2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

#23867 Uncertainty-Aware Multimodal Biomarker-Guided Treatment Response Prediction: Integrating FDG PET, T-cell Repertoire, and Inflammatory Cytokines with Conformal Uncertainty Quantification

Faisal Yaseen[1], Daniel S. Hippe[2], Sunan Cui[1], Jie Fu[1], Yejin Kim[1], Clemens Grassberger[1], Ting Ye[1], Paul Kinahan[1], John H. Gennari[1], Stephen R. Bowen[1,2]

1. University of Washington, Seattle 2. Fred Hutchinson Cancer center

ABSTRACT

Purpose: Uncertainty-aware multimodal biomarker-guided treatment response prediction in metastatic NSCLC remains a critical unmet need to support robust therapy selection and adaptation over time. We developed a multimodal framework integrating FDG-PET, T-cell receptor (TCR), and cytokine biomarkers with conformal prediction for uncertainty-quantified treatment response prediction.

Methods: Thirty-five metastatic NSCLC patients (19 female [54%], age: 33-84 years [median: 71 years]) receiving first-line platinum-doublet–pembrolizumab on the PET-BRIGHT trial (NCT04151940) from 2019-2023 underwent FDG-PET/CT and blood collection at baseline and week 3; 27 were evaluable (18 post-treatment responders, 9 non-responders). Multimodal biomarkers included FDG-PET metrics (standardized uptake value/total lesion glycolysis), T-cell receptor (TCR) diversity metrics, and inflammatory cytokines. Nested leave-one-out cross-validation with direction-aligned consensus feature selection identified a single biomarker per modality. Class-balanced logistic regression was implemented to ensure well-calibrated probabilities for conformal prediction. Response prediction performance was evaluated via AUCROC and Brier score. Full conformal prediction (80% coverage) generated sets of {responders, non-responders}, with certainty defined by average set size (singleton=definitive). Unimodal and multimodal fusion (early vs late) models of baseline-only and baseline+mid-treatment biomarkers for adaptation/consolidation were compared.

Results: At baseline, unimodal TCR was dominant (AUCROC 0.80, Brier: 0.19), outperforming PET (AUCROC: 0.76, Brier: 0.21) and cytokines (AUCROC: 0.55, Brier: 0.27). Late fusion PET+TCR improved discrimination (AUCROC: 0.85, Brier: 0.17) and reduced uncertainty (set size 1.11 vs 1.22), with increased singleton predictions (78% to 89%). When using baseline+mid-treatment biomarkers, unimodal cytokines improved relative to the baseline-only model (AUCROC 0.78, Brier 0.20), while PET was similar (AUCROC: 0.76) and TCR was attenuated (AUCROC: 0.67). Late fusion PET+cytokines achieved highest discrimination (AUCROC: 0.86, Brier: 0.17) with identical set size and singleton predictions at both time-points (1.15, 85%).

Conclusion: Multimodal late fusion of longitudinal biomarkers with conformal prediction enhanced response prediction while providing patient-level uncertainty quantification for clinical decision support in managing metastatic NSCLC.

CONTACT: fyaseen@uw.edu

INTRODUCTION

- Metastatic non-small cell lung cancer(mNSCLC) patient experience variable response to standard of care chemoimmunotherapy.
- . Many patients experience toxicity and limited clinical benefit (roughly 45-55%).
- Programmed cell death ligand 1 (PD-L1) and tumor mutation burden (TMB) are weak single modality predictors and miss the interplay of tumor burden, immunity and inflammation.
- FDG-PET and circulating immune biomarkers (TCR, cytokines) offer complementary view of tumor biology and treatment response.
- Multimodal models improve accuracy compared to unimodal but often mis calibrated [1]
- We utilized conformal prediction [2], a distribution-free, model agnostic method to deliver statistically valid per patient uncertainty estimate for trustworthy treatment selection and adaptation over time.

METHODS AND MATERIALS

Cohort & Design: PET-BRIGHT trial (NCT04151940); FDG-PET/CT + blood at baseline (PreTx) and week 3 (MidTx). Response by RECIST v1.1 at week 12 (responder = CR/PR = 0; non-responder = SD/PD = 1).
Multimodal biomarkers: PET: SUVmax / SUVpeak / TLG from primary MTV (lung + liver). TCR β-chain diversity & clonality (ImmunoSEQ). 14 circulating cytokines (Isoplexis / ELISA). One biomarker per modality via nested LOOCV.
Modeling: Class-balanced logistic regression; early fusion (concatenate features) vs late fusion (average modality probabilities); evaluated in baseline-only and baseline + MidTx settings.
Uncertainty & evaluation. Full conformal prediction (Jackknife+, 80% coverage, α = 0.2) → prediction set {0}, {1}, or {0,1}. Metrics: AUROC, Brier, coverage, set size, singleton rate; paired DeLong tests; 1000-resample bootstrap CIs.

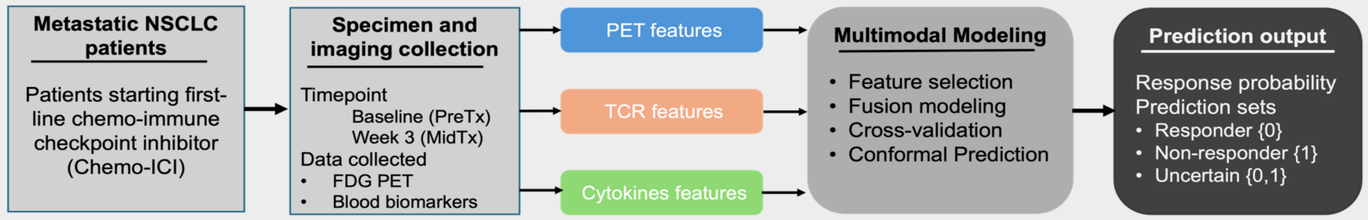

Fig 1. Study workflow: baseline & week-3 FDG-PET + blood → PET, TCR, cytokine features → multimodal fusion (nested LOOCV) with conformal prediction → per-patient response probability and prediction set.

RESULTS

- 35 mNSCLC patients enrolled; 27 evaluable (18 responders, 9 non-responders; 66.7% responders).
- Stable feature selection: SUVpeak selected in all folds; Pielou evenness dominant at baseline, IL-6 (MidTx) selected in 26/27 folds.
- Baseline: TCR was the best unimodal predictor (AUROC 0.80); PET+TCR late fusion improved discrimination to AUROC 0.85.
- Mid-treatment: PET+cytokine late fusion achieved the highest AUROC (0.86); DeLong gains vs. unimodal were exploratory (not statistically significant).

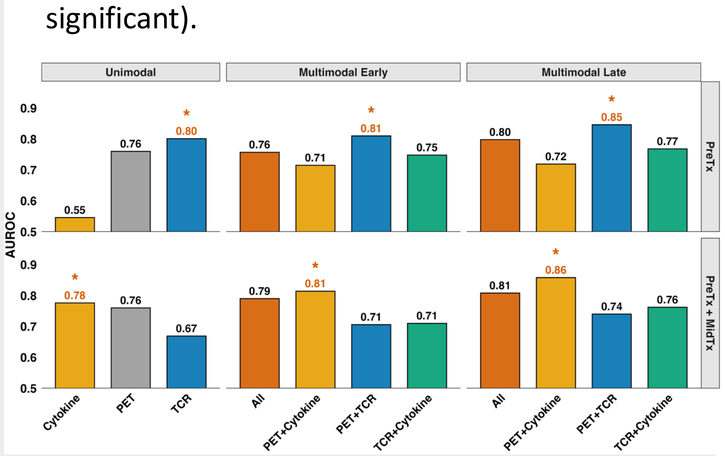

*Fig 2. AUROC for unimodal, early- and late-fusion models at baseline (PreTx, top) and combined PreTx+MidTx (bottom). * = best model per panel.*

- **Multimodal fusion improved calibration**: Brier score fell from 0.19 (unimodal TCR) to 0.17 (PET+TCR) at baseline, and reached 0.17 for PET+cytokine at mid-treatment
- **Fusion also increased decisiveness:** singleton (definitive) prediction rate rose from 78% to 89% at baseline and stayed high (85%) at mid-treatment.

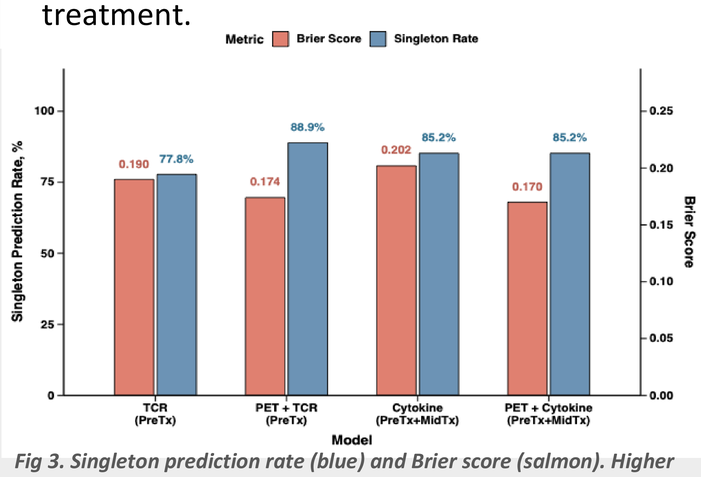

Fig 3. Singleton prediction rate (blue) and Brier score (salmon). Higher singleton = more confident; lower Brier = better calibration.

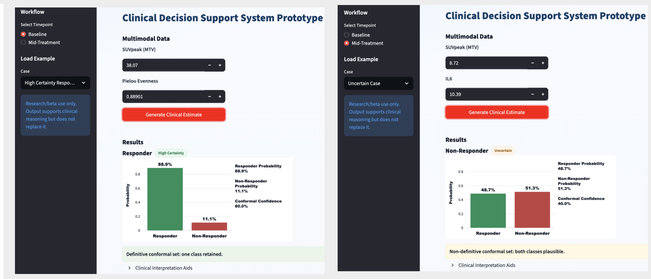

Fig 4. Decision-support prototype: (A) baseline definitive prediction; (B) mid-treatment uncertain prediction {0,1}.

DISCUSSION

- Late fusion consistently beat unimodal and early-fusion models, with a biological handoff from baseline TCR evenness to mid-treatment IL-6, averaging complementary metabolic and immune signals.
- Best AUROC (0.86) matches prior NSCLC studies [1],[3] (Vanguri 0.80; Captier). Fusion also improved calibration (Brier 0.19→0.17) and decisiveness (singleton 78→89%); {0,1} sets flag uncertain patients.
- **Limitations:** small single-institution cohort, no external validation, exploratory (NS) DeLong tests, and a proof-of-concept tool untested in workflow.

CONCLUSIONS

- Multimodal late fusion of longitudinal FDG-PET, TCR, and cytokines with conformal prediction improves early response prediction in metastatic NSCLC while quantifying per-patient uncertainty, with better discrimination, calibration, and decisiveness and a prototype tool showing clinical feasibility.
- Multicenter validation and prospective, workflow-based evaluation are the key next steps.

REFERENCES

1. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2021;22(2):114-126. doi:10.1038/s41568-021-00408-3
2. Angelopoulos AN, Bates S (2023), "Conformal Prediction: A Gentle Introduction". *Foundations and Trends in Machine Learning*, Vol. 16 No. 4 pp. 494–591, doi: https://doi.org/10.1561/2200000101
3. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer*. 2022;3(10):1151-1164. doi:10.1038/s43018-022-00416-8