Integrative multi-omics characterization and AI-driven biomarker discovery for NSCLC stage III outcomes

Katherina C. Chua[1]*, Yun-Ching Chen[1]*, Ariel Chen[1], Stewart Bates[1], Mehdi Pirooznia[1]^, Assieh Saadatpour[1]^
[1]*Johnson & Johnson, New Brunswick, NJ*
*KCC and YC are equal contributors and co-first authors ^MP and AS are senior co-authors
Research supported by: Johnson & Johnson

#1056

BACKGROUND

- The PACIFIC trial established consolidation durvalumab as standard of care for unresectable stage III non-small cell lung cancer (NSCLC) following concurrent chemoradiotherapy (CRT)
- Real-world studies such as PACIFIC-R[1] confirmed these benefits; however, comprehensive characterization of molecular features in this setting remains limited
- Advances in AI enable integration of heterogeneous genomic and transcriptomic data, opening new opportunities for biomarker discovery

Objective: Our study applies multimodal AI framework combining contrastive learning and foundation model representations to identify distinct molecular features associated with survival outcomes from pre-treatment tumor data

METHODS

Cohort Selection:

- Patients with diagnosed Stage III NSCLC disease, negative for EGFR/ALK alterations, and no evidence of resectable disease who were treated with CRT followed by durvalumab (CRT+D, N=281) or CRT alone (n=72) were identified from the Tempus AI multimodal real-world database
- All patients in cohort had tissue biopsies sequenced on Tempus targeted DNA and whole transcriptome RNA assays prior to CRT initiation

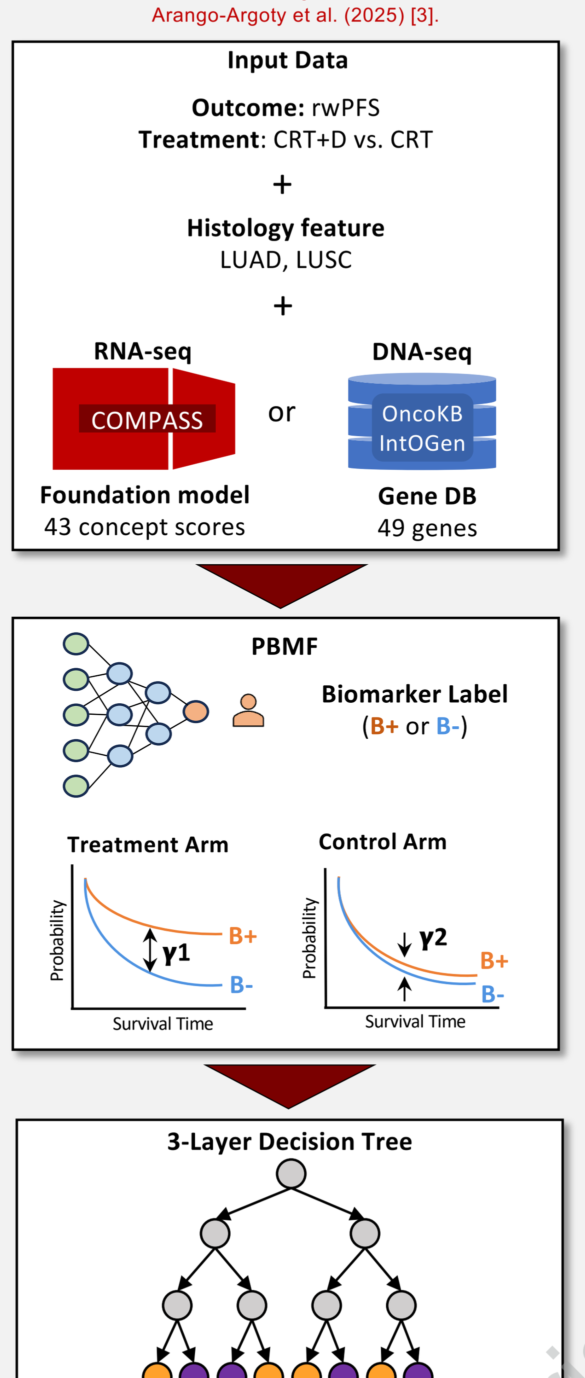
Molecular Identification and Characterization:

- Any patients identified with likely pathogenic/pathogenic single nucleotide variants and insertions/deletions, copy number loss (CN=0), copy number gain (CN>6), and/or presence of rearrangements as annotated by Tempus bioinformatic pipelines were considered positive for gene alteration
- Gene-level expression data were generated and normalized with Tempus bioinformatic pipelines. Principal component analysis were performed on the normalized log2-transformed expression data to assess sample relationships and potential bias
- Assessment of radiation and immunotherapy (IO)-related biomarkers were evaluated for impact on real-world progression-free survival (rwPFS)

AI-Integrated Framework for Biomarker Discovery:

- To identify candidate predictive biomarkers of treatment response to CRT+D, using CRT as the control arm for comparison, a multimodal AI framework was developed (Figure 1), integrating two recently published models:
  1) COMPASS[2], a foundation model encoding RNA profiles into interpretable immune cell states, tumor microenvironment, and signaling pathways
  2) Predictive Biomarker Mapping Framework (PBMF)[3], a deep learning model mapping molecular features to treatment outcomes
- Briefly, the framework utilized in this study was performed in following steps:
  1) Molecular features derived from whole-transcriptome data and targeted mutational data from NSCLC-associated genes[4-6] were used as inputs in individual PBMF models. All models included binary variables for LUAD and LUSC histology
  2) PBMF models generated identify predictive biomarker signatures associated with better outcomes (B+ classification) and worse outcomes (B- classification)
  3) Biomarker classification rules from PBMF models were extracted using 3-layer decision trees

Figure 1. AI-driven framework for biomarker discovery. Figure adapted from Arango-Argoty et al. (2025) [3].



DISCLOSURES: This study was funded by Johnson & Johnson. Authors KCC, YC, AC, SB, MP, and AS are employees of Johnson & Johnson

ACKNOWLEDGMENTS: The authors thank Tempus AI for providing collaborative data access and assistance, and Mu Sigma Inc., Bangalore, India for supporting analytics and providing design assistance for the poster.

REFERENCES: 1. Girard N, Bar J, Garrido P, et al. *J Thorac Oncol*. 2023;18(2):181-193. doi:10.1016/j.jtho.2022.10.003. 2. Shen W, Nguyen TH, Li MM, et al. Preprint. *medRxiv*. 2025;2025.05.01.25326820. Published 2025 May 5. doi:10.1101/2025.05.01.25326820. 3. Arango-Argoty G, Bikiel DE, Sun GJ, et al. *Cancer Cell*. 2025;43(5):875-890.e8. doi:10.1016/j.ccell.2025.03.029. 4. Martínez-Jiménez F, Muiños F, Sentís I, et al. *Nat Rev Cancer*. 2020;20(10):555-572. doi:10.1038/s41568-020-0290-x. 5. Chakravarty D, Gao J, Phillips SM, et al. *JCO Precis Oncol*. 2017;2017:PO.17.00011. doi:10.1200/PO.17.00011. 6. Suehnholz SP, Nissan MH, Zhang H, et al. *Cancer Discov*. 2024;14(1):49-65. doi:10.1158/2159-8290.CD-23-0467. 7. Ayers M, Lunceford J, Nebozhyn M, et al. *J Clin Invest.* 2017;127(8):2930-2940. doi:10.1172/JCI91190 8. Kaufman JM, Amann JM, Park K, et al. *J Thorac Oncol.* 2014;9(6):794-804. doi:10.1097/JTO.0000000000000173. 9. Eschrich SA, Pramana J, Zhang H, et al. *Int J Radiat Oncol Biol Phys.* 2009;75(2):489-496. doi:10.1016/j.ijrobp.2009.06.014

RESULTS

Patient and tumor characteristics

- Positive PD-L1 status was observed for >50% of cohort with most tumors (>80%) classified as TMB-Low (<10 mut/Mb)

Table 1. Patient counts for baseline characteristics.

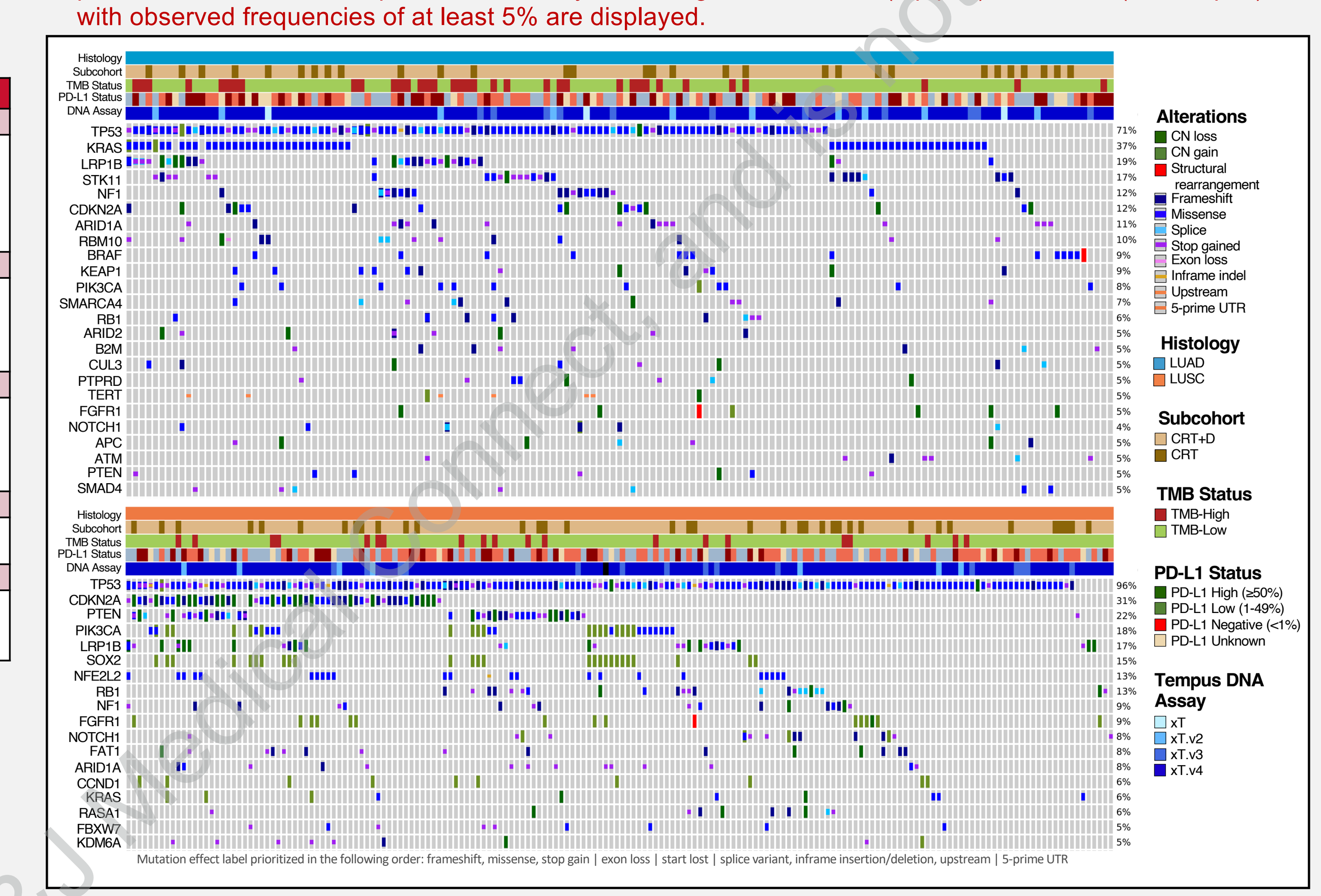
	CRT + D (% of 281)	CRT alone (% of 72)
Age		
≤59	64 (23)	8 (11)
60-69	119 (42)	29 (40)
70-89	98 (35)	35 (49)
Sex		
Female	113 (40)	27 (38)
Male	168 (60)	45 (62)
Race		
Asian	3 (1)	2 (3)
Black or Afr. Am.	32 (11)	7 (10)
Other Race	9 (3)	3 (4)
Unknown	50 (18)	13 (18)
White	187 (67)	47 (65)
ECOG Status		
0-1	133 (47)	30 (42)
2-3	14 (5)	7 (10)
Unknown	134 (48)	35 (48)
Smoking Status		
Current smoker	74 (27)	19 (26)
Former smoker	138 (49)	36 (50)
Never smoker	4 (1)	1 (2)
Unknown	65 (23)	16 (22)

	CRT + D (% of 281)	CRT alone (% of 72)
Overall Stage		
Stage 3A	125 (44)	36 (50)
Stage 3B	111 (40)	22 (31)
Stage 3C	25 (9)	5 (7)
Stage 3, NOS	14 (5)	7 (10)
Unknown	6 (2)	2 (2)
Histology		
Adenocarcinoma (LUAD)	120 (42)	30 (42)
Squamous cell carcinoma (LUSC)	142 (51)	37 (51)
Other	19 (7)	5 (7)
PD-L1 Status		
High	62 (22)	11 (15)
Low	90 (32)	26 (36)
Negative	93 (33)	28 (39)
Unknown	36 (13)	7 (10)
TMB Status		
High	49 (17)	13 (18)
Low	232 (83)	59 (82)
CRT Regimen		
Carboplatin-based	244 (87)	66 (92)
Cisplatin-based	34 (12)	5 (7)
Both	3 (1)	1 (1)

- TP53 was the most frequently mutated gene in LUAD and LUSC tumors. LUAD tumors featured *KRAS* (37%), *STK11* (15%), and *BRAF* (9%) variants while LUSC tumors notably displayed alterations in *PI3KCA* (18%), *SOX2* (15%), and *NFE2L2* (13%)

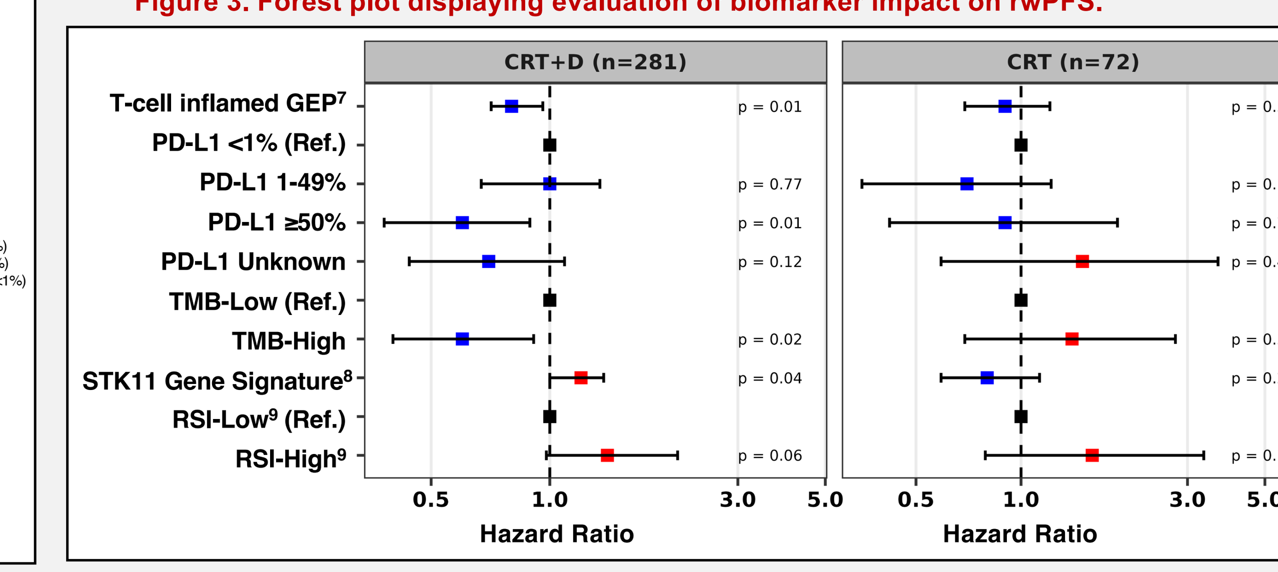
Molecular profiling and biomarker evaluation

Figure 2. Oncoprint displaying the mutational profile of cohort. Baseline characteristics for each patient are shown in the top bars. Commonly mutated genes for LUAD (top plot) and LUSC (bottom plot) with observed frequencies of at least 5% are displayed.



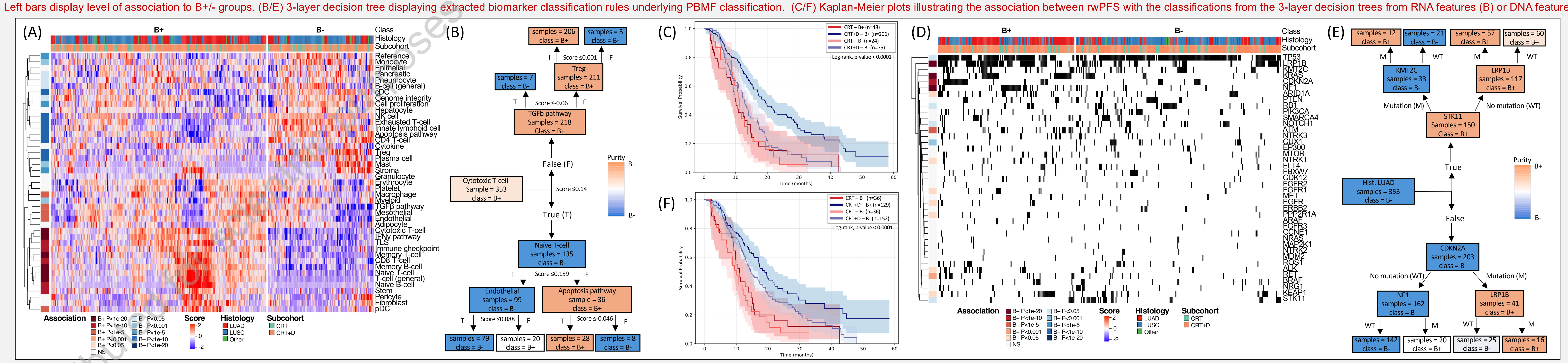
- In the CRT+D group, patients with high levels of biomarkers linked to IO response (i.e., high T-cell inflamed GEP[7], PD-L1 ≥50%, TMB-High, low STK11 gene signature[8]) showed improved survival rates
- High scores using radiosensitivity index (RSI) have been shown to indicate tumors with resistance to radiation[9]. Tumors with RSI-high signature trended toward diminished rwPFS benefit in both subcohorts

Figure 3. Forest plot displaying evaluation of biomarker impact on rwPFS.



AI-Integrated Framework for Biomarker Discovery

Figure 4. PBMF classification using 43 COMPASS concept scores encoded from RNA-seq profiles and somatic mutations in 49 driver genes as DNA features. (A/D) Heatmap exhibiting RNA (A) or DNA (D) features for each patient. Top bars display biomarker classification (B+/-), histology, and subcohort information. Left bars display level of association to B+/- groups. (B/E) 3-layer decision tree displaying extracted biomarker classification rules underlying PBMF classification. (C/F) Kaplan-Meier plots illustrating the association between rwPFS with the classifications from the 3-layer decision trees from RNA features (B) or DNA features (E).



- Our multimodal AI framework revealed additional molecular markers associated with improved rwPFS to CRT+D (**B+ classification**), including adenocarcinoma histology; mutations in *LRP1B*, *NF1*, *KRAS*, and *CDKN2A*; and enriched expression of immune activation pathways (cytotoxic T-cell, IFN-γ, tertiary lymphoid structure, and immune-checkpoint signatures)
- In contrast, tumors with certain molecular characteristics were associated with worse outcomes (**B- classification**), including *TP53*, *RB1*, *NOTCH1*, *CUX1*, and *STK11* mutations; elevated expression of T-reg, exhaustion, and stromal signals
- Cross-validation indicated generalizability for RNA features (P < 0.05) but not for DNA features, likely due to limited sample size. Evaluation in an independent cohort is warranted

CONCLUSION

- Our study utilized an AI-driven, foundation model-based analytical framework that demonstrated a scalable approach to integrate multi-omic features and extract rule-based molecular signatures specific to identify predictive biomarkers, offering a blueprint for interpretable, generalizable AI methods in translational oncology
- Future work could include expanding benchmarking across other real-world datasets, incorporating spatial and single-cell features, and refining explainable AI mechanisms for clinical translation

Scan the QR code

American Association for Cancer Research (AACR) Annual Meeting
April 17–22, 2026 | San Diego, California