0.001$, Figure 4A). While timepoint C appeared to have a notably greater association measured by HR with DFS than timepoint B, the difference resides primarily between the MRD-positive groups ($p = 0.018$), but not between the MRD-negative groups ($p = 0.692$). Given the convenience of blood collection at timepoint B and similar DFS

between the two MRD-negative patient groups identified at timepoints B and C, we propose a hypothetical landmark MRD testing strategy in clinic as such a patient is first assessed based on their plasma sample collected at timepoint B. If and only if the first MRD status is positive, another sample would be collected for reevaluation at timepoint C, and the landmark MRD status would be determined by the timepoint C result (Figure 4B). Hypothetical application of such a "sequential double-check" strategy in our study showed higher HR (10.88 vs. 5.31, Figure S3A) and positive predictive value (PPV, 78.6% vs. 53.6%, Figure 4C) than that of timepoint B, with 1-month plasma reevaluation at timepoint C only needed for 17.9% (28/156) patients (Figure 4A), hence enabling earlier decision for MRD negative patients.

In landmark MRD detection, it is critical to interpret how its predictive power for prognosis changes during the follow-up period. The dynamic change over time for the negative predictive value (NPV) of the PROPHET MRD test at timepoint C was plotted in Figure 4D, which was calculated as the percentage (*y*-axis) of patients who had not recurred at the corresponding length of time from the landmark test. In particular, the NPV was 99.3%, 95.1%, and 90.9% at 120 days, 330 days, and 600 days, respectively.

Similar to previous studies [12], there was no DFS benefit from adjuvant therapy (ADT) for landmark MRD-negative patients after PSM adjustment of clinical factors (Figure S3B and C). Due to insufficient patients in the non-adjuvant therapy MRD-positive subgroup, the analysis of ADT effectiveness in the MRD-positive group was not feasible (Table S5).

In comparison, the predictive power for DFS and OS by PROPHET was stronger than the fixed panel assays (Figure S3D–G). Seven patients were MRD-positive by both PROPHET and tumor-informed fixed-panel assay, while PROPHET identified an additional 12 MRD-positive patients at timepoint C (Figure S3H). Those patients MRD-positive in both assays had significantly worse DFS than MRD-negative patients determined by both assays (HR 9.37, $p < 0.001$; PROPHET positive vs. both

negative HR 21.12, $p < 0.001$), and MRD positives detected in both assays and PROPHET only showed no significant difference from each other ($p = 0.174$, Figure S3H).

To evaluate how much the non-canonical mutations (non-tier I & II) [16] contributed to MRD detection, we compared the prognostic value of MRD-positive patients with or without non-canonical mutations detected. Out of 117 patients who had at least one canonical mutation (tier I & II) [16] included in panel design, 17 were identified to be landmark MRD-positive, of which 8 had no canonical variants detected yet demonstrating significantly worse DFS than MRD-negative patients, indicating tracking non-canonical variants is also beneficial for MRD detection as tracking canonical variants (HR 2.19, $p = 0.143$, Figure 4E).

Previous studies presented an additive effect of the landmark MRD with clinical stage and tumor size on DFS prediction [14,15]. In multivariable Cox regression study, the landmark MRD positivity by PROPHET was the only independent factor associated with prognosis ($p < 0.001$, Figure 4F), accounting for 85.6% of the predictive power for DFS (Figure 4G), implying its dominating prognostic value over tumor size and TNM stage, which was not observed in the multivariable Cox regression of fixed-panel assays (Figure S4A and B).

Because of the outstanding prognostic power of the landmark MRD status by PROPHET, we attempted to define a new indicator, TNMB (tumor-node-metastasis-blood) classification for prognostic prediction (Figure 4H). In MEDAL, TNMB showed a better prognostic effect than the TNM stage (Chi-square test, $p < 0.001$, Figure 4I). Then, we have also validated the predictive power of our PKU (Peking University) TNMB indicator in two previously published datasets (LUNGCA-1 [14] and ChiCTR1900024656 [15]; Figure S4C–E), in which TNMB also demonstrated a more desirable prognostic power for DFS than TNM staging.

Dynamic measurement of longitudinal MRD

In total, 347 plasma samples from 110 patients, who have samples available at timepoint C and at least one follow-up visit, were included in the longitudinal analysis (Figure 5A and S5A). Patients were defined as longitudinally MRD-positive if they were MRD-positive at any timepoint. DFS was greatly significantly associated with PROPHET longitudinal MRD positive, with HR of 37.42 ($p < 0.001$, Figure S5B), sensitivity of 84.2% (16/19), and NPV of 96.6% (84/87), outperforming landmark MRD (Figure 5B). Similar trends were observed by other assays, though with lower HRs (Figure S5B and C). The median lead time of MRD detection by PROPHET to radiological recurrence was 299 days for 84.2% (16/19) relapsed patients, compared with 95 days for 31.6% (6/19) relapse patients in tumor-informed fixed-panel assays, and 129 days for 31.6% (6/19) relapse patients in tumor-agnostic assays, respectively (Figure S5D). The NPV of longitudinal MRD maintained at 99.4%, 98.4%, and 95.5% at 180 days, 360 days, and 720 days, respectively, (Figure 5C), which was calculated as the percentage of patients who had not recurred within a projected interval time from each longitudinal test.

Accounting for more high-risk clinical factors, the longitudinal MRD status was the dominant prognostic indicator with an 85.2% contribution in the multivariable Cox model (Figure 5D and E), with far greater significance than fixed-panel assays (Figure S5E). Stage I and stage II–III patients had similar DFS when grouped by longitudinal MRD status (MRD+ group HR 0.95, $p = 0.926$; MRD− group HR 7.36, $p = 0.103$, Figure 5F). Remarkably, while significantly fewer non-shedders were longitudinally MRD-positive than shedders (4.8% [3/63]) vs. 45.9% [17/37], $p < 0.001$, Figure 5G), those non-shedders who were MRD positive still had significantly worse outcomes (HR 48.4, $p = 0.002$). These results indicate that non-shedders could also benefit from MRD monitoring, and their preoperative undetectable ctDNA might be attributed to low-shedding rather than non-shedding.

We further explored the utility of longitudinal MRD for assessing the efficiency of ADT. Among 110 patients, 6 patients were MRD positive at landmark timepoint C and had additional MRD tests at post-ADT. All four patients with detectable MRD at

post-ADT eventually relapsed, while the other two patients with post-ADT MRD-negative remained disease-free by the end of the follow-up (Figure 5H). In addition, among seven patients with epidermal growth factor receptor (*EGFR*) Leu858Arg (L858R) mutation at the landmark, two stage III patients received standard chemotherapy. Both patients experienced recurrence during the follow-up (e.g. MEDAL-026). One patient (stage III, MEDAL-152) who received EGFR-tyrosine kinase inhibitors (TKIs) maintained disease-free till the last follow-up (Figure S5A).

ctDNA quantification for personalized disease monitoring.

To explore the utility of the capability of ctDNA quantification, we defined a dynamic changing rate (DCR) of ctDNA fraction between each pair of adjacent tests (see STAR Methods) in the longitudinal analysis. Among the MRD-positive patients, those who relapsed had a significantly larger DCR than those who remained disease-free, a trend not observed among MRD-negative patients, implying another dimension of prognosis prediction beyond the binary MRD status (Figure 6A). A joint Bayesian (JB) model was conducted to predict the patient-level time-dependent recurrence risk using all longitudinal plasma samples before a certain timepoint at t days post-operation (see STAR Methods). An example of such predicted risk curves was shown for patient MEDAL-010 at time t of 180, 360, and 540 days (Figure 6B), in which the cumulative risk of recurrence at t = 540 days was dramatically increased up to over 80%, providing additional information for clinical management. The JB model of respective tumor-informed and tumor-agnostic fixed-panel was also investigated and further compared with the PROPHET JB model. The latter still outperformed in terms of area under the curve (paired Wilcoxon, $p < 0.001$, Figure 6C).

Potential of using postsurgical ctDNA status to guide clinical decisions.

Finally, we explored whether ctDNA status could aid postsurgical clinical decisions, particularly at later lines of ADT. Longitudinal-positive MRD was detected among 59.3% (16/27) patients before local recurrence, including 50.0% (9/18) for lung recurrence and 77.8% (7/9) for mediastinal lymph node metastasis, and 81.0% (17/21) patients before distant metastasis. Of 20 patients with documented metastatic sites,

50.0% (2/4) of patients with brain-only metastasis, 80.0% (4/5) of patients with abdominal metastasis, and 100% (11/11) of patients with bone metastasis were longitudinal MRD positive (Figure 6D). Furthermore, post-recurrence blood samples were collected from 18 patients, of which 11 patients received treatments after recurrence and had PROPHET tests during or after treatment (Figure 6E). All seven patients with undetectable ctDNA either survived till the end of follow-up or died from other diseases. Among four patients with detectable ctDNA, three patients died from cancer. This indicates the potential utility of ctDNA positivity for predicting outcomes of later lines of treatment.

Radiographic imaging is frequently used to assess the efficacy of ADT and monitor disease progression. However, the definitive diagnosis from cross-sectional computed tomography (CT) imaging remains a challenge for some patients, as it is difficult to distinguish the tumor nodules from the pulmonary tissue inflammation or fibrosis during follow-up [19]. We evaluated the MRD findings adjacent to equivocal radiographic scans (Table S6) to study whether MRD detection might improve diagnostic accuracy. Among 110 patients in the longitudinal analysis, 15 patients had equivocal scans during follow-up with at least 1 longitudinal MRD test conducted within 90 days before or after the scan (Figure S6A). Among them, 5 patients were MRD-positive on their most adjacent test to the equivocal finding, all of whom recurred finally, hence the PPV was 100%. Conversely, among 10 patients with MRD-negative results within 90 days, 80.0% (8/10) of patients remained disease-free at the last follow-up. When enlarging the window from 90 to 180 days, the number of analyzable patients increased to 19 (Figure S6B), and the PPV and NPV of the most adjacent MRD detection were 87.5% (7/8) and 81.8% (9/11), respectively. For example, the first equivocal images for patient MEDAL-010 were observed at postoperative day-253, with MRD-positive at the same time. The patient was confirmed radiographic recurrence at postoperative day-552, with nearly 300 days of lead-time by MRD detection (Figure 6F). For relapse-free patient MEDAL-108, the

MRD status remained negative while the equivocal image result appeared at postoperative day-718 (Figure 6G).

Discussion:

The tumor-informed personalized approach has attracted considerable interest for its potential to maximize MRD detection sensitivity. We established an advanced MRD technology, PROPHET, based on a previously reported liquid-biopsy assay that we have built using high-efficiency library preparation, enhanced hybrid capturing, and UMI-based ultra-deep targeted sequencing, which demonstrated excellent sensitivity, specificity, and robustness in a quality metric alignment study by the Sequencing Quality Control Phase2 (SEQC2) workgroup at Food and Drug Administration (FDA) [20]. Therefore, PROPHET has the advantages of even coverage depth, low background noise, UMI-assisted molecular level recovery, accurate quantification, and a 0.004% LOD using 50 mutation markers with as low as 20 ng cfDNA input material. While tracking trunk mutations shared by most tumor cells in MRD testing is intuitive, the benefit of personalized design has not been thoroughly discussed. Besides the abundance of ctDNA fragments, the detectability of each mutation is mostly driven by the genomic region context of the intended mutation site and the background noise of specific base alteration in the assay. Thus, we developed a selection algorithm to favor the mutations that are with higher AF, hot/driver or non-synonymous mutation, and not located in repeat or homologous regions. In the MEDAL, 98.2% of all selected variants were non-canonical, 99.9% of which were unique to one patient. Nearly half of landmark MRD-positive samples would not have been identified if only canonical mutations were monitored. To our knowledge, our study is the first to present the contribution of the non-canonical mutations in prognosis prediction, thereby demonstrating the necessity of the personalized approach for MRD detection, especially when sensitivity is the primary concern, e.g. when MRD-negative patients are considered “cured” by surgery and exempt from adjuvant therapies.

Previous studies have demonstrated that ctDNA positivity is correlated with poor clinical outcomes in NSCLC [11,12,15,17]. In clinical practice, it would be important to

standardize the blood draw timepoint for landmark MRD testing. In the MEDAL, we proposed a "sequential double-check" strategy to achieve high predictive power while enabling earlier decisions for 82.1% of patients. In the longitudinal context, if MRD tests were to identify "clinically cured" patients who could be offered less or no treatment[12,21], it would become increasingly important to define an optimal testing frequency so that a high enough NPV can be maintained throughout the monitoring process. In the MEDAL, we conducted an exploratory analysis and found the NPV of longitudinal MRD monitoring achieved 99.4% at 180 days, thereby suggesting that when tested every 6 months, patients who are continuously MRD-negative could be considered as potentially cured with high confidence. However, it should also be noted that PPV and NPV depend on both assay accuracy and the overall occurrence rate of events. Therefore, they should be interpreted in the context of the recurrence rate of this cohort.

In the MEDAL, till the end of follow-up, PROPHET demonstrated a longitudinal sensitivity of 84.2%. The 3.45% "false-negative" cases could be biologically non-shedders or truly low-shedders when MRD tests were not frequent or sensitive enough to catch. For shedders, regardless of stage, landmark MRD-positive patients were 6–7 times more likely to experience relapse than those with negative landmark MRD. Meanwhile, non-shedders, with quite low odds to become longitudinally MRD-positive, also experienced significantly worse prognosis when relapsed. In our study, *TP53* mutation and overall TMB were only enriched among shedders after PMS balance of shedding-related variables, which is consistent with the previous study reporting that *TP53* mutation has high frequency in patients with the detectable ctDNA [12] In addition, P53 signaling pathway and transcriptional mis-regulation in cancer were also enriched in only shedders, calling for an in-depth exploration of genetic, epigenetic, and immune microenvironmental features of non-shedding tumors.

The immediate application of MRD tests is to facilitate clinical decisions on adjuvant therapies and follow-up strategies during the landmark window. In LUNGCA-1,

landmark MRD, TNM stage, and tumor size all showed significant association with recurrence-free survival (RFS) in the multivariable model, demonstrating an additive effect [14]. Another two prospective cohort studies showed significant prognosis prediction by longitudinal MRD, but not by landmark MRD [12,15]. While longitudinal MRD is valuable for relapse monitoring and potential adaptive treatment strategy, it may result in delayed receipt of effective treatments, leading to an incurable disease. In the MEDAL, we observed that landmark MRD status determined by PROPHET remained the only significant factor for prognosis prediction, overwhelming all clinical variables including the TNM stage, tumor size, and age. In the meanwhile, fixed-panel MRD assays could not achieve such dominating effect in multivariable analysis. If MRD tests become a part of future routine care, the combined "TNMB" stage integrating MRD landmark status and the hallmark cancer staging system (TNM) might provide a more informative clinical strategy for the prognostic stratification of patients. Our data presented that such a hypothetical TNMB staging system generated more significant separation on DFS compared with TNM stages. Based on our findings in the MEDAL, we have initiated a national prospective phase III randomized controlled trial (ChiCTR2200061108), in which adjuvant therapy would be guided by the integration of the MRD status and TNM stage.

Another possible application for MRD is to support the radiological diagnosis during surveillance. Guidelines for NSCLC patients recommend regularly scheduled surveillance, including chest CT scanning [22,23]. However, OS might not be significantly improved by the CT-based follow-up approach [2]. Particularly, in patients with equivocal findings, clinicians may make subjective judgments or consider repeated imaging to clarify the extent of the disease before making a treatment decision. This is a challenge that has plagued post-operative management for many years. Longitudinal MRD tests could serve as an arbiter of the equivocal radiological diagnosis (e.g. MEDAL-010 and -108) to avoid duplicate follow-ups, providing great promise to complement radiologic imaging for disease monitor. In MEDAL-010, MRD positivity by non-canonical mutations preceded MRD positivity determined by

canonical mutations by 187 days. In addition, we explored the sensitivity of MRD monitoring for recurrence at different metastatic sites. In the MEDAL, all the patients (11/11) with bone metastasis were longitudinally MRD-positive. Due to the lack of measurable lesions, determining the therapeutic responses in these patients has proposed a great challenge. Thus, MRD dynamic changes could provide a unique opportunity for guiding their disease assessment. In addition, our study extended the potential clinical utility of ctDNA detection after recurrence. Dynamic change of ctDNA continued to show a strong association with treatment response at later lines of treatment.

We demonstrated the quantitative change of ctDNA fractions could help improve prognosis prediction beyond the binary MRD status. A dynamic quantitative model was proposed to predict the time-dependent recurrence risk for each patient. Unlike the published quantitative model developed on fixed-panel assays, which required information such as T stage and baseline *TP53* status [15], our dynamic quantitative model built on PROPHET fits well without prior knowledge of baseline clinical factors and could be applied to all patients as a real-time and independent indicator to predict relapse risk after surgery. Further validation studies are required to warrant the clinical utility of this model.

Our study had certain limitations. First, due to the COVID-19 pandemic, routine follow-up was not strictly timed every 3 to 6 months after landmark tests. Second, PROPHET was designed based on patient-specific mutations of the primary tumor, and thus not suitable for early detection or diagnosis of new emerging tumors. Furthermore, the lack of transcriptomic and tumor microenvironmental data limits the in-depth analysis of the biological features of non-shedders. Finally, the clinical utility of PROPHET in NSCLC needs to be further validated in a multicenter prospective study.

In summary, patient-specific tumor-informed ctDNA-based postoperative monitoring assay supports the clinical utility of ctDNA testing for the detection of residual

497 disease and recurrence in NSCLC. This could be used as a sensitive tool for
498 identifying patients at high risk of relapse, helping patients with diagnostic
499 difficulties, and even monitoring treatment responses after clinical recurrence. Our
500 study highlights the potential of patient-specific tumor-informed MRD assay to
501 deliver precise stratification of patients for personalized therapy decision-making.

ACKNOWLEDGMENTS: We thank all the patients and their families for their participation. We also appreciate the contribution of Bing Li, Jianxing Xiang, Juan Lv, Shuai Fang, Wenjun Wu, Chenxi Li, Pengfei Zhu, Yangyang Liu, and Jiaochun Shi for their assistance with sample collection and reservation. We would also like to thank our other colleagues at the Department of Thoracic Surgery, Peking University People's Hospital, for their support during the study. We would also like to thank Dr. Huafei Lu and Dr. Guoqiang Wang for their helpful suggestions and comments. This work was supported by Research Unit of Intelligence Diagnosis and Treatment in Early Non-small Cell Lung Cancer, Chinese Academy of Medical Sciences (2021RU002), CAMS Innovation Fund for Medical Sciences (CIFMS) 2022-I2M-C&T-B-120, National Natural Science Foundation of China (No.92059203, No.92259303, No.82072566), Clinical Medicine Plus X - Young Scholars Project, Peking University, the Fundamental Research Funds for the Central Universities (PKU2023LCXQ008), Beijing Natural Science Foundation (L222021) and Peking University People's Hospital Research and Development Funds (RZ2022-03).

AUTHOR CONTRIBUTIONS: KC, ZZ, and SC conceptualized the study. ZZ, CW, and FQ developed the methodology. CW and HS curated and analyzed the data. CW wrote the statistical model. HS and CW performed the statistical analysis. KC and SC provided resources. KC, DP, FQ, and SW provided technical or material support. FY and JW supervised the study. KC obtained funding. KC, XL, SB, DC, OC, XZ, SB, and NK wrote the manuscript. All the authors reviewed and revised the manuscript for content.

DECLARATION OF INTERESTS: Z.Z., C.W., X.L., D.P., F.Q., S.W., X.Z., and S.C. are employment and stock/stock option ownership in Burning Rock Biotech. The remaining authors declare no potential conflict of interest.

Figure Legends:

Figure 1. Overview of MEDAL study design and MRD assay

Flow diagram of the study design. In total, 181 patients were enrolled, whose paired tumor and adjacent tissues were collected in the surgery. The plasma samples were collected at four types of timepoints, explicitly from 157 patients at pre-operative, 178 at landmark (3–30 days after surgery), 116 during follow-up, and 18 after recurrence. Genomic DNA was extracted from tumor tissues and paired with white blood cells. The whole-exome sequencing (WES) library prep was performed for tumor tissues and paired normal samples. For all variants called by WES data, up to 50 highly ranked markers were selected and the patient-specific probe panel was designed to cover the selected markers. The cfDNA library prep and enrichment process were performed and captured by the personalized panel and sequenced to design a personalized panel, PROPHET. The MRD status detected by PROPHET, tumor-informed fix panel, and tumor-agnostic fix panel were analyzed and compared at the abovementioned four types of timepoints to evaluate their performance in ctDNA variant detection, prognostic prediction, and potential clinical benefits from MRD tests for non-small cell lung cancer patients. Also see Table S1.

Figure 2. Proof-of-concept and analytical validation for PROPHET

(A) Simulation of the limit of detection (LOD) with different combinations of monitoring variant number (m) and effective sequencing depth (n) at background noise rate (q) of 0.0001%. (B–D) Proof-of-concept study to establish PROPHET using 5 cancer cell line-contrived samples. (B) Initial sensitivity assessment using serial diluted five cell-line contrived samples. Detection power as *y*-axis was estimated by random sampling x ($x = 10,\ 20,\ 30,\ 40,\ 50,\ 70$) variants from each cell line set. The average of the data of all five lines at *x* was used as the results for this point. For each *x* and each cell line contrived sample, 200 iterations were applied. The dot horizontal line represented a threshold of power = 0.95. (C) False positive rate assessment with different thresholds for sample-wise p-value from the PROPHET

algorithm using a panel designed for five cell lines. The dot horizontal line $y = 0.5$ represents the false positive rate of 0.5%. (D) Site-level sensitivity with different dilutions of each cell line in normal background GM24385. (E–I) Analytical validation study with H2009 diluted into GM24385 background and a personalized panel of 100 H2009-specific variants. For each sample, 50 variants were randomly selected from 100 variants to conduct MRD calculation, and the procedure was repeated 200 times to generate enough replications for statistical validity. (E) Unique fragment depth achieved with different combinations of DNA input amount and raw target sequencing depth. (F) LOD estimation with probit curves for 20 ng and 60 ng DNA input in library preparation. (G) Specificity assessment for different combinations of DNA input and dilution ratio. The dot horizontal line $y = 1.0$ represents the false positive rate of 1.0%. The background noise was recalculated by excluding the chosen 50 sites for each replication. (H) Estimation of ctDNA fraction with serial dilutions (0.0008%–0.1%) of the contrived sample at all DNA inputs. Error bars depict the median estimated fraction ± 95% CI (repeat times = 200), and the dotted line represents $y = x$. P value from Pearson correlation test. (I) Percentage of significant variants among all targeted variants with different dilutions of cell line for various DNA input amounts. Also see Table S4 and Figure S1.

Figure 3. Clinical validation with pre-operative plasma samples

(A) Sensitivity of pre-operative plasma from patients with different clinical stages by three MRD assays of PROPHET, tumor-informed fix panel, and tumor-agnostic fix panel. (B) Estimated ctDNA fraction by PROPHET for patients with different clinical stages. The red and gray dots represent PROPHET positive and negative, respectively. P value from Kruskal-Wallis test. (C) Correlation of quantitative results between PROPHET and tumor-informed fixed-panel assay. The statistics of Pearson correlation were calculated among paired MRD-positive samples defined by both assays. P value from Pearson correlation test. (D) MRD-positive samples identified by three MRD assays and ctDNA fractions level estimated by PROPHET for MRD-positive patients. P value from Wilcoxon test. (E) Pie chart of the proportion of

significant variants in personalized panels that were covered by the fixed panel. (F) Heatmap of the top 30 mutated genes from the WES data of 151 tumor tissues. Also see Figure S2.

Figure 4. Prognostic analysis with post-operative plasma at landmark timepoints

(A) Kaplan-Meier (K-M) curves for disease-free survival (DFS) analysis on post-operative landmark MRD by PROPHET at timepoint B (3–7 days) and C (30 days). The hazard ratios (HR) and p-values were estimated by Cox proportional hazards model. (B) Flowchart of the sequential double-check strategy in landmark MRD monitoring. (C) Statistics of PPV and NPV for recurrence prediction by PROPHET MRD at timepoint B, at timepoint C, and the sequential double-check strategy. (D) The dynamic change of NPV over time by PROPHET at landmark timepoint C. (E) K-M curves for DFS analysis for the 117 patients with canonical mutations covered in their personalized panels. The hazard ratios (HR) and p values were estimated by Cox proportional hazards model. (F–G) Coefficients of multivariable Cox model with clinical risk factors and landmark MRD by PROPHET. The contribution percentages were estimated using ANOVA for the multivariable Cox model. (H) The definition of TNMB stage (B for blood) by combining TNM stage and landmark PROPHET MRD status. (I) K-M curves for DFS analysis with TNMB and TNM classifications at the landmark timepoint C. The hazard ratios (HR) and p values were estimated by Cox proportional hazards model. Also see Table S5 and Figure S3 and S4.

Figure 5. Longitudinal data analysis with follow-up MRD tests

(A) Swimmer plots for 19 relapsed patients with results of longitudinal PROPHET MRD and adjuvant treatment (ADT). (B) Statistics of sensitivity (Sen), NPV, specificity (Spec), and PPV for recurrence prediction by PROPHET MRD of landmark timepoint C and longitudinal analysis. (C) The dynamic change of NPV over time for longitudinal PROPHET MRD. (D–E) Coefficients of multivariable Cox model with clinical risk factors and longitudinal MRD by PROPHET. The contribution percentages were estimated using ANOVA for the multivariable Cox model. (F) K-M curves for DFS analysis with longitudinal MRD by PROPHET for

different clinical stages. The hazard ratios (HR) and p values were estimated by Cox proportional hazards model. (G) K-M curves for DFS analysis with different groups of patients defined by longitudinal MRD by PROPHET and pre-operative ctDNA positivity (shedder/non-shedder). The hazard ratios (HR) and p values were estimated by Cox proportional hazards model. (H) The dynamic changes of ctDNA fraction estimated by PROPHET from landmark to post-ADT timepoint for 6 out of 110 patients. The lines were colored by the patient's DFS status, light blue color for disease-free and orange color for relapsed by the end of follow-up. The dots reflect the sample's ctDNA positivity at each timepoint, blue for PROPHET MRD negative and red for PROPHET MRD positive. Also see Figure S5.

Figure 6. Quantitative modeling for PROPHET and exploratory applications
(A) Dynamic changing rate (DCR) of ctDNA fraction by MRD status and DFS status for all patients in longitudinal study. P value from Wilcoxon test (B) Predictive dynamic recurrence risk ratio changes over time with accumulative MRD monitoring, log10(ctDNA fraction) as input for patient MEDAL-010. Blue lines for MRD status, red lines for predicted recurrence risk, and pink background for 95% CI of the risk. (C) The area under the curve (AUC) scores for different observation time with joint Bayesian (JB) models trained on three MRD assays. The final relapse status was treated as response. P value from paired Wilcoxon. (D) MRD-positive rate for different recurrence sites among 48 relapsed patients. (E) Swimmer plots for 11 relapsed patients with continuous MRD monitoring results and therapy after recurrence. The observed time (*x*-axis) began from recurrence till the last follow-up or death for each patient. (F) Results of longitudinal MRD tests and imaging scans for patient MEDAL-010 at different follow-up timepoints. Equivocal lesions by computed tomography scan are presented in light yellow circles, while the metastatic lesions are presented in red circles. Red dots in the curve represent MRD-positive detection. (G) Results of longitudinal MRD tests and imaging scans for patient MEDAL-108 at different follow-up timepoints. Equivocal lesions by computed tomography scan are presented in light yellow circles, while the metastatic lesions are

643 presented in red circles. Red dots in the curve represent MRD-positive detection.

644 Also see Table S6 and Figure S6.

645

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to Kezhong Chen (chenkezhong@pkuph.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The sequence data and the code and scripts used in PROPHET have been deposited at the National Omics Data Encyclopedia (NODE) and Github. The accession number is listed in the Key Resources Table.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Participants

Participants included in this study were enrolled in the MEDAL (MEthylation-based Dynamic Analysis for Lung Cancer) study (NCT03634826), of which the detailed study plan was reported [24]. The MEDAL study is a prospective study to confirm the value of circulating tumor DNA and its aberrant methylation in the longitudinal monitoring of surgical lung cancer patients.

In total, 387 participants who provided written informed consent and were suspected of stage IA to III lung cancer based on the radiology examination results and candidates for curative surgery, were consecutively enrolled at the Department of Thoracic Surgery of the Peking University People's Hospital from August 2018 to July 2019. Patients were been excluded in the following conditions: (1) pure ground-glass opacity (pGGO), (2) history of neoadjuvant therapy, including chemotherapy, radiotherapy, targeted therapy, or immunotherapy prior to surgery, (3) history of a

malignant tumor within 5 years, (4) multiple primary lung cancer, which cannot be completely resected, (5) other types of lung cancer rather than NSCLC confirmed by pathological diagnosis, (6) pathological stage IIIB-N3 or IV, (7) unqualified or failure to obtain blood or tissue samples, (8) refused or withdrawn consent, and (9) circumstances inappropriate for this study, such as blood transfusion, gestation, lactation, etc.

Ethical Approval and Consent to participate

This study was performed according to the Declaration of Helsinki in 1964 and its current amendments. The study protocol was approved by the Medical Ethics Committee of the Peking University People's Hospital (approval number: 2018PHB077–01). Written informed consent was obtained from every patient before study enrollment.

METHOD DETAILS

DNA extraction and quality control

Surgery or biopsy tumor tissue was prepared as formalin-fixed and paraffin-embedded (FFPE) sections, and tumor fraction for each sample was determined after hematoxylin and eosin staining. Genomic DNA (gDNA) was extracted from tumor tissues with pathological tumor cellularity of 30% or above using a QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany) or MagPure FFPE DNA Kit (High Pure) (Magen, Guangzhou, China). Matched genomic DNA was extracted from EDTA-anticoagulated peripheral whole blood or buffy coat samples using MagPure Universal DNA Kit (Magen, Guangzhou, China) according to the manufacturer's instructions. DNA concentration was measured by the Qubit™ dsDNA HS assay (Thermo Fisher, Carlsbad, CA, US). For circulating cell-free DNA (cfDNA), approximately 10 mL of peripheral blood sample was obtained and stored in Cell-Free DNA BCT tubes (Streck, La Vista, NE, US). The collected sample was centrifuged at 2,000 g and 4°C for 10 min within 72 h of collection. The supernatant was transferred to a 15 mL centrifuge tube and centrifuged at 16,000 g and 4°C for 10 min. The supernatant was then transferred to a new tube and stored at −80°C until further use.

cfDNA was recovered from 4 to 5 mL of plasma by using the QIAamp Circulating Nucleic Acid kit or QIAsymphony DSP Circulating DNA Kit (Qiagen, Hilden, Germany). Quantification of cfDNA was conducted with the Qubit™ dsDNA HS assay (Thermo Fisher, Carlsbad, CA, US).

Whole exome library preparation and sequencing

The whole-exome sequencing (WES) library prep was performed using Twist Human Core Exome kit (Twist Bioscience, South San Francisco, CA, US) according to the manufacturer's recommendations. Briefly, FFPE DNA samples and paired white blood cell (WBC) gDNA samples were sheared using M220 Focused-ultrasonicator (Covaris, Woburn, MA, US) to peak size 200 bp. Sheared DNA samples were end-repaired and dA-tailed, followed by ligation with Universal Adapters. After post-ligation cleanup, ligated products were polymerase chain reaction (PCR) amplified with index primers. The number of amplification cycles used varied during library preparation according to the manufacturer's recommendations. Exome capture was performed with Twist Fast Hybridization and Wash Kit, using a 33 Mb Human Core Exome panel and a customized-designed supplementary panel. Up to 8 libraries were multiplexed in one capture reaction, and 400 ng of each library was used as input. Final libraries of WES were quantified using Qubit™ dsDNA HS assay (Thermo Fisher, Carlsbad, CA, US). After library size-distribution determination by using a LabChip GX Touch System, the libraries were sequenced on NovaSeq 6000 sequencer (Illumina, San Diego, CA, US) with 2 × 151 bp pair-end reads with unique dual index, with a mean target coverage of 500× for tumor samples and 150× for paired normal samples.

WES data analysis platform

All the analyses on WES reads data were performed using Illumina DRAGEN Bio-IT Platform (Illumina, Inc., San Diego, CA, US) unless otherwise indicated. FastQ files were generated from raw BCL data using the DRAGEN Bcl Convert pipeline v3.7.4. Adapters were trimmed by using fastp v0.23.0 [25], and reads length <50 bp were discarded. Clean reads were mapped to the human reference genome (NCBI GRCh37;

hg19) [26], and PCR duplicates were marked for filtering in the downstream analysis. As a quality-control (QC) process for all samples captured by the Human Core Exome panel, tumor and paired normal alignments were checked for multiple QC parameters using the in-house software to assess capture efficiency, coverage uniformity, and library complexity. SNVs and INDELs were called if the variant supporting reads ≥ 5 and mutation VAF ≥ 3%. According to ExAC, 1000 Genomes, dbSNP, and ESP6500SI-V2 databases, variants with a population frequency > 0.5% in tumors were grouped as single-nucleotide polymorphisms (SNPs) and excluded from further analysis. In order to get real somatic mutations, for each variant, the ratio of tumor AF to paired-normal AF was calculated. Any variant called from a tumor with the ratio < 3 or both tumor and paired-normal AFs > 10% were excluded. All variants passing the applied filters were annotated with ANNOVAR [27] and SnpEff v3.6 [28].

Personalized Panel design

Patient-specific somatic variants were identified by analysis of the primary tumor and matched normal WES samples (Table S2). For a given set of variants, up to 50 highly ranked variants with a VAF ≥ 3.0% were selected for panel design. The selection of somatic variants followed proprietary rules developed in-house to prioritize high VAF and high-impact mutations. Briefly, variants falling in the repetitive regions, regions with a high GC content (> 75%), and homologous regions were filtered out. The biotinylated capture probe pool was produced in-house based on each personalized panel design.

cfDNA library preparation and sequencing

The library prep and enrichment process were performed using Burning Rock UHS unique molecular identifier (UMI) library preparation kit with minor modifications. Briefly, up to 60 ng extracted cell-free DNA samples were end-repaired and dA-tailed, followed by ligated with adapters containing UMI. Post-ligation cleanup was performed using SPRI beads and then ligated products were PCR amplified with universal adapter primers using 9 cycles. Libraries were cleaned up and quantified using Qubit™ dsDNA HS assay (Thermo Fisher, Carlsbad, CA, US). A minimum of

1 microgram of each library was hybridized and captured by a personalized biotinylated probe panel with a proprietary hybridization enhancer or the fixed panel LungPlasma for 12–16 hours. After affinity purification by streptavidin magnetic beads and stringency wash, the target enrichment library was further amplified with indexed primers using 16 cycles of PCR. Final libraries of cfDNA were cleaned up and quantified using Qubit™ dsDNA HS assay (Thermo Fisher, Carlsbad, CA, US). Ultra-deep UMI-based sequencing was performed on a NovaSeq 6000 platform (Illumina, San Diego, CA, US), with 2 × 151 bp paired-end reads and a target raw depth of 100,000× for PROPHET and 35,000× for LungPlasma.

Somatic mutation caller for tumor-agnostic and tumor-informed assay using fixed targeted panels

The fixed panel LungPlasma spanning 273 kb of the human genome, covering 168 genes that are frequently mutated in lung cancer and other common cancers, was described in our previous report [29]. The raw sequencing data analysis and tumor-agnostic somatic mutation calling were performed with the in-house bioinformatics pipeline reported before [18].

Considering paired tumor tissue calling results in the tumor-informed caller, the variants would be treated as false positive and filtered out if they were not detected in paired tumor tissue. Besides, a bunch of extra variants with low VAF could be reported when the mutation (short insertion/deletion) was supported by ≥ 4 paired reads or AF $\geq 0.1\%$ in the plasma sample and was reported in the paired tumor tissue sample.

PROPHET site-level variant calling and significance assessment

The PROPHET ctDNA libraries were prepared with custom adapters that contain 6 bp UMIs. The UMIs were extracted into read name with the structure "UMI1UMI2" using homebrew software. Sequencing reads were initially mapped to the human reference genome (NCBI GRCh37; hg19) by using BWA v0.7.10 [30]. The consensus reads were created using homebrew software based on UMI sequence and read

alignment position. The consensus reads were remapped to the hg19 reference using BWA v0.7.10. Read families were first created by a single-stranded consensus sequence (SCS) approach. After that, the duplex consensus sequence (DCS) was created upon availability. Finally, in order to reduce the noise, we only allowed variants with the following evidence to get into the further analysis: 1) has DCS support and at least one side SCS family has ≥ 2 members, and 2) only has SCS support, but the family size ≥ 3. For each personalized tumor-specific allele inspected, the wide type (WT) bases and the mutant (MT) bases occurrence were counted to calculate the variant allele frequency (VAF), where VAF = MT / (MT + WT). The assumption of a Poisson distribution was used to calculate the significance of each informed mutation in the PROPHET assay. In detail, each informed tracking site, based on genomic context, was categorized into a trimer model, such as C<u>A</u>G → C<u>G</u>G, where the middle base represents the actual mutation. The p-value of cumulative probabilities of a tracking site alteration was calculated based on the observed VAF values and the corresponding alteration background noise in the sample. For a given trimer (e.g., CAG) with an observed coverage depth of n, when the number of reads of its altered version (e.g., CGG) is x, the probability (p-value) of detecting this trimer by chance is：

$$p = 1 - \sum_{k=0}^{x-1} \frac{\lambda^k}{k!} e^{-\lambda}$$

where $\lambda = n * p(CAG \rightarrow CGG)$, and $p(CAG \rightarrow CGG)$ represents the background VAF in this sample. In this study, the site is determined as positive if the p-value < 0.05 (Table S3).

For each type of alteration including SNV and INDEL, the background noise estimation was described in the following two sections.

SNV background error estimation

The sample-specific background is estimated from the probe-targeted region except for selected patient-specific mutations. For SNV background estimation, we focused on the trimer alteration (e.g. C<u>A</u>G→C<u>G</u>G) where the middle base is altered. There are

64 types of trimer combinations and 3 alteration types for each trimer. The specific steps of sample-specific background noise calculation are:

1) For all base positions in the panel targeted region, identify corresponding trimers based on reference sequences.
2) Based on panel design, remove all intended tumor-informed somatic mutation sites and other sites that need to be excluded such as sites with low base quality.
3) For the remaining sites, calculate each trimer alteration's background noise level by site.
4) Remove any sites with background noise higher than 0.5% to avoid other potential interference.
5) Integrate each trimer alteration information by type and calculate the trimer alteration frequency for this sample.

The background noise for different SNV trimer alterations using the MEDAL cohort is shown in Figure S1A.

INDEL background error estimation

INDEL mutation can also be used in this method. When calculating the sample-specific INDEL background, it is divided into two categories: 1) random INDEL, and 2) INDEL in the short tandem repeat (STR). The calculation of the sample-specific background of INDEL is similar to that of SNV, specifically:

1) For all base positions in the panel targeted region, other than the informed tracking INDEL, based on their reference sequences, count the number of INDEL signals by alteration type and calculate mutation frequency.
2) Remove the site with a certain INDEL frequency greater than 0.5% to prevent other unintended nearby mutations from interfering with the background calculation.
3) Integrate the information and calculate the frequency of each mutation based on the INDEL alteration type, as the specific INDEL background noise level for this sample.

For random INDELs, it is divided into two categories: 1) the insertion or deletion of a single base, and 2) the two or more bases. The calculation of the background of these two categories is the same as the SNV as described above. For a single base INDEL, we combine the previous base with the INDEL base to count the relevant frequencies respectively. For INDELs with 2 or more bases, due to the excessive number of base combinations, only the length of INDELs instead of base composition was used to group and calculate noise estimation.

In the STR region, the noise usually appears as the insertion or deletion of the whole repeat unit. For certain STR, if an INDEL of the repeat unit(s) has been observed, we define the repeat unit length as m ($m = 1, 2, \ldots$), repeat times in REF as $n_0 (n_0 = 2, 3, \ldots)$, and the repeat times change in ALT as $n_1 (n_1 = 1, 2, \ldots, n_0)$, the corresponding background noise p follows,

$$p(n_0, n_1 | m) = \exp\{\alpha_0^{(m)} + \alpha_1^{(m)} * n_0 - \alpha_2^{(m)} * n_1\}$$

Where $\alpha_0^{(m)}, \alpha_1^{(m)}, \alpha_2^{(m)}$ are coefficients estimated by INDEL counts of the STR in the sample. The INDEL background estimation for different repeat types of spike-in cell-line samples were presented in Figure S1B, S1C, and S1D.

PROPHET sample-level MRD calling and ctDNA fraction estimation

The ctDNA fraction in plasma samples was estimated based on multiple loci using the maximum likelihood (ML) method. Assuming that the ctDNA fraction is π, the VAF of the i-th somatic mutation in the tumor tissue sample is q_i, and the corresponding background VAF in the plasma sample is p_i, then the expectation of VAF (f_i) in the plasma sample satisfies:

$$f_i = \pi q_i + (1 - \pi) p_i$$

$l_i(\pi)$ is the posterior probability of the i-th somatic mutation and the equation was as follows:

$$l_i(\pi) = \frac{(n_i f_i)^{x_i}}{x_i!} e^{-n_i f_i}$$

where, n_i represents the effective coverage depth of the i-th somatic mutation and x_i is the target sequencing depth of the i-th somatic mutation. The log-likelihood function of π is:

$$l(\pi) = \sum_i \log\big(l_i(\pi; x_i, n_i, p_i, q_i)\big)$$

$\hat{\pi}$ was obtained by performing ML estimation on π. The null hypothesis π = 0 was tested by the likelihood ratio test, and the likelihood ratio statistic is as follows:

$$LRT(\hat{\pi}) = -2 * \big(l(0) - l(\hat{\pi})\big) \sim \chi_1^2$$

The p-value was then calculated using the probability density function of the χ_1^2 distribution. For the PROPHET algorithm, the sample-level p-value threshold is 0.005%.

Theoretical derivation for the limit of detection

Following the model assumption described above, the maximum estimator $\hat{\pi}$ asymptotic follows normal distribution [31],

$$\hat{\pi} \dot{\sim} N(\pi_0, I_n(\pi_0)^{-1})$$

$$I_n(\pi_0) = E\left[-\sum_{i=1}^{n} \frac{\partial^2}{\partial \pi^2} l_i(\pi_0)\right]$$

Where $l_i(\pi)$ represents the likelihood function for mixed poison distribution. Therefore, the fisher information matrix $I_n(\pi_0)$ could be derived as follows,

$$I_n(\pi_0) = E\left[-\sum_{i=1}^{n} -\frac{x_i}{{f_i}^2}(p_i - q_i)^2\right] = \sum_{i=1}^{n} \frac{n_i}{f_i}(p_i - q_i)^2 = \sum_{i=1}^{n} \frac{N_i(p_i - q_i)^2}{\pi_0 p_i + (1 - \pi_0) q_i}$$

Where n_i denotes observed coverage depth, p_i denotes tumor tissue VAF, q_i denotes the background VAF. Assuming all mutation VAF p_i equals to 1 and all background VAF equals to q, in the nominal level of $(1 - \alpha)$, the rejection region for $H_a: \pi_0 >$ 0 was

$$\pi_0 > Z_{1-\alpha} \sqrt{\frac{\pi_0 + (1 - \pi_0) q}{mn(1 - q)^2}}$$

Which means that within the condition of n coverage, m personal mutation loci and q background error rate, the LOD value follows the equation below,

$$mn(1-q)^2LOD^2 - {Z_{1-\alpha}}^2(1-q)LOD - {Z_{1-\alpha}}^2 q = 0$$

The dynamic quantitative ctDNA change model utilizing PROPHET results

A dynamic changing rate (DCR) of ctDNA fraction between each pair of adjacent plasma points was defined as,

$$\rho(t_{12}) = \frac{\pi(t_2) - \pi(t_1)}{t_2 - t_1}$$

If $\rho(t)$ is positive, it means that the level of ctDNA in the patient's blood increases over time, suggesting that the patient is at risk of relapse, otherwise there is no risk. Furthermore, a joint Bayesian model was developed to predict the dynamic risk of recurrence around surveillance for each patient [32]. The joint model was presented as follows,

$$h_i(t) = h_0 \exp\{\beta * \varphi_i(t)\}$$

In the Cox proportional hazards sub-model, the hazard function was denoted as $h_i(t)$, and h_0 was the basic hazard ratio. We also took into consideration other baseline clinical factors for model construction, such as tumor-shedding and TNM stage, but it showed no benefits to the model prediction. In the time-dependent dynamic sub-model $\varphi_i(t)$, natural cubic splines with three degrees of freedom were used for both the fixed- and random-effects parts of the mixed model based on the shapes of the dynamic change curve of ctDNA fraction. An R package JMbayes2 (v0.2-8) was used to estimate the parameters of the joint Bayesian (JB) model and to predict the recurrence risk ratio in time *t* using the ctDNA fraction values of plasma samples drawn before time *t*.

Quantification and statistical analysis

Data were analyzed using the R software (version 4.1.0). A range of appropriate statistical hypothesis testing techniques were applied, including Fisher's exact test,

Wilcoxon signed-rank test, Pearson correlation test, or Spearman rank-order, to identify potential significant differences among the groups. Survival analyses were performed using the Kaplan-Meier method with log-rank test. Hazard ratios (HR) with corresponding 95% confidence intervals (CI) were calculated using univariable Cox proportional-hazards regression model. Multivariable analyses were also performed using Cox proportional-hazards regression model. ANOVA was applied to evaluate the contribution of variables in Cox proportional-hazards regression model. P-values < 0.05 were considered statistically significant.

Supplemental material legends

Figure S1. Proof-of-concept study of site-specific background noise estimation and analytical validation study with Seracare standards, Related to STAR methods and Figure 2

(A) The background noise estimations for all different trimers and alteration types in a proof-of-concept study. The width of the box represents the variance of all samples for each trimer and alteration type. (B–D) The background noise estimations with the number of repeat times for reference and the number (1–3 bp) of base insertion or deletion when the reference is the single-base repeat (B). The background error estimations with different repeat times (2–9 times) and different repeat bases (A/T/C/G) when the reference is the single-base repeat (C). The background error estimations with the number of repeat unit length for reference and the number of repeat times (2–9 times) when the reference is the tandem repeat (D). (E–F) Analytical validation study with Seracare standards Correlation between Prophet-predicted ctDNA fraction and Seracare ctDNA MRD standards blended ratio (0% [WT], 0.5%, 0.05%, and 0.005%). Error bars depict the median estimated fraction ± 95% CI (repeat times = 200) (E). Significant ratios of targeted variants for different diluted standards with three replicates in each dilution (F).

Figure S2. Clinical validation with pre-operative plasmas, Related to Figure 3

(A) MRD results of PROPHET and two fix-panels using pre-operative plasma samples in different histological subtypes of NSCLC, lung adenocarcinoma (LUAD, left), and lung squamous cell carcinoma (LUSC, right). (B) Estimation of ctDNA fraction with different subtypes of NSCLC by PROPHET. (C) Pearson correlation between tumor volume and ctDNA fraction detected PROPHET for all MRD-positive patients. (D) K-M curves for DFS analysis on pre-operative MRD by three assays, PROPHET (left), Tumor-informed (middle), and Tumor-agnostic (right). (E) K-M curves for OS analysis on pre-operative MRD by three assays, PROPHET (left), Tumor-informed (middle), and Tumor-agnostic (right). (F) Comparison of TMB score for tumor WES data between shedder and non-shedder with (left) and without (right)

PMS adjustment for shedding-related variables. (G) Correlation matrix between biological variables and ctDNA-shedding status. The correlation test was applied using Pearson correlation method for continuous variables and Kruskal-Wallis test for categorical variables. (H) Gene pathway enrichment analysis between shedder and non-shedder after PMS adjustment for TMB and clinicopathologic variables.

Figure S3 Prognosis prediction by the landmark MRD status, Related to Figure 4.

(A) K-M curves for DFS analysis for MRD status at timepoint B and by sequential double-check strategy. (B–C) Prognostic analysis with adjuvant therapy for landmark MRD-negative group. K-M curves for DFS analysis with adjuvant therapy for landmark MRD-negative group (B). K-M curves for DFS analysis with adjuvant therapy for landmark MRD-negative group after propensity score matching adjustment for baseline clinical factors. After PSM adjustment, 20 matched pairs of patients were selected for comparison including 6 in stage I, 11 in stage II, and 3 in stage III (C). (D–G) Prognostic analysis with landmark MRD status by three assays. K-M curves for DFS analysis of the landmark MRD by three assays at timepoint B (D) and timepoint C (E) by three assays. K-M curves for OS analysis of the landmark MRD by three assays at timepoint B (F) and timepoint C (G). (H) K-M curves for DFS analysis at timepoint B of patients MRD-positive identified by PROPHET and by both PROPHET and tumor-informed panel.

Figure S4. Multivariable Cox model with landmark MRD by fixed-panel assays and TNMB staging validation, Related to Figure 4.

(A–B) Coefficients of multivariable Cox regression model with clinical risk factors and MRD status by fixed-panel assays, tumor-informed (A) and tumor-agnostic (B). The contribution percentages were estimated using ANOVA for the multivariable Cox model. (C–E) K-M curves for DFS analysis with the landmark TNMB stage of published datasets from LUNGCA-1 study (C), ChiCTR1900024656 study (D), and combined data of these two studies with PROPHET study (E).

Figure S5. Prognostic analysis with longitudinal MRD results from three assays, Related to Figure 5

(A) Swimmer plots for all patients without relapse in this study. (B–C) K-M curves for DFS (B) and OS (C) analysis with longitudinal MRD status by three assays PROPHET (left), tumor-informed (middle), and tumor-agnostic (right). (D) The lead time of MRD detection by PROPHET, tumor-informed fixed panel, and tumor-agnostic fixed panel. (E) Coefficients of multivariable Cox model with clinical risk factors and longitudinal MRD status by fixed panel assay, tumor-informed (upper), tumor-agnostic (down). The contribution percentages were estimated using ANOVA for multivariable Cox model.

Figure S6. Complementary information provided by PROPHET MRD for patients with equivocal imaging findings, Related to Figure 6

(A) Swimmer plot for 15 patients with equivocal image scan and nearby MRD monitoring by PROPHET. The patients were selected with equivocal scans during their follow-up with at least one longitudinal MRD test conducted within 90 days before or after the scan. (B) Swimmer plot for 19 patients with equivocal image scan and nearby MRD monitoring by PROPHET. The patients were selected with equivocal scans during their follow-up with at least one longitudinal MRD test conducted within 180 days before or after the scan.

Table S1. Clinical and demographic characteristics of the study participants, Related to Figure 1

Table S2. Personalized panel design information for each patient, Related to STAR Methods

Table S3. Three MRD assays and bioinformatics analysis comparison, Related to STAR Methods

Table S4. Summary of Proof-of-concept study and validation studies, Related to STAR Methods and Figure 2

Table S5. Clinical and demographic characteristics correlations with adjuvant therapy, Related to Figure 4

Table S6. Summary of longitudinal imaging scan results, Related to Figure 6

Reference

1 Ettinger, D. S., Wood, D. E., Aisner, D. L., Akerley, W., Bauman, J., Chirieac, L. R., D'Amico, T. A., DeCamp, M. M., Dilling, T. J., Dobelbower, M. et al. (2017). Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw *15*, 504-535 https://doi.org:10.6004/jnccn.2017.0050

2 Westeel, V., Foucher, P., Scherpereel, A., Domas, J., Girard, P., Trédaniel, J., Wislez, M., Dumont, P., Quoix, E., Raffy, O. et al. (2022). Chest CT scan plus x-ray versus chest x-ray for the follow-up of completely resected non-small-cell lung cancer (IFCT-0302): a multicentre, open-label, randomised, phase 3 trial. Lancet Oncol *23*, 1180-1188 https://doi.org:10.1016/s1470-2045(22)00451-x

3 Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R. & Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nature Reviews Cancer *17*, 223-238 https://doi.org:10.1038/nrc.2017.7

4 Parkinson, C. A., Gale, D., Piskorz, A. M., Biggs, H., Hodgkin, C., Addley, H., Freeman, S., Moyle, P., Sala, E., Sayal, K. et al. (2016). 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 https://doi.org:10.1371/journal.pmed.1002198

5 Jee, J., Lebow, E. S., Yeh, R., Das, J. P., Namakydoust, A., Paik, P. K., Chaft, J. E., Jayakumaran, G., Rose Brannon, A., Benayed, R. et al. (2022). Overall survival with circulating tumor DNA-guided therapy in advanced non-small-cell lung cancer. Nat Med *28*, 2353-2363 https://doi.org:10.1038/s41591-022-02047-z

6 Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R. et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature *545*, 446-451 https://doi.org:10.1038/nature22364

7 Moding, E. J., Liu, Y., Nabet, B. Y., Chabon, J. J., Chaudhuri, A. A., Hui, A. B., Bonilla, R. F., Ko, R. B., Yoo, C. H., Gojenola, L. et al. (2020). Circulating Tumor DNA Dynamics Predict Benefit from Consolidation Immunotherapy in Locally Advanced Non-Small Cell Lung Cancer. Nat Cancer *1*, 176-183 https://doi.org:10.1038/s43018-019-0011-0

8 Wan, J. C. M., Heider, K., Gale, D., Murphy, S., Fisher, E., Mouliere, F., Ruiz-Valdepenas, A., Santonja, A., Morris, J., Chandrananda, D. et al. (2020). ctDNA monitoring using patient-specific sequencing and integration of variant reads. Sci Transl Med *12* https://doi.org:10.1126/scitranslmed.aaz8084

9 Abbosh, C., Frankell, A., Garnett, A., Harrison, T., Weichert, M., Licon, A., Veeriah, S., Daber, B., Moreau, M., Chesh, A. et al. (2020). Abstract CT023: Phylogenetic tracking and minimal residual disease detection using ctDNA in early-stage NSCLC: A lung TRACERx study. Cancer Research *80*, CT023-CT023 https://doi.org:10.1158/1538-7445.Am2020-ct023

10 Abbosh, C., Frankell, A. M., Harrison, T., Kisistok, J., Garnett, A., Johnson, L., Veeriah, S., Moreau, M., Chesh, A., Chaunzwa, T. L. et al. (2023). Tracking early lung cancer metastatic dissemination in TRACERx using ctDNA. Nature *616*, 553-562 https://doi.org:10.1038/s41586-023-05776-4

11 Gale, D., Heider, K., Ruiz-Valdepenas, A., Hackinger, S., Perry, M., Marsico, G., Rundell, V.,

Wulff, J., Sharma, G., Knock, H. et al. (2022). Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer. Ann Oncol *33*, 500-510 https://doi.org:10.1016/j.annonc.2022.02.007

12 Zhang, J. T., Liu, S. Y., Gao, W., Liu, S. M., Yan, H. H., Ji, L., Chen, Y., Gong, Y., Lu, H. L., Lin, J. T. et al. (2022). Longitudinal Undetectable Molecular Residual Disease Defines Potentially Cured Population in Localized Non-Small Cell Lung Cancer. Cancer Discov *12*, 1690-1701 https://doi.org:10.1158/2159-8290.Cd-21-1486

13 Yang, Y., Zhang, T., Wang, J., Wang, J., Xu, Y., Zhao, X., Ou, Q., Shao, Y., Wang, X., Wu, Y. et al. (2022). The clinical utility of dynamic ctDNA monitoring in inoperable localized NSCLC patients. Mol Cancer *21*, 117 https://doi.org:10.1186/s12943-022-01590-0

14 Xia, L., Mei, J., Kang, R., Deng, S., Chen, Y., Yang, Y., Feng, G., Deng, Y., Gan, F., Lin, Y. et al. (2022). Perioperative ctDNA-Based Molecular Residual Disease Detection for Non-Small Cell Lung Cancer: A Prospective Multicenter Cohort Study (LUNGCA-1). Clin Cancer Res *28*, 3308-3317 https://doi.org:10.1158/1078-0432.Ccr-21-3044

15 Qiu, B., Guo, W., Zhang, F., Lv, F., Ji, Y., Peng, Y., Chen, X., Bao, H., Xu, Y., Shao, Y. et al. (2021). Dynamic recurrence risk and adjuvant chemotherapy benefit prediction by ctDNA in resected NSCLC. Nat Commun *12*, 6770 https://doi.org:10.1038/s41467-021-27022-z

16 Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., Tsimberidou, A. M., Vnencak-Jones, C. L., Wolff, D. J., Younes, A. et al. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn *19*, 4-23 https://doi.org:10.1016/j.jmoldx.2016.10.002

17 Chaudhuri, A. A., Chabon, J. J., Lovejoy, A. F., Newman, A. M., Stehr, H., Azad, T. D., Khodadoust, M. S., Esfahani, M. S., Liu, C. L., Zhou, L. et al. (2017). Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. Cancer Discov *7*, 1394-1403 https://doi.org:10.1158/2159-8290.Cd-17-0716

18 Liang, N., Li, B., Jia, Z., Wang, C., Wu, P., Zheng, T., Wang, Y., Qiu, F., Wu, Y., Su, J. et al. (2021). Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. Nat Biomed Eng *5*, 586-599 https://doi.org:10.1038/s41551-021-00746-5

19 Kocak, Z., Evans, E. S., Zhou, S. M., Miller, K. L., Folz, R. J., Shafman, T. D. & Marks, L. B. (2005). Challenges in defining radiation pneumonitis in patients with lung cancer. Int J Radiat Oncol Biol Phys *62*, 635-638 https://doi.org:10.1016/j.ijrobp.2004.12.023

20 Deveson, I. W., Gong, B., Lai, K., LoCoco, J. S., Richmond, T. A., Schageman, J., Zhang, Z., Novoradovskaya, N., Willey, J. C., Jones, W. et al. (2021). Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. Nat Biotechnol *39*, 1115-1128 https://doi.org:10.1038/s41587-021-00857-z

21 Tie, J., Cohen, J. D., Lahouel, K., Lo, S. N., Wang, Y., Kosmider, S., Wong, R., Shapiro, J., Lee, M., Harris, S. et al. (2022). Circulating Tumor DNA Analysis Guiding Adjuvant Therapy in Stage II Colon Cancer. N Engl J Med *386*, 2261-2272 https://doi.org:10.1056/NEJMoa2200075

22 Postmus, P. E., Kerr, K. M., Oudkerk, M., Senan, S., Waller, D. A., Vansteenkiste, J., Escriu, C. & Peters, S. (2017). Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *28*, iv1-iv21

https://doi.org:10.1093/annonc/mdx222

23 Schneider, B. J., Ismaila, N., Aerts, J., Chiles, C., Daly, M. E., Detterbeck, F. C., Hearn, J. W. D., Katz, S. I., Leighl, N. B., Levy, B. et al. (2020). Lung Cancer Surveillance After Definitive Curative-Intent Therapy: ASCO Guideline. J Clin Oncol *38*, 753-766 https://doi.org:10.1200/jco.19.02748

24 Kang, G., Chen, K., Yang, F., Chuai, S., Zhao, H., Zhang, K., Li, B., Zhang, Z. & Wang, J. (2019). Monitoring of circulating tumor DNA and its aberrant methylation in the surveillance of surgical lung Cancer patients: protocol for a prospective observational study. BMC Cancer *19*, 579 https://doi.org:10.1186/s12885-019-5751-9

25 Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884-i890 https://doi.org:10.1093/bioinformatics/bty560

26 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921 https://doi.org:10.1038/35057062

27 Wang, K., Li, M. & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164 https://doi.org:10.1093/nar/gkq603

28 Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80-92 https://doi.org:10.4161/fly.19695

29 Li, Y. S., Jiang, B. Y., Yang, J. J., Zhang, X. C., Zhang, Z., Ye, J. Y., Zhong, W. Z., Tu, H. Y., Chen, H. J., Wang, Z. et al. (2018). Unique genetic profiles from cerebrospinal fluid cell-free DNA in leptomeningeal metastases of EGFR-mutant non-small-cell lung cancer: a new medium of liquid biopsy. Ann Oncol *29*, 945-952 https://doi.org:10.1093/annonc/mdy009

30 Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760 https://doi.org:10.1093/bioinformatics/btp324

31 Sweeting, T. J. (1980). Uniform Asymptotic Normality of the Maximum Likelihood Estimator. The Annals of Statistics *8*, 1375-1381, 1377

32 Rizopoulos, D., Molenberghs, G. & Lesaffre, E. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biom J *59*, 1261-1276 https://doi.org:10.1002/bimj.201600238

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Tumor biopsies and peripheral blood of lung cancer patients (MEDAL study)	Kang et al. [24]	https://clinicaltrials.gov/ct2/show/NCT03634826
Critical commercial assays		
HS UMI library prep kit	Burning Rock	Cat# RS07P-12
Human Core Exome panel	Twist Bioscience	Cat# 102027
Fast Hybridization and Wash Kit	Twist Bioscience	Cat# 104181
MagPure FFPE DNA Kit (High Pure)	Magen	Cat# D6323-02B
MagPure Universal DNA Kit	Magen	Cat# MD5105-02
QIAamp DNA FFPE tissue kit	Qiagen	Cat# 56404
QIAamp Circulating Nucleic Acid kit	Qiagen	Cat# 55114
QIAsymphony DSP Circulating DNA Kit	Qiagen	Cat# 937556
Qubit™ dsDNA HS assay	Thermo Fisher	Cat# Q32854
Mechanical Fragmentation Library Preparation Kit	Twist Bioscience	Cat# 101281
Universal Blockers	Twist Bioscience	Cat# 100767
Deposited data		
Raw data	This paper	https://www.biosino.org/node/project/detail/OEP004204
LUNGCA-1	Xie et al. [14]	N/A
ChiCTR1900024656	Qiu et al. [15]	N/A
Experimental models: Cell lines		
Human: H2009	ATCC	RRID: CVCL_1514
Human: H2126	ATCC	RRID: CVCL_1532
Human: HCC1395	ATCC	RRID: CVCL_1249
Human: H1437	ATCC	RRID: CVCL_1472
Human: HCC38	ATCC	RRID: CVCL_1267
Human: GM24385	ATCC	RRID: CVCL_1C78
Software and algorithms		
DRAGEN Bcl Convert pipeline (v3.7.4)	N/A	https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/downloads.html
Fastp (v0.23.0)	Chen et al. [25]	RRID:SCR_016962
ANNOVAR (20160201)	Wang et al. [27]	RRID:SCR_012821
SnpEff (v3.6)	Cingolani et al. [28]	RRID:SCR_005191
R package JMbayes2 (v0.2-8)	N/A	https://cran.r-project.org/web/packages/JMbayes2/index.html

BWA (v0.7.10)	Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2	RRID:SCR_010910
Code and scripts used in PROPHET	This study	https://github.com/bnr-cdx/prophet
Other		
Cell-Free DNA BCT tubes	Streck	Cat# 230244
M220 Focused-ultrasonicator	Covaris	Cat# 500295
LabChip® GX Touch™ Nucleic Acid Analyzer	PerkinElmer	Cat# CLS138162
Novaseq 6000 System	Illumina	Cat# 20012850
NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)	Illumina	Cat# 20028312

Figure 1

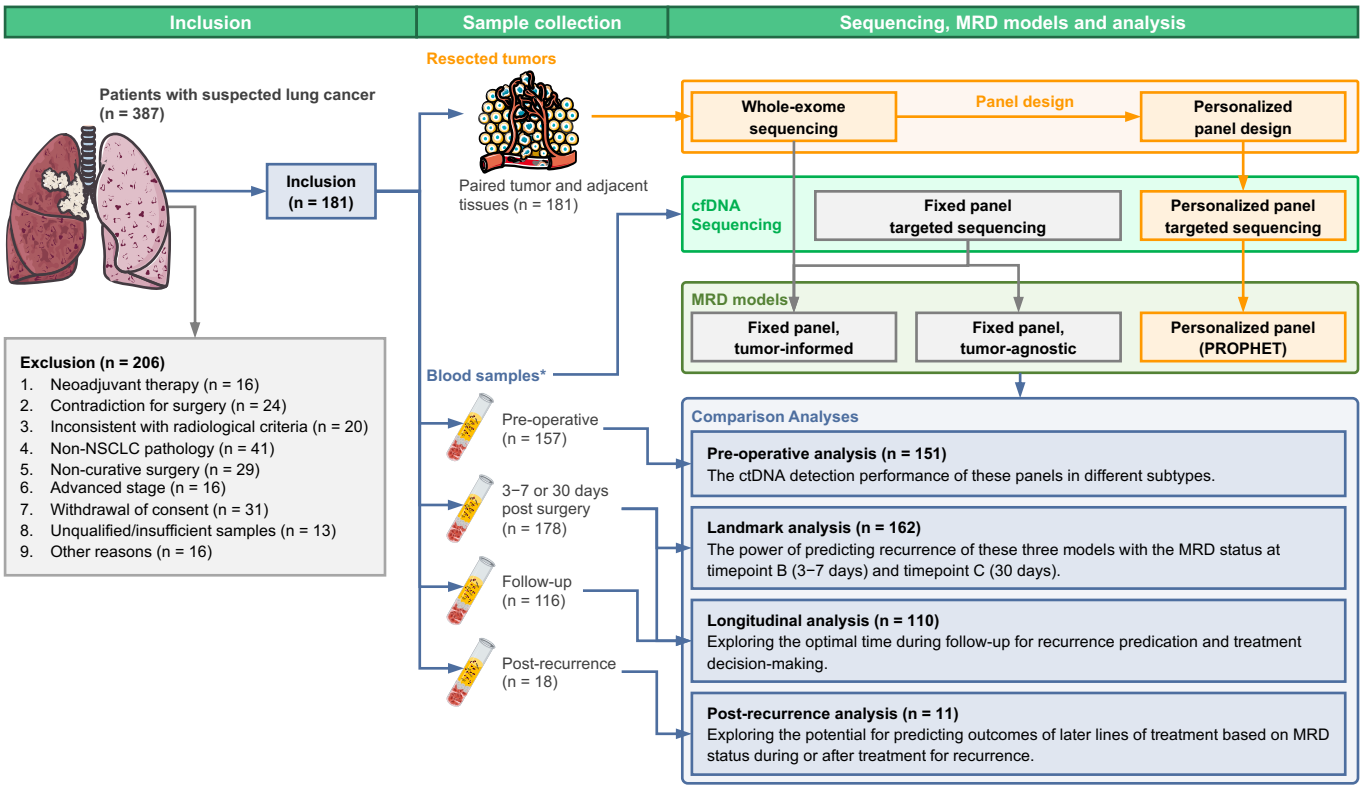

Inclusion
Sample collection
Sequencing, MRD models and analysis
Patients with suspected lung cancer (n = 387)
Inclusion (n = 181)
Exclusion (n = 206)
1. Neoadjuvant therapy (n = 16)
2. Contradiction for surgery (n = 24)
3. Inconsistent with radiological criteria (n = 20)
4. Non-NSCLC pathology (n = 41)
5. Non-curative surgery (n = 29)
6. Advanced stage (n = 16)
7. Withdrawal of consent (n = 31)
8. Unqualified/insufficient samples (n = 13)
9. Other reasons (n = 16)
Resected tumors
Paired tumor and adjacent tissues (n = 181)
Blood samples*
Pre-operative (n = 157)
3−7 or 30 days post surgery (n = 178)
Follow-up (n = 116)
Post-recurrence (n = 18)
Whole-exome sequencing
Panel design
Personalized panel design
cfDNA Sequencing
Fixed panel targeted sequencing
Personalized panel targeted sequencing
MRD models
Fixed panel, tumor-informed
Fixed panel, tumor-agnostic
Personalized panel (PROPHET)
Comparison Analyses
Pre-operative analysis (n = 151)
The ctDNA detection performance of these panels in different subtypes.
Landmark analysis (n = 162)
The power of predicting recurrence of these three models with the MRD status at timepoint B (3−7 days) and timepoint C (30 days).
Longitudinal analysis (n = 110)
Exploring the optimal time during follow-up for recurrence predication and treatment decision-making.
Post-recurrence analysis (n = 11)
Exploring the potential for predicting outcomes of later lines of treatment based on MRD status during or after treatment for recurrence.

Figure 2

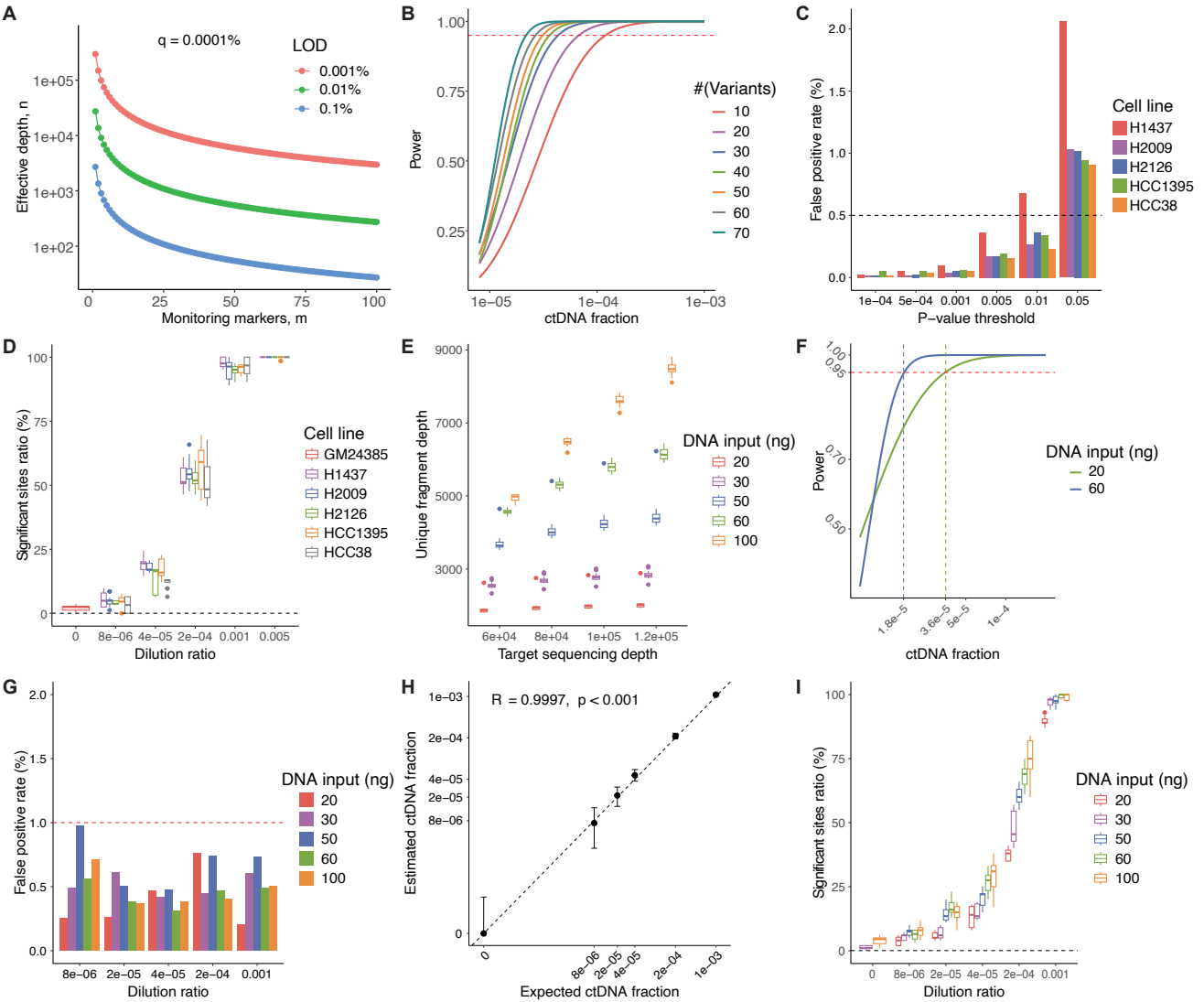

A
q = 0.0001%
LOD
0.001%
0.01%
0.1%
Effective depth, n
1e+05
1e+04
1e+03
1e+02
0
25
50
75
100
Monitoring markers, m
B
Power
1.00
0.75
0.50
0.25
#(Variants)
10
20
30
40
50
60
70
1e−05
1e−04
1e−03
ctDNA fraction
C
False positive rate (%)
2.0
1.5
1.0
0.5
0.0
Cell line
H1437
H2009
H2126
HCC1395
HCC38
1e−04
5e−04
0.001
0.005
0.01
0.05
P−value threshold
D
Significant sites ratio (%)
100
75
50
25
0
Cell line
GM24385
H1437
H2009
H2126
HCC1395
HCC38
0
8e−06
4e−05
2e−04
0.001
0.005
Dilution ratio
E
Unique fragment depth
9000
7000
5000
3000
DNA input (ng)
20
30
50
60
100
6e+04
8e+04
1e+05
1.2e+05
Target sequencing depth
F
Power
1.00
0.95
0.70
0.50
DNA input (ng)
20
60
1.8e−5
3.6e−5
5e−5
1e−4
ctDNA fraction
G
False positive rate (%)
2.0
1.5
1.0
0.5
0.0
DNA input (ng)
20
30
50
60
100
8e−06
2e−05
4e−05
2e−04
0.001
Dilution ratio
H
R = 0.9997, p < 0.001
Estimated ctDNA fraction
1e−03
2e−04
4e−05
2e−05
8e−06
0
0
8e−06
2e−05
4e−05
2e−04
1e−03
Expected ctDNA fraction
I
Significant sites ratio (%)
100
75
50
25
0
DNA input (ng)
20
30
50
60
100
0
8e−06
2e−05
4e−05
2e−04
0.001
Dilution ratio

Figure 3

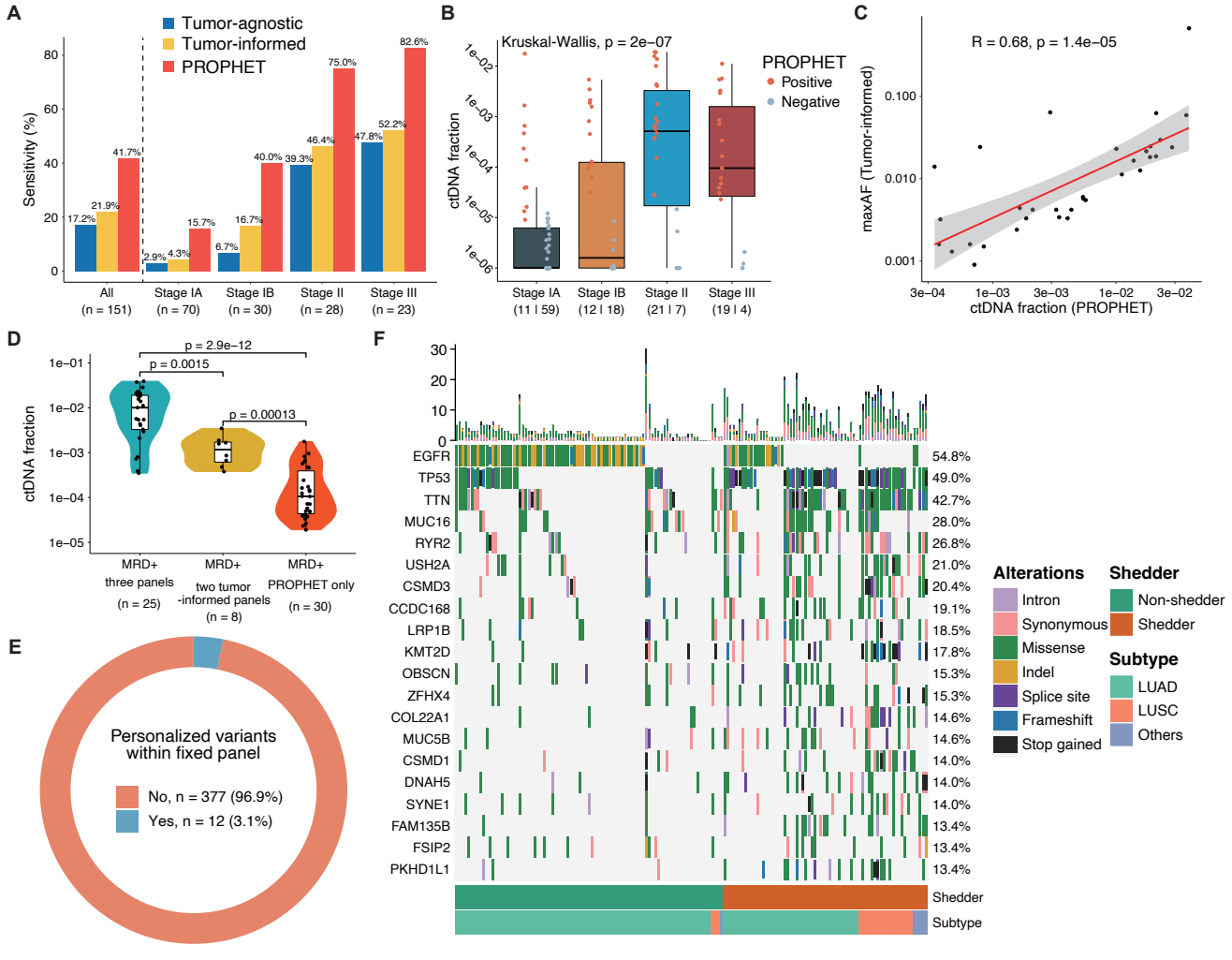

A
Tumor-agnostic
Tumor-informed
PROPHET
Sensitivity (%)
17.2%
21.9%
41.7%
2.9%
4.3%
15.7%
6.7%
16.7%
40.0%
39.3%
46.4%
75.0%
47.8%
52.2%
82.6%
All (n = 151)
Stage IA (n = 70)
Stage IB (n = 30)
Stage II (n = 28)
Stage III (n = 23)
B
Kruskal-Wallis, p = 2e−07
PROPHET
Positive
Negative
ctDNA fraction
Stage IA (11 | 59)
Stage IB (12 | 18)
Stage II (21 | 7)
Stage III (19 | 4)
C
R = 0.68, p = 1.4e−05
maxAF (Tumor-informed)
ctDNA fraction (PROPHET)
D
p = 2.9e−12
p = 0.0015
p = 0.00013
ctDNA fraction
MRD+ three panels (n = 25)
MRD+ two tumor -informed panels (n = 8)
MRD+ PROPHET only (n = 30)
E
Personalized variants within fixed panel
No, n = 377 (96.9%)
Yes, n = 12 (3.1%)
F
EGFR 54.8%
TP53 49.0%
TTN 42.7%
MUC16 28.0%
RYR2 26.8%
USH2A 21.0%
CSMD3 20.4%
CCDC168 19.1%
LRP1B 18.5%
KMT2D 17.8%
OBSCN 15.3%
ZFHX4 15.3%
COL22A1 14.6%
MUC5B 14.6%
CSMD1 14.0%
DNAH5 14.0%
SYNE1 14.0%
FAM135B 13.4%
FSIP2 13.4%
PKHD1L1 13.4%
Shedder
Subtype
Alterations
Intron
Synonymous
Missense
Indel
Splice site
Frameshift
Stop gained
Shedder
Non-shedder
Shedder
Subtype
LUAD
LUSC
Others

Figure 4

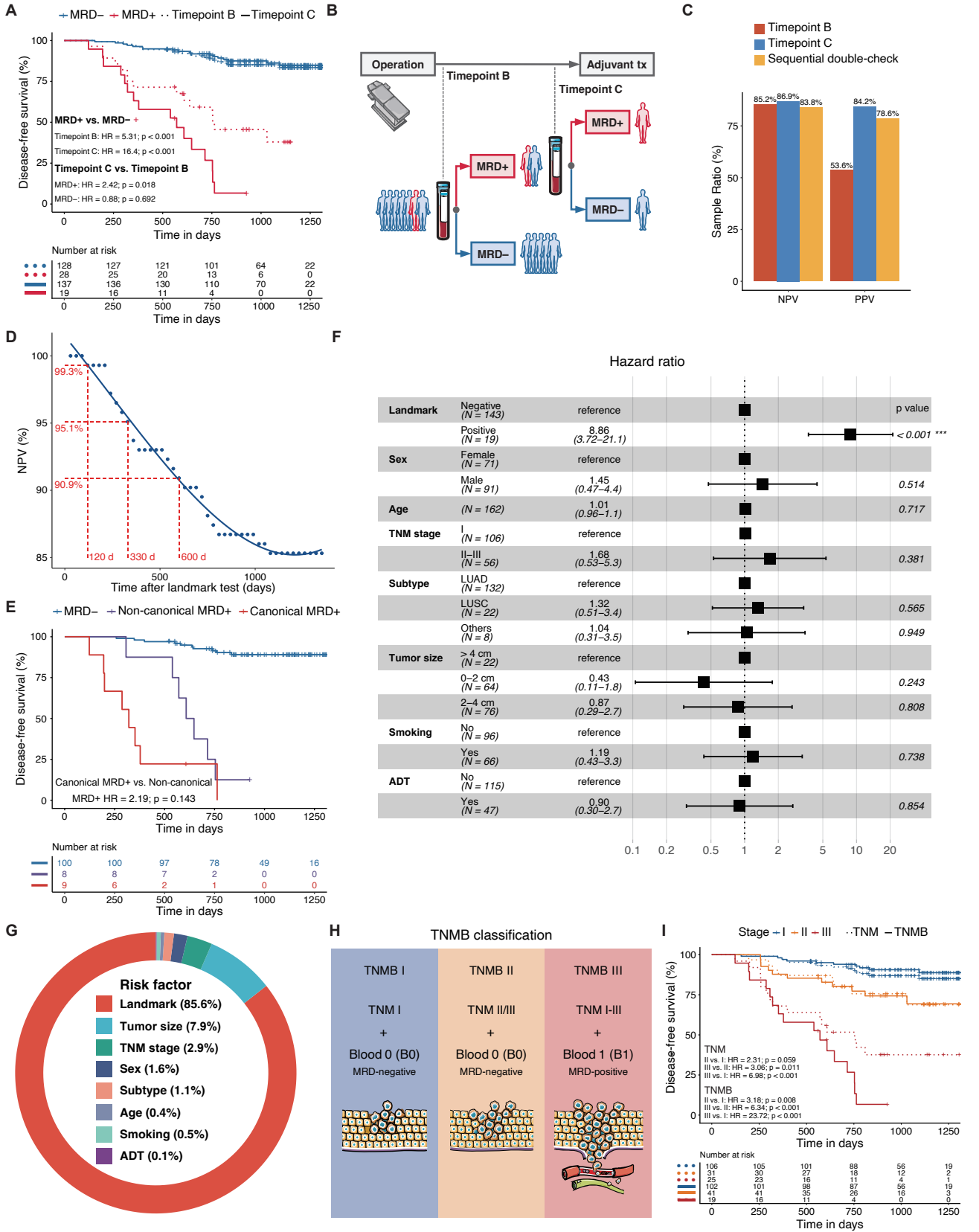

A
MRD− MRD+ Timepoint B Timepoint C
Disease-free survival (%)
MRD+ vs. MRD−
Timepoint B: HR = 5.31; p < 0.001
Timepoint C: HR = 16.4; p < 0.001
Timepoint C vs. Timepoint B
MRD+: HR = 2.42; p = 0.018
MRD−: HR = 0.88; p = 0.692
Number at risk
Time in days
B
Operation
Timepoint B
Adjuvant tx
Timepoint C
MRD+
MRD−
MRD+
MRD−
C
Timepoint B
Timepoint C
Sequential double-check
Sample Ratio (%)
85.2%
86.9%
83.8%
53.6%
84.2%
78.6%
NPV
PPV
D
NPV (%)
99.3%
95.1%
90.9%
120 d
330 d
600 d
Time after landmark test (days)
E
MRD− Non-canonical MRD+ Canonical MRD+
Disease-free survival (%)
Canonical MRD+ vs. Non-canonical
MRD+ HR = 2.19; p = 0.143
Number at risk
Time in days
F
Hazard ratio
Landmark
Negative (N = 143) reference
Positive (N = 19) 8.86 (3.72–21.1) < 0.001 ***
Sex
Female (N = 71) reference
Male (N = 91) 1.45 (0.47–4.4) 0.514
Age
(N = 162) 1.01 (0.96–1.1) 0.717
TNM stage
I (N = 106) reference
II–III (N = 56) 1.68 (0.53–5.3) 0.381
Subtype
LUAD (N = 132) reference
LUSC (N = 22) 1.32 (0.51–3.4) 0.565
Others (N = 8) 1.04 (0.31–3.5) 0.949
Tumor size
> 4 cm (N = 22) reference
0–2 cm (N = 64) 0.43 (0.11–1.8) 0.243
2–4 cm (N = 76) 0.87 (0.29–2.7) 0.808
Smoking
No (N = 96) reference
Yes (N = 66) 1.19 (0.43–3.3) 0.738
ADT
No (N = 115) reference
Yes (N = 47) 0.90 (0.30–2.7) 0.854
p value
0.1 0.2 0.5 1 2 5 10 20
G
Risk factor
Landmark (85.6%)
Tumor size (7.9%)
TNM stage (2.9%)
Sex (1.6%)
Subtype (1.1%)
Age (0.4%)
Smoking (0.5%)
ADT (0.1%)
H
TNMB classification
TNMB I
TNM I
+
Blood 0 (B0)
MRD-negative
TNMB II
TNM II/III
+
Blood 0 (B0)
MRD-negative
TNMB III
TNM I-III
+
Blood 1 (B1)
MRD-positive
I
Stage I II III TNM TNMB
Disease-free survival (%)
TNM
II vs. I: HR = 2.31; p = 0.059
III vs. II: HR = 3.06; p = 0.011
III vs. I: HR = 6.98; p < 0.001
TNMB
II vs. I: HR = 3.18; p = 0.008
III vs. II: HR = 6.34; p < 0.001
III vs. I: HR = 23.72; p < 0.001
Number at risk
Time in days

Figure 5

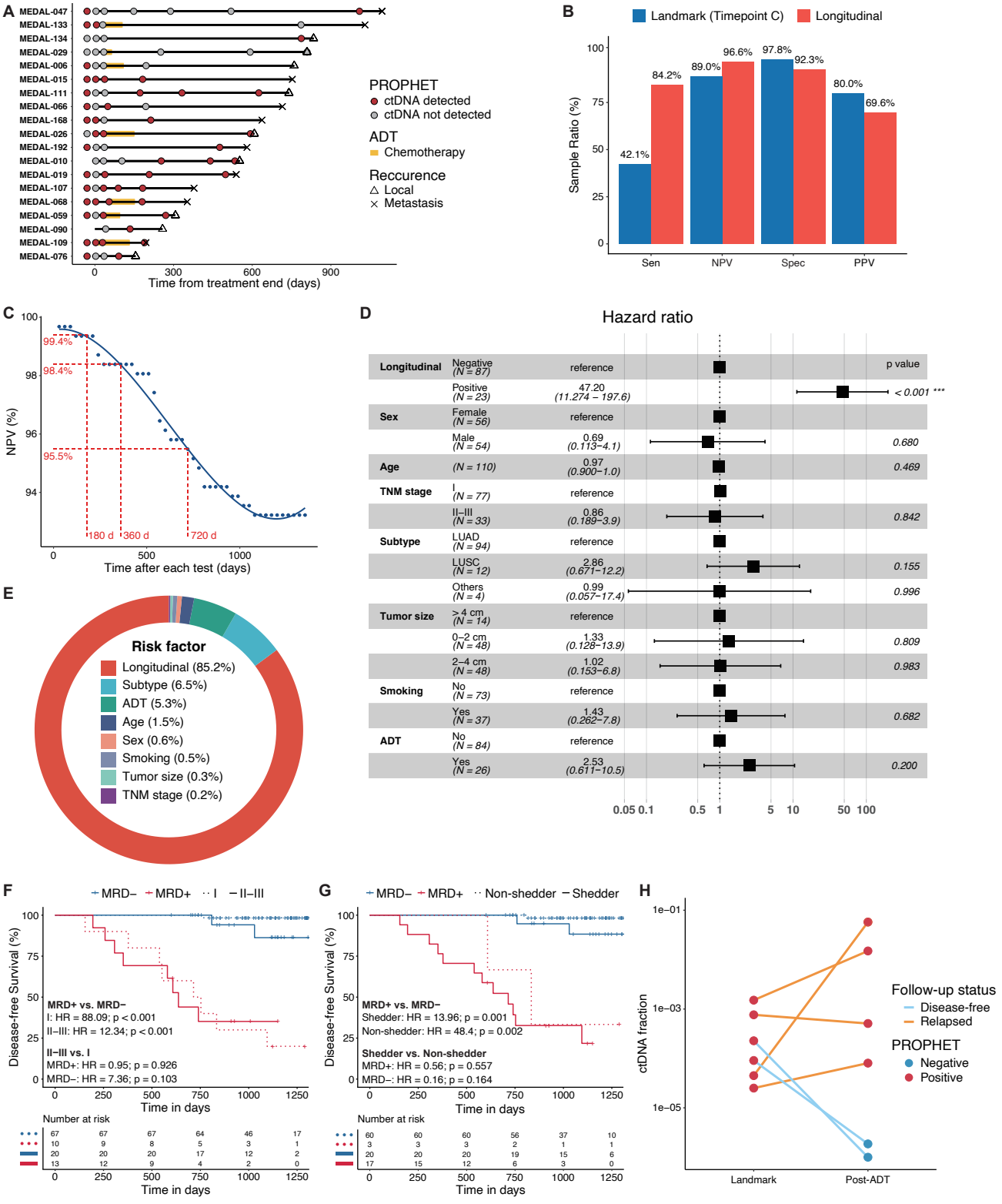
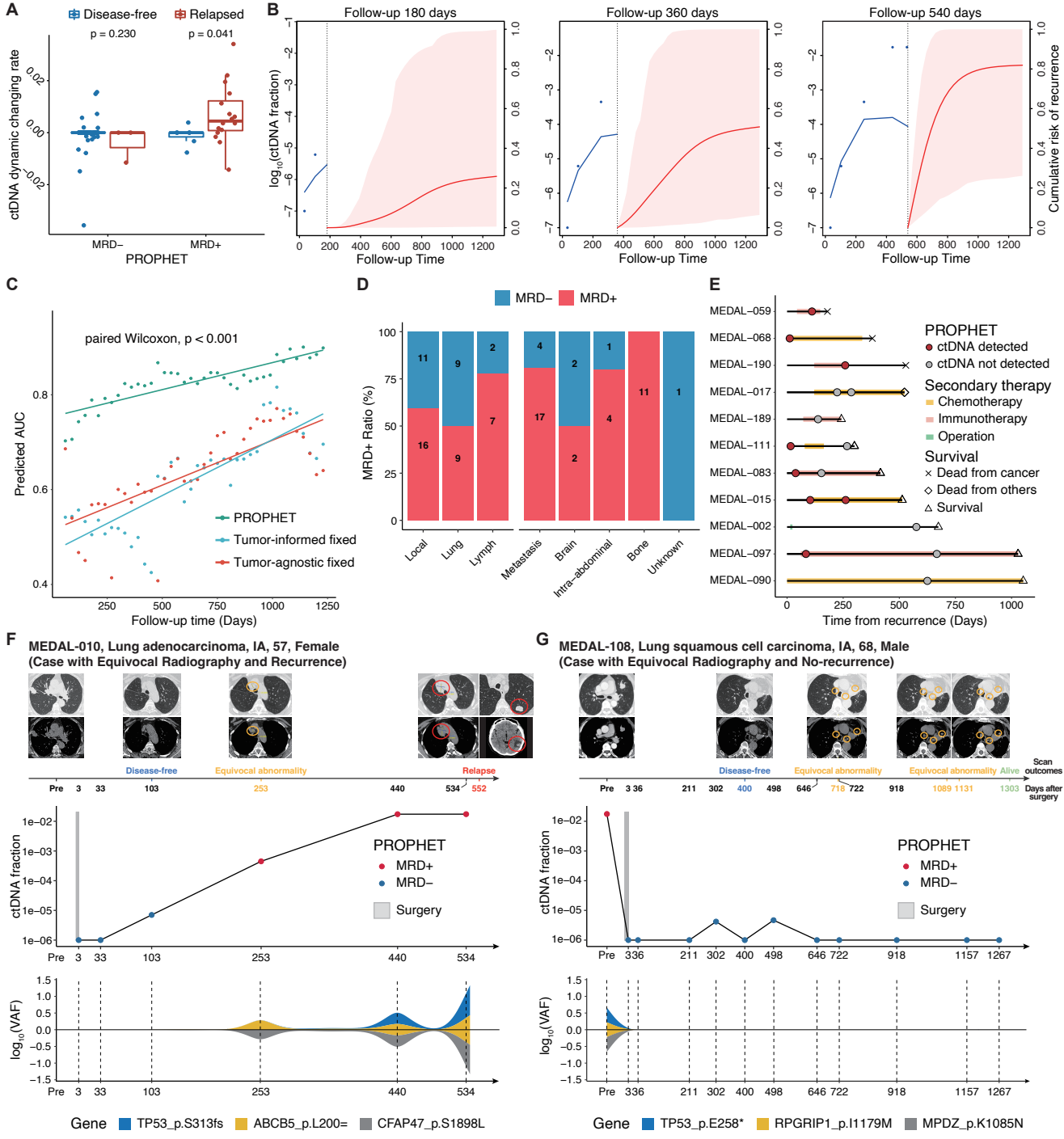

A
MEDAL-047
MEDAL-133
MEDAL-134
MEDAL-029
MEDAL-006
MEDAL-015
MEDAL-111
MEDAL-066
MEDAL-168
MEDAL-026
MEDAL-192
MEDAL-010
MEDAL-019
MEDAL-107
MEDAL-068
MEDAL-059
MEDAL-090
MEDAL-109
MEDAL-076
0
300
600
900
Time from treatment end (days)
PROPHET
ctDNA detected
ctDNA not detected
ADT
Chemotherapy
Reccurence
Local
Metastasis
B
Landmark (Timepoint C)
Longitudinal
Sample Ratio (%)
100
75
50
25
0
42.1%
84.2%
89.0%
96.6%
97.8%
92.3%
80.0%
69.6%
Sen
NPV
Spec
PPV
C
NPV (%)
100
98
96
94
99.4%
98.4%
95.5%
180 d
360 d
720 d
0
500
1000
Time after each test (days)
D
Hazard ratio
Longitudinal
Negative (N = 87)
reference
Positive (N = 23)
47.20 (11.274 – 197.6)
p value
< 0.001 ***
Sex
Female (N = 56)
reference
Male (N = 54)
0.69 (0.113–4.1)
0.680
Age
(N = 110)
0.97 (0.900–1.0)
0.469
TNM stage
I (N = 77)
reference
II–III (N = 33)
0.86 (0.189–3.9)
0.842
Subtype
LUAD (N = 94)
reference
LUSC (N = 12)
2.86 (0.671–12.2)
0.155
Others (N = 4)
0.99 (0.057–17.4)
0.996
Tumor size
> 4 cm (N = 14)
reference
0–2 cm (N = 48)
1.33 (0.128–13.9)
0.809
2–4 cm (N = 48)
1.02 (0.153–6.8)
0.983
Smoking
No (N = 73)
reference
Yes (N = 37)
1.43 (0.262–7.8)
0.682
ADT
No (N = 84)
reference
Yes (N = 26)
2.53 (0.611–10.5)
0.200
0.05
0.1
0.5
1
5
10
50
100
E
Risk factor
Longitudinal (85.2%)
Subtype (6.5%)
ADT (5.3%)
Age (1.5%)
Sex (0.6%)
Smoking (0.5%)
Tumor size (0.3%)
TNM stage (0.2%)
F
MRD−
MRD+
I
II–III
Disease-free Survival (%)
100
75
50
25
0
MRD+ vs. MRD−
I: HR = 88.09; p < 0.001
II–III: HR = 12.34; p < 0.001
II–III vs. I
MRD+: HR = 0.95; p = 0.926
MRD−: HR = 7.36; p = 0.103
0
250
500
750
1000
1250
Time in days
Number at risk
67 67 67 64 46 17
10 9 8 5 3 1
20 20 20 17 12 2
13 12 9 4 2 0
Time in days
G
MRD−
MRD+
Non-shedder
Shedder
Disease-free Survival (%)
MRD+ vs. MRD−
Shedder: HR = 13.96; p = 0.001
Non-shedder: HR = 48.4; p = 0.002
Shedder vs. Non-shedder
MRD+: HR = 0.56; p = 0.557
MRD−: HR = 0.16; p = 0.164
Time in days
Number at risk
60 60 60 56 37 10
3 3 3 2 1 1
20 20 20 19 15 6
17 15 12 6 3 0
Time in days
H
ctDNA fraction
1e−01
1e−03
1e−05
Landmark
Post-ADT
Follow-up status
Disease-free
Relapsed
PROPHET
Negative
Positive

Figure 6


A
Disease-free
Relapsed
p = 0.230
p = 0.041
ctDNA dynamic changing rate
MRD−
MRD+
PROPHET
B
Follow-up 180 days
Follow-up 360 days
Follow-up 540 days
log10(ctDNA fraction)
Cumulative risk of recurrence
Follow-up Time
C
paired Wilcoxon, p < 0.001
Predicted AUC
PROPHET
Tumor-informed fixed
Tumor-agnostic fixed
Follow-up time (Days)
D
MRD−
MRD+
MRD+ Ratio (%)
Local
Lung
Lymph
Metastasis
Brain
Intra−abdominal
Bone
Unknown
E
MEDAL−059
MEDAL−068
MEDAL−190
MEDAL−017
MEDAL−189
MEDAL−111
MEDAL−083
MEDAL−015
MEDAL−002
MEDAL−097
MEDAL−090
PROPHET
ctDNA detected
ctDNA not detected
Secondary therapy
Chemotherapy
Immunotherapy
Operation
Survival
Dead from cancer
Dead from others
Survival
Time from recurrence (Days)
F
MEDAL-010, Lung adenocarcinoma, IA, 57, Female
(Case with Equivocal Radiography and Recurrence)
Disease-free
Equivocal abnormality
Relapse
ctDNA fraction
PROPHET
MRD+
MRD−
Surgery
log10(VAF)
Gene
TP53_p.S313fs
ABCB5_p.L200=
CFAP47_p.S1898L
G
MEDAL-108, Lung squamous cell carcinoma, IA, 68, Male
(Case with Equivocal Radiography and No-recurrence)
Disease-free
Equivocal abnormality
Equivocal abnormality
Alive
Scan outcomes
Days after surgery
ctDNA fraction
PROPHET
MRD+
MRD−
Surgery
log10(VAF)
Gene
TP53_p.E258*
RPGRIP1_p.I1179M
MPDZ_p.K1085N

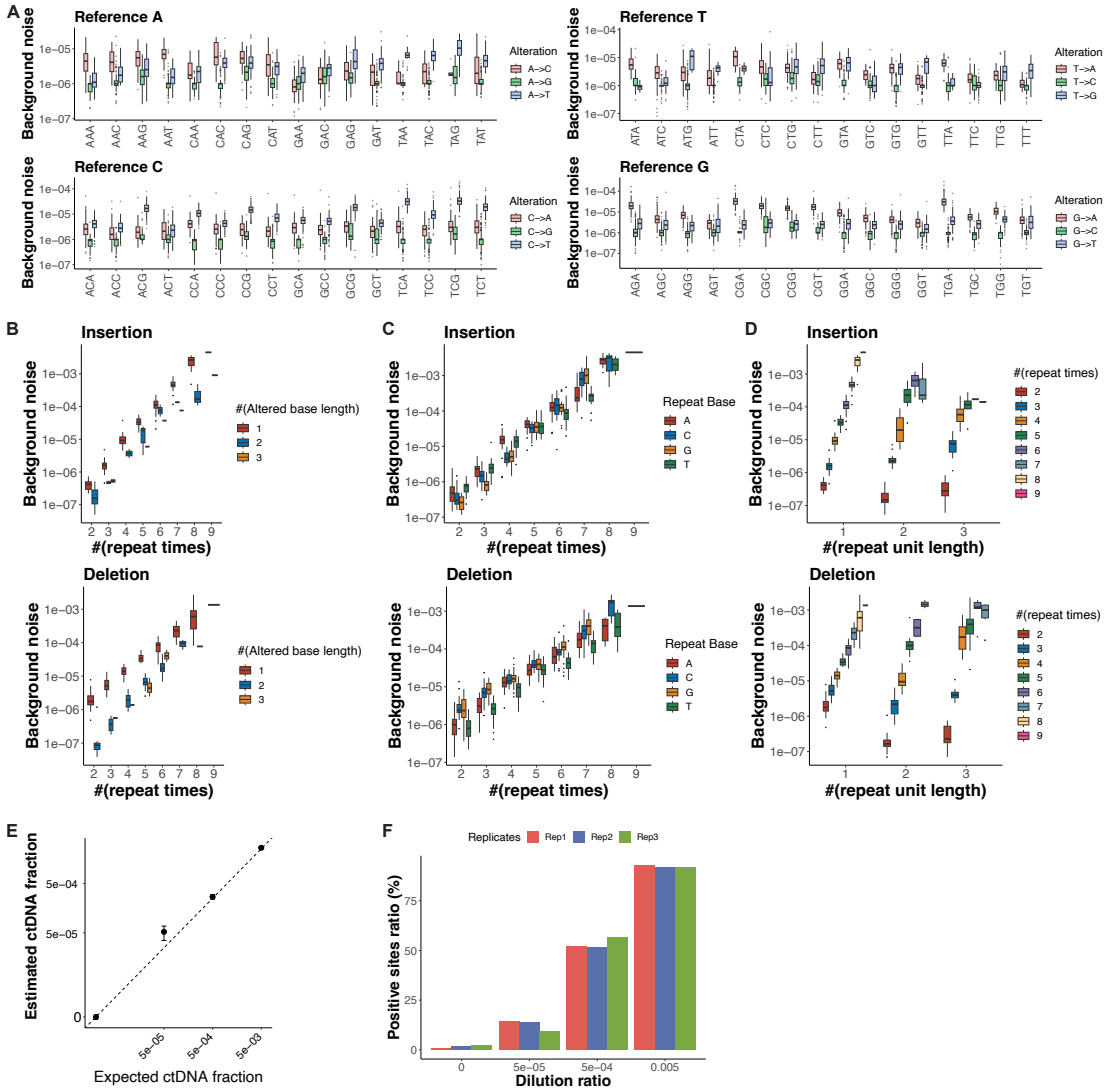


Figure S1. Proof-of-concept study of site-specific background noise estimation and analytical validation study with Seracare standards, Related to Figure 2

(A) The background noise estimations for all different trimers and alteration types in a proof-of-concept study. The box represents the variance of all samples for each trimer and alteration type. (B–D) The background noise estimations for all insertions (deletions) in genomic repeat regions. The background noise estimations with the number of repeat times for reference and the number (1~3 bp) of base insertion or deletion when the reference is the single-base repeat (B). The background error estimations with different repeat times (2~9 times) and different repeat base (A/T/C/G) when the reference is the single-base repeat (C). The background error estimations with the number of repeat unit length for reference and the

number of repeat times (2~9 times) when the reference is the tandem repeat (D). (E−F) Analytical validation study with Seracare standards Correlation between Prophet-predicted ctDNA fraction and Seracare ctDNA MRD standards blended ratio (0% [WT], 0.5%, 0.05%, and 0.005%). Error bars depict the median estimated fraction ±95% CI (repeat times = 200) (E). Significant ratios of targeted variants for different diluted standards with three replicates in each dilution (F).

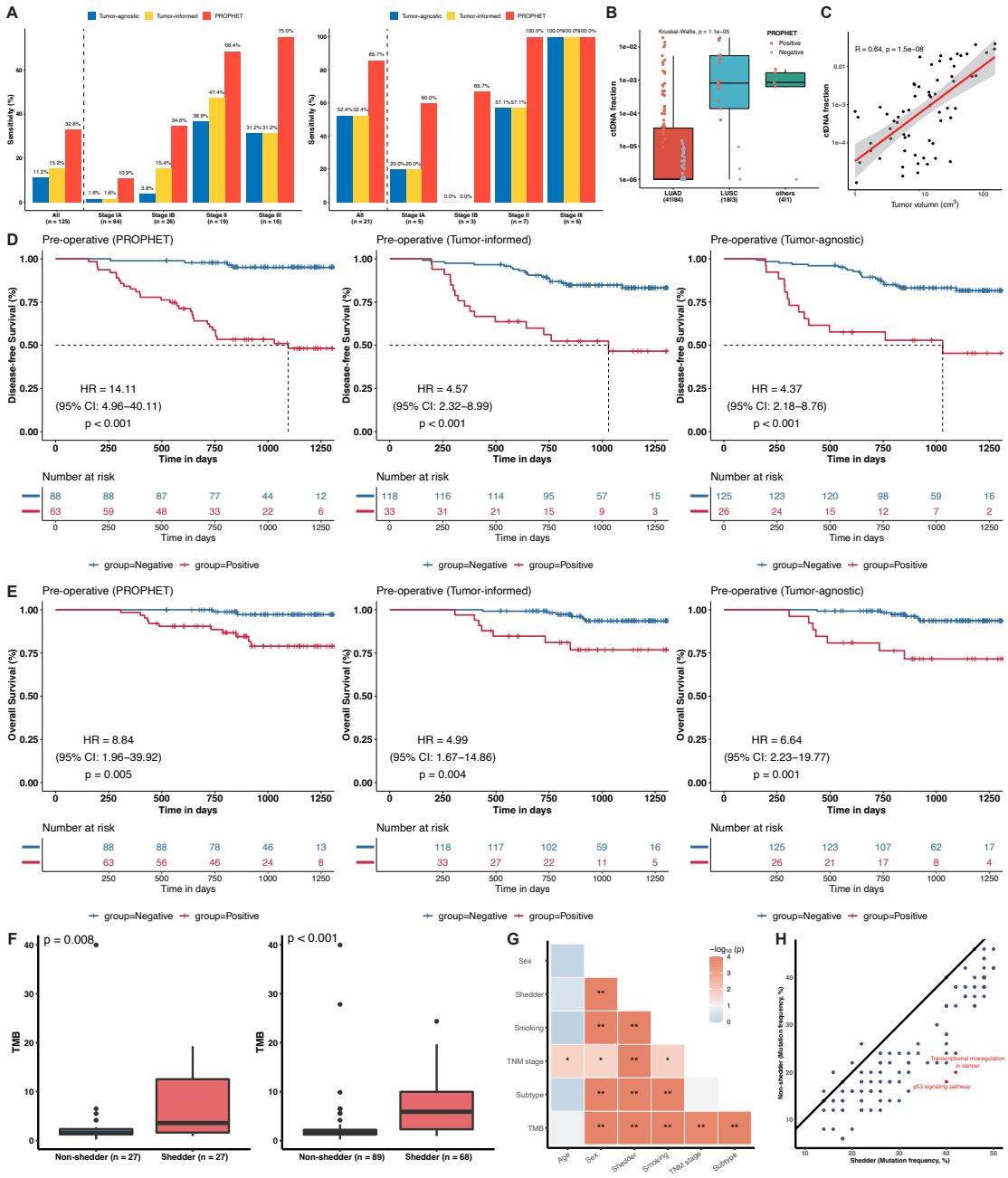


Figure S2. Clinical validation with pre-operative plasmas, Related to Figure 3

(A) MRD detection results of PROPHET and two fix-panels using pre-operative plasma samples in different histological subtypes of NSCLC, lung adenocarcinoma (LUAD, left) and lung squamous cell carcinoma (LUSC, right). (B) Estimation of ctDNA fraction with different subtypes of NSCLC by PROPHET. (C) Pearson correlation between tumor volume and ctDNA fraction detected PROPHET for all MRD-positive patients. (D) K-M curves for DFS analysis on pre-operative MRD by three assays, PROPHET (left), Tumor-informed

(middle), and Tumor-agnostic (right). (E) K-M curves for OS analysis on pre-operative MRD by three assays, PROPHET (left), Tumor-informed (middle), and Tumor-agnostic (right). (F) Comparison of TMB score for tumor WES data between shedder and non-shedder with (left) and without (right) PMS adjustment for shedding-related variables. (G) Correlation matrix between biological variables and ctDNA-shedding status. The correlation test was applied using Pearson correlation method for continuous variables and Kruskal-Wallis test for categorical variables. (H) Gene pathway enrichment analysis between shedder and non-shedder after PMS adjustment for TMB and clinicopathologic variables.

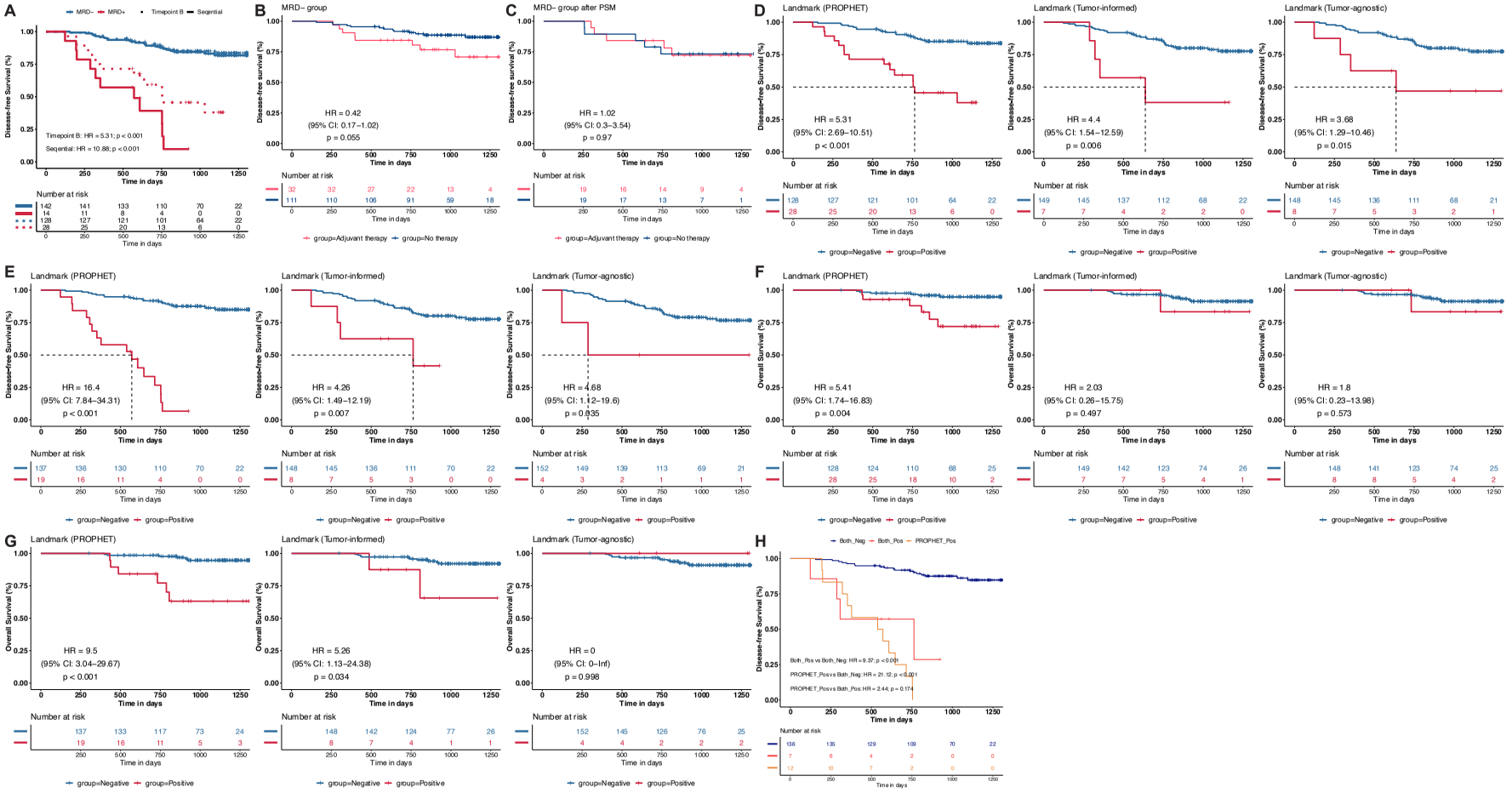

A
MRD−
MRD+
Timepoint B
Seqential
Disease-free Survival (%)
Timepoint B: HR = 5.31; p < 0.001
Seqential: HR = 10.88; p < 0.001
Time in days
Number at risk
B
MRD− group
Disease-free survival (%)
HR = 0.42
(95% CI: 0.17−1.02)
p = 0.055
Time in days
Number at risk
group=Adjuvant therapy
group=No therapy
C
MRD− group after PSM
Disease-free survival (%)
HR = 1.02
(95% CI: 0.3−3.54)
p = 0.97
Time in days
Number at risk
group=Adjuvant therapy
group=No therapy
D
Landmark (PROPHET)
Disease-free Survival (%)
HR = 5.31
(95% CI: 2.69−10.51)
p < 0.001
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-informed)
Disease-free Survival (%)
HR = 4.4
(95% CI: 1.54−12.59)
p = 0.006
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-agnostic)
Disease-free Survival (%)
HR = 3.68
(95% CI: 1.29−10.46)
p = 0.015
Time in days
Number at risk
group=Negative
group=Positive
E
Landmark (PROPHET)
Disease-free Survival (%)
HR = 16.4
(95% CI: 7.84−34.31)
p < 0.001
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-informed)
Disease-free Survival (%)
HR = 4.26
(95% CI: 1.49−12.19)
p = 0.007
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-agnostic)
Disease-free Survival (%)
HR = 4.68
(95% CI: 1.12−19.6)
p = 0.035
Time in days
Number at risk
group=Negative
group=Positive
F
Landmark (PROPHET)
Overall Survival (%)
HR = 5.41
(95% CI: 1.74−16.83)
p = 0.004
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-informed)
Overall Survival (%)
HR = 2.03
(95% CI: 0.26−15.75)
p = 0.497
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-agnostic)
Overall Survival (%)
HR = 1.8
(95% CI: 0.23−13.98)
p = 0.573
Time in days
Number at risk
group=Negative
group=Positive
G
Landmark (PROPHET)
Overall Survival (%)
HR = 9.5
(95% CI: 3.04−29.67)
p < 0.001
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-informed)
Overall Survival (%)
HR = 5.26
(95% CI: 1.13−24.38)
p = 0.034
Time in days
Number at risk
group=Negative
group=Positive
Landmark (Tumor-agnostic)
Overall Survival (%)
HR = 0
(95% CI: 0−Inf)
p = 0.998
Time in days
Number at risk
group=Negative
group=Positive
H
Both_Neg
Both_Pos
PROPHET_Pos
Disease-free Survival (%)
Both_Pos vs Both_Neg: HR = 9.37; p < 0.001
PROPHET_Pos vs Both_Neg: HR = 21.12; p < 0.001
PROPHET_Pos vs Both_Pos: HR = 2.44; p = 0.174
Time in days
Number at risk

Figure S3 Prognosis prediction by the landmark MRD status, Related to Figure 4.

(A) K-M curves for DFS analysis for MRD status at timepoint B and by sequential double-check strategy. (B−C) Prognostic analysis with adjuvant therapy for landmark MRD-negative group. K-M curves for DFS analysis with adjuvant therapy for landmark MRD-negative group (B). K-M curves for DFS analysis with adjuvant therapy for landmark MRD-negative group after propensity score matching adjustment for baseline clinical factors. After PSM adjustment, 20 matched pairs of patients were selected for comparison including 6 in stage I, 11 in stage II, and 3 in stage III (C). (D−G) Prognostic analysis with landmark MRD status by three assays. K-M curves for DFS analysis of the landmark MRD by three assays at timepoint B (D) and timepoint C (E) by three assays. K-M curves for OS analysis of the landmark MRD by three assays at timepoint B (F) and timepoint C (G). (H) K-M curves for DFS analysis at timepoint B of patients MRD-positive identified by PROPHET and by both PROPHET and tumor-informed fixed-panel.

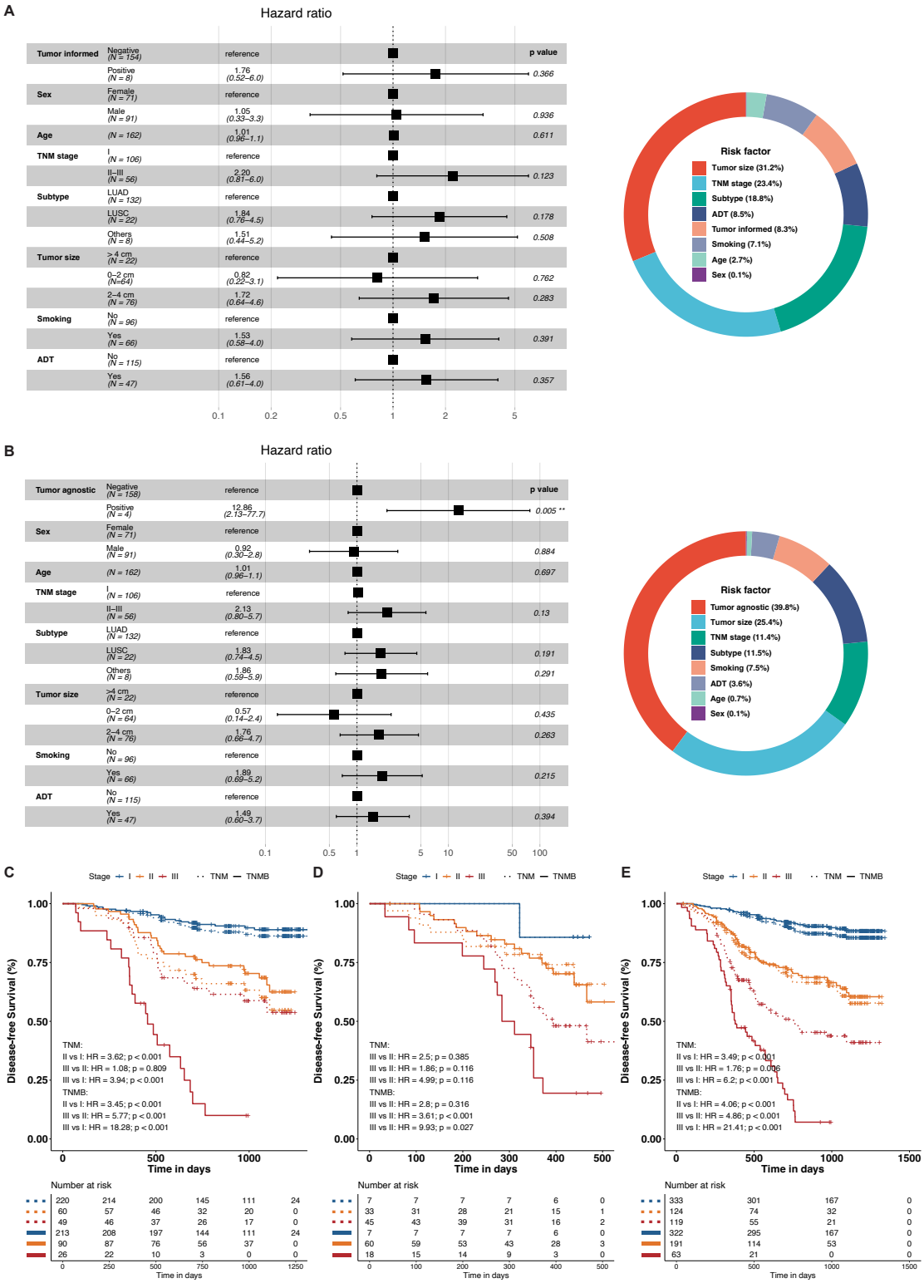


Figure S4. Multivariable Cox model with landmark MRD by fixed-panel assays and TNMB staging validation, Related to Figure 4.

(A−B) Coefficients of multivariable Cox regression model with clinical risk factors and MRD status by fixed-panel assays, tumor-informed (A) and tumor-agnostic (B). The contribution

percentages were estimated using ANOVA for the multivariable Cox model. (C−E) K-M curves for DFS analysis with the landmark TNMB stage of published datasets from LUNGCA-1 study (C), ChiCTR1900024656 study (D), and combined data of these two studies with PROPHET study (E).

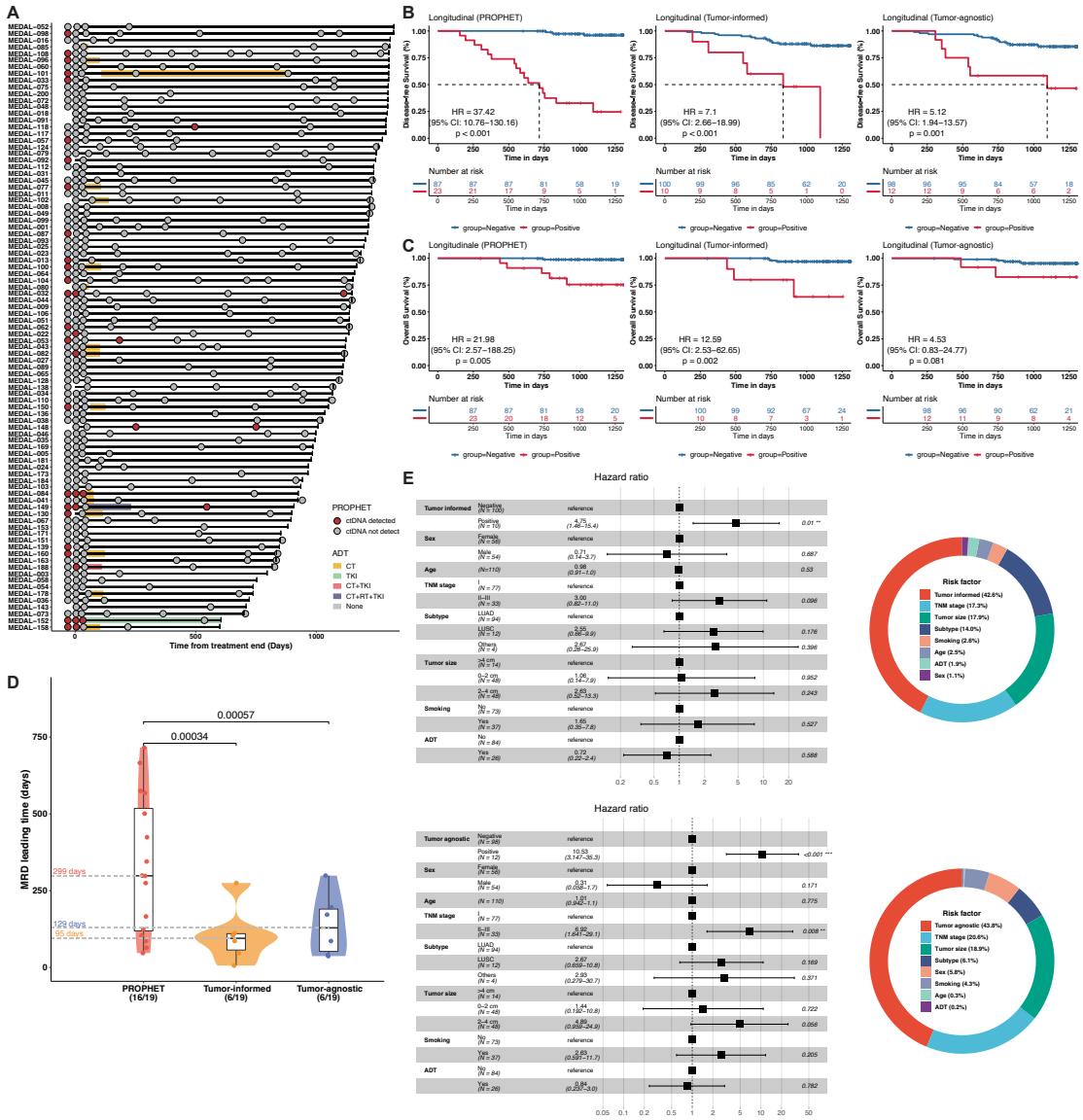


Figure S5. Prognostic analysis with longitudinal MRD results from three assays, related to Figure 5

(A) Swimmer plots for all patients without relapse in this study. (B−C) K-M curves for DFS (B) and OS (C) analysis with longitudinal MRD status by three assays PROPHET (left), tumor-informed (middle), and tumor-agnostic (right). (D) The lead time of MRD detection by PROPHET, tumor-informed fixed panel, and tumor-agnostic fixed panel. (E) Coefficients of multivariable Cox model with clinical risk factors and longitudinal MRD status by fixed panel assay, tumor-informed (upper), tumor-agnostic (down). The contribution percentages were estimated using ANOVA for the multivariable Cox model.

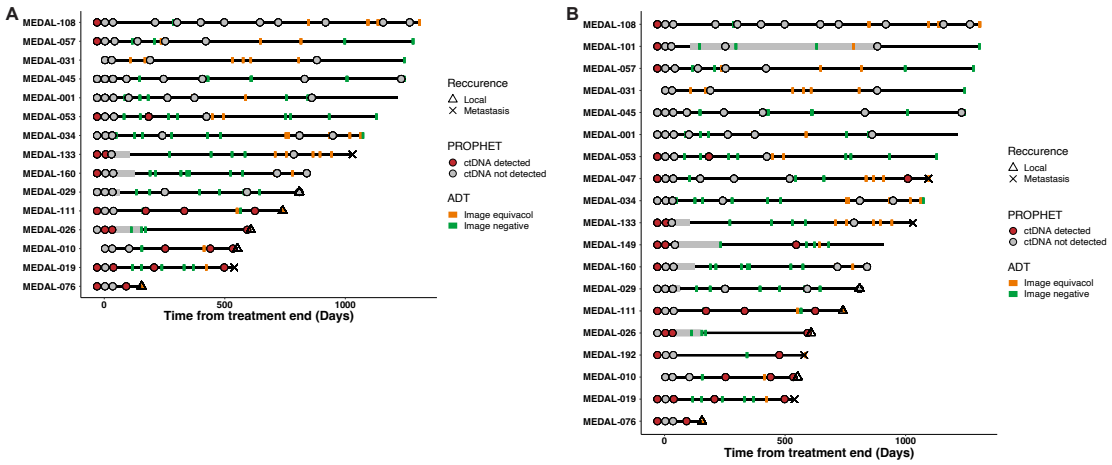


Figure S6 Complementary information provided by PROPHET MRD for patients with equivocal imaging findings, Related to Figure 6

(A) Swimmer plot for 15 patients with equivocal image scan and nearby MRD monitoring by PROPHET. The patients were selected with equivocal scans during their follow-up with at least one longitudinal MRD test conducted within 90 days before or after the scan. (B) Swimmer plot for 19 patients with equivocal image scan and nearby MRD monitoring by PROPHET. The patients were selected with equivocal scans during their follow-up with at least one longitudinal MRD test conducted within 180 days before or after the scan.