Molecular subtypes identified in lung adenocarcinoma using a large-scale, real-world dataset

Akul Singhania¹, Prerna Jain¹, Ezgi Rizvi¹, Radia M. Johnson¹, Justin Guinney¹ ¹ Tempus AI, Inc., Chicago, IL

INTRODUCTION

- Despite advances in targeted and immuno-oncology (IO) therapies, lung adenocarcinoma (LUAD) remains a high unmet clinical need due to its complex molecular landscape.
- Previous studies have focused on early-stage disease with smaller patient populations to identify molecular subtypes in LUAD.
- This study characterized the molecular profiles and clinical features of LUAD patient tumors from a large real-world dataset (RWD) containing early and late-stage patients as well as primary and metastatic sites.

METHODS

- Molecular subtyping was performed using non-negative matrix factorization (NMF) on Tempus RWD RNA-seq data from primary lung site tumors (n=3,975; Stage III/IV: 68%).
- A random forest subtype classifier, trained on tumor-intrinsic features, was applied to primary non-lung and metastatic site tumors in Tempus RWD (n=3,981; Stage III/IV: 99%), as well as RNA-seq data from TCGA, patient-derived tumor organoids, and CCLE cell lines.
- Real-world progression-free survival (rwPFS), defined as the time from the start of first metastatic therapy to the first progression event (recurrence, progressive disease, change of treatment, or death from any cause), was assessed in LUAD patients with available data (n=1,267;17.06%). Clinical confounders were controlled using multivariate Cox proportional hazards (CoxPH) analysis.
- Weighted Gene Co-expression Network Analysis (WGCNA) was performed to identify gene modules associated with clinical features and molecular subtypes.
- Mutation over-representation analysis was performed to calculate z-scores by comparing observed mutation counts to a null model from 1,000 bootstrap resampling iterations.
- Immune cell proportions were estimated using quanTIseq.



Molecular subtype identification and characterization using de-identified clinical and genomic records from LUAD patients profiled with Tempus' targeted DNA and whole-transcriptome RNA assays.

ACKNOWLEDGMENTS

We thank Vanessa Nepomuceno from the Tempus Science Communications team for poster development.

SUMMARY

- and new treatment modalities.

RESULTS

Figure 1. Identification of six distinct molecular subtypes in LUAD

а	Te		empus RWD				TCGA			
Frequency		800 800 800 800 800 800 800 800 800 800	8 C4 (1720 64 064 064		C1 C2 C3 -C4 C5		P		
С		Quarall	C 1	C 2	<u> </u>	CA	<u>C</u>	6/		
Total Sa	eature	Overall 7429	2361	1398	<u>038</u>	522	1720	<u> </u>	p-val	
Age at biopsy (mean (SD))		68.35 (10.09)	68.14 (10.15)	68.23 (10.35)	69.61 (9.69)	67.73 (10.11)	68.25 (9.70)	68.33 (11.03)	0.003	
Biological Sex;		3513	1166	594	344	233	937	239	<0.001	
Male (n	(%))	(47.3)	(49.4)	(42.5)	(36.7)	(44.6)	(54.5)	(48.8)		
Stage at biopsy (n (%))	Stage 1	515 (8.8)	(4.1)	(6.9)	(38.1)	51 (12.3)	48 (3.5)	(5.4)	<0.001	
	Stade 2	427	87	56	137	47	76	24		
	Stage Z	(7.3)	(4.5)	(5.3)	(21.5)	(11.4)	(5.5)	(5.9)		
	Stage 3	936 (16.0)	303 (15.6)	168 (15.8)	133 (20.9)	68 (16.5)	229 (16.5)	35 (8.5)		
	Stade 1	3969	1470	766	124	247	1033	329		
		(67.9)	(75.8)	(72.1)	(19.5)	(59.8)	(74.5)	(80.2)		
Biopsy site (n (%))	Primary: Lung	3975	961 (40.7)	658 (47.1)	918 (97.9)	341 (65.3)	824 (47.9)	(55.7)	<0.001	
	Primary:	94	36	8	1	3	42	4		
	Non-Lung	(1.3)	(1.5)	(0.6)	(0.1)	(0.6)	(2.4)	(0.8)		
	Lymph	1224	578	335	7	54	222	28		
		(16.5) 173	(24.5)	(24.0)	(0.7)	(10.3)	(12.9) 171	(5.7)		
	Met: Brain	(6.4)	(7.1)	(9.8)	(0.2)	(4.6)	(8.2)	(0.4)		
	Met: Liver	385	103	74	1	12	182	13		
		(5.2)	(4.4)	(5.3)	(0.1)	(2.3)	(10.6)	(2.7)		
	Met: Bone	(3.8)	64 (2.7)	36	1 (0.1)	25	89	64 (13.1)		
	MatiOthar	999	452	150	8	63	220	106		
	Met:Uther	(13.4)	(19.1)	(10.7)	(0.9)	(12.1)	(12.8)	(21.6)		
TMB; Hi	igh (≥10)	1156 (15.6)	522 (22 1)	185 (13-2)	47 (5.0)	94 (18 0)	269 (15.6)	39 (8 0)	<0.001	
PDL1 status (n (%))	Negative	(15.0)	(22.1)		(5.0)	(10.0)	(15.0)	(0.0)	<0.001	
	(TPS <1)	(40.7)	363 (15.4)	(54.2)	467 (49.8)	(40.4)	(58.0)	231 (47.1)		
			740		200	100		100		
	Low positive $(\geq 1 \& < 50)$	2580 (34.7)	(30.5)	534 (38.2)	398 (42.4)	(38.1)	542 (31.5)	188 (38.4)		
	High positive	1822	1279	106	73	112	181	71		
	(1PS≥50)	(24.5)	(54.2)	(7.6)	(7.8)	(21.5)	(10.5)	(14.5)		

Figure 1. (a) Distribution of LUAD subtypes in the Tempus RWD. (b) Comparison of LUAD subtypes mapped to TCGA RNA-seq data versus the original LUAD subtypes identified by the Cancer Genome Atlas Research Network (Nature 2014). (c) Clinical characteristics of the molecular subtypes in Tempus LUAD RWD.

• This study highlights the heterogeneity of LUAD, revealing patient states with distinct mutational landscapes, tumor microenvironment (TME) profiles, and clinical outcomes in a real-world setting. • Future work is needed to explore how these subtypes may inform strategies for disease management

Figure 2. The molecular C5 subtype exhibits the poorest prognosis among LUAD patients



Figure 2. (a) Kaplan-Meier survival curves showing rwPFS from the start of index treatment (metastatic line of therapy [LOT] 1) to the first progression event. (b) Distribution of LOT1 therapy groups across LUAD subtypes. (c) Forest plot showing multivariate CoxPH analysis, controlling for clinical confounders.

Figure 3. Distinct transcriptomic and mutational landscapes characterize LUAD molecular subtypes



Figure 3. (a) Heatmap of the top 50 genes identified from NMF clustering, showing row-scaled log, TPM values. (b) Key biological pathways enriched in the C5 molecular subtype-associated gene modules identified using WGCNA. (c) Over-representation analysis showing z-scores for enriched mutations. Size indicates significance, and color reflects magnitude of enrichment. (d) Oncoplot of alteration prevalence in the C5 molecular subtype.

"I"EMPUS

Abstract Number # 4589

Figure 4. LUAD molecular subtypes exhibit unique TME profiles



Figure 4. (a) Heatmap of GSVA scores for 29 gene signatures grouped into four TME subtypes (Bagaev *et αl.*, 2021). (b) Single-cell RNA-seq analysis (Salcher *et al.*, 2022) showing expression of top 50 C5 molecular subtype genes in healthy and cancerous lung cells (UMAP plot) and epithelial cells (violin plot). (c) Proportion of 10 immune cell types across clusters estimated from bulk RNA-seq using quanTISeq.

Figure 5. *In vitro* models reflect LUAD molecular subtype-specific characteristics

Figure 5. Distribution of LUAD subtypes mapped to RNA-seq data from (a) patient-derived tumor organoids (PDO) and (b) CCLE cell lines. Scatterplots showing GSVA scores for the C5 signature and the four C5 subtype WGCNA gene modules with the prediction probability for the C5 molecular subtype assignment in (a) PDOs and (b) CCLE cell lines. Spearman correlation coefficients (r) and p-values are shown.