Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prognostic value of circulating H3K27Me3-nucleosomes in newly diagnosed lung cancer patients

Aristotelis Kotronoulas
Belgian Volition https://orcid.org/0000-0001-9695-0876

Emmanuel Grolleau
Acute Respiratory Disease and Thoracic Oncology Department, Lyon Sud University Hospital, Hospices Civils de Lyon, Pierre Bénite, France https://orcid.org/0000-0002-5064-4715

Julie Candiracci
Belgian Volition https://orcid.org/0000-0003-3811-4612

Gaelle Lescuyer
Medical Biology Multisite Laboratory (LBMMS), Department of Molecular Biology and Biochemistry, Lyon Sud Hospital, Hospices Civils de Lyon, Lyon, France https://orcid.org/0000-0002-5946-9474

Laurent Villeneuve
Center for Innovation in Cancerology of Lyon (CICLY), EA 3738, Lyon 1 University, Oullins, France
https://orcid.org/0000-0002-8713-7821

Michael Duruisseaux
Center for Innovation in Cancerology of Lyon (CICLY), EA 3738, Lyon 1 University, Oullins, France
https://orcid.org/0000-0003-1495-0966

Patrick Merle
Department of Pneumology, University Hospital Gabriel Montpied, Clermont-Ferrand, France
https://orcid.org/0000-0002-9435-9641

Patrick Mas
Department of Pneumology, Les Portes du Sud Hospital, Vénissieux, France

Sébastien Larivé
Department of Pneumology, Hospital Center of Mâcon, Mâcon, France

Sébastien Couraud
Acute Respiratory Disease and Thoracic Oncology Department, Lyon Sud University Hospital, Hospices Civils de Lyon, Pierre Bénite, France https://orcid.org/0000-0002-4362-0950

Marielle Herzog
Belgian Volition https://orcid.org/0000-0003-1311-101X

Marie Piecyk
marie.piecyk@chu-lyon.fr

Medical Biology Multisite Laboratory (LBMMS), Department of Molecular Biology and Biochemistry, Lyon Sud Hospital, Hospices Civils de Lyon, Lyon, France https://orcid.org/0009-0000-3765-9472

Léa Payen-Gay

lea.payen-gay@chu-lyon.fr

Medical Biology Multisite Laboratory (LBMMS), Department of Molecular Biology and Biochemistry, Lyon Sud Hospital, Hospices Civils de Lyon, Lyon, France

---

Research Article

Keywords: Lung Cancer, circulating nucleosome, H3K37Me3, risk stratification

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9880337/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare potential competing interests as follows: AK, JC and MH are employees of Belgian Volition, SRL. MH holds shares of VolitionRx Limited. Belgian Volition SRL has patents covering Nu.Q® technology and are developers of Nu.Q® assays. LP received financial support for consumable from VolitionRx and AstraZeneca. MP received VolitionRx support for congress participation. All other authors declare no financial or non-financial competing interests.

Prognostic value of circulating H3K27Me3-nucleosomes in newly diagnosed lung cancer patients

Aristotelis Kotronoulas[1*], Emmanuel Grolleau[2,3*], Julie Candiracci[1*], Gaelle Lescuyer[4], Laurent Villeneuve[5], Michael Duruisseaux[6,7], Patrick Merle[9], Patrick Mas[10], Sébastien Larivé[11], Sébastien Couraud[2,3], Marielle Herzog[1], Marie Piecyk [4,5Ψ], Léa Payen-Gay[4,5Ψ]

1.Belgian Volition SRL, 22 rue Phocas Lejeune, 5032 Isnes, Belgium

2. Acute Respiratory Disease and Thoracic Oncology Department, Lyon Sud University Hospital, Hospices Civils de Lyon, Pierre Bénite, France

3. RESHAPE Unit (Research on Healthcare Performance), INSERM U1290, Lyon 1 University, Lyon, France

4. Medical Biology Multisite Laboratory (LBMMS), Department of Molecular Biology and Biochemistry, Lyon Sud Hospital, Hospices Civils de Lyon, Lyon, France.

5. Center for Innovation in Cancerology of Lyon (CICLY), EA 3738, Lyon 1 University, Oullins, France.

6. Respiratory Department, Lyon East University Hospital – Louis Pradel, Hospices Civils de Lyon, Lyon, France

7. Oncopharmacology Laboratory, Cancer Research Center of Lyon, INSERM UMR 1052 CNRS 5286, Lyon, France

9. Department of Pneumology, University Hospital Gabriel Montpied, Clermont-Ferrand, France

10. Department of Pneumology, Les Portes du Sud Hospital, Vénissieux, France

11. Department of Pneumology, Hospital Center of Mâcon, Mâcon, France

*These authors contributed equally to this work

ΨCorresponding author: lea.payen-gay@chu-lyon.fr and marie.piecyk@chu-lyon.fr

Abstract

Lung cancer remains a leading cause of cancer-related mortality, with heterogeneous outcomes, underscoring the need for robust prognostic biomarkers. Circulating tumor DNA (ctDNA) enables detection of actionable mutations and may provide prognostic information; however, many patients lack detectable alterations, limiting its utility for molecular-based risk stratification. We analyzed plasma from 617 treatment-naïve patients using targeted next-generation sequencing for ctDNA and immunoassays to quantify circulating H3K27Me3-nucleosomes. Survival analyses included 514 patients with ≥8-month follow-up. Elevated H3K27Me3-nucleosome levels were independently associated with poorer overall survival. Patients with levels <56 ng/mL had significantly longer survival than those with higher concentrations (26 vs 12 months; $p<0.0001$), including in ctDNA-negative patients (29 vs 12 months; $p<0.0001$). A multivariable Cox model integrating H3K27Me3, stage, mutation, and age showed good performance (C-index 0.80) and was validated in an independent cohort. Circulating H3K27Me3-nucleosomes enable clinically relevant risk stratification at diagnosis, supporting more precise patient management in lung cancer.

Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, accounting for approximately 2.5 million new cases each year and 1.8 million deaths [1]. The disease is frequently diagnosed at an advanced stage. Data from the American Cancer Institute (NCI) statistics indicate that nearly 75% of patients present with locally advanced (~21%) or metastatic (~53%) disease at the time of diagnosis [2]. This late presentation is associated with a poor prognosis, as five-year overall survival in advanced stages typically ranges from approximately 12% to 55% and may fall below 10% in the most aggressive forms [3-6], largely driven by clinical stage and other factors such as tumor histology and treatment-modalities.

In light of these poor outcomes and the persistent unmet medical need, advances in the understanding of the lung cancer biology and molecular heterogeneity have enabled the development of innovative treatments, ushering in the era of precision oncology and revolutionizing patient care [7]. Tissue-based biomarkers, including PD-L1 expression, mutations in RAS, ERBB2, EGFR and BRAF; and ALK and ROS1 rearrangements, have profoundly transformed treatment selection by enabling immunotherapies and targeted agents [8]. Consequently, molecular profiling is now fully integrated into the routine management of lung cancer to guide therapeutic decisions for ultimately improving clinical outcomes. More recently, liquid biopsy approaches offer additional prospects compared to tissue biopsy, as measuring circulating biomarkers provides a minimally invasive and cost-effective procedure that can be easily repeated over time. Circulating tumor DNA (ctDNA) analysis by Next-Generation Sequencing (NGS) or Polymerase Chain Reaction (PCR), have emerged as powerful tools in lung cancer management, complementing or reducing the need for invasive tissue sampling while enabling diagnosis [9], guiding treatment selection [10] (e.g. ctDNA profiling can detect EGFR and ALK alterations to inform the use of appropriate tyrosine kinase inhibitors (TKIs) in EGFR- or ALK-mutant non-small cell lung cancer (NSCLC) patients), and can capture emerging resistance mechanisms during therapy [11,12] and minimal residual disease (MRD) monitoring [13]. While

potential benefits of ctDNA analysis are promising, they are not without limitations. Among these, negative ctDNA results represent an inability to detect circulating tumor material, either because of the low sensitivity of the technique or because the ctDNA does not contain any detectable mutations. Indeed, about 30-50% of NSLC patients don't have common, actionable mutations when tested with comprehensive molecular profiling methods [14]. Therefore, while predictive biomarkers have transformed therapeutic decision-making [15], the identification of reliable prognostic markers is still lacking. Such tools are essential to better stratify patients by risk, anticipate clinical evolution, and tailor interventions beyond current standards [16]. Prognostic biomarkers could guide dynamic clinical decision-making by identifying low-risk patients suitable for watch-and-wait strategies and high-risk individuals who may benefit from early imaging, treatment escalation, or clinical trial enrollment. This risk-adapted approach advances precision oncology toward proactive, biology-informed care supported by clinical decision-support frameworks.

Epigenetic biomarkers offer a powerful means to capture additional layers of tumor biology beyond genomic alterations, thereby addressing the complexity and heterogeneity that limit the current precision oncology approaches. ctDNA methylation assays have demonstrated the ability to identify the tissue of origin and to classify tumors based on CpG island methylation profiles [17]. Beyond DNA methylation, other epigenetic biomarkers have emerged. Histone post-translational modifications (PTMs) carried by nucleosomes (DNA wrapped around histone core proteins) represent critical regulators of chromatin structure and gene expression. Among these, the tri-methylation of lysine 27 on histone H3 (H3K27Me3), is a well-established epigenetic mark associated with gene silencing and transcriptional regulation [18,19]. This mark plays a crucial role in tumorigenesis, cell cycle progression, and dysregulated proliferation [20]. H3K27Me3 is catalyzed by the Polycomb repressive complex 2 (PRC2), whose catalytic subunit, Enhancer of zeste homolog 2 (EZH2), mediates the trimethylation of histone H3 at Lysine 27 and deposits this repressive mark on the promoters or enhancers of target genes [21,22]. The resulting modification

recruits PRC1, leading to chromatin compaction and transcriptional repression [23,24]. Dysregulation of specific PTMs, such as methylation patterns of lysine 9 and 27 on histone H3 have been associated with tumor heterogeneity and resistance to targeted therapies [25-27]. Histone PTMs are increasingly recognized as clinically relevant biomarkers, while also representing attractive pharmacological targets, thereby expanding the molecular toolkit for lung cancer management. Importantly, the integration of epigenetic biomarkers into personalized medicine frameworks holds promise for refining patient stratification and treatment selection. When combined with longitudinal molecular and clinical data, these biomarkers may further support to the development of emerging "digital twin" models, enabling dynamic representation of the tumor biology to guide individualized therapeutic, immunotherapeutic, and surveillance strategies.

In a previous study, circulating H3K27Me3-nucleosome was demonstrated to complement ctDNA analysis by enhancing the reliability of negative molecular results at the time of diagnosis [28]. In this real-life study, we evaluated the potential of the baseline level of circulating H3K27Me3-nucleosomes as a prognostic biomarker for overall survival and risk stratification of patients with lung cancer. The measurement of H3K27Me3-nucleosomes provides prognostic information that complements ctDNA analysis. Furthermore, we developed and validated a Cox proportional hazards (CPH) model that integrates these two biomarkers with key clinical covariates (age and cancer stage) to predict survival which is suitable for implementation in routine clinical practice.

Results

Circulating H3K27Me3-nucleosomes associate with cancer stage in a well-defined real-world study cohort

All 617 analyzed samples came from treatment-naïve patients, either with suspected or confirmed cancer at the time of inclusion, and for whom both ctDNA profiling and H3K27Me3-nucleosome measurement were available (Figure 1).

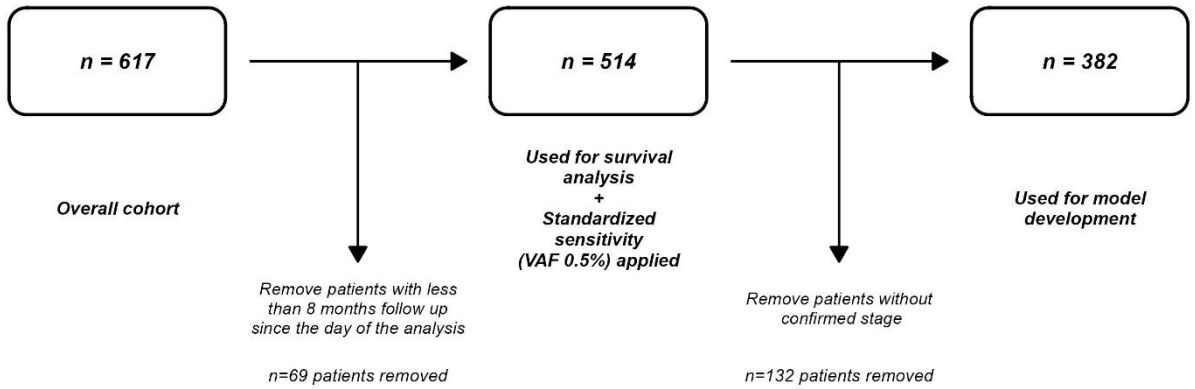


Figure 1: Flowchart of the study population analysis. *The association between H3K27Me3-nucleosome levels and disease severity was assessed on the overall cohort. Survival analyses were conducted on 514 patients with a minimum follow-up of 8 months. The Cox proportional hazards model was developed on 382 lung cancer patients with available cancer stage.*

Within the overall cohort, the distribution of patients according to clinical stage of the disease at the time of inclusion is as follows: (1) suspicious imagery (n = 34, 6%), (2) localized to locally-advanced cancer (stage I–III; n = 130, 21%), (3) metastatic cancer (stage IV; n = 331, 54%), and (4) cancer with undetermined stage (n = 122, 20%) (Table 1). Patient distribution by NGS technology: Plasma-SeqSensei (n=231, 37%), a custom comprehensive panel (n=189, 31%), and a custom targeted ultra-deep assay (n=197, 32%) was equivalent and corresponded to the sequential implementation of NGS technologies in the laboratory during the recruitment period. By the end of the follow-up period, 369 patients (60%) were alive (overall survival [OS], 21.8 months), among the entire cohort.

Table 1: Study population characteristics and distribution by disease stage and NGS type.

Overall patients (n=617)		
Age at inclusion (mean +/-SD)	73.3 (± 12.5)	
Disease stage	**n**	**%**
Suspicious imagery	34	6%
Stage I-III	130	21%
Stage IV	331	54%
Cancer with undetermined stage	122	20%
NGS type	**n**	**%**
PlasmaSeqSensei, Sysmex	231	37%
Custom comprehensive panel, SOPHiA Genetics	189	31%
Custom targeted ultra-deep technique, SOPHiA Genetics	197	32%
Mean follow-up (mean +/- SD)	12.6 (± 10.0)	
Survival - Yes (n, %)	369	60%
Restricted mean survival times (RMST in months)	21.8	

We examined the association between H3K27Me3-nucleosome levels and disease severity (Figure 2). Median H3K27Me3-nucleosome concentrations increased across groups: Healthy (n=201; median 11.5 ng/mL), Suspicious Imagery (n=34; median 12.9 ng/mL), Stage I–III (n=130; median 16.0 ng/mL), Stage IV (n=331; median 28.7 ng/mL), and cancer with undetermined stage (n=122; median 39.1 ng/mL). Between-group differences were assessed on log-transformed H3K27Me3-nucleosome levels and were significant overall (Kruskal–Wallis: $H = 342$, $df = 4$, $P < 1e$-

04, $\varepsilon^2 = 0.416$). Post hoc pairwise comparisons confirmed a progressive increase with advancing stage, with significantly higher levels in Stage I–III than in Healthy controls ($p = 2.5 \times 10^{-14}$), and were further increased in Stage IV relative to both Stage I–III ($p = 3.4 \times 10^{-8}$) and Healthy controls ($p = 2.9 \times 10^{-61}$). The suspicious imaging group showed levels comparable to those observed in patients with stage I–III disease, whereas the undetermined stage group exhibited levels similar to those of stage IV patients (all comparisons not statistically significant).

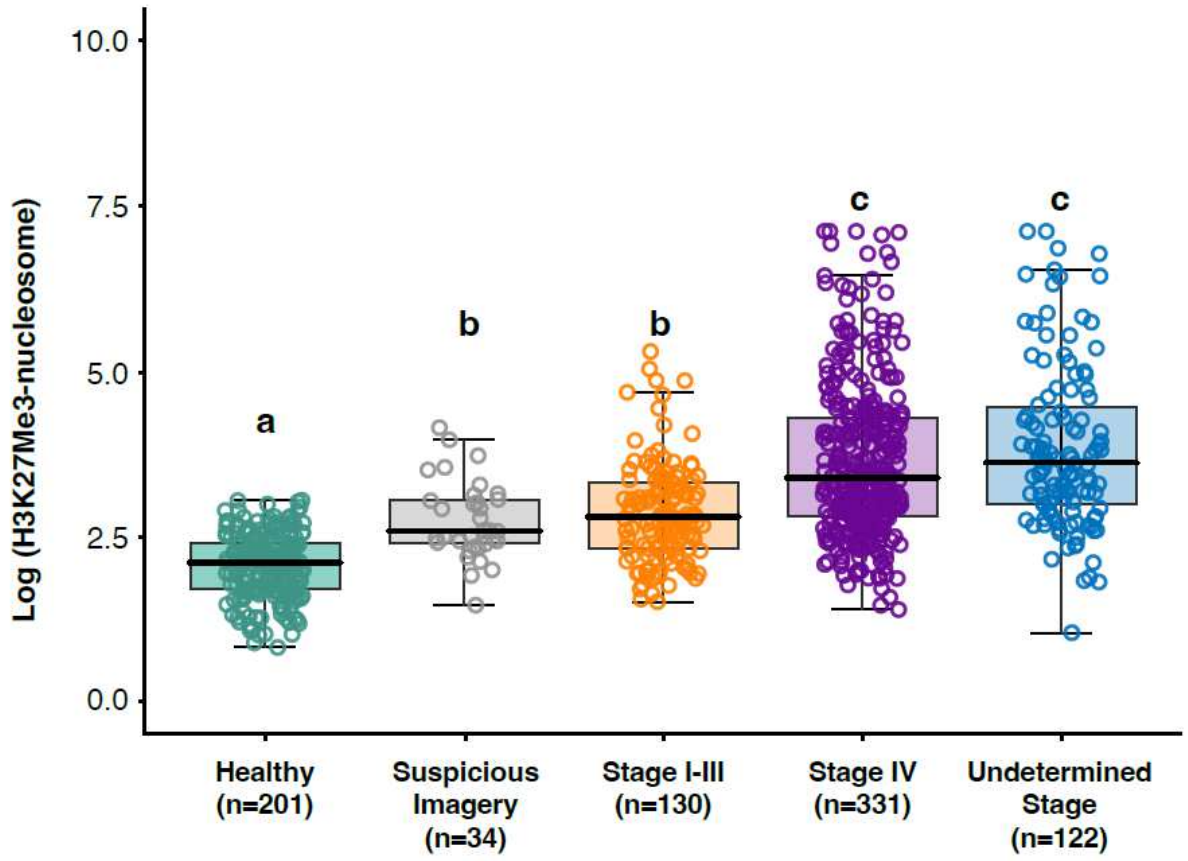


Figure 2: Circulating H3K27Me3-nucleosome levels increase with lung cancer stage. *Box plots show histone H3 lysine 27 trimethylation (H3K27Me3)-nucleosome levels increasing across five groups: Healthy, Suspicious Imagery, lung cancer Stage I–III, lung cancer Stage IV, and lung cancer with Undetermined stage. Groups sharing the same letter are not significantly different, whereas groups with different letters differ at $p < 0.05$ (multiple comparison adjusted). Box plots display the median (center line) and interquartile range (IQR; box), with whiskers extending to 1.5× the IQR.*

We also compared H3K27Me3-nucleosome levels in function of the ctDNA status of the patients. Patients were classified as ctDNA negative (ctDNA−) in the absence of detectable pathogenic somatic alteration, and as ctDNA positive (ctDNA+) when at least one somatic pathogenic alteration was identified. ctDNA- patients had significantly lower H3K27Me3-nucleosome levels than those who were ctDNA+ (median = 20.4 vs 29.5 ng/mL respectively, p-value = $3.99*10^{-6}$) (Sup. Figure S1).

Survival analysis demonstrates that circulating H3K27Me3-nuclesomes can be a powerful prognostic and risk-stratification biomarker

Survival analysis was performed on 514 evaluable patients with a minimum follow-up of 8 months (Figure 1). To facilitate cross-method comparability, a uniform sensitivity threshold of 0.5% variant allele frequency (VAF) was imposed across all the NGS methods. This harmonization does not reflect the higher analytical sensitivity of the custom ultra-deep assay (0.1% VAF), but it allowed results to be pooled and used for survival analyses. In the Plasma-SeqSensei cohort, TP53 was not assessed and was therefore considered ctDNA-. Overall, among the 514 evaluable patients, 330 (64%) were classified as ctDNA-, whereas 184 (36%) harbored at least one somatic alteration and were classified as ctDNA+ (Table 2). Stage distribution differed according to ctDNA status but showed substantial overlap across subgroups (Table 2). Among patients with localized or locally advanced disease (stage I–III, n=106), 86.8% were ctDNA- (92/106). Metastatic disease (stage IV) was present in both ctDNA groups: 51% of stage IV patients were ctDNA- (141/276), and 49% were ctDNA+ (135/276). When examining stage composition within each ctDNA subgroup, all disease stages were represented. In the ctDNA- group (n=330), patients were distributed across all disease stages, including 42.7% with stage IV. A comparable stage distribution was observed in the ctDNA+ group (n=184), although the proportion of metastatic cases was higher (73.4%) (Table 2).

Table 2: Characteristics of the survival analyses cohort, stratified by ctDNA status and ctDNA mutation group.

Survival analysis set (n=514)		Age at inclusion (mean +/- SD)	Disease stage				Mean follow-up (+/- SD)
			Suspicious Imagery	Stage I-III	Stage IV	Cancer with undetermined stage	
ctDNA - (n=330)		73.2 (± 12.5)	27	92	141	70	16.1 (± 10.3)
ctDNA + (n= 184)		72.6 (± 12.3)	-	14	135	35	10.5 (± 9.2)
ctDNA+ (classified by mutation type)*	**EGFR (n=33)**	73.5 (± 10.9)	-	-	23	9	17.0 (± 10.3)
	KRAS (n= 61)	69.7 (± 13.4)	-	4	48	10	9.9 (± 9.5)
	TP53 (n= 54)	73.1 (± 11.7)	-	6	41	7	7.5 (± 5.7)
	Other (n= 36)	76.5 (± 11.6)	-	4	23	9	10.1 (± 9.7)

**Patients harboring multiple ctDNA mutations were assigned to a single category according to the following priority:*

EGFR → KRAS → TP53 → Other

The ctDNA+ group was subsequently subdivided into four molecular subgroups based on predefined priority order (EGFR, KRAS, TP53, and other mutations). The proportion of patients with stage IV disease was maintained across these subgroups. Notably, no cases of localized or locally advanced cancer (stage I-III) were observed in the EGFR subgroup.

Building on these findings, we assessed the ability of H3K27Me3-nucleosomes to stratify patients by mortality risk. To this end, we systematically scanned potential concentration cut-off and applied log-rank tests across different subgroups of our cohort. In the patients eligible for survival

analysis (n=514), an optimal cut-off value of 56 ng/mL of H3K27Me3-nucleosome significantly separated patients into two groups: a group below the cut-off *(orange; n = 382)* and a group equal or above it *(green; n = 132) (*log-rank P = $9.4*10^{-23}$; HR = 3.43, 95% CI 2.62–4.47) (Figure 3A). This H3K27Me3-nucleosome threshold was associated with marked survival differences as reflected in RMST values: 21.9 months for the below 56 ng/mL group and 10.9 months for the ≥56 ng/mL group, while the unstratified cohort (grey) showed an RMST of 19.1 months (Figure 3A).

We next examined survival by ctDNA status (Figure 3B). ctDNA− patients (blue; n = 330) had higher survival and longer RMST (21.5 months) than ctDNA+ patients (red; n = 184; RMST = 14.4 months), with a significant log-rank difference (log-rank P = $1.9*10^{-10}$; HR = 2.30, 95% CI 1.77–2.99). Interestingly, when we applied H3K27Me3-nucleosome optimal cut-off method within each ctDNA stratum, subgroups with significant different survival probabilities were identified (Figure 3C and D). Among ctDNA− patients, an optimal cut-off of 57 ng/mL H3K27Me3-nucleosome identified two groups with significantly different survival probabilities: patients with circulating H3K27Me3-nucleosome concentrations below 57 ng/mL have significantly better survival (light blue; n = 265; RMST = 24.2 months) compared to the group of patients with a H3K27Me3-nucleosome levels equal or above 57 ng/mL (dark blue; n = 65; RMST = 10.9 months, log-rank P = $1.0*10^{-20}$, HR = 5.07 (95% CI 3.48–7.38) (Figure 3C). Notably, the group ≥ 57 ng/mL comprised 19.7% of the total ctDNA− patients, indicating that H3K27Me3-nucleosome marker can uncover a substantial "high-risk" subset within an otherwise favorable ctDNA− category. Both risk groups included patients from all cancer stage groups as shown in Supplemental Table S1.

In parallel, an analysis in ctDNA+ patients identified an optimal cut-off of 123 ng/mL that separated "low-risk" (< 123 ng/mL, light red; n = 154; RMST = 15.8 months) and "high-risk" (≥ 123 ng/mL, dark red; n = 30; RMST = 7.3 months) subgroups with log-rank P = $1.1*10^{-05}$ and HR = 2.65 (95% CI 1.69–4.17) (Figure 3D).

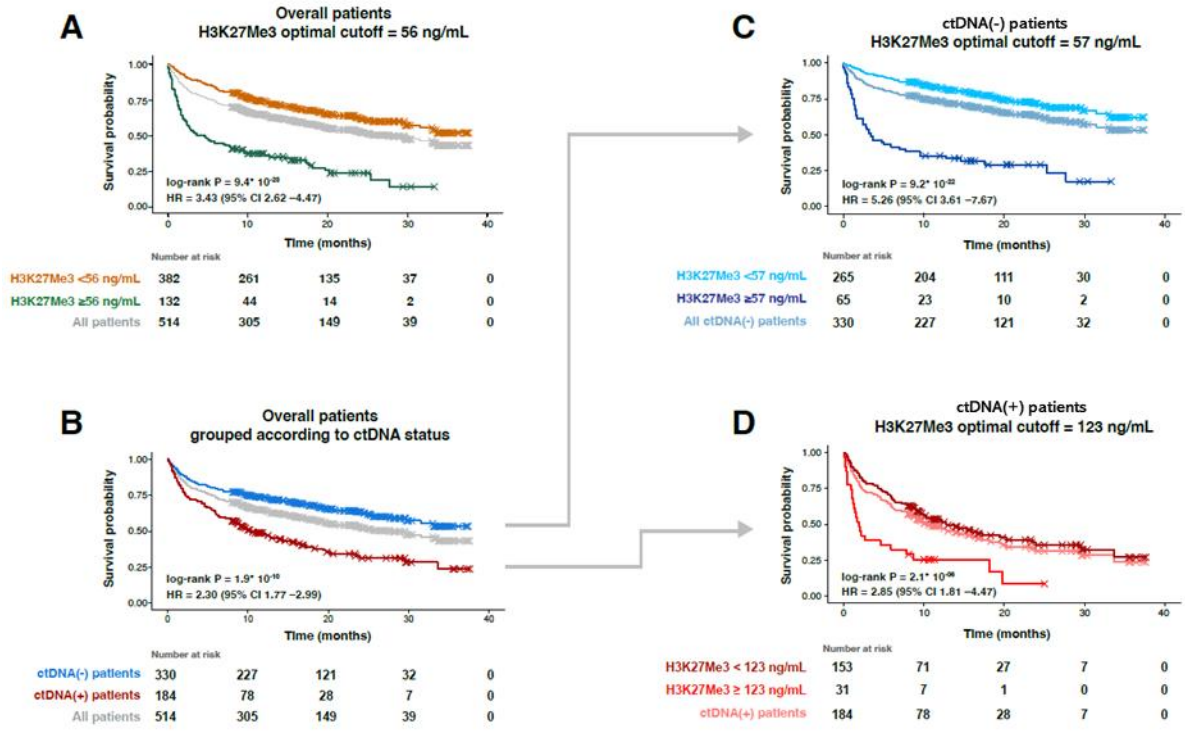


Figure 3: Analysis of H3K27Me3-nucleosomes cut-offs enhances survival stratification across ctDNA status groups. ***(A) Optimal H3K27Me3-nucleosome cut-off value of 56 ng/mL classifies overall patients in two different mortality risk groups:*** *patients with H3K27Me3 levels below the optimal cut-off (orange) have higher survival probabilities (restricted mean survival time [RMST] =21.9 months) than patients with H3K27Me3 levels above the optimal cut-off (green; RMST= 10.9 months). In grey, Kaplan-Meier (K-M) for all patients independent of the H3K27Me3-nucleosome levels (RMST= 19.1 months).* ***(B) Distinct survival probabilities between ctDNA- and ctDNA+ patients:*** *significant higher survival probabilities were observed in ctDNA- patients (blue; RMST= 21.5 months) compared to ctDNA+ patients (red; RMST= 14.4 months). In grey, K-M for all patients independent of the ctDNA status (RMST= 19.1 months).* ***(C) Optimal H3K27Me3-nucleosome cut-off value of 57 ng/mL classifies ctDNA- patients in two groups with different survival outcomes****: ctDNA- patients with H3K27Me3 levels below the optimal cut-off (light blue) have higher survival probabilities (RMST= 24.2 months) compared to patients above the optimal cut-off (deep blue; RMST= 10.9 months). In grey blue, K-M for all ctDNA- patients, independent of the H3K27Me3-nucleosome levels (RMST= 21.5 months).* ***(D) Optimal H3K27Me3-nucleosome cut-off value of 123 ng/mL classifies ctDNA+ patients in two groups with different survival outcomes:*** *Higher survival probabilities (RMST= 15.8 months) were observed for ctDNA+ patients with H3K27Me3 levels below the optimal cut-off (light red) compared to those above (deep red; RMST= 7.3 months)). In pink, K-M for all ctDNA+ patients, independent of the H3K27Me3-nucleosome levels (RMST= 14.4 months). For each panel, the between-group comparison (log-rank test p value, hazard ratio [HR], and*

95% confidence interval [CI]) and the number of patients at risk are reported. Tick marks indicate censored individuals.

We next evaluated clinically defined subgroups, establishing optimal H3K27Me3–nucleosome cut-off values according to cancer stage and ctDNA mutation (EGFR, KRAS, TP53, and other mutations) (Sup. Figures S2 and S3). As observed previously, across all clinical and molecular subgroups, H3K27Me3–nucleosome levels above the cut-off were systematically associated with significantly worse outcomes.

Taken together, these results demonstrate that circulating H3K27Me3-nucleosome levels provide clinically meaningful risk stratification for overall survival in the study cohort and retain significant prognostic value within the ctDNA-negative subgroup.

H3K27Me3-based Cox model stratifies risk at diagnosis and predicts survival probabilities at 30 months

Given the strong prognostic value of circulating H3K27Me3-nucleosome, we developed a Cox proportional hazards (CPH) model to estimate survival at diagnosis. Although the optimal cut-off varied across molecular and clinical strata (Figure 3; Sup. Figure S2, and Sup. Figure S3), we used the cohort-wide cut-off (56 ng/mL; ≥ vs <) to ensure model interpretability and facilitate clinical implementation. The final model included four routinely available covariates: (i) H3K27Me3-nucleosome concentration, (ii) cancer stage (I–III vs IV), (iii) ctDNA molecular profile (ctDNA- vs EGFR vs KRAS vs TP53 vs Other mutations) and (iv) age (per year).

The model development and validation are summarized in Figure 4A. Analysis included patient with complete data and ≥8 months of follow-up (n=382) randomly split into training set (70%; n = 267, events = 125) for model development and held-out test set (30%; n = 115, events = 47) for final evaluation, maintaining strict train/test separation throughout. Model training used nested, stratified 5-fold cross-validation, and subsequent calibration were performed as described in Methods. Full model parameters are provided in Supplemental Table S2.

With the model fixed, performance was evaluated on the held-out test set. Time-dependent discrimination was consistently high, with AUC(t) values of 0.879, 0.888, 0.841, 0.876, and 0.777 at 6, 12, 18, 24, and 30 months, respectively (Sup. Figure S4A). Calibration remained acceptable over time, with an integrated Brier score of 0.146 between 3 and 30 months (Sup. Figure S4B).

Crucially, these characteristics translated into answers to two routine clinical questions at diagnosis: i) Can I stratify a new patient in low- or high-risk mortality group? ii) Can I predict his survival probabilities in different time points? Using the pre-specified calibrated 30-month mortality risk threshold (calculated during development), patients were stratified into Low- (n=55) or High-risk (n=60) of dying groups. These groups showed clearly separated K-M curves (P = 2.8×10-9; HR = 7.39, 95% CI 3.43–15.90), distinct restricted mean survival times (Low-risk = 27.4 months; High-risk = 13.4 months) and a Harrell's C-index of 0.80 (Figure 4B).

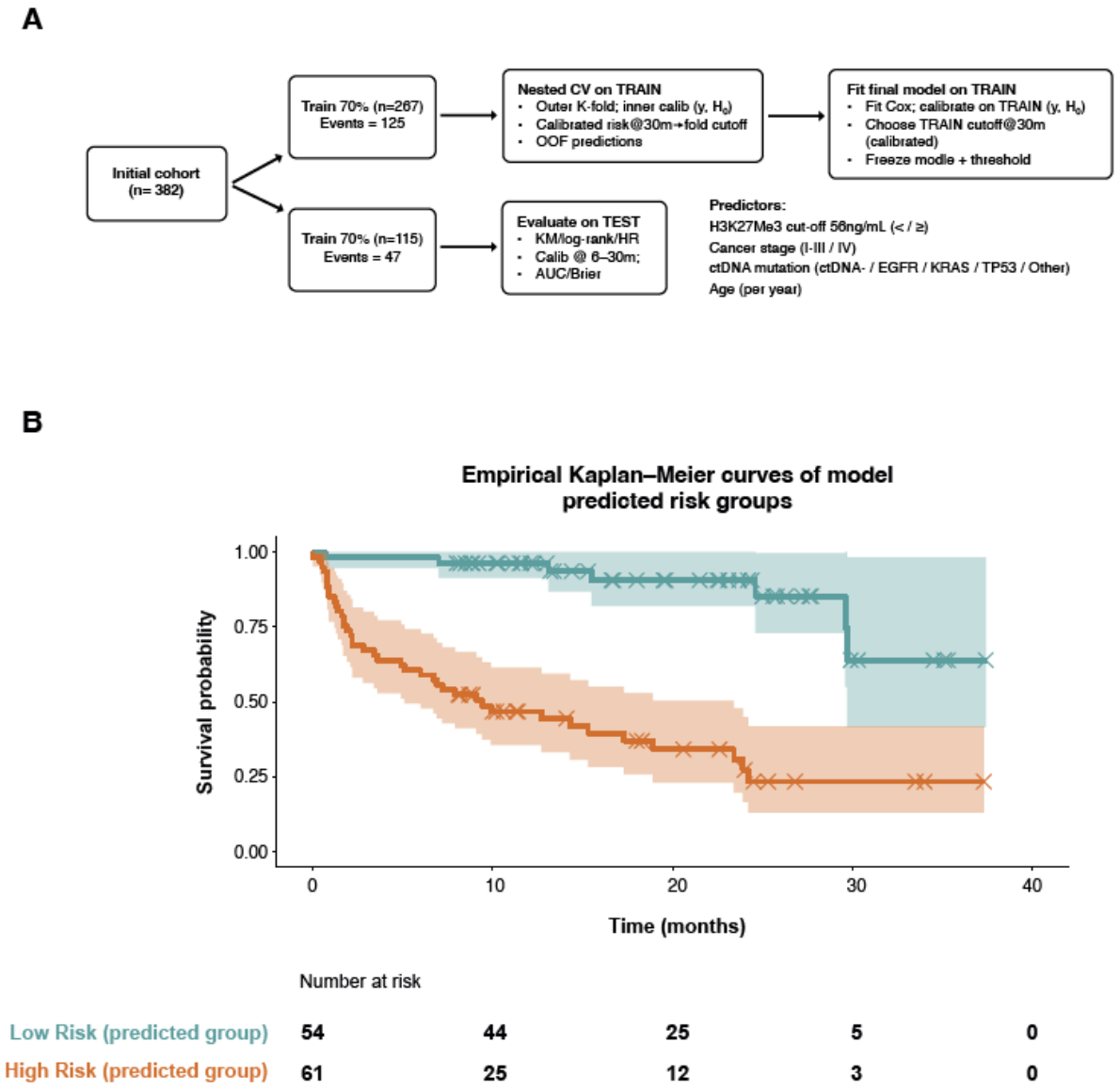


Figure 4: Cox proportional hazards model development description. ***(A) Overall workflow and predictors used for the model development. (B) Model performance on the held-out test set indicated accurate risk stratification at diagnosis and reliable prediction of 30-month survival probabilities.*** *Kaplan-Meier (K-M) survival curves for predicted Low-risk (turquoise; restricted mean survival time [RMST] = 27.4 months) and High-risk (salmon; RMST= 13.4 months) groups in the held-out test set, showing significantly different survival probabilities (log-rank P = 2.8*10^{-9}, hazard ratio [HR] = 7.39, 95% confidence interval [CI]: 3.43–15.90). The number of patients at risk is reported below the plot. Tick marks indicate censored individuals.*

The model also provided patient-level survival probabilities at clinically relevant time points (6-30 months), which closely matched empirical K-M estimates within each risk group (difference of mean predicted and empirical probability ≤ ~0.12 across horizons) (Table 3).

Table 3: Predicted vs empirical survival probabilities at fixed horizons times (6-30 months).

Group	Time (months)	n	S predicted median	S predicted mean	S empirical K-M	Diff mean	Risk pred	Risk KM
Low Risk	6	54	0.92	0.92	0.98	-0.07	0.08	0.02
High Risk	6	61	0.63	0.71	0.59	0.04	0.37	0.41
Low Risk	12	54	0.87	0.88	0.96	-0.09	0.13	0.04
High Risk	12	61	0.51	0.59	0.47	0.04	0.49	0.53
Low Risk	18	54	0.8	0.81	0.91	-0.11	0.2	0.09
High Risk	18	61	0.36	0.42	0.37	-0.01	0.64	0.63
Low Risk	24	54	0.76	0.77	0.91	-0.14	0.24	0.09
High Risk	24	61	0.31	0.35	0.27	0.04	0.69	0.73
Low Risk	30	54	0.73	0.74	0.64	0.09	0.27	0.36
High Risk	30	61	0.27	0.3	0.24	0.03	0.73	0.77

For each group (Low/High) and time point (6, 12, 18, 24, 30 months), the table reports the mean and median of the individual predicted survival probabilities (S predicted), alongside the calculated Kaplan–Meier survival estimate (S empirical K-M, 95% CI) at the same time horizon; diff mean is the difference between (S predicted mean) – (S empirical K-M); Risk pred and Risk K-M are the predicted and empirical cumulative event probability given by the formula 1−(S predicted mean) and 1−(S empirical K-M) respectively.

Together, these results show that the model maintains clear risk separation at every horizon while providing numerically faithful survival estimates.

External validation of the locked model confirms preserved discrimination and clinical risk stratification

In the external validation cohort, the locked CPH model stratified patients into predefined Low- and High-risk groups with significantly different survival across the entire follow-up (Figure 5). K–M analysis demonstrated lower survival probabilities in the High-risk group (log-rank $P = 7.0 \times 10^{-04}$; HR = 2.1, 95% CI 1.36–3.36). At 30-months, survival time was reduced in High-risk compared to Low-risk patients (RMST = 8.7 vs 14.8 months). Discrimination based on the continuous linear predictor yielded a Harrell's C-index of 0.62 (95% CI 0.56–0.67). Importantly, between-group differences remained statistically significant, confirming the transportability of the model to an external cohort.

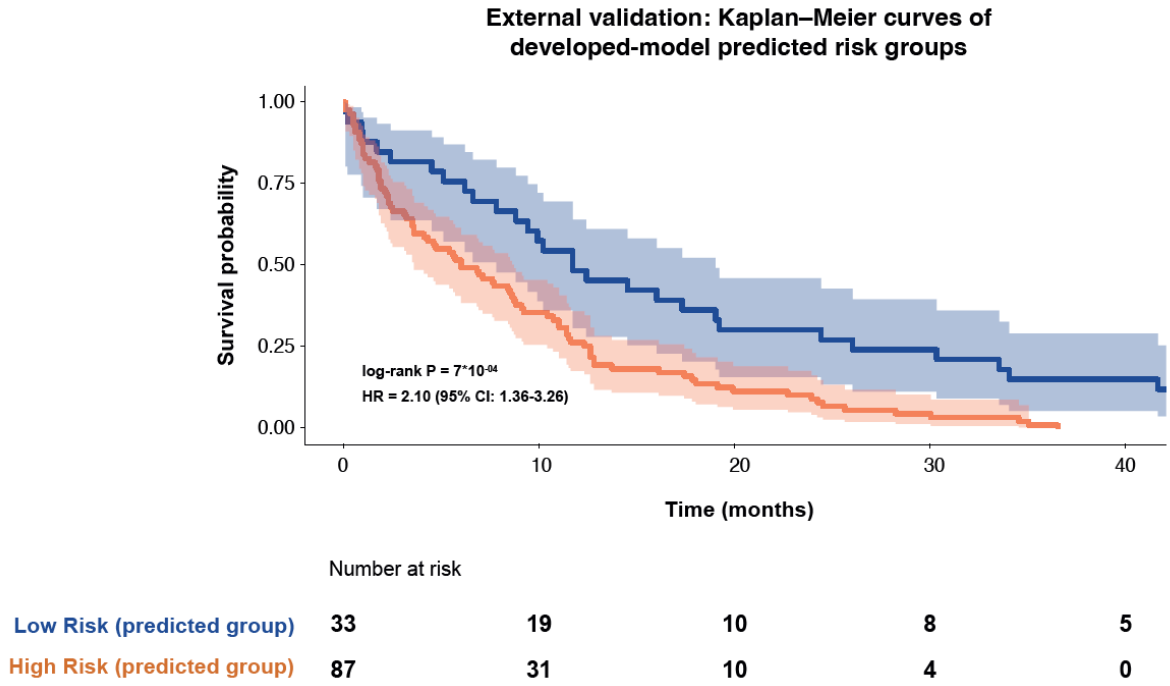


Figure 5: External validation of the locked Cox model. *External validation of the locked Cox model. Kaplan–Meier survival curves in the independent validation cohort for patients classified as low risk (blue; restricted mean survival time [RMST] = 14.8 months) and high risk (orange; RMST = 8.7 months), showing clear separation between the two groups (log-rank P = 7.0*10^{-04}, hazard ratio [HR] = 2.1, 95% confidence interval [CI]: 1.36-3.26). The number of patients at risk is reported below the plot. Tick marks indicate censored individuals.*

Overall, although discrimination was attenuated compared to internal validation, clinically meaningful risk stratification was preserved, with predefined risk groups remaining well separated and exhibiting significant differences in overall survival and restricted mean survival time.

Translation of the multivariable Cox model into a clinically applicable nomogram

Our multivariable Cox proportional hazards model was translated into a points-based nomogram for clinical usage. The nomogram assigns points to each predictor in direct proportion to its Cox coefficient and scales the largest partial effect to 100 points. Continuous variables (e.g. age) are displayed on linear axes, so that moving from the lower to upper bound contributes up to 100 points. Categorical variables (H3K27Me3_concentration, Cancer stage, ctDNA mutation group) are shown as tick marks; the reference level yields 0 points, and each non-reference level contributes points according to its coefficient sign and magnitude. For any patient, points from all predictors are summed on the 'Total Points' ruler and then projected downward to (i) the linear predictor (LP) scale and (ii) survival-probability rulers at 6, 12, 18, 24 and 30 months. Thus, a patient with H3K27Me3-nucleosome below the cut-off (0 points), Stage=IV (33.2 points), Mutation=EGFR (38.3 points), and Age=85 years (76.1 points) totals 147.6 points, corresponding to LP= 1.00799 and predicted survival probabilities of S(6)=79.5%, S(12)=69.7%, S(18)=55.2%, S(24)=48.8% and S(30)=43.6%. Figure 6 includes the individual predictor axes, the total-points ruler, the LP ruler, and time-specific survival rulers to facilitate bedside use.

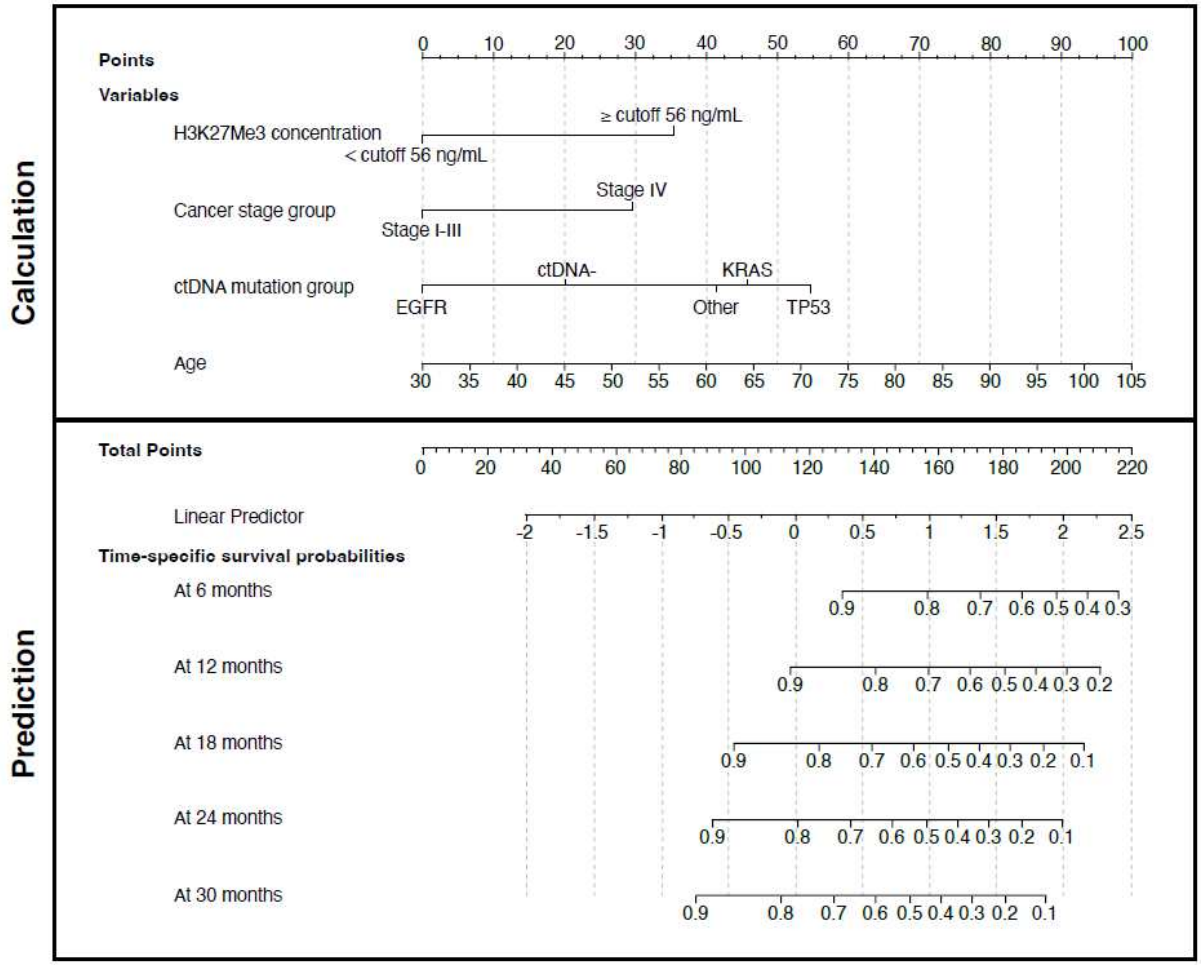


Figure 6: Nomogram for predicting the overall survival probability up to 30 months at personnel level.

Discussion

Lung cancer is characterized by marked biological heterogeneity, underscoring the urgent need for robust prognostic biomarkers capable of refining risk stratification and guiding clinical decision making beyond current clinicopathological standards. Conventional stratification tools—such as TNM classification or cancer stage, and age—combined with molecular profiling have significantly advanced prognostic modeling but still fall short of capturing inter-patient variability in tumor behavior and microenvironmental complexity [29]. Consequently, there is sustained interest in biomarkers capable of integrating multi-dimensional biological information,

to achieve truly personalized patient management. While ctDNA-based assays have emerged as valuable prognostic tools in NSCLC, their clinical utility remains limited by incomplete representation of tumor heterogeneity, challenges in interpreting negative results, and substantial technical and cost-related constraints [30]. These limitations highlight the need for complementary, biologically informative, and clinically feasible circulating biomarkers.

In this context, we investigated circulating H3K27Me3-nucleosomes as a prognostic circulating biomarker that bridges biological relevance and clinical feasibility. As circulating carriers of both genetic and epigenetic information, H3K27Me3-nucleosomes reflect key features of tumor biology, including cellular turnover, and tumor burden. Importantly, their measurement using a standardized, immunoassay-based approach enables cost-effective, rapid, and reproducible assessment, making them particularly attractive for integration into routine clinical workflows. Previous study has demonstrated the added value of H3K27Me3-nucleosomes when combined with molecular profiling, notably for the detection of circulating tumor material at diagnosis and for minimal residual disease monitoring [28]. Our findings extend this concept by demonstrating that H3K27Me3-nucleosome levels provide incremental prognostic value at diagnosis, beyond established clinical and molecular variables, supporting their role as complementary biomarker within a multimodal prognostic framework.

Importantly, circulating H3K27Me3-nucleosomes levels were associated with both cancer stage and mortality, independently of molecular profiling. This observation suggests that H3K27Me3-nucleosomes capture biological features not fully reflected by current genomic assays, potentially integrating signals related to tumor burden, chromatin dynamics, or microenvironmental processes. Rather than duplicating genomic information, H3K27Me3-nucleosome appears to add a distinct layer of prognostic stratification. From a clinical perspective, the consistent association observed across clinical and molecular subgroups supports the potential utility of H3K27Me3-nucleosomes as a broadly applicable prognostic biomarker, including in settings where comprehensive genomic testing is limited or inconclusive.

Consistent with this concept, H3K37Me3-nucleosome levels were associated with overall survival in newly diagnosed lung cancer patients and improved risk discrimination and model performance when combined with molecular profiling. Indeed, patients with H3K27Me3-nucleosome levels below 56 ng/ml exhibited significantly longer mean survival compared with those with higher concentrations. More interestingly, this survival advantage persisted and appeared even more pronounced among ctDNA-negative patients. In patients without detectable actionable driver mutations found in blood or tissue analysis, therapeutic decision-making currently relies exclusively on histopathological classification and disease stage, as no validated biomarkers are available to further refine risk stratification. Notably, this population represents a substantial proportion of lung cancer patients: in metastatic disease, approximately 50% of patients lack identifiable actionable driver alterations, depending on the population studied and the comprehensiveness of molecular testing, and similar proportions are likely observed in stage III disease, although comprehensive NGS profiling is not systematically performed in routine practice at this stage [31,32]. Consequently, standard treatment approaches are applied uniformly within stage-defined categories, including perioperative chemo-immunotherapy for resectable stage III NSCLC and chemotherapy with or without radiotherapy for inoperable disease, as well as adjuvant immunotherapy for patient treated with chemoradiotherapy. In this context, our findings position H3K27Me3-nucleosome as a novel biomarker that provides additional prognostic discrimination beyond conventional clinicopathological parameters, with the potential to refine risk stratification and support more personalized treatment decisions. Notably, quantification of H3K27Me3-nucleosome levels enables further stratification of ctDNA-negative patients - a subgroup often considered analytically uniform due to the absence of detectable plasma-based molecular alterations. However, the lack of detectable ctDNA does not necessarily reflect biological uniformity, and our results suggest that epigenetic circulating biomarkers may uncover clinically relevant heterogeneity within this population. Elevated H3K27Me3-nucleosome levels in this group were associated with significantly worse survival

outcomes, independent of disease stage, highlighting the potential of this biomarker to identify high-risk patients who may benefit from treatment intensification and/or closer longitudinal monitoring. Consistent trends were also observed in the ctDNA-positive subgroup, which overall exhibited higher H3K27Me3-nucleosome levels. Within this group, patients with nucleosome concentrations above 123ng/mL experienced particularly unfavorable outcomes, further supporting the prognostic relevance of H3K27Me3-nucleosome levels across distinct molecular contexts.

While H3K27Me3-nucleosome alone demonstrates meaningful prognostic value, its utility can be further contextualized when interpreted alongside established clinical and molecular parameters. Integrating these variables into multivariate Cox proportional hazards models, together with clinicopathological features, enables more refined risk stratification and supports increasingly personalized management strategies in lung cancer. In this framework, H3K27Me3-nucleosome was consistently identified as a key independent prognostic biomarker for overall survival. Beyond statistical significance, this survival modeling approach facilitates translation of H3K27Me3-nucleosome measurements into clinically actionable information at the time of diagnosis. The model integrates H3K27Me3-nucleosome with routinely available parameters, including disease stage, ctDNA mutational status, and age, to estimate patient prognosis and guide management decisions. First, it enables robust risk stratification of newly diagnosed patients into low- and high-risk mortality groups based on their estimated hazard potentially informing treatment intensity, follow-up schedules, and selection for intensified or experimental therapeutic approaches. Second, it allows estimation of individualized survival probabilities across clinically relevant time horizons by combining the effect of H3K27Me3-nucleosome with baseline hazard estimates. Survival probabilities can thus be predicted at specific time points (e.g., 6, 12, up to 30 months survival) providing a more nuanced and patient-centered prognostic assessment than categorical risk classification alone and may facilitate patient counseling and shared decision-making at diagnosis. The ability of H3K27Me3-nucleosomes to inform both

relative risk (hazard ratio) and absolute survival probabilities underscores their potential value as a clinically meaningful and interpretable biomarker.

Several limitations of this study should be acknowledged. First, biomarker assessment was performed at a single time point at diagnosis, precluding evaluation of longitudinal changes during treatment and follow-up. While this baseline-focused design reflects current clinical practice for initial risk stratification, future studies incorporating repeated measurements will be essential to investigate the dynamic prognostic value of H3K27Me3-nucleosome over the disease course. Second, inclusion in the multivariable prognostic model required complete data for all four parameters, leading to the exclusion of patients with missing information. This requirement may represent a practical limitation, as comprehensive staging may still be ongoing at diagnosis, and ctDNA molecular profiling is not systematically performed in routine practice. In addition, ctDNA detectability may vary depending on the sequencing panel and analytical approaches used, particularly with respect to sensitivity and specificity. Importantly, however, H3K27Me3-nucleosome levels alone demonstrated meaningful prognostic value across patient subgroups, supporting clinical utility even in settings where complete molecular and clinicopathological information is not available. Third, the absence of detailed clinical and pathological data restricted comprehensive adjustment for potential confounders and preclude full assessment of the representativeness of the study population. In particular, incomplete staging information resulted in a subset of patients with undetermined stage, thereby preventing precise evaluation of stage distribution. However, patient recruitment across centers may mitigate concerns regarding potential selection bias. Nevertheless, external validation in independent cohorts and longitudinal studies will be essential to confirm the robustness and clinical utility of this biomarker within a precision medicine framework.

Finally, the nomogram developed in the present study provides a user-friendly and clinically interpretable tool for estimating individualized survival probabilities by integrating H3K27Me3-nucleosome levels with routinely applied prognostic factors, including disease stage, ctDNA

mutational status, and age. While currently designed for use at diagnosis, this tool could be further refined into a dynamic prognostic model enabling risk recalibration throughout follow-up, supporting efforts toward data-driven personalized medicine. Such an approach would enable clinicians to monitor temporal changes in prognostic classification, detect shifts in risk in response to therapy, and better capture disease evolution. Beyond its immediate clinical applications, the consistent prognostic performance of circulating H3K27Me3-nucleosome suggests that this biomarker could represent a promising candidate for integration into future computational patient-modeling frameworks. Within such evolving approaches, circulating nucleosomes may contribute to more adaptive and individualized decision-support strategies in lung cancer care.

In this study, we demonstrate that H3K27Me3-nucleosome levels measured at diagnosis provide incremental prognostic informationfor overall survival in patients with lung cancer. This biomarker offers clinically relevant information to refine risk stratification—particularly among ctDNA-negative patients—and may support informed clinical decision-making. Beyond baseline prognostication, integration of H3K27Me3-nucleosome into nomogram-based models offers a pragmatic framework for individualized and potentially dynamic risk prediction. From a precision medicine perspective, such adaptive prognostic tools may help guide treatment strategies, enable longitudinal risk reassessment during follow-up, and ultimately support morepersonalized management of lung cancer.

Methods

Population Study

We included 617 patients followed between 2022 and 2024 at the Pulmonology Department of Lyon University Hospital and collaborating centers across France. This non-interventional study was conducted as a sub-study from CIRCAN II study approved by the Scientific and Ethics Committee of Hospices Civils de Lyon (Etude RNIPH CIRCAN II n°25-5099) and was conducted

in accordance with the Declaration of Helsinki. Blood samples were collected at initial clinical evaluation from treatment-naïve patients. Healthy control samples (n=201) were provided by the French blood agency "Etablissement Français du Sang," as previously described by Grolleau et al. [28]. An independent dataset including 120 samples from patients recruited from the same centers between 2017 and 2022 (Etude RNIPH NUCLEO_CIRCAN n°22-5065) was used as an external validation cohort for the Cox Proportional Hazards (CPH) model. Survival was measured according to dates of death from publicly available records of the French National Statistics Institute, INSEE.

Blood sampling and pre-analytical processing

All blood samples were collected in K2-EDTA tubes and centrifuged within 6 hours after blood collection for 10 min at 1600x g. The supernatant was centrifuged a second time at 6000 g for 10 min. The plasma was then aliquoted and stored at −80 °C until cfDNA extraction and nucleosome analysis.

Cell free DNA extraction and library preparation

Cell-free DNA (cfDNA) extraction was performed following the manufacturer's instructions using either the QIAamp Circulating Nucleic Acid Kit (ref 55114, Qiagen, Valencia, CA, USA) with a Qiagen vacuum manifold, or Maxwell cfDNA plasma kit (ref AS1840, Promega, Charbonnières-Les-Bains, France). CfDNA samples were quantified using Qubit™ dsDNA HS Assay Kit (32854, Invitrogen, Carlsbad, CA, USA) and Qubit™ 4 Fluorometer (ref Q33238, Invitrogen, Carlsbad, CA, USA).

DNA sequencing

For PlasmaSeq-Sensei panel, an optimal cfDNA input of 36ng (min 5 -max 100 ng) was used to generate libraries with Plasma-SeqSensei™ NSCLC RUO kit (Sysmex Inostics GmbH, Hamburg,

Germany) as described previously [33]. The assay covers somatic variants in BRAF, EGFR, KRAS, and PIK3CA genes (see Supplemental Information). The sensitivity is 0.1%.

For the custom comprehensive panel, libraries were manually prepared with an optimal cfDNA input of 50ng (min 6 – max 150 ng) using a custom capture-based technology provided by SOPHiA GENETICS (Lausanne, Switzerland) according to the manufacturer's instructions as previously described [33]. The custom comprehensive panel covers 86 genes involved in cancer (including EGFR, TP53 or KRAS, see Supplemental Information). The sensitivity is 0.5%.

For the custom ultra-deep targeted panel, an optimal cfDNA input of 25ng (min 3 – max 32 ng) was used to manually prepare libraries according to manufacturer's instructions with a custom capture-based technology provided by SOPHiA GENETICS (Lausanne, Switzerland). The ultra-deep targeted panel covers 27 genes and enables the detection of somatic alterations and structural variants (see Supplemental Information). The sensitivity is 0.2% or more than 30 unique molecules.

All libraries were sequenced on short read Illumina technology (Illumina Technology, San Diego, CA, USA) in 2 × 150 paired-end runs for libraries prepared with SOPHiA GENETICS' chemistry, and in single end (1 x 148) for PlasmaSeq-Sensei's libraries.

Identification of the somatic alterations was performed using Sysmex Inostics software (Plasma SeqSensei IVD Software v1.1.7) and the Sysmex Inostics' bioinformatics pipeline for PlasmaSeq Sensei assay, with a limit of detection of VAF at 0.1%. For the two other custom assays provided by SOPHiA GENETICS, the bioinformatics analyses were performed using the SOPHIA DDM platform. Respectively, the limits of detection validated internally were VAF 0.5% and 0.1% for the custom comprehensive panel and custom ultra-deep targeted assay.

Quantification of H3K27Me3-nucleosomes

All samples were tested using Nu.Q® H3K27Me3 ChLIA immunoassay (Belgian Volition SRL, Isnes, Belgium) according to the manufacturer's instructions. This sandwich immunoassay,

based on chemiluminescence technology, was performed using the IDS-i10 automated immunoanalyzer system (Immunodiagnostic System Ltd., Boldon, UK). Briefly, K2-EDTA plasma samples were centrifuged at 14 000g for 2 min and 50 µL of plasma were incubated with acridinium ester labeled anti-nucleosome detection antibody. Magnetic particle beads coated with anti-histone modification H3K27Me3 monoclonal capture antibody were added. After washing, trigger solutions were added, and the light emitted by the acridinium ester was measured using the luminometer. The results were reported in relative light units, and the concentration (in ng/mL) was extrapolated using four-parameter logistic regression based on a reference standard curve.

Statistical Analysis

All analyses were performed in R (version 4.4.1; R Foundation for Statistical Computing). If not otherwise mentioned, tidyverse (v2.0.0) was used for data wrangling and visualization; ggplot2/ggpubr (v4.0.0/v0.6.1) for box plots and p-value annotation; rstatix (v0.7.2) for non-parametric tests and label positioning; FSA (v0.10.0) for Dunn post-hoc tests; multcompView (v0.1.10) for compact letter displays; and survival (v3.8.3)/survminer (v0.5.1) for time-to-event analyses. Two-sided p-values < 0.05 were considered statistically significant unless otherwise specified. Where applicable, 95% confidence intervals (CI) are reported.

Data preprocessing

Continuous variables were inspected for outliers and distributional properties. Unless otherwise noted, continuous predictors were summarized as median (interquartile range, IQR) and compared between groups using non-parametric tests. Categorical variables were summarized as counts (percentages). Missing data were handled by complete-case analysis (i.e., observations with missing values for the variables involved in a given analysis were excluded from that specific analysis). When applicable, biomarker values were log-transformed (log10) prior to analysis for visualization and rank-based testing.

Box plots and group comparisons

For two-group comparisons, the Wilcoxon rank–sum test (Mann–Whitney U) was computed with rstatix::wilcox_test(). For comparisons involving three or more groups, the Kruskal–Wallis test was first applied; when the global test was significant, post-hoc pairwise comparisons were conducted using Dunn's test with Bonferroni adjustment for multiple testing (FSA::dunnTest()). For multi-group displays, Bonferroni-adjusted pairwise p-values were summarized as compact letter displays (multcompView::multcompLetters())

Survival analysis (Kaplan–Meier) and Optimal cut-point exploration

Time-to-event outcomes were analyzed using Kaplan–Meier (K-M) estimators. Survival time was defined as the interval (in months) from [index time point] to [event/censoring]. The data-cut date for censoring was [15-09-2024], to include patients with at least 8 months of follow up period from the day of the analysis. Patients without event by the data cut date were right censored at their last follow up.

K-M curves were generated for predefined clinical subgroups and visualized with survminer::ggsurvplot(). Between-group differences were assessed with two-sided log-rank tests using survival::survdiff() [34-36]. Median survival (95% CI) and restricted mean survival time (RMST; prespecified horizon) were reported when estimable using survRM2::rmst2. The RMST horizon was fixed at 30 months and chosen a priori based on follow-up distribution (truncation at the minimum of the largest observed event time across groups) [37,38].

Optimal cut-off was identified using maximally selected log-rank statistics (maxstat::maxstat.test) [39-41]; patients were then classified as low/high risk subgroups and K-M curves reestimated. Effect sizes were quantified via univariable Cox models (CPH), reported as hazard ratios (HR) with 95% CI. Median survival with 95% confidence intervals was reported when estimable. Proportional-hazards' assumptions were assessed using Schoenfeld residuals (cox.zph); no material violations were detected for the reported univariable contrasts, or results

were interpreted descriptively where diagnostics were inconclusive. RMST over a prespecified horizon was additionally summarized for interpretability.

Predictive model development and validation

Model development

Reporting of the multivariable prediction model followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (see Supplemental Information). We implemented a fully reproducible, calibrated CPH modeling pipeline with strict separation between training and held-out test data. The prespecified model included four predictor covariates: (i) H3K27Me3-nucleosome level (≥56 vs <56 ng/mL), (ii) cancer stage (I–III vs IV), (iii) blood mutation status (ctDNA-negative, EGFR, KRAS, TP53, or other), and (iv) age (per year). Patients with undetermined cancer stage were excluded, yielding a final cohort of 382 patients for model development.

The dataset was randomly split (70/30) into a training set and an independent held-out test set using a fixed random seed. Reference levels for categorical predictors were defined in the training set and propagated unchanged to the test set. The training set met predefined requirements for sample size (n = 267; required n = 248), number of events (ne = 125; required ne = 116), and events per predictor (EPP = 17.9; required EPP = 16.6), as determined using pmsampsize(type = "s"). The held-out test set comprised n = 115 individuals with ne = 47 events, providing sufficient information to estimate model performance with acceptable precision.

The prespecified CPH model was evaluated via 5-fold stratified outer cross-validation on training data only. In each outer split, the Cox model was fit, the baseline cumulative hazard $\mathrm{H}_0(\mathrm{t})$ was derived, and calibrated survival was obtained using a one-parameter slope recalibration γ (from $\mathrm{coxph}\{\mathrm{Surv} \sim \eta\}$, where η is the linear predictor) together with a recalibrated baseline $\mathrm{H}_0^{\mathrm{cal}}(\mathrm{t})$, yielding $\hat{S} = exp\{-H_0^{cal}(t)\, exp(\gamma\, \eta)\}$. Risk at $\mathrm{t}_0 = 30$months defined a fold-specific high/low threshold via surv_cutpoint (minimum group proportion 0.20; quantile fallback). After

cross-validation, a final Cox model was fit on all training data, recalibrated on a fixed 40% training subset to estimate γ_{final} and $\mathrm{H}_0^{\text{final}}(\mathrm{t})$, and a single training-based cut-point at t_0 was re-estimated for classification [42-44].

On the held-out test set (no refitting), time-dependent discrimination and calibration were quantified using *riskRegression::Score* function applied to the final recalibrated Cox model with metrics = c("AUC","Brier"), a grid from 3 to 30 months in 3-month steps, K-M censoring (cens.model = "K-M"), and no resampling (split.method = "none") [45-47]. The area under the time-dependent Receiver Operating Characteristic (ROC) curve $\mathrm{AUC}(t)$ and the Brier score $\mathrm{Brier}(t)$ were summarized over time; the integrated Brier score (IBS) was computed by trapezoidal integration of $\mathrm{Brier}(t)$ over the evaluated domain, normalized by its length. K-M curves for the final high/low groups were displayed with a risk table. Discrimination was assessed by a log-rank test, univariable Cox hazard ratios (HRs, 95% CIs; reference = low-risk) and Harrell's C-index. RMST was computed at a common, data-bounded horizon $\tau = \min\{t_0, \min_g \max(\text{follow-up in group } g)\}$, using survRM2::rmst2 and log-rank p, HR (95% CI), and RMST were annotated. Agreement between predicted and empirical survival was summarized by comparing calibrated $\hat{S}(t)$ with $S_{\text{K-M}}(t)$ 6, 12, 18, 24, and 30 months. Independent seeds controlled the split, outer-fold allocation, inner/final calibration sampling, and cut-point search; altering the calibration or cut-point seeds changes γ, $H_0(t)$, and/or the threshold, but not the composition of the held-out test cohort

External validation

External validation was conducted in an independent cohort not used at any stage of model development, tuning, calibration, or cut-point selection. The finalized Cox proportional hazards model, including fixed coefficients, baseline hazard, and risk threshold, was applied without refitting. Model transportability was primarily assessed through predefined risk stratification, using K-M survival analysis to compare high- and low-risk groups. Between-group discrimination was evaluated using the log-rank test, univariable Cox hazard ratios with 95% confidence

intervals (reference: low-risk group), and Harrell's C-index. RMST was estimated at 30 months for both risk groups. Calibration was examined by comparing predicted survival probabilities with empirical K-M estimates at clinically relevant time points.

Data availability

Due to patient privacy and institutional ethical regulations, the datasets from this study are not publicly available but can be requested from the corresponding author upon a reasonable request.

Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Acknowledgments

The authors thank Dr. Andrew Retter for careful reading and constructive comments on the manuscript. We would like to thank AstraZeneca for their financial support in developing and validating new prognostic and predictive biomarkers for use in routine settings. The funder plays no role in the study design, analysis and interpretation of data, or writing of the manuscript.

Author contributions

Conception and design of the work: LV, SC, MH, MP, LP; Patient inclusion, diagnosis and follow-up: EG, MD, PMe, PMa, SL, SC; Data acquisition: GL, MP; Data analysis: AK, JC, MP; Data

interpretation: AK, EG, JC, MH, MP, LP; Model development: AK; Manuscript writing: AK, JC, MH, MP, LP. All authors read and approved the final manuscript.

Funding

This research did not receive any grant funding. The nucleosome kits were provided free of charge by VolitionRx company to LP. AstraZeneca contributed to staff funding for MP.

Competing Interests

AK, JC and MH are employees of Belgian Volition, SRL. MH holds shares of VolitionRx Limited. Belgian Volition SRL has patents covering Nu.Q® technology and are developers of Nu.Q® assays. LP received financial support for consumable from VolitionRx and AstraZeneca. MP received VolitionRx support for congress participation. All other authors declare no financial or non-financial competing interests.

Additional Information

This manuscript is complemented by 1 additional file “Supplemental_Information.docx”. This additional file compiles:

- Supplemental Figures S1, S2, S3, and S4.
- Supplemental Tables S1 and S2.
- Supplemental Information regarding Material and Methods.

References

1 Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229-263 (2024). https://doi.org:10.3322/caac.21834

2 National Cancer Institute (NCI), S., Epidemiology, and End Results (SEER) Program. *SEER Stat Facts: Lung and Bronchus Cancer*, <https://seer.cancer.gov/statfacts/html/lungb.html> (

3 De Angelis, R. *et al*. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. *Lancet Oncol* **15**, 23-34 (2014). https://doi.org:10.1016/S1470-2045(13)70546-1

4 Francisci, S. *et al*. Survival patterns in lung and pleural cancer in Europe 1999-2007: Results from the EUROCARE-5 study. *Eur J Cancer* **51**, 2242-2253 (2015). https://doi.org:10.1016/j.ejca.2015.07.033

5 N Howlader, A. N., M Krapcho, D Miller, A Brest, M. YU, J. Ruhl, Z. Tatalovich, A Mariotto, D. LEWIS, H Chen, E. Feuer, K(Cronin, A-M Noone, D Miller, M. Yu, D.R. Lewis, D. MILLER, AM Noone, K. BISHOP, SF Altekruse, CL Kosary, DR LEWIS, HS Chen, K Bishop, C.L KOSARY, J Garshell, DR Lewis, EJ Feuer, KA Cronin. SEER Cancer Statistics Review, 1975–2016. (National Cancer Institute, Bethesda, MD, 2019).

6 Li, J. *et al*. Prognostic factors of survival in patients with lung cancer after low-dose computed tomography screening: a multivariate analysis of a lung cancer screening cohort in China. *BMC Cancer* **25**, 646 (2025). https://doi.org:10.1186/s12885-025-14036-9

7 Paranjpe, I., Salter, A. I., Chen, K. & Das, M. Molecular Markers in the Era of Precision Care in Lung Cancer. *Semin Respir Crit Care Med* **46**, 396-408 (2025). https://doi.org:10.1055/a-2649-9199

8 Hofman, P. *et al*. Current challenges and practical aspects of molecular pathology for non-small cell lung cancers. *Virchows Arch* **484**, 233-246 (2024). https://doi.org:10.1007/s00428-023-03651-1

9 Marinello, A. *et al*. Circulating tumor DNA to guide diagnosis and treatment of localized and locally advanced non-small cell lung cancer. *Cancer Treat Rev* **129**, 102791 (2024). https://doi.org:10.1016/j.ctrv.2024.102791

10 Saarenheimo, J., Eigeliene, N., Andersen, H., Tiirola, M. & Jekunen, A. The Value of Liquid Biopsies for Guiding Therapy Decisions in Non-small Cell Lung Cancer. *Front Oncol* **9**, 129 (2019). https://doi.org:10.3389/fonc.2019.00129

11 Ku, B. M. *et al*. Role of Circulating Tumor DNA Profiling in Patients with Non-Small Cell Lung Cancer Treated with EGFR Inhibitor. *Oncology* **100**, 228-237 (2022). https://doi.org:10.1159/000516813

12 Cicin, I. *et al*. ALK TKI therapy in patients with ALK-positive non-small cell lung cancer and brain metastases: A review of the literature and local experiences. *Crit Rev Oncol Hematol* **180**, 103847 (2022). https://doi.org:10.1016/j.critrevonc.2022.103847

13 Boukouris, A. E. *et al*. A comprehensive overview of minimal residual disease in the management of early-stage and locally advanced non-small cell lung cancer. *NPJ Precis Oncol* **9**, 178 (2025). https://doi.org:10.1038/s41698-025-00984-9

14 Serna-Blasco, R. *et al*. Comprehensive molecular profiling of advanced NSCLC using NGS: Prevalence of druggable mutations and clinical trial opportunities in the ATLAS study. *Lung Cancer* **204**, 108550 (2025). https://doi.org:10.1016/j.lungcan.2025.108550

15 Penault-Llorca, F. & Socinski, M. A. Emerging molecular testing paradigms in non-small cell lung cancer management-current perspectives and recommendations. *Oncologist* **30** (2025). https://doi.org:10.1093/oncolo/oyae357

16 Provencio-Pulla, M. *et al*. Identification of non-actionable mutations with prognostic and predictive value in patients with advanced or metastatic non-small cell lung cancer. *Clin Transl Oncol* **26**, 1384-1394 (2024). https://doi.org:10.1007/s12094-023-03362-8

17 Chen, Y., Peng, S. & Wang, C. The advances of DNA methylation in the liquid biopsy for the detection of lung cancer. *Front Oncol* **15**, 1547797 (2025). https://doi.org:10.3389/fonc.2025.1547797

18 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011). https://doi.org:10.1038/nature09784

19 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007). https://doi.org:10.1016/j.cell.2007.02.005

20 Simon, J. A. & Lange, C. A. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat Res* **647**, 21-29 (2008). https://doi.org:10.1016/j.mrfmmm.2008.07.010
21 Cao, R. *et al*. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002). https://doi.org:10.1126/science.1076997
22 Czermin, B. *et al*. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-196 (2002). https://doi.org:10.1016/s0092-8674(02)00975-3
23 Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009). https://doi.org:10.1038/nrm2763
24 Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-649 (2015). https://doi.org:10.1038/nrm4067
25 Bajbouj, K., Al-Ali, A., Ramakrishnan, R. K., Saber-Ayad, M. & Hamid, Q. Histone Modification in NSCLC: Molecular Mechanisms and Therapeutic Targets. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms222111701
26 Zhang, G., Wang, Z., Song, P. & Zhan, X. DNA and histone modifications as potent diagnostic and therapeutic targets to advance non-small cell lung cancer management from the perspective of 3P medicine. *EPMA J* **13**, 649-669 (2022). https://doi.org:10.1007/s13167-022-00300-6
27 Dai, J. *et al*. NNMT promotes acquired EGFR-TKI resistance by forming EGR1 and lactate-mediated double positive feedback loops in non-small cell lung cancer. *Mol Cancer* **24**, 79 (2025). https://doi.org:10.1186/s12943-025-02285-y
28 Grolleau, E. *et al*. Circulating H3K27 Methylated Nucleosome Plasma Concentration: Synergistic Information with Circulating Tumor DNA Molecular Profiling. *Biomolecules* **13** (2023). https://doi.org:10.3390/biom13081255
29 Liu, L. *et al*. A molecular and staging model predicts survival in patients with resected non-small cell lung cancer. *BMC Cancer* **18**, 966 (2018). https://doi.org:10.1186/s12885-018-4881-9
30 Dhillon, P., Du, S. & Cheng, H. Circulating Tumor DNA as a Prognostic and Predictive Biomarker in Lung Cancer. *Cancers (Basel)* **17** (2025). https://doi.org:10.3390/cancers17203327
31 Al-Qaisi, T. S., Abumsimir, B., Sughayer, M. & Kasmi, Y. Actionable Mutations and Survival Rates in Non-Small Cell Lung Cancer. *World J Oncol* **16**, 161-172 (2025). https://doi.org:10.14740/wjon2531
32 Song, Z., Lv, T., Zhang, Y. & Song, Y. Frequencies of actionable mutations and survival in variants of invasive adenocarcinoma of lung. *Translational Cancer Research* **6**, 1162-1169 (2017).
33 Barthelemy, D. *et al*. Paired Comparison of Routine Molecular Screening of Patient Samples with Advanced Non-Small Cell Lung Cancer in Circulating Cell-Free DNA Using Three Targeted Assays. *Cancers (Basel)* **15** (2023). https://doi.org:10.3390/cancers15051574
34 Netskar, H. *et al*. Pan-cancer profiling of tumor-infiltrating natural killer cells through transcriptional reference mapping. *Nat Immunol* **25**, 1445-1459 (2024). https://doi.org:10.1038/s41590-024-01884-z
35 Li, M., Su, Y., Gao, Y. & Tian, W. ReCIDE: robust estimation of cell type proportions by integrating single-reference-based deconvolutions. *Brief Bioinform* **25** (2024). https://doi.org:10.1093/bib/bbae422
36 Liu, T. *et al*. Identification of immune patterns in idiopathic pulmonary fibrosis patients driven by PLA2G7-positive macrophages using an integrated machine learning survival framework. *Sci Rep* **14**, 22369 (2024). https://doi.org:10.1038/s41598-024-73625-z
37 Mekontso Dessap, A. *et al*. Comparison of Prophylactic and Therapeutic Doses of Anticoagulation for Acute Chest Syndrome in Sickle Cell Disease: The TASC Double-Blind Controlled Randomized Clinical Trial. *Am J Respir Crit Care Med* **211**, 832-841 (2025). https://doi.org:10.1164/rccm.202409-1727OC
38 Mayers, J. R. *et al*. A metabolomics pipeline highlights microbial metabolism in bloodstream infections. *Cell* **187**, 4095-4112 e4021 (2024). https://doi.org:10.1016/j.cell.2024.05.035

39 Roider, T. *et al.* Multimodal and spatially resolved profiling identifies distinct patterns of T cell infiltration in nodal B cell lymphoma entities. *Nat Cell Biol* **26**, 478-489 (2024). https://doi.org:10.1038/s41556-024-01358-2

40 Fan, G. *et al.* Single-cell and spatial analyses revealed the co-location of cancer stem cells and SPP1+ macrophage in hypoxic region that determines the poor prognosis in hepatocellular carcinoma. *NPJ Precis Oncol* **8**, 75 (2024). https://doi.org:10.1038/s41698-024-00564-3

41 Liu, Z. L. *et al.* Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer. *Nat Commun* **15**, 2595 (2024). https://doi.org:10.1038/s41467-024-46912-6

42 Rahman, M. S., Ambler, G., Choodari-Oskooei, B. & Omar, R. Z. Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Med Res Methodol* **17**, 60 (2017). https://doi.org:10.1186/s12874-017-0336-2

43 Vonk, J. M. J., Greving, J. P., Gudnason, V., Launer, L. J. & Geerlings, M. I. Dementia risk in the general population: large-scale external validation of prediction models in the AGES-Reykjavik study. *Eur J Epidemiol* **36**, 1025-1041 (2021). https://doi.org:10.1007/s10654-021-00785-x

44 Wang, J. *et al.* Development and external validation of a prognostic model for survival of people living with HIV/AIDS initiating antiretroviral therapy. *Lancet Reg Health West Pac* **16**, 100269 (2021). https://doi.org:10.1016/j.lanwpc.2021.100269

45 Zhang, Z. *et al.* Overview of model validation for survival regression model with competing risks using melanoma study data. *Ann Transl Med* **6**, 325 (2018). https://doi.org:10.21037/atm.2018.07.38

46 Hartz, S. M. *et al.* A flexible modeling approach for biomarker-based computation of absolute risk of Alzheimer's disease dementia. *Alzheimers Dement* **19**, 1452-1465 (2023). https://doi.org:10.1002/alz.12781

47 Ozenne, B., Sørensen, A. L., Scheike, T., Torp-Pedersen, C. & Gerds, T. A. riskRegression: Predicting the Risk of an Event using Cox Regression Models. *The R Journal* **9**, 440-460 (2017). https://doi.org:10.32614/RJ-2017-062

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalInformation.docx